Year : 2023

Fakes, Forgeries, and Statistics in the Art Market:A Case Study of the Works of Félix Vallotton

Wuffli Liliana

Wuffli Liliana, 2023, Fakes, Forgeries, and Statistics in the Art Market:A Case Study of the Works of Félix Vallotton

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_840A3FF12A7D3

UNIL | Université de Lausanne
Faculté de droit, des sciences criminelles
et d'administration publique

UNIVERSITÉ DE LAUSANNE FACULTÉ DE DROIT, DES SCIENCES CRIMINELLES ET D'ADMINISTRATION PUBLIQUE
ÉCOLE DES SCIENCES CRIMINELLES

Fakes, Forgeries, and Statistics in the Art Market: A Case Study of the Works of Félix Vallotton

THÈSE DE DOCTORAT

présentée à la

Faculté de droit, des sciences criminelles et d'administration publique de l'Université de Lausanne

pour l'obtention du grade de

Docteure en criminologie

par

Liliana Wuffli

Directeur de thèse
Prof. Marcelo F. Aebi

Co-directeur de thèse
Prof. André Berchtold

Jury
Prof. Franco Taroni (Université de Lausanne) : Président
Prof. Stefano Caneppele (Université de Lausanne) : Expert interne
Dr. Nicolas Galley (Université de Zurich) : Expert externe

LAUSANNE 2023

UNIL | Université de Lausanne
Ecole des sciences criminelles
bâtiment Batochime
CH-1015 Lausanne

IMPRIMATUR

A l'issue de la soutenance de thèse, le Jury autorise l'impression de la thèse de Madame Liliana Wuffli, candidate au doctorat en droit en criminologie et sécurité, intitulée :

**« Fakes, Forgeries and Statistics in the Art Market :
A Case Study of the Works of Félix Vallotton »**

Professeur Franco Taroni
Président du jury

Lausanne, le 26 septembre 2023

ACKNOWLEDGMENT

This PhD dissertation was conducted at the School of Criminal Sciences of the University of Lausanne. I would like to express my heartfelt gratitude to all those without whom this work would not have been possible.

I am extremely grateful to my supervisors, Dr. Marcelo Aebi, Vice Director of the School of Criminal Sciences at the University of Lausanne, and Dr. André Berchtold, Director of the Institute of Social Sciences at the University of Lausanne. Professor Aebi gave me the opportunity to pursue a doctoral degree and agreed to supervise this research as my principal advisor. I thank him for his guidance, invaluable feedback, insightful comments, precious advice on theoretical matters, and skilful resolution of complex situations. I would also like to extend my sincere gratitude to Professor Berchtold for his enthusiasm, unwavering guidance, and willingness to engage in numerous enriching discussions. Without his support in the statistical analyses, this thesis would not have been possible. As my background in law and criminology only provided me with a basic knowledge of statistics, the advanced methods required for this experimental PhD study, such as the classification tree, necessitated the assistance of a statistical specialist. Professor Berchtold assumed this role along with his responsibilities as my thesis supervisor.

I would like to offer special thanks to Dr. Sarah Burkalter, Head of the Romandy Branch at the Swiss Institute for Art Research in Lausanne, for her interest and trust in my research, enriching discussions, advice in the field of art, and insightful support. She served as my guide to the art world, enabling access to information on Vallotton's counterfeits without which this project could not have been undertaken.

This Ph.D. project owes much to the cooperation and gracious support provided by the Vallotton Foundation. My deepest appreciation goes to the team at the Félix Vallotton Foundation: Katia Poletti and Marina Ducrey, both curators and authors of Félix Vallotton's catalogue raisonné, and Prof. Dr. Yves Noël, President of the Foundation. Our collaboration was vibrant with discussions, advice, and exchanges of opinions. The results of these discussions can be found throughout the pages of this thesis. I am

profoundly grateful for their trust in this Ph.D. research. They showed keen interest in the subject and welcomed me to their institution for data collection.

I extend my sincerest appreciation to the members of my dissertation committee. I would like to thank Prof. Franco Taroni (Committee President) from the School of Criminal Sciences at the University of Lausanne, Prof. Dr. Stefano Caneppele (Internal Expert) from the School of Criminal Sciences at the University of Lausanne, and Dr. Nicolas Galley (External Expert) from the Executive Master of Art Market Studies at the University of Zurich for their agreeing to be a part of the jury and for their interest in this research. I am grateful for their comments, questions, and suggestions, which greatly contributed to the completion of this study.

Many individuals provided assistance during this research study, and I am deeply indebted to them all. I extend my thanks to Lynda Albertson, Chief Executive Officer of the Association for Research into Crimes against Art, and Alexandre Von der Muhll, Inspector of the Vaud Cantonal Police, for their advice and help in searching for counterfeit data. I am grateful to Dr. Valentina Locatelli, an independent curator and writer, for her keen interest in the subject and invaluable knowledge of authentication issues. I thank Julie Janet, Julie van der Meulen, Nadia Bauer, Christopher Michael Hauck, Magdalena Mecweld, Barbara Meier Kuhn, Benedicte Chatelan, and Dr Elsa Reymond for their guidance and encouragement.

Finally, I would like to express my heartfelt appreciation to my daughter for her unwavering patience and to my husband for his daily support and enthusiasm for my work.

If it's good, it's mine. If it's bad, it's a fake

Pablo Picasso

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAM	American Association of Museums
CART	Classification and regression tree
CR	Catalogue raisonné
D	Categorical dichotomous variable
MD	Missing data
N	Numerical variable
NOM	Categorical nominal variable
PCA	Principal component analysis
PCs	Principal components
PCS	Police crime statistics
PR	Prison records
SNA	National Union of Antiquarians
SEM	Scanning electron microscope
SIK-ISEA	The Swiss Institute for Art Research
UV	Ultraviolet light
XRD	X-ray diffraction

GLOSSARY

Chromatography – the process in which a chemical mixture carried by a liquid or gas is separated into components as a result of the differential distribution of the solutes as they flow around or over a stationary liquid or solid

Commonplace book – a large handwritten notebook by an artist

Craquelure – the network of fine cracks in the paint surface

Machine learning – a computer science subfield focused on the generation of algorithms and statistical models to recognise patterns within data without using explicit instructions

Infrared radiation – the portion of the electromagnetic spectrum with frequencies just below those of visible light

Pastiche – a literary, artistic, musical, or architectural work that imitates the style of a previous work.

Pentimenti – a name for the traces of graphics and paintings that show that corrections were made to an image during production or even that other paintings were painted

Relined artwork – a canvas has been cut or cropped

Relining of painting – the replacement of a lining (for which relining is a common misnomer) originally applied to the back of an oil painting on canvas to strengthen it.

Underdrawings – the artist's preliminary sketch made before the paint layers are applied.

ABSTRACT

Various art authentication measures can be used to prevent the circulation of fakes and forgeries in the art market. However, owing to the complexity of their techniques, they are mainly used by professional art players (experts and art dealers). In this context, this Ph.D. thesis aims to develop a tool that can provide arguments regarding the risks of art counterfeiting in terms of probabilities. Although these arguments should not be considered evidence, they give an early *alert* to potential problems related to authenticity. This thesis adopts an original position by postulating that counterfeit artworks can be distinguished from original works by their standard characteristics such as subject matter, size, date, and place of signature. In support of this thesis, this doctoral research provides a comparative analysis of these characteristics using a data-mining model.

A practical advantage of the model lies in its ability to be applied by people without statistical training or the assistance of an art expert. This result was achieved owing to the readable features used for the model's creation. The database was built according to the characteristics of the *catalogue raiso*nné and archival documents on counterfeit works. Although data on forgeries are generally confidential, our methodology permits public access to this knowledge without revealing the owners' names.

The study was conducted using a series of experiments and case studies based on a dataset comprising information on 1,704 originals and 210 counterfeits related to the works of the Swiss and French artist Félix Vallotton (1865–1925). The accuracy of the model in classifying counterfeits in most tests is 70–91%. Despite these promising results, the model's probabilities can only be applied to the work of Vallotton. However, the sampling process and techniques used in this PhD study are standard. Hence, the methodology can be replicated to analyse the probabilities related to another artist. This can be achieved by using a new dataset incorporating the characteristics of the forgeries and the original works of the artist on consideration

RÉSUMÉ

Diverses mesures d'authentification des œuvres d'art peuvent servir de filtre pour empêcher la circulation de faux tableaux dans le marché de l'art. Or, elles sont toutes utilisées principalement par des acteurs professionnels (experts, marchands d'art) en raison de la complexité de leurs techniques. Dans ce contexte, cette thèse de doctorat vise à développer un outil capable de fournir des arguments sur les risques de contrefaçon d'œuvres d'art en termes de probabilités. Bien que ces arguments ne doivent pas être pris comme des preuves, ils constituent une *alerte* anticipée à un problème potentiel lié à la question de l'authenticité. La thèse de doctorat adopte une position originale en postulant que les faux peuvent être distingués des œuvres originales par leurs caractéristiques standards, telles que le sujet, la taille, la date et le lieu de la signature. À l'appui de cette thèse, la recherche doctorale présente une analyse comparative de ces caractéristiques à l'aide de modèles d'exploration de données.

Un avantage pratique du modèle réside dans sa capacité à être appliqué par des personnes sans formation statistique et sans l'assistance d'un expert en art. Ce résultat est obtenu grâce aux caractéristiques lisibles que nous avons utilisées pour la création du modèle. La base de données est construite selon la description des caractéristiques du catalogue raisonné et les documents d'archives sur les faux tableaux. Bien que l'information sur les faux soient des renseignements généralement confidentiels, notre méthodologie permet de rendre ces connaissances accessibles au public sans révéler les noms des propriétaires.

L'étude a été réalisée à l'aide une série d'expériences et de cas pratiques basés sur un ensemble de données comprenant l'information sur 1,704 originaux et 210 faux tableaux liés aux œuvres du artiste franco-suisse Félix Vallotton (1865–1925). Dans la plupart des tests, le modèle a démontré une précision allant de 70 % à 91 % lors de la classification des contrefaçons. Bien que ces résultats soient prometteurs, ces probabilités ne sont applicables qu'aux œuvres de Félix Vallotton. Cependant, l'échantillonnage et les techniques de l'étude doctorale sont standards. Ainsi, la méthodologie pourrait être reproduite pour analyser les probabilités faites pour un

autre peintre sur la base d'un nouvel ensemble de données incorporant les caractéristiques des faux et des œuvres originales de l'artiste en question.

INTRODUCTION

The art market is complex and diverse. Art counterfeiting, also known as *fakes* or *forgery*, is a major problem in this field. To combat art counterfeiting, it is necessary to establish preventive and control measures to ensure the authenticity of the works of art sold in the market. This may include enforcing strict laws and regulations to suppress counterfeiting. Such measures would involve lengthy and challenging legislative changes. Currently, several authentication methods are used in practice. How do they function, and which are the most appropriate to prevent the circulation of fakes and forgeries in the art market?

The stylistic connoisseurship method involves a visual examination performed by a *trained eye*, which includes two interdependent functions. One is to identify features corresponding to the artist's originality, whereas the other detects various incompatibilities and errors committed by counterfeiters. Similar to stylistic analysis, historical documentation analysis operates in two ways to prove the authenticity of an artwork by reconstructing the chain of ownership. In contrast, finding incoherence in a chronological sequence can reinforce doubts about authenticity. Another method used by art experts to confirm or refute doubts regarding authenticity is technical/scientific analysis. For example, laboratory examinations detect inconsistencies between the composition of the materials and pigments used in the production of the work and the materials available during an artist's lifetime. In these cases, when technical analysis does not identify suspicious elements, the evidence is insufficient to confirm authenticity without the help of other methods.

Among the various technical methods, computerised art classification stands out as operating similarly to stylistic connoisseurship. This approach enables the identification of distinct characteristics commonly associated with both genuine and forged artworks. Interestingly, while several studies have primarily focused on developing methods to classify original artworks (e.g., studies by Sablatnig, Kammerer, & Zolda, 1998; Lombardi, 2005; Saleh, Abe, Arora, & Elgammal, 2014; Trochim, Donnelly, & Arora, 2016; Chen, Chen, Zou, Huang, & Li, 2017), only a few published studies have explored methods capable of classifying originals as well as

counterfeits (e.g., studies by Teegen, 2002; Montani, 2015; Elgammal et al., 2018). These methodologies hold significant potential as detection tools that can rival traditional technical methods in effectiveness.

Hence, one of the critical focuses of this study is to understand the factors contributing to the limited utilisation of computational methods for counterfeit detection in the art market. By gaining insight into these reasons, our objective is to propose innovative solutions that can be effectively applied to prevent counterfeiting in the art market. Through this study, we aim to bridge the gap between computerised methods and counterfeit detection, ultimately ensuring the integrity of the art market.

Furthermore, while the methods currently employed to authenticate art objects, whether independently or in combination, act as barriers impeding the circulation of counterfeit works in the art market, the continuous and evolving nature of counterfeit production calls for novel solutions that surpass the limitations of traditional authentication methods. Accurately assessing the real risk of encountering a counterfeit work of art should be at the forefront of such solutions.

From this perspective, the analytical phase of the thesis begins with the following questions: Is it possible to use statistical methods to create an effective tool for the early assessment of the probability of fake and forgery? If the answer is in the affirmative, what would be the methodology for constructing such a tool? How can this tool be operationalised and implemented?

The methodology employed in this study is grounded in an original premise that art counterfeits can be discerned from original works through their inherent standard characteristics, including subject matter, size, date, and location of the signature. To substantiate this thesis, this study presents a comparative analysis of these characteristics between original artworks and counterfeits, supplemented by experiments and case studies.

The ultimate goal is to create an alert tool that can effectively assess the risk of counterfeiting in the art market while remaining user-friendly for individuals with

varying levels of expertise. Two strategies have been explored to address this challenge. The first is to use features that are readily accessible to all users, such as those found in textual documents. The second approach involves employing algorithms that generate easily interpretable results, thus ensuring accessibility for all users. Therefore, we opted for a data-mining model. Using probabilistic indicators, we aim to determine whether a work of art exhibits suspicious characteristics.

Statistical analyses were carried out using a database built from the extensive records found in the catalogue raisonné, together with archival documents relating to forgeries[1] produced by various individuals, generously provided by the Foundation of the Swiss and French painter Félix Vallotton (1865–1925).

The reader will see that the methodology presented in this study can be used as a basis for developing new comparative models for other artists. However, the research findings may not be universally applicable. Therefore, appropriate modifications and adaptations are required to ensure their relevance and effectiveness in different contexts.

This manuscript presents a doctoral study, its findings, and the insights it provides in the following structure. The first part 'The state of the art' describes the research field and various facets of the art world in which originals and fakes coexist. Chapter 1 focuses on providing introductory remarks and an overview of the terms and concepts used in this study. Chapter 2 discusses topics related to the art market's organisation and its actors' functions. It also explains how the artistic legacy of the artist Félix Vallotton was protected against attempts at fraud. Chapter 3 presents state-of-the-art methods gathered under the concept of *art authentication*. This review discusses the theoretical issues concerning these methods and cases related to their practical applications.

The second part 'Data and Methods' includes tree chapters. Chapter 4 defines the objectives, hypotheses, and working questions. It also provides an overview of the main

[1] Counterfeit data often contain sensitive information related to the identities of the owners. However, with a carefully devised methodology, it was possible to make this knowledge publicly accessible without compromising confidentiality.

statistical techniques used in the analysis of our data: principal component analysis and classification trees. Chapter 5 investigates the general issues related to the data sources used for this research, explains how we obtained access to the archive on forgery, outlines the sources used, and defines the sampling and extraction phases of the research. Chapter 6 provides a comprehensive understanding of each technique and describes experiments conducted using a specially designed dataset.

The third part 'Findings' consists of three chapters. Chapters 7 and 8 describe the statistical analysis results, outline the performance of the models, and evaluate their validity. Chapter 9 explains the issues of variable interpretation and the explanatory scheme with guidelines for applying the model in practice and finally tests the model with two case studies.

The fourth part corporates discussions and conclusions that stem from the findings of the PhD research, and constitutes the final part of the thesis. Chapter 10 outlines introductory remarks. Chapter 11 discusses the differences and affinities of our model compared with other methods and methodological issues, outlines the limitations and particularities of the database, and proposes recommendations for their refinement in future research. Chapter 12 explains the nature of the data mining model as a tool for the early detection of counterfeits. It presents a series of formal recommendations regarding appropriate strategies for developing alternative models for other artists and suggests ways to implement the model from the perspective of current art market practices.

PART 1 – THE STATE OF THE ART

1 INSIGHTS AND DEFINITIONS: REMARKS TO PART 1 AND KEY TERMINOLOGY IN THE FIELD OF ART CRIME

Counterfeits pose significant challenges, leading to detrimental effects on the markets for specific artists and art schools. Even lesser-known artists whose artistic heritage may be well protected are not exempt from being targeted by counterfeiters. Conducting a literature review on the structure of the art market and the roles played by its participants will provide valuable insights into how counterfeit work can infiltrate the art market. Given that our doctoral research focuses on the Swiss and French artist Félix Vallotton, we address issues regarding the preservation of his artistic legacy.

In the initial phase of this study, we explore the concept of art authentication. This crucial analysis serves as a solid foundation, enabling readers to gain a deeper understanding of the models that will be developed presently. This model, which is rooted in the fundamental principles of art authentication and the controversies surrounding the identification of forgeries, holds significant implications.

Before delving into topics concerning the organisation of the art market and the methods of art authentification, it is necessary to establish clear definitions of frequently used terms within the context of this doctoral work. The field of art crime is rife with terminology that often carries ambiguous or conflicting meanings. In 1994, John E. Conklin defined art crimes as *'[…] criminally punishable acts that involve works of art*' (Conklin, 1994, p. 3). While no universally accepted definition of art crime exists in the literature, Conklin's formulation has gained widespread recognition among researchers (for example, Tijhuis, 2009; Bazley, 2010; Durney & Proulx, 2011; Balcells, 2013). This doctoral study aligns with Conklin's definition, using the term 'art crime' to mean 'the field of crime that includes punishable criminal acts involving works of art'. By clarifying these foundational concepts, we can effectively navigate the complex realm of art crime in our research.

More concretely, art crime includes a wide spectrum of acts, such as theft, the faking and forging of works of art, the looting of antiquities, art fraud, vandalism, illicit

excavations, and export of antiquities or archaeological resources. This list is approximate and could be extended. In addition, art crimes can be classified according to the object concerned, where there are two main groups of objects: antiques and fine arts (Dobovsek, Charney, & Vucko, 2009). Given these classifications, we note that the scope of our research extends only to the faking or forging of fine art (i.e. paintings).

The terms 'fake' and 'forgery' have different meanings. Notably, a 'fake' is a work that is honestly created but is altered for fraudulent purposes to increase its value (Craddock, 2009; Charney, 2015; Rapley, 2015). One example is a painting in which a fake signature is added. A forgery is a new work of art created from scratch with the intention of misleading others (Craddock, 2009; Charney, 2015; Rapley, 2015). These terms clearly refer to two schematic approaches to falsification. However, in a more general sense, both terms refer to falsified artwork. Thus, to avoid any misinterpretation, particularly in our experimental work, we use the terms 'fake' and 'forgery' to refer to falsified works of art[2].

Alternatively, the term 'forgery' may refer not only to a falsified object but also to the *act of manufacturing* a falsified art object. For example, the *Encyclopedia of White-Collar and Corporate Crime*[3] specifies that *'[f]orgery is the fake making or altering of handwritten or electronically produced documents, artwork, or cultural artefacts with the intent to deceive or defraud*' (Fenoff, 2013, p. 360). In contrast, Rapley (2015) defines a forgery as *'[…] an object made in fraudulent imitation of an existing item […]*' (Rapley, 2015, p. 33).

There is some confusion between the terms 'forgery' and 'art fraud'. Notably, Conklin alleged that '*[t]he mere production of a work that resembles an existing one is not a crime, but intentionally and deceptively passing it of as someone else's work is forgery, a type of fraud*' (Conklin, 1994, p. 48). Therefore, Conklin uses the word 'forgery' to refer to a type of fraud. In contrast, the Encyclopedia states that *'[a]rt fraud is a unique*

[2] In most cases, we use the term 'fake(s)/forgery(ies)', except in quoted sentences.
[3] Henceforth, we use the term 'the Encyclopedia' to refer to the *Encyclopedia of White-Collar and Corporate Crime* (Salinger, 2013).

type of forgery […][4] (Theodorakis, 2013, p. 54). Theodorakis (2013) thus defines art fraud as a form of forgery. According to Bazley (2010), the category 'art fraud' encompasses the notion of forgery[5]. Conklin (1994) stressed that art fraud, which comprises *'[…] the production and sale of counterfeit art, is made possible by the trust essential to art world transactions, a trust that is sometimes abused by collectors, dealers, action houses, and museums* (Conklin, 1994, p. 87). Furthermore, he distinguished between different forms of art fraud according to the offenders involved. For example, he proposed a classification of art fraud by collectors (e.g. insurance fraud, tax fraud, and corporate collector fraud), dealers (e.g. fraud against artists and other dealers), auction houses (e.g. erroneous appraisals and sales of stolen or counterfeit works), and museums (e.g. buying stolen items and abuse by collectors). The confusion between the terms 'art fraud' and 'forgery' is not surprising, because the term 'forgery' has no legal status. For this reason, when art forgery is prosecuted in the United States, fraud is often invoked as the charge because United States law does not specify a crime of 'forgery' (Bazley, 2010). Under Swiss law, the falsification of artwork is prohibited by Article 146 (1), paragraph 1 of the Swiss Criminal Code (1937), which stipulates that '*any person who with a view to securing an unlawful gain for himself or another wilfully induces an erroneous belief in another person by false pretences or concealment of the truth, or wilfully reinforces an erroneous belief, and thus causes that person to act to the prejudice of his or another's financial interests, is liable to a custodial sentence not exceeding five years or to a monetary penalty*'. In addition, Article 28 (F) of the Swiss Code of Obligations (1911) defends a party who is the victim of fraud: 'a party induced to enter into a contract by the fraud of the other party is not bound by it even if his error is not fundamental'[6]. The Uniform Crime Report (UCR)[7] defines forgery as *'[t]he altering, copying, or imitation of something, without authority or right, with the intent to deceive or defraud by passing the copy or*

[4] '*Art fraud is a unique type of forgery, usually defined as the creation and attempted sale of an object falsely purporting to have the history of production requisite for the origin of the work*' (Theodorakis, 2013, p. 54).

[5] It seems that there is a long history of difficulty in interpreting this term. In particular, Polk and Chappell (2009) presented the example of a trial that occurred in 1897, during which the question of the application of the term 'forgery' to works of art was raised in a tribunal (previously, this term only applied to the forging of documents or writings).

[6] This is a free translation from the Swiss government portal of Article 28 (F), paragraph 1 of the Swiss Code of Obligations: La partie induite à contracter par le dol de l'autre n'est pas obligée, même si son erreur n'est pas essentielle.

[7] This resource is a U.S. compendium of crime reported to the police.

thing altered or imitated as that which is original or genuine; or the selling, buying, or possession of an altered, copied, or imitated thing with the intent to deceive or defraud' (Bazley, 2010, p. 64)

However, confusion of terms exists not only in the literature but also in practice. For example, the Scotland Yard police usually record art fraud with the qualification 'fraud' without indicating that the fraud relates to artwork (Rapley, 2016). Consequently, the terms' fake', 'forgery', 'copy', and 'misattribution' are rarely used in investigations of art fraud (Rapley, 2016).

Another confusion is related to the illicit nature of the art object. In principle, the art object is not criminal, but the intentions and actions of its author or executor may be criminal (Polk & Chappell, 2009). For example, copying a pre-existing work does not constitute a criminal act. An artist can make an imitation without intending to deceive the public. Such an object may then be used for malicious purposes by others, for example, by replacing their authentic signature with that of more famous artists whose style resembles their own, or falsifying authentication certificates. In such cases, the copy is considered a forgery[8].

Given the confusion and ambiguities mentioned above, we have opted to simplify their use in the text of this thesis. Because the terms 'art fraud', 'art forgery', and 'art counterfeiting' refer broadly to the process of making a counterfeit work with the intention to deceive or defraud, we will use them as synonyms. Likewise, we will not distinguish between the terms 'forgery' and 'fake' by using them as synonyms or by applying them together as 'fake/forgery'. The term 'dubious/problematic work' refers to an artwork with a significant level of uncertainty based upon a comparison with the original works. This classification indicates a significant level of uncertainty regarding the authenticity of the artwork, suggesting that it shares resemblances or characteristics with other confirmed forgeries. In our research, this term implies a forgery that is technically and stylistically flawless but, for some reason, part of the evidence to confirm attribution to an artist is missing. Because they bear the apparent

[8] The term 'misattribution' signifies an error in authentication. Similarly to the issue of a copy, such a mistake is not considered an art fraud, provided the error was committed without dishonest intent (Rapley, 2016).

designation of a concrete artist (such as a signature), they are considered doubtful forgeries.

2 THE ART MARKET FOR ORIGINAL AND FAKE WORKS

2.1 Navigating the art market: structure and principal actors

Art objects have long been traded on the market. Nevertheless, the current organisation of the art market has not changed much over the years (Gérard-Varet, 1995). In the late twentieth century, the art market became an important sector of economic activity. The art market exists within the general economic concept of a predictive market. However, this type of market is different from the *supply and demand* approach of the market for commodities: art is bought and sold not only on the basis of the perceived cultural value of the work but also on the market value in the past and its projected value in the future (Gérard-Varet, 1995). These two market types represent a hierarchy. A primary art market is characterised by new art objects that come to the market for the first time, a secondary art market for objects that have been sold at least once in the primary market (Gérard-Varet, 1995). The particularity of the primary market stems from the absence of a history of sales that would allow for the analysis of art prices. Therefore, the '*supply and demand'* rule affects mainly the secondary market (Plattner, 1998).

Private and public collectors are central buyers of art. For some collectors, the search or negotiation for an artwork is part of the pleasure of collecting. They buy through all types of intermediaries and regularly resell (Moureau, Sagot-Duvauroux, & Vidal, 2016). Others are loyal to a few galleries with which they have established relationships of trust and rarely resell (Moulin, 2003). Galleries have an important place in the art market, as they deal with the primary market (Moureau & Sagot-Duvauroux, 2016). Moreover, galleries invest in and promote the work of specially chosen artists. This promotion serves as an indicator of buyer quality. There are various approaches to selling artworks through galleries. One of the most famous systems is 'the Castelli model', which relies on the interaction between an artist and dealer (Russell, 1999; Molin, 2003; Winkleman, 2015; Moureau & Sagot-Duvauroux, 2016)[9]. Trust between artists and dealers plays a central role in the primary market,

[9] Leo Castelli, the New York art dealer, was one of the most influential American art dealers of the late 1950s. He played a major role in shaping contemporary American art and in promoting the international acceptance of painters. His managing method was not absolutely new; he rather optimized the best experience in art dealing from others (Winkleman, 2015) This model's

whereas for the secondary market, it is the trust between the dealer and the collector that is at stake. A gallery's reputation is an influential element for art buyers, who can even provide a kind of authenticity presumption. Indeed, during the trial of the 'Knoedler & Co' fraud case, most of the victims claimed that they believed that the Knoedler Gallery could only sell authentic works (Adam, 2018). On the one hand, a gallery's reputation inspires confidence. However, a tendency in the art world that is often unknown to the public can undermine this confidence. Noa Charney[10] said in an interview with Georgina Adam (Adam, 2018, p. 125) that *'[t]here is strange psychology in the art world. Art dealers tend to accept work even if they have some suspicions. It is easy to claim that one has made a mistake; it is complicated to prove the intention to defraud*'[11]. Furthermore, when a problem arises, collectors prefer to pursue alternative solutions or civil lawsuits to reporting the matter to law enforcement (Pryor, 2016). This may be related to the close relationship between the dealer and collector, or to their fear that the police may seize the object during an investigation and that judges may order their destruction in some jurisdictions[12]. Accordingly, the secondary market is at the forefront of the infiltration of fakes.

Auction houses are the most important intermediaries in the secondary art market. Sotheby's and Christie's are the dominant art players[13]. Many smaller firms occupy specific niches and set their prices under the shadow of the dominant art players. Over the past decade, high art prices have made the market more attractive to forgers. Even art market players, such as major auction houses, have decided to become proactive by developing additional measures to prevent forgeries from entering their sphere of

particularities are consignment rather than buying and reselling, paying artists a percentage of sales, sustained promotion and regular exhibitions, managing artist's career, releasing the artists from administrative tasks, and having more time for the creation process (Winkleman, 2015).

[10] Founder & Trustee of ARCA, the Association for Research into Crimes against Art.

[11] This quote is a free translation: '*Il y a une étrange psychologie dans le monde de l'art. Les marchands d'art ont tendance à accepter une oeuvre même s'ils ont quelques soupçons. Il est facile de prétendre que l'on s'est trompé; il est très difficile de prouver l'intention de fraud*er' (Adam, 2018, p. 125).

[12] For example, the seized forgeries in the affairs of Joan Miró and Marc Chagall (Valentin, 2013; BBC News, 2014).

[13] Sotheby's was founded in 1744 by Samuel Baker. In 1977, the company was listed on the stock exchange and was bought by Alfred Taubman, who applied modern management methods, using marketing, and acting as a banker to sellers and buyers. Christie's was founded in 1766 by James Christie. In 1998, the firm was bought by businessman François Pinault, a renowned collector of contemporary art (Moureau, & Sagot-Duvauroux, 2016).

activity. For instance, Sotheby's established a scientific research department to offer clients an unparalleled level of provenance and market expertise (Sotheby's, 2017). Furthermore, Christie's, after a scandal involving a fake provenance, claimed rules on provenance evidence (Harris, 2020). Nevertheless, such preventive measures are rarely applied by small-market players such as galleries or auction houses. They do not systematically take significant precautions to check their clients' backgrounds or call in expert assistance (Pryor, 2016), mainly because of the limited funds available for expertise and because they rely on trusting relationships with their clients. Moreover, when auction houses refer to the authenticity guarantee of outside experts, their identities are not systematically stated in the catalogue of sale. According to Bandle (2015), this is related to the fact that the expert may be exempt from legal action when the auction house decides to keep his name anonymous. Auction houses often include in terms of sale an explicit clause stating that the established attributions of the consigned property are merely an expression of their opinion. Action houses seek to extend the scope of their guarantees of authenticity and limit liability risks (Bandle, 2015).

In the late 1990s, online auctions grew, and specialised sites multiplied, including those devoted to art sales. Web sales platforms have become art players. It is interesting to note that at the beginning of their integration into the art market, they tried to ensure quality art goods and had a label recognised in the art world through acquisition or partnership with traditional auction houses[14]. Unfortunately, the sustainability of these efforts has been short. Moulin (2003) pointed out that the proliferation of sales on the Internet mostly concerns medium- and low-quality objects. Trade professionals insist on the lack of serious selection of the objects offered and the insufficient guarantee of their authenticity and financial assessment. Moreover, Moulin (2003) argues that the participation of multinational actors and the transaction's location may complicate or even prevent the application of international and national regulations. These legal loopholes open the possibility of fraud. Loll (2016, p. 69) pointed out that '*this segment is especially vulnerable to organised fraud.*

[14] For instance, in 1999 Ebay.com, the emerging online private sales market leader, acquired Butterfield and Butterfield, the third-largest American auction house (Moulin, 2003).

Forgery factories operate openly and freely and buyers have little recourse when they have been victimised'. The complexity of the art market's current structural and systemic situation requires greater transparency, better regulation, innovative technologies, and cooperation among art market players.

2.2 Vallotton's legacy in the art market: a look at originals and forgeries

Although a wide range of artists are affected by forgery, some names appear more often than others. The *Art News* published a list of the ten most faked artists: Giorgio de Chirico, Jean-Baptiste-Camille Corot, Salvador Dalí, Honoré Daumier, Vincent van Gogh, Kazimir Malevich, Amedeo Modigliani, Frederic Remington, Auguste Rodin, and Maurice Utrillo (Esterow, 2005). Likewise, *MutualArt* (2018) proposed ratings for the top six faked artists: Pablo Picasso, Jackson Pollock, Amedeo Modigliani, Andy Warhol, Vincent van Gogh, and Pierre-Auguste Renoir. According to the *Times*, Marc Chagall is one of the most frequently forged painters of the last 100 years (Moody, 2018). These examples show that the most common denominator of the referenced artists was worldwide fame. Nevertheless, other artists who are less famous to the large public but appreciated by museums and private collectors do not escape forgery. The Swiss-French painter Félix Vallotton (1865–1925) is an example of such an interest.

Félix Vallotton was an important artist of his generation. Born in Lausanne in 1865 and dying in France in 1925, Vallotton created a legacy that belongs to both countries. A portraitist in his early years, after 1890, he became involved in wood engraving. His revival of this traditional technique quickly earned him an international reputation as an artist at the cutting edge of modernity. He became friends with Vuillard, Bonnard, and Denis, joined the Nabis group and returned to his earlier vocation of painting (Félix Vallotton Foundation, n.d.). His place in the modern movement was confirmed by his participation in the Salon des Indépendants, Salon d'Automne, and Secession in Munich, Vienna, and Berlin, and in the Armory Show in New York (Cogeval, Cahn, Ducrey, & Poletti, 2013). Currently, his paintings can be seen in different museums worldwide, such as the Museé d'Orsay in Paris, the Musée Cantonal des Beaux-Arts in Lausanne, the Kunsthaus in Zurich, and the Hermitage Museum in Saint Petersburg.

A grand retrospective of Félix Vallotton was organised at the Musée d'Art Moderne in Paris in 1966, then in 1979 at the Petit Palais, and in 1997 at the Musée Maillol around his nudes. In 2013, a new retrospective study at Grand Palais in Paris, followed by a reduced exhibition at the Van Gogh Museum in Amsterdam and then at the Mitsubishi Ichigokan Museum in Tokyo, showed a growing public interest in his works. The recent exhibitions in 2019 at the Royal Academy of Arts in London and the Metropolitan Museum of Art in New York affirm that even though Félix Vallotton is well known among art lovers, especially in Switzerland and France, his works are gaining new audiences with great success in other countries; for example, one million visitors attended the retrospective event mentioned above (Cogeval et al., 2013). His work can be found in galleries, such as Ackerman's Fine Art in New York, the Sylvan Cole Gallery in Spain, and the Bailly Gallery in Switzerland. Although Vallotton is less well-known to the general public than Picasso or Monet, his art is highly appreciated by collectors worldwide. His art sells very well on the market, especially his works of the Nabi period. The record price of CHF 2.4 million held since 2000 for the painting *Sur la plage* was surpassed in 2013 when his *Au marché* sold for CHF 3.5 million.

CR No. 269: At the beach (*Sur la plage*), 1899.

CR No. 196: At the market (*Au marché*), 1895.

In the same year, at Sotheby's Zurich, a *Bord de Seine à Tournedos, effet gris* approached the CHF 1 million. Another painting, *Neva, léger brouillard*, made in Russia in 1913, was sold at Sotheby's in 2017. Estimated at between 1 and 1.5 million Swiss francs, its value has risen to 1.33 million.

CR: No. 1428 Edge of the Seine at Tournedos, grey effect (*Bord de Seine à Tournedos, effet gris*), 1921.

CR: No. 967 The Neva, light mist (*Neva, léger brouillard*), 1913

These examples illustrate the highest prices for the most desirable period (the Nabi) and topic (landscape). However, the average price range for less-favoured works may vary considerably. For example, one of Vallotton's paintings sold for CHF 16,520[15], another ten times more (CHF 162,500)[16] (Artnet Price Database, n.d.).

The artistic legacy of Vallotton is well preserved. Indeed, Félix Vallotton maintained accurate records of his artworks from his youth. These notes have become essential authentication tools for expertise and have notably contributed to the creation of the catalogue raisonné by the Félix Vallotton Foundation in 2005. It enjoys an excellent reputation in the art world and provides valuable reassurance to art stakeholders. The Foundation is the centre for documentation and research on the artist's life. It also oversees the enrichment of its archives, conducts expert appraisals, and supports exhibitions and publications. Similar to other foundations, the Vallotton Foundation is a private institution with no obligation to detect or report fakes. In most cases, it is an art player who asks the Foundation for its expertise. In exceptional cases, for

[15] According to the price database of the Artnet: Félix Vallotton Title 'Femme nue couchée sur le sable', Medium Oil on Canvas laid on Cardboard Year of Work 1918 size height 4.9 in.; width 8.5 in. / height 12.5 cm.; width 21.5 cm. Misc. Stamped Sale of Sotheby's Zürich: Monday, May 29, 2000 [Lot 00117] Schweizer Kunst Estimate 15,000–20,000 CHF (8,885–11,847 USD) Sold For 16,520 CHF Premium (9,786 USD)

[16] According to the price database of the Artnet: Félix Vallotton Title 'Groseilles et capucines' medium, oil on canvas, year of work 1919, size height 21.5 in.; width 25.8 in. / height 54.5 cm.; Width 65.5 cm. Misc. Signed Sale of Sotheby's Zürich: Tuesday, December 1, 2015 [Lot 00040] Schweizer Kunst / Swiss Art Estimate 150,000–200,000 CHF (145,687–194,250 USD) Sold For 162,500 CHF Premium (157,828 USD).

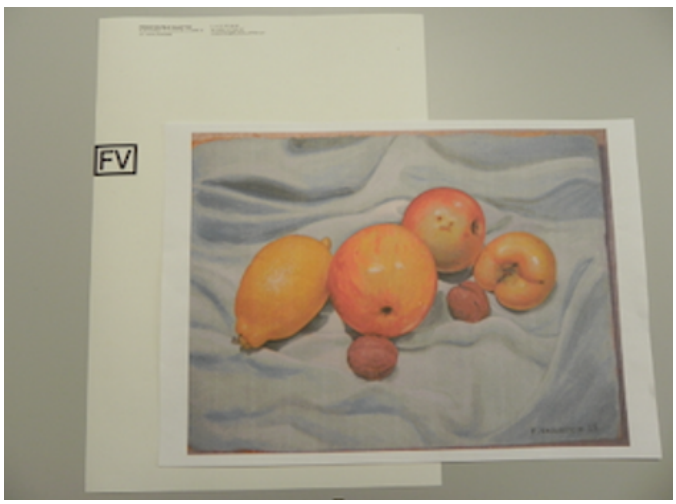
example, when a forgery appears at a trusted art dealer (such as an auction house or renowned gallery), experts may take the initiative to give their opinions.

The images below are examples of artworks that experts have classified as forgeries. Numbers 1 and 2 are paintings signed with an imitation of Vallotton's signature 'Vallotton' and the monogram 'FV', which are 'poor' forgeries in terms of stylistic affinity compared to Vallotton's works. Hypothetically, these paintings could have been created without any intention of cheating, but the signature imitating Vallotton's was added later.

1. Artwork without title with signature 'Vallotton' (fake)

2. Artwork without title with monogram 'FV' (fake)

The other category includes works of art with doubtful defects. Such works are the most 'successful' forgery attempt. For example, Number 3 is classified as a dubious forgery. This talented painting is stylistically very close to Vallotton's aesthetic, but the composition of the fruit has left experts perplexed because it includes a persimmon. While Vallotton painted many still lifes with fruits that are repeated throughout his works, it seems strange and suspicious that he chose to paint the persimmon, which had never been used in a composition before. This is compounded by the lack of provenance. Therefore, the painting was classified as doubtful. Nevertheless, in the future, the historical documentation of its owners will be established and a plausible explanation for the presence of the persimmon will be found, which might radically change its classification.



3. Still life with persimmons (doubtful work)

Doubtful pictures correspond to artworks that contain few (or even no) fundamental errors. They could not be identified as 'original' because, for example, their provenance is suspicious or unknown. Such works comprise most of the 'successful' falsification attempts.

The next example (No. 4) is a landscape without Vallotton's signature, which is classified as a doubtful work.

4. Landscape (doubtful work)

In 1978, an auction house approached the Foundation's experts for their opinions on this unsigned landscape, which they believed to be by Vallotton. For the Vallotton experts, there was insufficient conclusive evidence to attribute this to Vallotton. Consequently, the painting was noted in their records as 'doubtful'.

There is also a category of artwork without a signature or monogram which were hypothetically assigned or presumed by the owners to be Vallotton's paintings but were not ultimately authenticated as Vallotton's paintings. These types of artwork were wrongly attributed to Vallotton in the past. This attribution could be honest or malicious. Therefore, experts have grouped these works of art into a separate category of forgeries. For example, the painting below (No. 5) was previously attributed to Vallotton.

The artwork (No. 6) appeared in 1962, then in 1977 at Christie's, in 1987 at Sotheby's, in 1997 at the London Gallery, and in 2007 again at Christie's (in Switzerland, France and the UK).

5. Landscape (hypothetical work)

6. Artwork without title with signature 'Vallotton' ('hot potatoes')

Such an astute strategy is logical: ten years is long enough for the specialists who appraised this painting before to have left the auction houses. One might assume that vendors would change the country every ten years to blur the traces of previous attempts to treat the painting[17]. However, the auction houses were vigilant, each time asking the Foundation's experts to confirm its authenticity. The signature was an imitation of Vallotton's. What is also interesting in this case is that experts did not exclude the possibility that this picture was painted by Jean Edouard Vuillard or another artist in the Nabi group. Perhaps this painting was not initially signed by the

[17] Such a strategy is called '*hot potatoes'* (see Appendix I.)

artist and was subsequently wrongly attributed to Vallotton. It is also possible that a false signature was added to overcome the difficulty of its (lack of) provenance.

As the Foundation is the exclusive institution that provides a certificate of authenticity for Vallotton's artwork, it is problematic to trade the artwork if the Foundation refuses to issue a certificate. Nevertheless, the risk that they may be sold on an Internet platform is high. The Vallotton Foundation has meticulously preserved information on forgeries. Their archives contain information on several hundred fake paintings, drawings, and prints. Because we obtained permission to use these documents (letters, photos, and correspondence), we will exploit them in the case study of the experimental part of this thesis.

2.3 Summary of Chapter 2

In a broad sense, this section delineates the central antagonism dominating the art market. Confidentiality has various repercussions on the art world. Institutions that can authenticate art objects tend to withhold information on the owners of fakes and prohibit entry into their databases, thus undermining transparency. The confidentiality and trust between art players (buyers and sellers) do not relieve this situation, but reduces the due diligence of both parties. There is mutual interest in not breaking this type of relationship; buyers fear losing a good deal, and sellers risk damaging their reputation. Legal liability obliging art actors to exercise due diligence could be a potential measure to make the entry of fraudulent objects into the art market less vulnerable.

The art market is a 'free zone' compared to other markets. Some art market players have developed internal guidelines to minimise the risks in their business activities. Nevertheless, implementing such programmes is difficult because not all competitors conduct their activities in accordance with similar standards. The need for legislative reform has become increasingly imminent. Such measures would involve lengthy and challenging legislative changes. Furthermore, transparency and control can make art-trading activities less lucrative.

In summary, the unconditional trust and, at the same time, the antagonism between the players, the dichotomy between art market practices, and court judges on the issue of authentication of artwork make this sector vulnerable to counterfeiting. In this context, Bandle (2015) proposed promoting alternative researchers and facilitating access to expertise and knowledge to improve authentication quality. At the same time, facilitating access can improve the detection of counterfeits and prevent the victimisation of art market participants. Which current methods may provide such support?

The following chapter outlines competing approaches to authenticating originals and detecting forgeries. Theoretically, these methods are complementary and can operate simultaneously. However, in practice, their application is complex. We will inquire into how the roles of these methods function depending upon the examiner's objectives and technical means.

3 ART AUTHENTICATION

3.1 The role of authenticity in shaping cultural value

The term 'authenticity' may be applied in different contexts; it can refer to one's own personality, spirit, or character (Merriam-Webster, n.d.). In this section, we will examine 'authenticity' relating to the art object, which, according to Dutton (2003), can be distinguished under two main definitions.

First, 'authenticity' is simply defined as the correct identification of an object's authorship or provenance (Dutton, 2003). In this way, authenticity can be directly (or implicitly) associated with the value of artwork (Charney, 2015). In fact, the price of a work of art fluctuates considerably depending on the reliability of its authorship. For example, the painting 'Salvator Mundi' was attributed to Leonardo da Vinci. In 1900, this painting appeared in the hands of an English collector. In 1958, it was catalogued as a copy of a work by Boltraffio, one of Leonardo's most gifted students and sold for 45 pounds. Almost 50 years later, the painting reappeared at a regional auction in the United States. It was sold in May 2005 for USD 1,175 as a copy of Leonardo da Vinci's disciple. Its new owners decided to move forward with the cleaning and restoration of the painting, extensive research and documentation, and careful examination of its authenticity. After many expert appraisals, the painting was finally attributed to Leonardo da Vinci and was presented for the first time in an exhibition dedicated to the artist at The National Gallery in London in 2011. Since then, the value of the small wooden panel has dramatically increased. In 2013, it was sold for USD 80 million; four years later, when the celebrated author Walter Isaacson included a chapter on 'Salvator Mundi' in his biography of Leonardo da Vinci, it was auctioned by Christie's New York for USD 450 million[18]. However, Dutton (2013) stresses that the meaning of authenticity serves the function of understanding the history of art in terms of human value, as well as the fundamental philosophical and spiritual ideas of an epoch, which is more important than maintaining a simple market value of an art object. Consequently, he proposed a second meaning of 'authenticity' that evokes the

[18] Journalists have provided extensive media coverage of this subject, resulting in a number of different retellings of the story. Regardless of any discrepancies in the details, the discussion in this section is based only on research from Ben Lewis's book *The Last Leonardo* and the attribution chronology report from the official website of Christie's, the house that participated in the sale (Christie's, 2017; Lewis, 2019).

character of an object as a true expression of the values and beliefs of an individual or a society. Modern society tends to appreciate the aesthetic value of a work of art more than its functionality (Johnson, 1973). At the same time, art and money are inseparable (Savage, 1963; Flynn, 2016). As a result, for some art purchasers, a work of art's aesthetic qualities always outweigh its economic benefits, whereas other purchasers who are more motivated to invest are mainly interested in the work's economic profile (Flynn, 2016). Nevertheless, we cannot conclude that people in the first category do not care about the economics of their art purchases and that investment-oriented buyers are not indifferent to the aesthetic qualities of the art. Instead, Flynn (2016) argues that the intersection of borders between these categories is the most objective consideration. The value of a work of art, which is expressed as an economic equivalent, is determined not only by its aesthetic qualities but also by the emotions and associations that it inspires (Tietze, 1948). In general, a work of art symbolises a personality, epoch, or culture. Interest in art is often linked to an artist's personality and their beloved and appreciated expressions. Tietze (1948) remarked that a mediocre painting by a genius artist like Rembrandt remains a work of art that belongs to that artist; a forgery, even if it is an excellent work in that artist's name, is always a forgery. He also stated that when a work of art turns out to be fake, emotions based on erroneous assumptions provoke frustration and a loss of interest (Tietze, 1948).

Art counterfeiting creates serious problems and causes real harm. At the same time, the ethical consideration regarding countrifies within the cultural and philosophical traditions of certain countries remain ambiguous. For example, the culture of the Far East is unfamiliar with the concepts of origin and identity in art. Notably, the Chinese idea of the original artwork is determined not by a unique act of creation but by an unending *process*, and not by a definitive identity but by constant *change* (Han, 2017). In contrast, a fundamental Western concept is that of *being*, that is, something that resembles only itself and tolerates no reproduction outside itself. Chinese art has a functional rather than mimetic relationship with nature. Thus, the goal is not to represent nature in the most realistic way possible, but rather to act exactly like nature (Han, 2017). In his book *Shanzai: Deconstruction in Chinese* (2017), Byung-Chul Han, a professor of philosophy and cultural studies at the University of Arts in Berlin, hypothesised that the root of the Far Eastern concept came from the Buddhist

philosophy of the endless cycle of life. He stresses that the Western idea of *being* gives rise to a cyclical process that includes death and decomposition. In contrast, the Buddhist notion of the infinite excludes anything unique, original, singular, or definitive. Copying (and even forgery) has a different meaning in China as well. For a Chinese painter, putting the forgery of an old master into the collection of a well-known connoisseur means gaining recognition for one's skills (Fong, 1962; Pomfret, 1999; Diego, 2012). For example, during an exhibition of Chinese art masterpieces held at the Museum of Asian Art in Paris in 1956, it became clear that the pictures in the exhibition were forgeries. The sensitive issue was that the forger was the most famous Chinese painter of the twentieth century, Zhang Daqian (or Chang Dai-chien), whose works were simultaneously displayed at the Museum of Modern Art. Fong (1962) cites a considerable number of historical names and facts from the time when the greatest Chinese painters and connoisseurs were simultaneously master 'forgers'.

In terms of the legal or moral issues connected to this perception in China, Fong (1962, p. 99) points out that '*the legal or ethical problems of an "honest business transaction" never entered into the picture'.* Ethical and cultural tact imposes protecting the owner of a false art object as much as possible from learning the truth.[19]

Hence, art is a testimony to our shared human experience, embodying our history, values, and diverse cultural perspectives. It represents a rich palette of expression that crosses, bridging our past, present, and future. Each society cultivates its artistic traditions, aesthetics, and interpretations, shaping the perception and evaluation of works of art within its cultural context. Ethical considerations also come into play, prompting reflection on cultural appropriation, provenance, and authenticity.

Within this subtle mosaic, trying to disentangle the value of art as a human heritage from individualistic consumer desires, cultural subtleties, and ethical dilemmas becomes an arduous undertaking.

[19] The aim is to ensure that the individual remains shielded from the knowledge that the object they possess is, in fact, a forgery (Fong, 1962).

3.2 Traditional and modern methods for art authentication

Art authentication uses a set of methods, either separately or in combination, to provide strong evidence that a work of art is genuine. One of their functions is to act as a filter to prevent the infiltration of counterfeits into the art market (Bazley, 2010). A theoretical and historical overview provides an understanding of the stylistic connoisseurship method and provenance research. The role of provenance research in the detection of fakes and forgeries is discussed and illustrated using several cases. Technical and scientific approaches and computer-assisted authentication methods are then reviewed and compared with other methods.

3.2.1 A critical examination of stylistic connoisseurship: strengths and limitations

Stylistic connoisseurship is a visual inspection by a trained eye; that is, '*a method of judgment based on an informed perception and interpretation of the form and the facture specific to the artist who created it'* (O'Connor, 2004, p. 6). O'Connor (2004) found similarities between this approach and the methods used by other scientific professionals. Notably, he compared stylistic connoisseurship with signature analyses. A handwriting analyst examines the formal characteristics of the signature in question relative to an authentic signature to detect the absence of relevant characteristics of an authentic signature (Montani, 2015)[20]. Similar to a handwriting analyst, an art expert *'[…] has seen hundreds, maybe thousands, of works by the artist in question, and has absorbed into visual memory the artist's characteristic form – shapes, compositional devices, linear rhythms, typical colors, and habits of facture – to the extent that such an expert can tell, […], that the work presented is authentic or fake*' (O'Connor, 2004, p. 8).

A knowing eye requires training and practice, sensitivity to quality, and visual reconstruction of the artist's art. Visual memory provides a means to internalise comparisons guided by perception and the ability to categorise objects (Ebitz, 1988). Therefore, these skills are strongly conditioned by an expert's intellectual disposition and individual experience. Indeed, these results depend entirely on subjective judgements and can lead different experts to different conclusions. Any tool intended to assess a specific phenomenon can be considered an indicator of it. In the natural

[20] See Section 3.2.3.1.

sciences, where phenomena are studied directly or indirectly, an indicator usually has a proven validity. In contrast, the social sciences are most often concerned with abstract concepts, and scholars regularly question the validity of their indicators. Consequently, comparisons between different indicators are delicate and the findings must be interpreted with prudence (Aebi, 2006). Similarly, it can be assumed that the 'trained eye', as an instrument for judging authenticity, should be subjected to standardisation measures with great diligence. Technical aids and other scientific means of examination can thus add objectivity to stylistic connoisseurship by increasing the validity and reliability of the method.

Nevertheless, most authors have traditionally considered stylistic connoisseurship to be '*the primary and most programmatic method*' (Neuhaus, 2014, p. 67) or '*to be the gold standard*' (Bazley, 2010, p. 184). O'Connor (2004) pointed out that connoisseurship is the most important method for the issue of authenticity and that this fact is the most difficult for non-experts to understand. He stressed that '*the ability of the connoisseur to perceive the rightness of a work usually precedes the need for the lab or the archive*' (O'Connor, 2004, p. 6). However, some authors do not share this perception. For example, Amore (2015) pointed out that science now provides the important ability to stop even the most intelligent and formidable fraudster and proposed applying science to a rigorous interdisciplinary analysis of historical materials and technology by combining it with in-depth training in the field of fine arts. Friedlander (1944), an art expert and writer, was also not entirely convinced that connoisseurship should be prioritised unconditionally. He argued that '*[s]tyle-criticism inevitably reckons with probabilities, builds up hypotheses. In order to make fruitful use of such sensitive and delicate means, it is necessary to possess imagination and sincerity, a quality which is often unavailing*' (Friedlander, 1944, p. 172). Therefore, he proposed three objective criteria to guide expert verdicts: signatures and monograms, documentary information, and objective data of different forms, referring to the method developed by Giovanni Morelli[21]. Morelli's method was a well-known theory of art attribution in the nineteenth century (i.e. the study of similar forms) that attempted to systematise the visual judgment of expert authentication (Morelli &

[21] Giovanni Morelli (1816–1891) was an Italian doctor of natural science, art critic, and political figure.

Anderson, 1994). Morelli's idea was that artists who work in strict stylistic and classical traditions differ from each other in the ways in which they paint minor details (e.g. nails, fingers, or ears), which, in his opinion, expresses the originality of the artist's personal observations (Morelli & Anderson, 1994). At that time, art experts did not unanimously accept this method. For example, Friedlander (1944) criticised the principles of Morelli's method, arguing that attribution is more a matter of intuition and corresponds not to careful observation of details but rather to an overall vision (Friedlander, 1944). However, this opinion is itself open to criticism. For instance, according to Kahneman (2012), experts are not unconditionally confident in their intuitions. He concluded that experts' intuitions are justified as long as they are formed in a context in which situations can be reproduced under similar circumstances and in which the individual receives immediate and clear feedback (Kahneman, 2012).

Approximately half a century later, the Dutch scholar Maurits Michel van Dantzig further explored Morelli's method. Van Dantzig compared the differences between the works of art of the great masters and their fakes. His examination concluded that '*all pictures contain so many basically unconscious personal characteristics that each work betrays the identity of its maker more or less in the same way that handwriting does that of the writer*' (Van Dantzig, 1973, p. 20). On the basis of these findings, he developed a new method named 'Pictology' based on comparing specific visual features typical of an artist with the features apparent in a suspect painting. He argued that an original is created by spontaneous strokes, whereas a forgery is recognisable by the effort with which this stroke signature is mimicked (Van Dantzig, 1973). According to his discovery, an original work of art has a high level of spontaneity despite its difficulty and a forgery presents many inhibitions. He was convinced that it was possible to describe and analyse the quality of both spontaneity and inhibition. This method allows for the listing of constant personal characteristics. Each is scored as (+) or (−) in the examination, depending on whether it is typical of the artist. The artwork can then be numerically evaluated using the percentage of each score obtained (Craddock, 2009). For example, the list of characteristics of Vincent Van Gogh comprises several hundred features, like 'tragic atmosphere', 'brushstroke from left to right', 'natural growth', 'signature at the bottom', or 'darkest shadows: dark grey, blanc' (Van Dantzig, 1973). Lee (1981, p. 144) qualified the Van Dantzig approach as '*[…] the first serious attempt to systematise and quantify the process of artistic*

evaluation for attribution purposes' (Lee, 1981, p. 144). In 2017, scientists at Rutgers University inspired by 'Pictology' developed methods to quantify strokes. They attempted to rationalise Van Dantzig's method using a machine to facilitate the statistical analysis of many features (Elgammal, Kang, & Leeuw, 2018).

3.2.2 Tracing authenticity: historical documentation analysis as a tool for art authentication

Another authentication approach involves establishing the history of a work of art from its creation to the present day. In the literature, this approach is known as the analysis of the documentation of art history (or provenance; Clark, 2004; Craddock, 2009; Bazley, 2010; Bandle, 2015).

Provenance is a type of reconstruction of an artwork's history, which includes '*[…] oral and written evidence, with the letter consisting of such documentary items as bills of sale, recorded testimony, gallery/museum catalogues, and catalogues raisonnés […]*' (Bazley, 2010, p. 185). The dictionary definition of the term 'provenance' is '*the history of the ownership of a work of art or an antique, used as a guide to authenticity or quality; a documented record of this*' (Oxford English Dictionary, n.d.). This description is essentially valid for works of art, whereas for other objects such as books, wine, and archaeological objects, the term 'provenance' has different meanings.

Provenance may be interpreted more broadly than just the facts of ownership and transfer; instead, it may '*explore ideas and narratives about origins and itineraries of objects, consider the historical uses of provenance information and draw attention to the transformative power of ownership*' (Feigenbaum & Reist, 2013, p. 1). The term 'provenance' comes from the French word *provenir*, meaning 'to originate', and dates to 1785 in its first known use (Feigenbaum & Reist, 2013). At that time, interest in provenance began to grow, owing to the rapid development of the art market. In particular, in France, the concept of provenance in the 18th century was closely related to the refinement of the collection and associated reflections on the catalogue (Raux, 2013). Between the 1730s and the 1780s, the sale of paintings increased more than sixty times compared to previous periods, and consequently, the need to guarantee the authenticity of works of art entering the art market increased (Raux, 2013). Hence,

provenance has become a way to indicate the pedigree of a work of art[22] (Raux, 2013). In other words, the provenance of a work of art is considered a chain that goes back to the moment of its creation and contributes to the establishment of its authenticity (Feigenbaum, 2013).

3.2.2.1 The role of provenance gaps in authentication

Ideally, the provenance of a work of art should include a chronology of all its owners, from the time of its creation by the artist until the most recent owner, the methods of transference or sale, and its locations until the present day (Bamberger, 2017). Yeide, Walsh, and Akinsha (2001, p. 9) claimed that the provenance of an object only traces its ownership and location and can help establish legal evidence. Additionally, Shindell (2016) specified that the legal title of property is not the same as physical possession, saying '*provenance is based on physical possession (location), [..], which is not the same as ownership or legal title, though information on the prior location and possession of an artwork can help to shed light on legal title*' (Shindell, 2016, p. 408).

Provenance information can be found in different sources such as archival collections, photographic archives, transcriptions of oral histories, and a variety of media (Yeide et al., 2001). It is customary to distinguish between primary sources, such as receipts and bills of sale that document transfers of ownership, and secondary sources, such as scrap catalogues and artist monographs. The logic for obtaining this information is as follows: First, one should search for the names of the owners and the places and dates of ownership. Then, one must examine the terms and conditions of transfers between owners. Once all relevant data have been collected, gaps in known ownership become evident. We can clearly observe such gaps even in the catalogue raisonné. For example, the catalogue raisonné of Félix Vallotton includes the provenance of painting No. 18, 'Portrait of Mr. Duponchelle' (Ducrey & Poletti, 2005, p. 14, vol. 2):

> 18. Portrait of Mr. Duponchelle, 1885
> Current location unknown

[22] The term 'pedigree' is used in the sense of provenance. For example, Pergam (2013) refers to this term to explain the position of art dealers on the close conceptual link between the provenance of a portrait and its authenticity.

> Provenance: Collection Duponchelle, Paris (commissioned from the artist in 1885)[23].

We can see that the data contain information only about the first owner of this painting, the Duponchelle Collection. The data also note that its initial location was in Paris and that the portrait was ordered by the owner from the artist in 1885. Notably, when this catalogue was published, the painting was no longer included in the collection.

Yeide et al. (2001) stressed that there are many reasons for gaps that establishing an unbroken chain of ownership is rarely possible. A gap assessment is recommended when a work of art has a hole in its known ownership during the period 1933–1945, when the Nazis systematically looted the art collections of Jewish families across Europe (Yeide et al., 2001). Omissions in provenance do not necessarily imply a hidden problem, but can indicate potential problems with authenticity. Moreover, Hoving (1997) argues that convenient provenance is a necessary condition for the successful realisation of a fake/forgery. He referred to Anthony Grafton, a faker, who said, '*they [forgers] must […] provide a plausible explanation of where the piece came from and how it fits into the jigsaw puzzle of other surviving works by the same artist as well as similar or comparable works of the same period*' (Hoving, 1997, p. 23). In the same context, Rousseau (1968) pointed out that there are two types of forgeries: bad (or medium) forgeries, which have been detected, and good forgeries, which still hang on the walls. Incredibly smart fakes require a well-trained critical eye to be detected (Tietze, 1948; Savage, 1963; Hoving, 1997). However, bad and good fakes may contain careless mistakes, as in the case of Beltracchi. Beltracchi was a meticulous forger. However, on the hill of his 'success' period of fabrication fakes, he decreased his vigilance and used a wrong pigment. It was discovered when a buyer ordered a chemical analysis of a work presumed by Heinrich Campendonk, presented it as an unknown work made in 1914, and sold it at a Lempertz auction house. It became immediately clear that the type of titanium dioxide pigment used by Beltracchi appeared only after 1930 and was not yet available in 1914, the date of the painting (Beltracchi, 2015).

[23] Free translation: '18. Portrait de Monsieur Duponchelle, 1885 ; localistion actuelle inconnue ; provenance: collection Duponchelle, Paris (commandé à l'artiste en 1885)'.

Another type of mistake is deliberate errors, called a 'time bomb' in the literature. For example, it may be text written in lead white under a painting that can be viewed by X-rays. It may be an anachronism, such as a 20th-century object inserted into a 17th-century painting (Charney, 2015). Why do falsifiers use this technique? If we summarise the opinions from the literature review, two parallel reasons shed light on this phenomenon. First, the time bomb offers a subtle clue to confirm the inauthenticity of a painting. It is a type of self-protection that defends against accusations of fraud once discovered. Indeed, a deliberate error can be invoked to argue that a painting was created without attempting to deceive because it contains an obvious mistake (Keating, Norman, & Norman, 1977). The second is a manifestation of vanity towards the art world. These hoaxes are intended to embarrass individuals or institutions by demonstrating experts' incompetence: '*they [errors] are time-bombs that last just long enough to be taken seriously before blowing up in an expert's face*' (Groom, 2007, p. 1634).

Rousseau (1968) points out that good forgeries are exceptional at any time. To create a good forgery, its creator must be more than an art historian. He must have the knowledge of a restorer, that is, must be a painter with scientific knowledge (Foradini, 2018). If a good forger has knowledge of art, we can hypothesise that he or she can identify 'white zones' in an artist's work, such as missing, destroyed, or stolen paintings. In other words, a faker can recognise and appropriate these gaps to create a convincing forgery. In view of these points, several examples are presented below in which a forger exploits such information to his advantage.

The Dutch artist and forger Van Meegeren prospered just before and during the Second World War. He is the creator of the largest scenic forgery of all time, 'Christ and the Pilgrims at Emmaus', which was attributed (wrongly) to Vermeer. His fraud was discovered only because of his court confessions (Lopez, 2009). Van Meegeren was arrested and accused of collaborating with the Nazis by looting the national art heritage. During the interrogations, Van Meegeren remained silent, which the investigators interpreted as evidence of guilt. Soon afterwards, however, the artist made the sensational confession in court that he had forged the paintings in question (Lopez, 2009). His revelation was later confirmed. The main reason for the success of

Van Meegeren's fake/forged Vermeer is that he did not try to imitate Vermeer's classic style, which is well known to experts. Instead, Van Meegeren, who studied art history, knew that some of Vermeer's early works were missing and that a small number of his extant works are not representative of his late style (Rousseau, 1968). Rousseau then emphasised, '*Van Meegeren saw this gap and decided that he would fill it. He wouldn't copy Vermeer: he would create the missing part of Vermeer's oeuvre*' (Rousseau, 1968, p. 247).

Zhang Daqian, one of the most famous and prodigious Chinese artists of the twentieth century, was known as a faker (see Section 3.1). He often focused on descriptions of lost paintings in catalogues to create fake or forged paintings based on these references (Fu, Chung, & Start, 1991). This decision was made to mislead potential collectors who might have believed that they had discovered a rare masterpiece by chance. For instance, to fill the gaps in the provenance of his forgeries, Zhang attached his pictures to antique Japanese-patterned silk, which was fixed at the base using gilded copper and silver rollers. This design gave the impression that the manuscript was part of a prominent Japanese collection, which would explain its absence from Chinese documents of a particular period (Callaghan, 2008).

In another example, a faker named John Drewe falsified documentation in art archives. He was convicted of conspiracy to defraud, falsify, steal, and intentionally use a false instrument. Drewe hired artist John Myatt to make forgeries of famous painters. *The New York Times* said that Drewe's '*real genius lay in his ability to authenticate Myatt's works through bogus provenances […]*' (Landesman, 1999, para. 6). Drewe systematically visited some of the most protected art archives to modify the historical data of authentic paintings to establish 'new' histories that included his fakes/forgeries. He spiked collections with fake documents providing paintings with true legends. '*Alan Bowness, former head of the Tate and the son-in-law of Ben Nicholson, was fooled into authenticating two of Myatt's fake/forgery Nicholsons, not because the pictures were good – in fact, the general consensus was that they were unimpressive at best – but because the provenance was flawless*' (Landesman, 1999, para. 44).

Finally, we can mention the case of one of the most recent high-profile forgers, who, like Van Meegeren, exploited information about the unknown or lost works of famous artists. Wolfgang Beltracchi copied paintings that had gone missing during the Second World War and painted new works in the style of famous artists (Friedrichsen, 2011 cited by Hufnagel & Chappell, 2016). He invented convincing and sophisticated legends of provenance for faked collections.[24]

3.2.2.2 Exploring provenance research for information about art objects

The link between provenance and authenticity was not immediately given approval, and the provenance approach has been extensively challenged (Raux, 2013). It was only in the second half of the 19th century that provenance research was finally included in the process of analysing the authenticity and historical importance of a work of art (Von Stockhausen, 2013).

To assist international institutions in conducting provenance research, the American Association of Museums created *The AAM Guide to Provenance Research*, which includes a precise description of current methodologies, resource indices, inventories, and databases. This Guide was written to address the serious necessity of improving the situation regarding the restitution of looted art objects (Yeide et al., 2001). One part of the guide includes the specific issue of the search for provenance between 1933 and 1945, whereas another part deals with basic provenance research. Nevertheless, despite the existing guidelines, this area of research still faces many uncertainties. According to Gail Feigenbaum, associate director of the Getty Research Institute, and Inge Reist, chief of research collections and programmes of the Center for the History of Collecting, '*provenance, firmly entrenched though it may be as a standard part of art-historical research today, is neither stable as a concept nor constant as an instrument*' (Feigenbaum & Reist, 2013, p. 1).

Feigenbaum (2013) examined whether provenance could be determined from the objects themselves to understand how ownership could manifest otherwise. To summarise, there are forms and codes of provenance which '*are indispensable*

[24] The fake story was that Werner Jägers, who had died in 1992 and was a member of the Beltracchi family, had bought the pieces before the Second World War and hidden them in the German Eifel region during the war (Hufnagel & Chappell, 2016).

indicators in exploration of the concept of provenance' (Feigenbaum, 2013, p. 7). Notably, visible signs of ownership may include coats of arms, seals, collectors' stamps, labels, and inscriptions (Feigenbaum, 2013; Wieseman, 2010). These markers can be interpreted in various ways. For example, wax seals were often marked on the back canvas or a panel of a work of art when it entered a royal or state collection, or when such collections were moved or reinstalled. If a work of art is placed in a specially designed frame, it becomes recognisable. Similarly, if a work of art has inventory numbers, it means that it was once incorporated into a collection. Thus, inventory numbers may restore the link between an object and its history (Feigenbaum, 2013). Consequently, it is sometimes possible to identify an object's owner and even its affiliation with a collection through such indicators without documentation (Feigenbaum, 2013).

To obtain 'hidden' information about an art object, the authors of the AAM Guide recommend first paying attention to the details of the object, such as its style, subject, signature, materials, dimensions, and frame (Yeide et al., 2001). Below, we indicate the guidelines' recommendations for collecting data that can help reveal latent identifying information about a work of art. Consequently, we use these guidelines by analogy in the experimental part of this study. The list below contains the key points for examining an art object to trace its history, as proposed by the AAM Guide (Yeide et al., 2001):

Front of the painting	Determine its support (canvas, wood panel, board, or other support) and material (oil, tempera, acrylic, or other medium). Painted surface: removed and transferred to another support. Measurements: height preceding width. Record any inscriptions or distinctive marks that are visible (e.g. signature, date, or inscription). Record where the signature appears and how it reads (signatures and dates are usually found on the front of a painting in the lower corner). Record any numbers or other marks (numbers written in bold white or red paint in the lower corner are inventory numbers).

Back of the painting	For panel painting, investigate bevelling and panel reductions. For canvas painting, investigate reductions in size or relining. For painting on metal, investigate surface damage caused by abrasion or warp. Check labels, marks, tamps, and inscriptions; handwritten and stamp numbers; and words written directly on the object.
Sources	Curatorial files (usually organised by the artist and containing scholarly research and correspondence). Registrar's files (containing information on the legal aspects of a work of art, such as its acquisition, loan, or sale). These files may give information about previous owners and exhibitions. Conservation files (containing X-rays, infrared photographs, and any technical reports). Institutional archives, which may include information about gifts and exhibitions.

3.2.2.3 Catalogues raisonnés and their authors

The results of stylistic and provenance research are often included in the catalogue raisonné, the most important published research document on an artist's work (Spencer, 2004b). The catalogue raisonné is an inventory of all works known by an artist and is generally written in chronological order. It contains details such as dates, supports, sizes, references, provenance, and sometimes exhibition histories. However, there is no official standardisation of forms and countenance for a catalogue raisonné (Spencer, 2004b). Generally, a catalogue raisonné includes a detailed descriptive analysis of an artist's works that supports attribution through a discussion of relevant issues (Findlay, 2004). '*Today, the catalogue raisonné is a standard reference tool, used by scholars and critics as well as dealers and collectors to help determine or verify attribution or provenance and other matters relating to the business and study of an artist's production*' (Findlay, 2004, p. 55).

As mentioned above, there are no national or international regulations, sanctions, or standards for catalogues raisonnés. Consequently, the art world may reject a catalogue raisonné as unreliable. Findlay (2004) believed that the criteria to be applied to

determine the viability of a catalogue raisonné are a complex issue, but the context of its production, particularly its authors and sponsors, should be addressed first. In his opinion, *'[t]he most respected catalogues raisonnés are those created by individuals or teams with access to the artist's archives and long experience examining all the works, as curators, scholars, or dealers'* (Findlay, 2004, p. 59).

Stebbins (2004) mentioned that art experts are often the authors of recent or upcoming publications on an artist, which may range from short articles to catalogues raisonnés. These experts are typically on the lists of major auction houses. However, 'art expert' is not a profession regulated by Swiss law. (Neuhaus, 2014). In the art world, an art expert is a person or institution recognised by the art market as an authority over a particular artist (Lemoine, 1992, cited by Neuhaus, 2014). Neuhaus (2014) defined two types of authentication authorities in Switzerland. The first is individuals, including art historians (often involved in the publication of a catalogue raisonné), art dealers (often the artist's main dealer), and members of the artist's family. The second is a group of experts who collectively make judgments on authenticity, including foundations established by artists which authenticate the works of a particular artist or are in the process of producing a catalogue raisonné (e.g. the Robert Mapplethorpe Foundation, *Comité Marc Chagall*, and Félix Vallotton Foundation). This group also includes institutions. For example, the Swiss Institute for Research in Art (SIK-ISEA), founded in 1951 as a public-interest fund, is active in the research and documentation of Swiss artists.[25]

A good catalogue raisonné is usually a work in progress, meaning that its authors should periodically revise it (Findlay, 2004). Stebbins (2004) provided an example of the testimony provided by an expert and author of two catalogues raisonnés of the American painter Martin Johnson Heade (1819–1904), who stated that in 1981 he was asked to authenticate an unknown work by Heade that did not appear in his previously published catalogue raisonné and that was for sale at Sotheby's. This expert then wrote to Sotheby's to inform them that he attributed this work to Heade and would include it in the next edition of his catalogue raisonné. Sotheby's added the work to its

[25] SIK-ISEA is involved in the publication of various catalogues raisonnés, including those of Ferdinand Hodler, Cuno Amiet, and Félix Vallotton.

catalogue of American paintings, with a note indicating that it would be included in the next catalogue raisonné. However, at some point before the auction, the expert had doubts because his first assessment was based only on an 8 × 10-inch colour transparency.[26] He travelled to New York on the day of the sale to examine the painting in person, using stylistic analysis techniques. Unfortunately, his doubts were confirmed despite the solidity of the artwork's provenance. The expert asked Sotheby's to remove the painting from sale, and insisted that if they nevertheless sold the painting, the auctioneer should announce that the expert had removed his previous attribution. Ultimately, Sotheby's decided not to sell this work and, later, an art historian, curator, and scientist with the Museum of Fine Arts in Boston examined the painting. A meticulous stylistic and visual examination of the painting's surface, canvas, and frame indicated that it was a modern fake or forgery (Stebbins 2004). This example illustrates an almost ideal behavioural model for an art expert, as a competent professional would update his/her catalogue and inform the public about it. Indeed, expert assessments sometimes require interpretation, which is a subjective process subject to human error. At the same time, his professional independence affords him a certain freedom and confidence to correct his mistakes.

In practice, art experts in the US are generally protected when they give their opinions in court because they are considered to serve the public interest (except in cases of negligence; Spencer, 2004b). However, art experts today seem just as concerned about their legal liability for their opinions on authenticity as their predecessors were, and their fear is not unreasonable (Spencer, 2004a). Stebbins (2004) explained this fear by stating that the stakes have increased significantly owing to the enormous growth in the art market in recent years. In the event of litigation, millions of dollars could easily be at stake, compared to a few thousand dollars in the 1960s. Stebbins (2004) also stressed that collectors in the 1960s collected art because they were passionate and wanted to share it with the public. Today's collectors often purchase art for various reasons, including investment and prestige. Therefore, collectors are becoming increasingly contentious. Finally, Stebbins (2004, p. 136) stated that in general art experts and collectors 60 years ago were '*much more likely to be social and economic*

[26] The expert noted that in his practice, he always tried to see original works but, like almost all of his colleagues, has also made decisions based on photographs (Stebbins, 2004, p. 138).

peers than they are now'. Experts currently have modest academic and museum salaries. Therefore, they may have considerable economic reasons to worry about being sued.

According to Spencer (2004a), experts who determine the authenticity of a work of art in different contexts (including the publication of a catalogue raisonné) sometimes define their authenticity determinations as opinions.[27] Spencer (2004a) highlights two main reasons for doing so. The first is '*the nature of what they are rendering: their judgment, evaluation, or deduction, based upon an interpretation of existing facts which they have collected and analysed, and to which they have applied their learning and experience*' (Spencer, 2004b, p. 181). The second is that they try '*to limit or avoid legal liability in the event their conclusion is wrong*' (Spencer, 2004b, p. 180). Although the experts' concerns may seem unfair at first, the question of legal liability is complex and in practice depends on several factors. Notably, American courts reject the viewpoint that providing an 'opinion' does not imply that the expert has a legal responsibility (Spencer, 2004b).

Ideally, each catalogue raisonné should contain specific information about suspicious works of art, regardless of whether they are innocent copies of the artist's style or works created with the intention of deception. Integrating such information into a catalogue raisonné would raise awareness among actors in the art world. Jackson Pollock's catalogue raisonné is an excellent example (Spencer 2004a). This catalogue raisonné contains not only a section called '*Problems for Study*'[28] but also a section called '*False Attributions*', which includes malicious works or works wrongly attributed to Pollock. The most significant section (especially for scholars) is '*Works for Further Study'*, which includes works for which the catalogue's authors could not give an opinion owing to insufficient evidence for authentication (O'Connor & Thaw, 1978). Spencer (2004a) remarked that this section was included for the purpose of leaving the task of authentication to future art specialists. In most cases, data similar

[27] This term suggests the existence of specific facts and communicates the author's judgment on or interpretation of these facts (Spencer, 2004b).
[28] These are works that have not been authenticated or rejected by a consensus of opinion (Spencer, 2004a).

to those mentioned above are retained in experts' archives and are almost never disclosed to the public.

In the present study, considering all these points, it is presumed that a catalogue raisonné exploited in the experimental section constitutes credible sources of information whose contents indicate the attributions to the artists provided by their authors, recognised as leading art experts by the art world.

3.2.3 Technical/scientific analysis

As mentioned above, an expert who determines the authenticity of an art object can complement the connoisseurship approach through a technical or scientific examination. Furthermore, Neuhaus (2014) posits that integrating such a method may protect experts from harmful litigation. For example, she explains that in Switzerland, the authenticator must comply with legal obligations under the agency contract. Thus, given the duty of due diligence in authentication, incorporating technical/scientific methods, especially for valuable works of art, is the basis of the expert's precaution (Neuhaus, 2014).

This approach is relevant to authentication owing to the increasing sophistication of the technology available, both for expertise and for manufacturing forgeries [29] (Craddock, 2009). However, it should be not ignored that most technical/scientific methods are costly and time-consuming (Bazley, 2010; Neuhaus, 2014). For example, according to our sources, the price of an analysis may be as high as 25,000 euros. Certainly, low-cost scientific methods, such as ultraviolet light examination, microscopic examination, and visual inspection, can be used by art experts without the help of scientific experts. These basic technical examinations may raise suspicions, but both experts must be involved in interpreting and understanding the evidence to conduct a more sophisticated analysis (Bazley, 2010).

[29] '*New techniques of copying, exemplified by the ink-jet printer, and new materials, exemplified by epoxy resins, have made it much easier to produce copies that are visually convincing*' (Craddock, 2009, p. 2).

Furthermore, some art experts believe that objective technical/scientific analysis reassures non-specialists and leads to definitive conclusions in only a minority of cases. According to Stebbins (2004), '*collaboration between conservators and scientists is necessary only in the most difficult cases of questioned authenticity*' (Stebbins, 2004, p. 139). However, Levy (1991) argues that experts who do not use scientific examination but rely entirely on visual examination may be accused of a failure of due diligence if an error is subsequently discovered. (Levy, 1991). Neuhaus cited several authors (Reeves, 2011; Kallier, 2012 cited by Neuhaus, 2014, p. 67) in noting that art experts consider technical/scientific analysis impractical because they believe that it can clearly help detect forgeries but cannot establish authenticity (with some rare exceptions). She also clarified that the process of interpreting data collected through various technical/scientific analyses is inherently subjective and therefore subject to human error. Furthermore, she outlined that technical/scientific analysis is complicated, as it relies on a comparative examination of other works by the same artist/period; however, these references are often unavailable (Neuhaus, 2014). To further this debate, it is necessary to become familiar with the various methods and approaches based on their functions. Below, we provide selected examples of the techniques and methods currently used by scientists at the National Gallery in London. These examples demonstrate the technical and scientific examinations that provide specific information on paintings that cannot be determined otherwise.

3.2.3.1 Analysing art works through science and technology: various techniques

The speed of technological progress is currently high, indicating that technical and scientific methods and approaches are constantly evolving. Within this framework, Craddock (2009) developed a comprehensive guide which allows the reader not only to better understand the technical and scientific methods of authentication for a wide range of antiquities and works of art, but also to create better coordination between art historians and scientists.

Craddock (2009) pointed out that, in his experience, most museums are poorly equipped to handle scientific analysis. Nevertheless, there are encouraging examples

in which a leading museum or national gallery[30] has been able to combine the skills of scientists, art historians, and restorers to conduct competitive technical research using sophisticated tools. Advanced technical research helps detect potential problems, such as reasons for physical object changes or information about past owners (Wieseman, 2010).

The first step in a technical examination is the assessment of documentary evidence, followed by visual examination. Visual examination aims to obtain initial information about an unknown object (Craddock, 2009; Wieseman, 2010). In particular, a visual examination can provide information about any apparent anomalies and visible damage to the painting surface or the particularities of the brushwork and craquelure[31], coarsely ground pigments, and pigment mixtures (Wieseman, 2010). One example is the consistency of the cracks with the age of the work (Charney, 2015). Such inspections may be performed using a microscope or even a hand lens. Typically, scientists who conduct visual examinations aim to determine how an object was manufactured and provide evidence regarding its use, modification, and repair and the sequence of events (Craddock, 2009; Weismann, 2010). As explained in Section 3.2.2.2, the back and side of a painting may be important sources of information because they can clarify the work's construction, original format and function, geographic origin, dating, and provenance. The wood from which a panel was made can reveal its approximate age and place of origin. The painting should be checked for signs of potential damage, such as wormholes in the wood, fungus, patches of moisture, and places where the pigment is peeling from the support. For example, until the eighteenth century, wooden panels for painting were finished manually. Later, machines replaced the manual process so that the wood had a more uniform appearance than the handmade ones. Therefore, Italian oak back panels have a more refined and neater finish. This principle of distinction can also be applied to canvas supports (Wieseman, 2010). Old works on canvas almost always need to be reinforced, because the canvas can deteriorate. Consequently, old canvas that had never been reinforced is suspect. Charney (2015) drew attention to the fact that visual examinations have different consequences depending on who conducts the test. In

[30] Examples include the National Gallery in London and the Van Gogh Museum in Amsterdam.
[31] Craquelure is '*the network of fine cracks in the paint surface'* (Wieseman, 2010, p. 11).

particular, a collector or dealer is only likely to order further tests if a work still appears suspicious after an initial visual examination, whereas a conservator may apply any number of forensic tests, even if there is no visible sign of forgery after the initial test. Additional tests may be necessary, because direct visual examination is limited and cannot always provide complete information about the creation and physical history of a painting. To discover what is hidden under the visible surface of a painting, scientists can use different analyses of materials and production methods (Wieseman, 2010). Depending on specific questions about the art object, they can choose an appropriate method of examination. The following examples show how experts choose concrete methods in practice.

For instance, Van Gogh's self-portrait *Portrait with Straw Hat* was examined using X-rays. Scientists found an underpainting[32] that had been made two years earlier. If this painting was a forgery, the resulting question would be why a forger would bother to paint an original van Gogh and then cover it up, as it was impossible to discover it at the time, because X-rays were rarely used before 1930 (Conklin, 1994, p. 62). Despite this assumption, the Swiss scholar and merchant Walter Feilchenfeldt harboured suspicions about the authenticity of this painting because of gaps in its provenance and because he assumed that the forger had changed his mind and decided to repaint his first version because the first one was not good enough to be considered 'authentic' (Conklin, 1994).

Placing a paint sample under ultraviolet (UV) light in a dark room induces a fluorescent reaction in some paint materials (O'Connor, 2004; Wieseman, 2010). For example, experts at the London National Gallery discovered that the painting *Virgin and Child* attributed to Jan Gosseart and dated 1527 was a forgery. They examined the painting under UV light and observed fragments of an old varnish with greenish fluorescence. Another fragment exhibited bright pink fluorescence, indicating the presence of a red lake pigment that was used only in the 19th century (Wieseman, 2010).

[32] '*The artist's preliminary sketch made before the paint layers are applied*' (Wieseman, 2010, p. 22).

Infrared radiation (IR)[33] is used to penetrate deep into paint. Scientists typically use this method to study underdrawings and reveal the pentimenti[34] or areas of damage. The principle underlying this method is that certain pigments become more transparent under IR irradiation. Carbon black, which is used for underdrawing, absorbs IR light and thus appears dark in the IR images. The hidden lines become visible, and comparing the initial drawing to the final painting can determine whether the painting is original and not a copy (Wieseman, 2010).

Ultraviolet (UV) fluorescence microscopy can be used to identify organic materials. Under UV light, some materials exhibit characteristic fluorescence. A scanning electron microscope (SEM) can be useful for more detailed investigations, particularly for examining the characteristic topography and other minor constituents on fracture surfaces (Craddock, 2009). In highly complex cases, SEM is insufficient. X-ray diffraction (XRD) techniques are then exploited to identify pure samples of crystalline pigments; 'whereas chemical analysis will give the amounts of the various elements present, XRD can identify the molecules' (Craddock, 2009, p. 53). For example, XRD analysis of a copy after Poussin's *The Plague at Ashdod* helped establish that it was '*an effectively done copy produced by Poussin's contemporary, Angelo Caroselli*' (Wieseman, 2010, p. 46).

It is also necessary to mention the option of precisely characterising organic matter using a gas chromatography technique that separates all organic components and thus identifies a particular type of oil or resin. The presence of specific pigments or materials may be appropriate for one painting but unacceptable for another. Such a discovery could raise serious doubts about a painting's authenticity or date. For example, in Italy, ancient painters used *gesso* (calcium sulphate) for the ground layer, whereas, in northern Europe, they used calcium carbonate (natural chalk). Experts from the National Gallery examined *River Landscape*, based on a composition by Flemish artist Pieter Bruegel the Elder. This painting was catalogued as a sixteenth-

[33] IR is the portion of the electromagnetic spectrum with wavelengths just longer than those of visible light (Wieseman, 2010).

[34] This term is used for a change made by an artist during the process of painting which becomes visible if the subsequent paint layers become more transparent over time (Wieseman, 2010).

century Dutch picture, but they discovered that the panel was prepared using a gesso ground, indicating that the painting was made in Italy and not in the Netherlands.

Counterfeiters usually use materials available to artists whose works they are forging; however, they sometimes make mistakes[35]. For instance, Swiss authorities seized around 200 works of Russian artist Mikhail Larionov based on evidence that the two pigments used in the pastels did not exist at the time of their supposed execution (Conklin, 1994).

Scientists can determine the age of a work of art by evaluating the radioactivity of its carbon content and the chemical composition of its material (Conklin, 1994). The National Gallery acquired the portrait *A Man and Two Children* (*Portrait Group*). It was considered a fifteenth-century Italian painting. However, scientists identified the presence of a modern nineteenth-century synthetic material that had been used to create a misleading older appearance. In another case, scientists studying *The Virgin and Child with an Angel* of Francesco Francia, dating from 1490, discovered the presence of yellow chrome, a synthetic pigment from the nineteenth century. Together with other decisive evidence from this examination, they concluded that it was a skilful replica made in Italy during the second half of the nineteenth century '*by someone with an intimate knowledge of the original and some understanding of Italian Renaissance painting techniques. The re-use of an old wood panel and meticulously applied "cracks" suggest that it was made with a deliberate attempt to deceive*' (Wieseman, 2010, p. 41).

One specific method, handwriting or signature authentication, includes several of the aforementioned techniques and scientific analyses. The authentication of a painter's signature usually begins with visual observation, exploiting different types of illumination (Montani, 2015). Depending on the results of the visual examination, scientists may, if necessary, use macroscopic observations (Craddock, 2009). '*Handwriting analysts, for instance, when judging the authenticity of a signature, look to its formal characteristics – the shape of its letters, their angle in respect to a baseline, their loopings above and below that line, and others – in comparison with an authentic signature*' (O'Connor, 2004, p. 7). Based on their experience in perceiving such matters,

[35] See the example of the Beltracchi case in Section 5.1.

experts can see that a forged signature lacks the essential features of a genuine signature. In addition, current studies have highlighted that this method is well known in forensic science and for this reason has been widely accepted by the courts. Thus, identifying an artist's signature may be crucial for authenticating a work of art during the authenticity litigation process (O'Connor, 2004; Montani, 2015).

3.2.3.2 Innovative scientific approaches to art authentication

The current era of computer technology has also impacted the authentication of artwork. Computer science researchers are attempting to improve authentication processes using machine learning[36] and other computer technologies. This type of technical/scientific examination is called computer-assisted art authentication or classification (Bazley, 2010).

Scholars worldwide are attempting to develop methods for classifying artists' styles. For example, a study using computer-assisted classification based on facial recognition allowed semiautomatic classification to identify an artist (Sablatnig et al., 1998). Several machine learning techniques have been explored for visualising style relationships (Lombardi, 2005). Scholars from Rutgers University conducted a comparative study of different classification methodologies for automated classification (Trochim et al., 2016). Another study that performed a similar comparative analysis examined the influence and relationships between artists (Saleh et al., 2014). Another study raised the question of how to determine the era in which a painting was created. To resolve this issue, researchers have proposed a novel computational method that uses multi-view local colour features extracted from paintings (Chen et al., 2017).

All these studies based on stylistic analyses provide methodologies using global characteristics that '*mainly capture the composition of the painting*' (Elgammal et al., 2018, p. 2). Only a few studies have proposed models that facilitate authentication by

[36] Machine learning is a subfield of computer science involving the generation of algorithms and statistical models to recognize patterns within data without using explicit instructions. It is seen as a subset of artificial intelligence (Surden, 2014; Ben-Ari, Frish, Lazovski, Eldan, & Greenbaum, 2016).

detecting forgeries. A recent study based on a large mixed dataset of imitated forgeries and originals was conducted by a group of researchers at Rutgers University, USA, using machine learning. Their process is inspired by '*pictology*'[37] but uses statistical analysis by machines instead of the human eye (Elgammal et al., 2018). Their data collection included 297 digitised works from books, downloaded digitised images from different sources, and a fake drawing dataset consisting of 87 drawings similar to those of Picasso, Matisse, and Schiele created by five artists specifically engaged for this purpose. Researchers developed a new algorithm that compares different handcrafted works and learns deep neural network features. Overall, the authors of this paper argue that developing computer vision and artificial intelligence could provide alternative tools for analysing works of art that lie outside the scope of other techniques.

Wolf-Rüdiger Teegen (Teegen, 2002, cited by Craddock, 2009) proposed a method for distinguishing authentic Roman fibulas from copies. In 1863, archaeologists found 300 antique fibulas, and it was subsequently decided to make copies of them to sell as souvenirs to tourists. However, the copies quickly became confused with the originals, as the originals and their imitations were difficult to differentiate. Thus, Teegen chose measurable parameters (e.g. dimensions, weight, composition, degree of corrosion, evidence of manufacture, repairs, and documentary evidence) and analysed them using principal component analysis.

The Art and Artistic Legacy Protection Service recently launched a project in Washington to create an algorithm for predicting fraud in an online marketplace based on information provided in auction listings. First, a team of art historians and art fraud researchers collected a large dataset of suspected, faked, and forged works of art. They then developed a method to identify the types of artwork that are most vulnerable to forgery (Loll, 2016). They focused on the most vulnerable categories in the online art market according to their study, that is, 'particular medium (paintings, prints and drawings), price points (within the range of $1,000–$5,000) and artists (top-tier artists)' (Loll, 2016, p. 69). For their analysis, they used key indicators combined with a review of advanced image recognition and documentation in an attempt to

[37]Van Dantzig's methodology is described earlier in Section 3.2.1.

develop a predictive model which *'[…] could serve as a fraud prevention filter for the online marketplace in the near future*' (Loll, 2016, p. 71).

Over the past five years, there has been growing interest in the art world community in AI adapted for art authentication. For instance, several companies have recently emerged exploiting probabilistic authentication models. For example, in Switzerland, the start-up MATIS invented a multispectral camera and an image processing algorithm to help interpret information hidden in paintings (Wurlod, 2021). Another example is the start-up Art Recognition, which verifies the authorship of artworks using AI to learn the key characteristics of the artist from a set of images of original paintings (Art Recognition, n.d.). The Lausanne-based company Artmyn has developed a special scanner that photographs artworks from different perspectives. The images are then processed using an algorithm to obtain a three-dimensional view of the object that could be used for authentication (Bloch, 2017). While their active role in the art world demonstrates the potential need for novelty, their methodologies are not transparent because they are commercial platforms. The methodology used in this thesis is publicly available and can be used as an academic reference for future research in this area.

With the explosion of AI technology, it is unsurprising that Interpol launched an ID-Art mobile application that uses an algorithm (ID-Art mobile app, n.d.). The application identifies stolen cultural goods. It is connected to their database of stolen artworks and affords the general public access to mobile devices. This search can be performed manually or visually. In the first case, the researcher must enter the characteristics of an art object, such as type, medium, technique, title, artist's name, or country of origin. In the second case, ID-Art uses cutting-edge image recognition software to compare photos with database objects.

3.3 Spotting inconsistencies in authentication: a case study

To gain deeper insight into how stylistic connoisseurship analysis, provenance research, and scientific/technical analysis work together in the context of authentication, we present a case study from the BBC One documentary series 'Fake or Fortune?' Each episode examines the provenance and attribution of notable

artworks. The information is presented by art historian Philip Mould and journalist Fiona Bruce based on forensic analysis, archival research, and stylistic analysis conducted by various art specialists (BBC One, 2011). Therefore, our objective is to define the different stages of the process and highlight and explain the logic underlying the research problem. In addition, we focus on the links between the steps to understand why and under what circumstances researchers use these approaches.

We examine Series 1 (2011), concerning a painting signed by Claude Monet entitled *Bords de la Seine à Argenteuil*. The owner commissioned the programme to examine the authenticity of the painting. Previously, the painting had been excluded from the Monet catalogue raisonné, edited by the Wildenstein Institute. In general, each authentication case is specific and depends on the confluence of various factors; thus, it may be necessary to consider different paths throughout the process. Accordingly, in one case, three approaches may be proposed, whereas in another case, only one approach is sufficient to make a decision. Readers were invited to follow these three steps, as if it were an optimal study model. In practice, these steps can be performed either in parallel or sequentially.

Stage 1: scientific/technical analysis

In our example, the team of investigators decided that the first stage should be conducted in an art research laboratory. This choice was likely motivated by the presumption that they would find information hidden within the object; thus, based on their suspicions, they would be able to create the initial hypotheses. In the laboratory, a scientist used high-resolution infrared and X-ray photographs to identify clues inside the canvas. The painting was then scanned using a powerful camera that provided images with unparalleled resolution and colour accuracy[38]. This type of camera uses 13 different light filters, from ultraviolet to infrared, to not only see the surface of the painting in detail, but also see through each layer of paint to reveal the artist's technique. As mentioned in Section 3.2.3, one of the functions of technical/scientific analysis is to raise suspicion. However, this was not the case in our example. Scientific analysis only allowed the researchers to prepare the groundwork

[38] This camera can provide images with more than 240 million pixels, whereas a standard digital camera provides images with about 12 million pixels (BBC One, 2011).

for documentation and the work of specialists in stylistic analysis. The labels and inscriptions on the back of the canvas were accentuated, and the scientist enlarged the painting image to facilitate a more accurate visualisation of the layers for stylistic connoisseurship analysis.

Stage 2: provenance research

The initial hypotheses were developed during the provenance research. The general objective of provenance research is to determine the history of a work of art from its creation to the present (see Section 3.2.2). The team highlighted several successive claims that needed to be proven. One was to find documentary evidence of the painting's existence during Monet's lifetime. Claude Monet died in 1926. Thus, if the team could prove that the painting in question was documented before that date, it could support its originality. The team linked this statement with the argument that art dealers could send suspicious pictures to Monet in order to confirm their originality. In other words, evidence that the painting existed before Monet's death could imply that it was not a fake or forgery, because otherwise Monet would have pointed it out.

As discussed in Section 3.2.2.2, the provenance may be evident within an object itself. For example, labels and inscriptions may constitute a visible sign of ownership. Their interpretation and compatibility with documented evidence are fundamental for research. Consequently, the following hypotheses were developed based on labels and inscriptions in provenance research: As mentioned in the previous paragraph, scientific analysis has created the grounds for such an examination. On the back of the canvas, there were three stamps. One stamp read 'Latouche'. Through historical documentation, the team discovered that 'Latouche' was a colourman who supplied canvases to artists in Monet's circle. Thus, to prove that Monet may have used canvases with such a stamp, it was necessary to prove that Latouche supplied the canvases directly to Monet. The second stamp was a railway baggage label from Paris to Argenteuil[39]. The team decided that this step could help place the painting in the correct geographical area. They attempted to demonstrate this step by finding a place in Argenteuil resembling the painting's image. The third label is a dealer's stamp with

[39] A village in the Paris region.

the stock number of a sales catalogue. With this strategy, the team first expected to identify the previous owners before 1926 and hoped to then set up a chain of owners from the last owner to the owner who bought the picture from Monet.

Stage 3: stylistic connoisseurship analysis

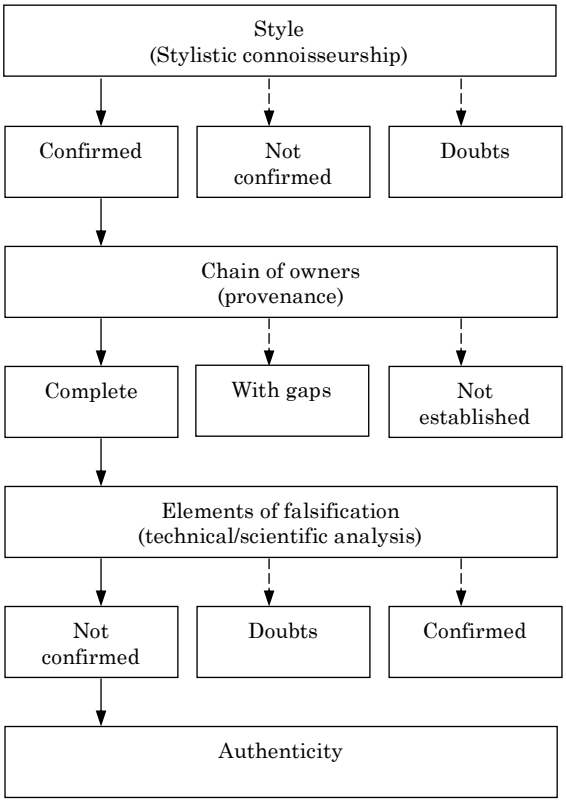
Stylistic connoisseurship is a visual inspection by a knowing eye based on the perception and interpretation of the form and facture that are specific to the artist who created a work of art (see Section 3.2.1). The team invited a connoisseur who was a specialist in Monet. In his opinion, Monet's style was recognisable in this painting. In particular, he said that the appearance of the brushwork and the way the paint was handled were convincing. He pointed out that the paint was extraordinarily varied and that he observed the simultaneous combination of freedom, spontaneity, and control in terms of style, which in particular is familiar with Monet's style. He also positively assessed the signature on the painting, based on the conclusion that the flow rhythm of the handwriting was identifiable. The second expert was a legal authority at the Cologne Museum who has used scientific techniques to detect whether a Monet is a fake or forgery. This expert examined the brush marks and confirmed that the painting techniques were typical of Monet. She also stated that the signature was written in green rather than the usual black or brown of other painters, which was also typical of Monet, as he loved using colour signatures in harmony with his painting.

Finally, a team of investigators collected solid evidence. They found no suspicious elements during the scientific and technical analysis, and found documentary evidence that Latouche supplied canvas to Monet. They were able to reconstruct the chain of owners, and they found documentary clues that the painting existed up to 1926. Despite all of this evidence, however, the verdict of the authority that made the final decision was negative – 'not by Monet'. Authorities argued that no paintings in Monet's oeuvre had been executed in the distinctive style of this river landscape (BBC One, 2011).

This case study illustrates the complexity of authentication. Although all relevant evidence was collected, it was not sufficient to change the initial opinion of the authorities who had previously excluded this painting from Monet's catalogue

raisonné. Second, it demonstrates a situation in which the significant results of all three methods were counterbalanced by the argument of stylistic incoherence. Supposedly, because Wildenstein Institute experts had different opinions on stylistic issues than the team, the results of the provenance research and scientific analysis were not recognised as relevant in this case. This observation leads us to the assumption that when scientific/technical analysis reveals suspicious elements, these discoveries may impact the conclusion of a fake/forgery; however, the absence of such elements is not always a decisive factor in proving the authenticity of a work. Furthermore, our example shows a situation in which authenticity is not confirmed even though scientific analysis does not reveal suspicious elements.

By describing an episode of the BBC programme, we can see the ultimate goal of each of the three methods and better understand the nuances of their relationships. Indeed, the objective of the stylistic connoisseurship method in the BBC enquiry was to find similarities between the style in question and the original style, and the objective of provenance research was to find historical documentation confirming the existence of a connection between the current owner of the painting and Claude Monet. Consequently, these two methods provided evidence of authenticity. The objective of the scientific analysis was to check for the presence of suspicious elements, that is, to detect falsification. Given the divergence of opinions in the current literature on the role of technical/scientific analysis (see Section 3.2.3), the following observations can be made regarding this example: when the findings of the technical/scientific analysis are negative, they support authenticity, and when they are positive, they point to falsification. However, an unequivocal assessment of this issue would be overstated. Drawing an overall conclusion is not appropriate because the situation can vary considerably in each case depending on the initial context of the examination, that is, whether there were any suspicious elements at the beginning of the examination, whether the painting was attributed to the artist before the examination, whether the style is irreproachable but there are gaps in the provenance, or other situations. The schematic diagram illustrates the potential interactions between the three approaches.



Schematic diagram: A theoretical schema in which → (vector A) indicates the paths that lead to the establishment of authenticity and ↛ (vector B) indicates that an option could exist but is not taken into consideration in this model.

It should be noted that this illustration is purely hypothetical and presents only a single model. Other likely models derived from this scheme can be designed by considering options (vector B). Overall, the situation can vary considerably in each case depending on the initial context of the examination, for example, as pointed out above, whether there are any suspicious elements at the beginning of the examination, whether the painting is attributed to the artist before the examination, or whether there are gaps in the provenance.

PART 2 – DATA AND METHODS

4 METHODOLOGY

4.1 Objectives, assumptions, and working questions

Forgery, as a counterfeiting offence, undermines art heritage and violates ethical and moral standards. No fundamental difference exists between art forgers and those seeking to deceive or defraud in other areas. The social judgment of immorality established by general criminal law against fraud also applies to art fraud (Merryman, 1992).

Practice shows that proving the intention to mislead during an investigation and prosecution is extremely difficult (Conklin, 1994). Polk and Chappell state that in Australia only three cases of art fraud have been successfully prosecuted over the last thirty years (Polk & Chappell, 2009). Furthermore, when art fraud involves actors from several countries, the victims and offenders may reside in different jurisdictions. In this case, criminal investigations require additional resources and international cooperation among police forces, not to mention the aspects related to the dichotomy between the practice of the art market and that of courts in authenticating artworks. Bandle (2015) stated that a significant discrepancy exists between courts' examination methods and the conclusions of art experts on a work's authenticity [40] . Moreover, civil art fraud disputes are challenging to decide (Amineddoleh, 2015). Based on several civil court cases in the United States, Amineddoleh (2015) explained that the main difficulty was the plaintiff's obligation to prove that the seller knew that the artwork in question was not authentic. At the same time, in our days *'[m]ore people than ever appear to be buying and investing in art. Many will have little or no knowledge of the art commodity, or of the way in which the art market operates* (Rapley 2016, p. 40).

Given the aforementioned factors, the risk of becoming a victim of falsification is not negligible. At this point, the first step in minimising such risks is prudence. Being able to determine when to increase one's vigilance involves issues of awareness and

[40] '*On the one hand, if a court decides a work's authenticity in disobedience of art market standards, the market is very unlikely to accept that authenticity ruling. On the other hand, if a court adheres to art market practices, it may interpret and apply them wrongly, and validate those practices that are detrimental to a healthy market*' (Bandle, 2015, p. 392)

knowledge. This approach was expressed by Fattah (1971), who stated that awareness, guidance, and education initiatives are effective against forms of recklessness and lack of critical judgment that often lead to victimisation. The ultimate objective of awareness is not only to reduce negligence and carelessness, but also to stimulate and reinforce a sense of alertness (Fattah, 1971). This vigilance can be perceived as a barrier that protects actors in the art market from forgers' malicious intentions. Consequently, we suppose that awareness of the probability of procuring a fake/forgery can stimulate vigilance, especially among non-professionals in the art market, and thus, can motivate them to be more objective to avoid possible errors. It can be supposed that whether an art market participant realises that a work of art is, for example, 90% likely to be an art counterfeit, his or her behaviour may differ from the case where such a likelihood is 10%. Hence, the question is which method can establish the probability of a painting being fake.

Most decisions in various spheres of life are made under conditions of uncertainty (Haigh 2012), where probability is the key to making decisions (Haigh, 2012). This is illustrated in the context of judicial decision-making. Common law and Continental European civil law know the civil standard called balance of probabilities (Schweizer, 2016). According to the case law of the Swiss Federal Supreme Court, the level of probability is reached when the motives for accepting an assumption as true are so compelling (from an objective point of view) that other hypothetical possibilities are not seriously considered[41]. In authenticity disputes, Swiss courts determine an art object's actual attribution based on the preponderance standard (Bandle, 2015). Indeed, '*courts will produce a decision according to the attribution that they consider "more likely than not" accurate, based on the experiences of life and objective factors*' (Bandle, 2015, p. 389). Under this doctrine, the degree of belief in the truthfulness of the claims is a probability of 50% +1[42].

In mathematical terms, probability is a number expressing the extent to which a given event is likely to occur under certain conditions which may recur an unlimited number

[41] For Swiss law: Federal Court Rulings 140 III 610, para. 4.1 (para. 612).

[42] If the decision-maker believes that one version of events is more likely than the other, this event would only be established at the level of the balance of probabilities when it would be greater than a 50% probability level.

of times. Theoretically, probability can be measured by the relative frequencies at which events occur and can be expressed as a number between zero and one (Encyclopedia of Mathematics, n.d.; Oxford English Dictionary, n.d.). The question of how to quantify the criteria of doubt that influence human decision-making is one of the main issues in this research. Thus, we must develop a methodology that provides sufficiently reliable estimates of various probabilities to avoid accepting fakes/forgeries as authentic. Concerning the different studies in this field, we conclude that scholars who develop scientific/technical authentication methodologies tend to use data on original and/or simulated fakes made by artists specifically commissioned for their experience. Furthermore, scholars often retain only basic analytical criteria that can be applied at the expert level. These criteria may include stroke features, such as their shape or length, and the physical characteristics of an object, such as its weight or degree of corrosion. Following previous research methodologies, we applied a statistical approach. However, unlike previous studies, our target was to use operating criteria that are fully visible and understandable by everyone, not just experts, and are derived from factual rather than simulated fake/forgery data. The resulting questions were as follows: What sources can provide such data? Which operating criteria should be used?

In a hypothetical situation in which a person is interested in purchasing a work of art, we can ask what information the potential buyer may seek to make a final purchase decision. Outside of consultation with experts, this preliminary information can be found in the available written documents, which include all types of catalogues, such as catalogues raisonnés, auction catalogues, and exhibition catalogues. These sources may contain many characteristics of a work of art such as its subject, size, date, method of signature, place of signature, or confirmation of its provenance. As a potential buyer can use these data in the decision-making process, we can also use them in our future statistical models.

To decide whether there is a danger of procuring a fake/forgery, it is necessary to analyse previous experiences in dealing with fakes/forgeries. In this regard, philosopher and theologian Joseph Butler (1740), reflecting on demonstrative and probable evidence, stated that, when a reasonable man decides whether an event has occurred or will occur, he thinks of other similar events for which he has some

knowledge. The greater the similarity between known events and the event being judged, the more certain he can be in his judgment. Daniel Kahneman, a Nobel laureate in economics, approached this topic from a different perspective in his book *Thinking, Fast and Slow*: *'[t]he best we can do is a compromise: learn to recognize situations in which mistakes are likely and try harder to avoid significant mistakes when the stakes are high*' (Kahneman, 2012, p. 28). Thus, we can assume that if a hypothetical buyer has knowledge of previous fakes/forgeries to which the artist in question has been subjected, his/her ability to make an appropriate decision improves. Following this logic, in a potential statistical model, data on the original artwork should be compared with data on fakes/forgeries. However, data relating to fakes/forgeries are extremely rare in a catalogue raisonné (e.g. Pollock's CR, see Section 3.2.2.3) and are mostly found in the archives of the foundations of famous artists, police departments that hold sequestered fakes/forgeries, and in documents about them. Thus, the objective of this study is to establish a database of the works of a famous artist who has often been faked and whose works have been published in a catalogue raisonné. The most challenging step is to gain access to the focal artist's fake/forged archives with the aim of integrating them into our database. It is important to stress that given the limited time and resources available, we must restrict the analysis to one artist. Furthermore, at this stage of research, it was not possible to determine whether our predictive model could be generalised to other artists. However, if the subsequent results are significant, the potential generalisation of certain rules within our statistical model may be conceivable. Therefore, the inevitable uncertainty must be communicated in the most effective form so that the public can understand the limitations of our model.

Thus, the ultimate objective of this study was to develop a model and build an alert tool. The resulting model is expected to help people make proper decisions with the utmost certainty to avoid mistakes. We adopt an approach based on characteristics accessible to non-specialists in art (at least for works that are considered authentic). Theoretically, there is no constraint on including characteristics that can only be discerned by specialists (such as brushstrokes) in the database. The choice to restrict the database to characteristics that everyone can assess is motivated by considerations of (a) greater accessibility and (b) greater objectivity. Another reason for our interest in such characteristics is that although our research procedure is exploratory, it

assumes that the fabrication of fakes/forgeries is not random; on the contrary, they are grouped in niches whose characteristics are accessible even to non-specialists. Hence, we aim to identify these niches statistically. Practically speaking, our model allows users to compare a particular work of art with genuine or false works in a database. For example, the model may predict that a hypothetical work by artist *A* which is small, untitled, without provenance, and made during period *Z* has a 90% chance of being fake or forged. When a buyer discovers in the sales documents that the work of art in which he or she is interested refers to the niche specified by the model (e.g. no provenance, untitled, made in period Z, and small in size), it may indicate the necessity for increased vigilance, as the probability of acquiring a fake/forgery is high (90%). The reader should note that this example is only an outline of the ideal scenario in which to apply the model.

We begin the analytical phase of this doctoral thesis by posing the question of whether statistical methods can be utilised to develop an instrument that can be easily employed by all stakeholders in the art market for the timely detection of counterfeit and forged items. Assuming that the answer is in the affirmative, the subsequent enquiry is about how to build such a tool. Additionally, if a tool is successfully developed, the issue is whether it can perform reliably. If the results of the tests are positive, then the final matter to consider is how to implement the tool in practice.

Developing a solid methodology for our analysis is essential, and it must be built on a consistent and well-thought-out basis. The contributions of different specialists, such as art and statistics experts, to data processing will help deliver efficient results. The following questions were established to guide the experimental analysis in the second and third parts of this study:

- What is the optimal structure for our database?
- How should the operating characteristics be collected?
- Should the variables be divided into groups according to their common meanings?
- How should the operating characteristics be codified?
- How should the variables for our statistical analysis be selected?
- Would a statistical distribution be useful for our analysis?

- ∞ What is the optimal model for measuring the probability of being a fake/forgery or of being an original?
- ∞ Which statistical methods can be used to better present and understand the relationships within the raw data?
- ∞ Which statistical methods can reduce a large set of variables to a smaller set?
- ∞ Which statistical methods can be applied to understand the real association between variables to facilitate the description, categorisation, and generalisation of a large amount of data?
- ∞ Which model can predict the value of a target variable based on several input variables?

4.2 Main statistical techniques used in analysis

This Section provides an overview of the main statistical techniques used in the analysis of our data: principal component analysis and classification trees. We explain their applications, benefits, and limitations and demonstrate how they can be effectively utilised to extract meaningful insights and patterns from complex datasets. Understanding these techniques is crucial for interpreting our results.

As an initial step in our analysis, we tested the effectiveness of principal component analysis (PCA) in uncovering underlying patterns and reducing the dimensionality of our dataset. We can provide a simple visual representation of the data by reducing the dimensionality of our database through a transformation of the data into a new set of variables called principal components (PC), which are not correlated or ordered (Jolliffe, 2002). PCA makes it possible to visualise the existence of groups of variables with similar distributions by studying the positions of the variables in the space defined by the main PCs.[43] Furthermore, the field of statistics provides a variety of interpretations of what can be considered a model in the context of PCA (Jollife, 2002). Jollife (2002, pp. 59–61) concluded that *'although PCA is a largely descriptive tool, it can be argued that building a model gives a better understanding of what the technique*

[43] In an example mentioned in Section 3.2.3.2, PCA was adopted by a research study to distinguish between genuine and reproduction Roman fibulas.

does, helps to define circumstances in which it would be inadvisable to use it, and suggests generalizations that explore the structure of a data set in a more sophisticated way. Therefore, as part of our experiment, we used the PCA model in two ways.

1. *PCA with all significant components,* which includes all components with variance greater than one. This method is used to understand the relationships between variables. The PCs can be interpreted by answering the question of how variables that are highly correlated with the same component share information.
2. *PCA with two components*, which includes the first two components. This method is used for visualisation by summarising the data in a PC score plot. The plot can reveal patterns in the data, such as clusters, that may not be apparent in the raw data (Koch, 2014). Often, the first two PCs exhibit the main structure of the data; therefore, it is advisable to consider a two-dimensional score plot. Nevertheless, it should be noted that an interesting structure (i.e. a split into clusters) may not always appear (Koch, 2014).

Following the PCA, our next step involved the implementation of a classification tree algorithm. The classification tree provides a powerful tool for predictive modelling and decision-making. The application of the classification tree is a critical step in our analysis as it enables us to determine the optimal criteria for differentiating between original and fake works. This statistical technique is particularly valuable because of its ability to handle complex data and identify nonlinear relationships, thus providing insights that would otherwise be difficult to detect. The decision-tree method addresses this issue by examining explanatory variables and collapsing them into groups with similar predicted outcome values using a recursive partitioning process (Breiman, Friedman, Olshen & Stone, 1984).

By employing both principal component analysis and classification tree techniques, we aimed to extract meaningful insights from our data, identify significant predictors, and gain a comprehensive understanding of the underlying patterns and relationships within our dataset. PCA and decision tree methods are both data-mining techniques. '*Data mining'* is a general term used for a variety of statistical techniques developed to analyse massive quantities of data (Strobl, 2013, p. 1). Strobl (2013) stressed that

the different techniques that comprise *data mining* are algorithmic, in the sense that they rely on computer programmes to comprehensively identify functional forms (models).

When large amounts of data need to be analysed, classical statistical methods such as linear regression models[44] may be too simple to describe the true, complex associations between the variables (Koch, 2014). The decision tree method is a nonparametric regression method, meaning that decision trees make no assumptions regarding the space distribution and classifier structure. The tree structure enables both regression (predicting continuous values) and classification (assigning data points to categories) tasks (Breiman et al., 1984). Unlike regression models, decision trees, like classification trees, are not penalized by possible multicollinearity between variables. Indeed, variables are considered one after the other by the model, and if one variable allows an efficient division of the data into two groups, another variable with the same information will simply not be used subsequently. In other words, the algorithm for constructing classification tree aims to maximise the purity of the resulting groups at each division. This naturally leads to the selection of variables with high discriminative power for the classification task, without being significantly influenced by correlations with other variables. Consequently, this approach reduces the dependence on multicollinear interactions (Chowdhury, Lin, Liaw, & Kerby, 2022). Thus, we decided to continue our examination using classification tree analysis[45] (CART). The aim of this method is to create a model that predicts the value of a target or response variable Y based on several input variables $X_1, X_2, \ldots X_n$ (Stoble, 2013).

To construct the classification tree, we used the following four approaches.

1. *Original vs. Fake/forgery* is based on taking all variables of the experimental set as explanatory variables and using them to explain a dichotomous variable distinguishing between original and fake/forgery artwork.

[44] '*In a linear regression model, the functional form of the association between the predictor variable × and the response variable Y is assumed to be linear. Because of this assumption, the association can be described in a very simple way by means of two values […]*' (Strobl, 2013, p. 2).

[45] When the predicted outcome is the class (discrete) to which the data belongs, the method is called classification tree analysis, and when the predicted outcome can be considered a real number, it is called regression tree analysis (Breiman et al., 1984).

2. *Original vs. Fake/forgery and Doubtful (Fake/forgery/Doubtful)* is based on taking all variables in the experimental set as explanatory variables and using them to explain a dichotomous variable distinguishing original from fake/forgery and doubtful artwork.
3. *Original vs. Fake based on PCs* is based on taking the PCs obtained from the PCA and using them to explain a dichotomous variable distinguishing between original and fake/forgery artwork.
4. *Original vs. Fake/forgery/Doubtful based on PCs* is based on taking PCs obtained from the PCA and using them to explain a dichotomous variable distinguishing original from fake/forgery and doubtful artwork.

Through this approach, we aimed to uncover the key variables and their thresholds that best separate the data and make accurate predictions. The CART model can be visualised using diagrams that resemble upside-down trees.

Figure 1 shows that a sample can be divided into separate groups according to a predictor variable, which are further divided into additional groups. The principle of a classification tree algorithm is to divide a set of observations step-by-step according to available variables. At each step, a variable that helps to better differentiate between two categories is identified. One branch of the tree is subdivided into two sub-branches according to the observed values of this variable. This process is repeated until it is no longer possible to improve the classification using additional variables (Nakache & Confais, 2005). This method helps to generate a set of rules by building a predictive model.

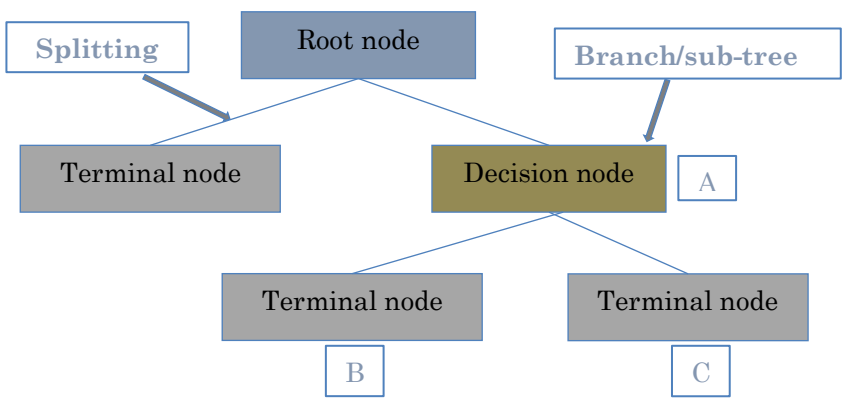


Figure 1: Example of a classification tree

The fundamental concepts associated with the classification tree explain the basic terms related to the decision/classification tree method (Analytics Vidhya, 2016):

- ∞ The root node represents the entire sample, and is divided into two or more homogeneous sets.
- ∞ Splitting is the process of dividing a node into two or more sub-nodes.
- ∞ A decision node is a sub-node split into further sub-nodes.
- ∞ A terminal (or leaf node) is a sub-node which does not split further.
- ∞ A branch (or subtree) is a subsection of an entire tree.
- ∞ A child is a node that originates from another node. The inverse relationship is that of a parent node. For example, if B and C are children of A, then A is the parent of B and C (Figure 1).

For a full understanding of each technique, we provide detailed explanations and experiments using a specially designed dataset (Chapter 6). In addition, this dataset was created to test and evaluate the feasibility and effectiveness of each technique. Each technique is then tested using the working datasets to gain valuable insights and draw meaningful conclusions about the patterns and features that distinguish genuine pieces from counterfeits.

5 DATA SOURCES

In the field of criminal justice, crime analysis encompasses a range of methodologies that combine qualitative and quantitative approaches to uncover the underlying meanings and patterns of relationships by examining non-numerical data (Boba, 2005). As we enter the field of art counterfeiting, a complex subject infrequently investigated within the academic world, it becomes essential to consider relevant data. This prompts us to examine the reliable indicators and identify publicly accessible data sources that can serve as valuable resources for our study.

The following sections examine the general issues related to the data sources, explain how we obtained access to the archive on forgery, outline the sources used, and define the research sampling and extraction phases.

5.1 Sources of data on art crime

Official or national crime statistics, particularly the police crime statistics (PCS), may provide data on art crime statistics. The PCS records all crimes detected by or reported to police. The main advantage of these data is that they provide a summary report of information on trends in illegal behaviour known to the police in different parts of a country; however, the disadvantage is that certain minor or specific types of crime (e.g. art fraud) are not classified separately in some countries' PCS (Aljumily, 2016). National police departments dealing with such specific crimes may maintain statistical data on art crimes. In Italy, the Carabinieri Command for the Protection of Cultural Heritage has an extensive file on stolen art and cultural heritage throughout the country that is accessible to the public[46]. In the United States, some data on art theft are available from the Los Angeles Police Department and the FBI, as they have specific units devoted to art crime (Chappell & Hufnagel, 2016, p. 38). In Switzerland, the police of the canton of Vaud[47] hold data on art forgery that are not accessible to the public.

[46] http://tpcweb.carabinieri.it/SitoPubblico/search

[47] The police of the canton of Vaud does not have a special department related to art crimes; the police include two investigators in charge of art crimes, and their knowledge is regularly exported for investigations in other Swiss cantons (Cutruzzolà, 2011).

Surveys are another way of measuring crime. A survey is a research tool in which variables are measured as a function of responses to questions asked of a large group of respondents using a questionnaire. These may be surveys of self-reported crime, where the source of information is the perpetrators, or surveys of victimisation, where the source of information is the victims (Aebi, 2006). For instance, victim and self-report surveys have been developed to record crimes that are not documented by the police or reported to the authorities; thus, these surveys show 'hidden' crimes committed in a society (Aljumily, 2016). Bazley (2010) noted that art crime is not a separate category in either the National Crime Victimization Survey (U.S.) or international crime collection surveys undertaken by the United Nations. This is logical because these surveys address the general population, while victims of art crimes usually belong to a specific sector of the population.

There are several other databases available to the public, such as the Interpol Stolen Works of Art database, Central Registry of Information on Looted Cultural Property, and London Stolen Art Database, which record stolen or lost art objects identified by governments, police, and private organisations[48]. According to Balcells (2016), most data collection efforts on art crime focus on theft. However, even this level of commitment does not imply systematic statistical data entry (Polk & Chappell, 2009). Bazley (2010, p. 14) stressed that '*Interpol strived to distinguish art-related crime but has found that, on average, only 60 of its 186 members have sent reports on art theft, with some incomplete or missing reports*'. This could be explained by the fact that most countries record such crimes by the type of theft (such as burglary and robbery), and not by the nature of the stolen object (Bazley, 2010). Furthermore, in 2021, Interpol created the ID-Art mobile application (see Section 3.2.3.2) connected with their database of stolen objects. The coordinator at the Works of Art Unit at Interpol, Corrado Catesi, explained that the development of a database for counterfeit art would be possible only for counterfeit items recognised by the last instance of judicial authority as counterfeit art. However, years are required to collect such information. Therefore, Interpol does not possess a database of art counterfeits (The Art Law Foundation, 2021).

[48] See, for example, the Interpol Stolen Works of Art database (Interpol, n.d.).

Usually, the first source of public information on discoveries of fakes/forgeries in the art market is the media. Neuhaus (2014, p. 60) asserted that '*[r]ecent high-profile scandals have provided evidence of the huge number of forgeries circulating in the art market'*. Here, we provide two noteworthy examples. Contemporary cases: The Beltracchi affair[49] and the Knoedler & Co. gallery affair[50]. Although there are some discrepancies in the number of Beltracchi forgeries in the art market identified by various media outlets, the overall numbers are spectacular. For instance, according to different media sources, approximately 100 cases of fraud were uncovered during an investigation, whereas another expert estimated that the number could be as high as 200 fakes/forgeries (Hufnagel & Chappell, 2016). The latest report by the principal police investigator in the Beltracchi affair, Chief Inspector René Allonge, announced that his team had discovered approximately 170 forged paintings (Hufnagel & Chappell, 2016). In the case of Knoedler & Co., the New York Times stated that the gallery sold 40 counterfeits through Knoedler & Company that took about USD 63 million from their sale (Cohen & Rashbaum, 2013).

Specialists who are (or have been) involved in the authentication process as part of their work can also stipulate information on the extent of fakes/forgeries in the art market. However, such testimony is extremely rare. For instance, in the current literature, authors like Conklin (1994), Polk and Chappell (2009), and Bazley (2010) refer to the findings of Thomas Hoving[51] (1997), who claimed that during fifteen years at the Metropolitan Museum of Art, he examined 50,000 works of art across various categories and found that 40% of them were forgeries. Robyn Sloggett[52], an art conservation expert at Australia's Melbourne University, said in an interview that 'about 10% of paintings on the market, both in Australia and internationally, are

[49] On 27 October 2011, Wolfgang Beltracchi, the painter of the forged works; his wife Helene Beltracchi; her sister Jeanette Spurzem, who helped him in various ways; and 'logistical expert' Otto Schulte-Kellinghaus were accused of the 'most spectacular' art forgery case in German post-war history (Hufnagel & Chappell, 2016).

[50] One of New York's oldest and most respected art galleries, the Knoedler & Co. Gallery, was allegedly involved in the sale of more than 30 fakes brought to the gallery by Glafira Rosales (Moynihan, 2016).

[51] Director of New York's Metropolitan Museum of Art from 1967 to 1977.

[52] Sloggett evaluates about twenty suspicious paintings per year. Usually, these evaluations are performed at the request of a gallery, merchant, or private owner. In addition, she leads the Australian Art Authentication Course, which involves collectors, curators, and police officers from Asia and other regions (Porter, 2007).

generally conceded to be cases of mistaken identity' (Porter, 2007). Finally, Vernon Rapley, who headed Scotland Yard's Art and Antiques Unit, said in an interview that law enforcement agencies estimated that around 20–40% of the art market consisted of fakes/forgeries (Bruce & Mould, 2010).

Although a wide range of data sources can theoretically provide reliable information on art crime, their accessibility remains a challenge. They are subject to restrictions and data protection obligations that vary according to national laws and confidentiality agreements. The following section details the chronology of the search for the data used in this study.

5.2 Accessing data for doctoral research

The literature review demonstrates that only a few studies are based on the analyses of real fakes/forgeries. Indeed, this is not surprising because researchers' permission to collect such information is limited, mainly because of confidentiality issues. We spent more than 18 months searching for relevant data for this study.

At the centre of our initial target was one of the most famous artists of the twentieth century, Marc Chagall. *Marc Chagall Committee* (foundation) protects and promotes Marc Chagall's artwork worldwide and holds the exclusive power to declare works purported to be his as genuine or forged. Our official request to the Committee explained that our research was strictly related to the details of paintings and did not involve any personal information about the owners or other compromising information which could breach confidentiality. However, the Committee definitively rejected our request and even recommended that we cease conducting research on fakes/forgeries of Chagall. The specific reason that the Committee invoked was the confidentiality of all information. Meanwhile, we wrote to *Freemanart Consultancy*, an investigation agency which published pictures of Marc Chagall's forgeries on its website and claimed to possess a large amount of information about fakes and forgeries of Chagall's works. However, the agency refused to cooperate for the same reason as the Committee. Theoretically, our search for Chagall's fakes and forgeries was not exhaustive. However, the most important sources were contacted and it took approximately 12 months to receive only negative responses.

Consequently, we decided to extend our research to other artists and approached different organisations via the *Association for Research into Crimes against Art* (ARCA). All our requests were ignored. For instance, several institutions and agencies, including the *Art & Antiques Unit Metropolitan Police* (Great Britain), *Italian Carabinieri Comando Tutela Patrimonio Culturale* (Italy), *Crime Unit of the Dutch National Police*, and the *K2 Intelligence Agency* (USA), did not respond to our demands for cooperation. Furthermore, we asked for help from the *Police Cantonal Vaudoise* (Switzerland), which handled the investigation of a case concerning the art forgery of another artist. The cantonal police were very cooperative. Nevertheless, police specialists explained that the paintings seized during the investigation were primarily very poor-quality fakes. According to them, such low-quality data could generate misleading results. Therefore, we were unable to use this information.

Finally, we contacted the *Swiss Institute for Art Research* (SIK-ISEA), which works with different foundations of Swiss artists. In particular, the catalogue raisonné of Félix Vallotton was published in collaboration with SIK-ISEA and the *Foundation Félix Vallotton*, located in Lausanne, Switzerland. Félix Vallotton met our objectives for three reasons: (1) his art is often the target of forgers; (2) information on forgeries and counterfeits is preserved in the archives of the Félix Vallotton Foundation; and (3) his catalogue raisonné is published and has an excellent reputation in the art world. With the support of SIK-ISEA, the Foundation Félix Vallotton kindly agreed to collaborate with us. The Foundation's experts opened access to their archives, and confidentiality issues were resolved through agreement.

5.3 Sources used for the development of the database

To develop a model that predicts the likelihood of being an original or a fake work, it is first necessary to create a database that includes two sets: one with the original works of the artist and one with known fakes and forgeries.

Accordingly, each dataset was collected from two sources. The first is the catalogue raisonné of Félix Vallotton's artwork. This catalogue raisonné is the result of twenty years of research by Marina Ducrey, a renowned expert on the artist, and Katia Poletti, the art historian and curator of the Foundation. The catalogue raisonné was published in March 2005 and includes a first monographic volume, followed by two

volumes devoted to the catalogue itself. The catalogue raisonné won the SNA[53] Prize in Paris, and Marina Ducrey was nominated for the rank of *Chevalier de l'ordre des art et des lettres* (Foundation Félix Vallotton, n.d.). The second source is the archived documentation of the falsification of Vallotton's works preserved by the Vallotton Foundation.

5.3.1 Valloton's catalogue raisonné

The catalogue raisonné (CR)[54] of Vallotton's painted works includes 1,704 works of art, including oil paintings and works executed in pastel, tempera, and gouache, identified to this day. The publishers of the CR respect the following order in the description of each work: 1) numbering and reproduction; 2) title and date of execution; 3) technique, support, and dimensions in centimetres; 4) signature; 5) current location; 6) commonplace book; 7) provenance; 8) exhibitions; 9) bibliography; and 10) commentary.

We focus on several aspects noted in the CR that may be pertinent for data sampling. The first concerns support. Félix Vallotton primarily used three types of support: wooden panels, cardboard, and canvas. Two oil paintings were executed on paper, and one painting on paper used oil combined with gouache.[55] The other oil paintings were executed on canvas. Small temperas made between 1923 and 1925 were generally executed on cardboard. His early works were almost exclusively painted on canvas, although this support was expensive for a young artist. Schematically, we can distinguish between the three periods characterised by different uses of support. The first was his youth, when Vallotton painted on canvas; the second was between 1888 and 1889, when he painted on wood because of the functionality of this material during travel; and the third was during the influence of the Nabis between 1890 and 1906, when cardboard was his favourite support. Outside these periods, his priority in choosing supports was not so strict.

[53] *Syndicat National des Antiquaries* (Paris).
[54] Henceforth, we use the abbreviation 'CR' to refer to the *catalogue raisonné* of Félix Vallotton's artwork.
[55] Specifically, Nos. 10, 828, and 291 in the CR.

Another focal aspect is the size of each painting in relation to its support. The CR contains a table of formats corresponding to the standard dimensions in the market during the artist's time. Vallotton adhered to these standards. He indicated in his commonplace book numbers corresponding to the numbers in the table. If he used oversized dimensions, he marked them differently by indicating the width and height in centimetres. Thus, using these notes, it is possible to identify which paintings have non-standard dimensions. In particular, most wooden panels used have standardised formats. Additionally, some paintings that could not be physically examined when the CR was created were described in their documentation as having been painted on wood. Nonetheless, their sizes are much larger than those of standard paintings. This discrepancy could be explained by the fairly common confusion between wood and cardboard supports, especially when the cardboard is thick and hard or reinforced with wood because soft cardboard tends to warp over time. For example, the support of the painting '*Au marché*' (CR No. 196) was considered to be wood until it was later identified as cardboard (Ducrey & Poletti, 2005). The experts examined all paintings on cardboard stored at the Beaux-arts Museum of Lausanne. Consequently, these paintings are considered representative of the types of cardboard used by the artist in the years from 1890 to 1900, and the conclusions of their examination apply to other works on cardboard from the same period (Ducrey & Poletti, 2005).

Another aspect that we would like to highlight is the artist's signatures, stamps, and monogram. Félix Vallotton signed and dated most of his paintings, but 350 were unsigned. However, many of these paintings contain handwritten inscriptions behind their frames. For instance, such inscriptions as the date, title, or location are often attached to his handwritten recommendation '*do not varnish*' (Ducrey & Poletti, 2005, vol. 1, p. 205). Among his unsigned paintings, approximately 330 were marked with a stamp imitating his signature that was made by the Vallotton family after his death. Thus, at present, approximately twenty works remain in their original unsigned states: some old paintings, some paintings donated by the artist to French museums before the stamp was made, and some small panels that were sold or perhaps donated long ago without Vallotton deeming it useful to sign them (Ducrey & Poletti, 2005, vol. 1).

There are two particularities in the timing of his painting signatures: Ducrey and Poletti pointed out, '*The painter generally signs his works only when they leave his studio, either because they have been sold, exhibited, or entrusted in commission to some intermediary. Contemporary paintings from even the same series may therefore bear different types of signatures because one may have been exhibited or sold shortly after its execution, with the other exhibited or sold much later. This observation implies that, with rare exceptions, the works that remained unsigned at Vallotton's death were not exhibited during his lifetime*'[56] (Ducrey & Poletti, 2005, vol. 1, p. 206). Second, the young Vallotton signed some of his paintings in the manner of the old masters. Designed as an ornament of the motif, these early signatures differed considerably from his later ones[57]:

Moreover, in his decorative portraits of famous men at the beginning of the 20th century, Vallotton introduced the identity of the model in the form of a dedication accompanied by his signature and the date of execution. Prior to 1885, when he started his commonplace book[58], most of his paintings remained unsigned, '*[…] since nothing was intended to be exhibited […]*'[59] (with rare exceptions, to which the signatures were probably added later; Ducrey & Poletti, 2005, vol. 1, p. 208). With regard to the monograms, it should be noted that the artist rarely signed his paintings with a monogram, which he usually used in woodcuts and illustrations. However, at the beginning of his career, a few monograms appeared in his paintings, two of which were enclosed in a cartouche as they appeared on his woodcut plates and illustrations. As

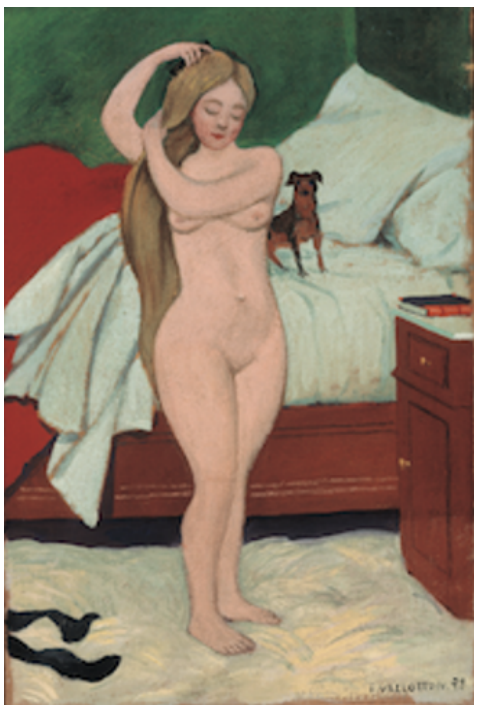
[56] This quote is a free translation of *'[...] que le peintre ne signe en général ses œuvres qu'à l'instant où elles quittent son atelier, soit parce qu'elles ont été vendues, soit pour être exposées ou confiées en commission à quelque intermédiaire. Des peintures rigoureusement contemporaines, voire provenant d'une même série, peuvent de ce fait porter un modèle de signature distincte, l'une ayant été exposée ou vendue peu après son exécution, l'autre beaucoup plus tard (fig. 228, 227). Cette constatation a pour corollaire qu'à de rares exceptions près les œuvres restées non signées à la mort de Vallotton n'ont pas été exposées de son vivant.*' (Ducrey & Poletti, 2005, vol. 1, p. 206).

[57] For example, CR No. 60.

[58] The issue of the commonplace book is discussed later in this section.

[59] This quote is a free translation of '*[…] puisque rien ne les destinait à être exhibées […]*'.

for stamps, the artist's heirs decided to stamp hundreds of his unsigned works of all types and periods. With this decision in mind, the CR includes an annotation of each stamped painting. Nevertheless, we did not distinguish stamps from signatures in our database for standardisation, as it is very difficult for amateurs to distinguish between stamps and signatures and to better interpret the analytical results.[60] An example is a signed painting.



CR No. 220 'Naked woman with a dog', c. 1897

The next below is a stamped painting:

[60] As a preamble to the following chapter, it should be mentioned that the variable containing information on whether a work of art was signed by signature or stamp is called '*signature*'.

CR No. 221 Woman in the bath doing her hair, c. 1897

In both paintings from the same period, the left signature below is the hand's signature, and the right is the stamp.

Handwritten signature and stamp example

As noted earlier, provenance refers to the ownership history of a precious work of art. In the CR, works' origins are established primarily based on different sources of information[61]. Valuable information was collected from artists' archives and the archives of public institutions. The heirs of some of Vallotton's friends, colleagues, and great collectors have made it possible to access letters and photographic documents that provide rich information. Several art merchants have helped to clarify the paths of works that passed through their hands (Ducrey & Poletti, 2005, vol. 1). The most

[61] Examples include the artist's commonplace book, account book, and correspondence, information from the backs of the works, inventory books of the artist's estate, copies of invoices, annotated photographs, and exhibition and auction catalogues (Ducrey & Poletti, 2005).

valuable source of information was Vallotton's commonplace book.[62] Vallotton maintained a commonplace book from 1885 until his death that took the form of a large, handwritten, bound notebook with the explicit title '*List of my works, paintings and engravings, made in chronological order, from 1885 onwards'*. Ducrey and Poletti (2005, vol. 1, p. 1) pointed out that '*scrupulously precise, his commonplace book goes beyond simple memory assistance for personal use; it reveals the desire to control production intended for posterity from beyond the grave and, thus, the artist's confidence in the durability of his art. [...] The commonplace book gives the measure of work that is all the more considerable as its realization is concentrated over only forty years and as Vallotton has exercised his talents in many disciplines other than painting*.'[63] However, more than a hundred paintings are not mentioned in the commonplace book. The CR's publishers explain that the omissions are sometimes inexplicable (Ducrey & Poletti, 2005). One example is the large portrait of Gertrude Stein[64], which Vallotton wanted to exhibit at the *Salon d'automne* shortly after its completion. Nevertheless, it was not mentioned in his commonplace book. These paintings were included in the chronology of the CR by referring to other available sources (Ducrey & Poletti, 2005, vol. 1).

Furthermore, it should be noted that the locations of several works of art were unknown at the time of the CR's creation, but they were noted and described by the artist in the commonplace book[65], and were then included in the CR. In addition, Vallotton sometimes mentioned a series of paintings using a common description without specifying the number of paintings in the series (Ducrey & Poletti, 2006, vol.

---

[62] The Oxford dictionary characterises a commonplace book as '*a book in which "commonplaces" or passages important for reference were collected, usually under general heads; hence, a book in which one records passages or matters to be especially remembered or referred to, with or without arrangement*' (Oxford English Dictionary, n.d.).

[63] This quote is a free translation of '*Celui-ci se présente sous la forme d'un gros cahier manuscrit et relié dont l'intitulé est explicite: «Liste de mes œuvres, peintures et gravures, faite dans l'ordre chronologique, à partir de 1885». D'une précision scrupuleuse, il dépasse le simple aide-mémoire à usage personnel; il révèle la volonté de contrôler au-delà de la tombe une production destinée à la postérité et donc la confiance de l'artiste en la pérennité de son art. [...] Le Livre de raison donne la mesure d'une œuvre d'autant plus considérable que sa réalisation est concentrée sur une quarantaine d'années seulement et que Vallotton a exercé ses talents dans maintes autres disciplines que la peinture*' (Ducray & Poletti, 2005, vol. 1, p. 1).

[64] '612. Gertrude Stein, 1907; Livre de raison: Non cité' (Ducrey & Poletti, 2005, vol. 2, p. 369).

[65] An example is '149. *Baigneuses au clair de lune*, ca. 1893. Current location unknown' (Decrey & Poletti, 2005, vol. 2, p. 81).

1). For example, he noted *'Various subjects sea bathing. Etretat'*. For paintings in this category, the CR reserved only one unit per series[66]. Paintings that were physically destroyed[67] or stolen[68] were also included in the CR if information about them was noted in the commonplace book.

5.3.2 The Vallotton Foundation's archive

The second source of information is the Vallotton Foundation's[69] archive. This archive contains approximately 200 files[70] with documents describing works of art deemed either doubtful or fake/forged works of Vallotton. The archive files are classified into nine folders, each of which includes documents grouped by subject and technique as follows: 1) interiors; 2) portraits; 3) still lifes; 4) landscapes; 5) various scenes; 6) nudes; 7) drawings (male or female nudes); 8) drawings (female nudes); and 9) objects, drawings, watercolours, and techniques: paintings, watercolours, gouaches, and pastels. To ensure coherence with our first source and compare the works from both sources, we selected and codified information for only those fakes/forgeries that corresponded to the techniques mentioned in the CR. For example, the CR does not include Vallotton's drawings. Consequently, we did not use data on fake/forgery drawings, even though the archive files contain these data. For the same reason, we analysed the data in the files by analogy with the order of the CR sections mentioned above. Importantly, owing to gaps in certain information in both sources, some of the collected data refer to the artwork in either the CR or the archive files, as explained in Appendix II.

As indicated in Section 3.2.2.1, the term 'provenance' in the legal sense is based on physical possession (i.e. location; Shindell, 2016). Consequently, our experimental

[66] An example is '*Livre de raison Lrz 408: Divers sujets bain de mer. Etretat'*.

[67] An example is Portrait d'Henry Martin, created in 1887. According to the model's grandson (oral communication), Blanche Martin destroyed the portraits of her brother and sister-in-law in the 1930s following an outburst of anger (Ducrey & Poletti, 2006, vol. 2, p. 33).

[68] An example is '123. Le lac vu de Chexbres, 1892. Musée cantonal des beaux-arts, Lausanne. This painting is 'acquired in 1959 from the Department of the Interior, succession Marie-Louise Moreillon. Stolen in Mesola, Italy, August 31, 1988, at the 1988 exhibition, Bellinzona' (Decrey & Poletti, 2005, vol. 2, p. 67).

[69] Henceforth, we use the term 'the Foundation' to refer to the *Vallotton Foundation.*

[70] Henceforth, we use the term 'archive files' to refer to the archives of the Vallotton Foundation.

work is based on the presumption that the possessor of a work of art has the legal right to detain it[71]. In fact, to create our database, we use information from archive files that provide evidence about the locations of works of art but does not in itself satisfy the burden of proof of ownership. For example, the archive files contain emails and letters with details about the place of acquisition, previous owners' names and addresses, and descriptions and photos of paintings provided by people who asked the Foundation to authenticate their artwork. Likewise, any catalogue raisonné cannot prove a priori that the provenance that it describes is free of factual mistakes or is complete because this general information is partly or entirely received from third parties (Shindell, 2016).

5.3.3 Variables selected for the database

The characteristics of the works of art were sampled in the following manner. First, we identified eighty-one variables. Each variable represents a characteristic of the observed artwork. Depending on the needs of the statistical study, different sets of modalities can be conceived for the same statistical variable[72], and the appropriate modality set depends on the research objective (Ouellet, 1998). The variables were sorted according to their categorisation in eleven groups. Second, the categorical variables[73] were recoded using dummy coding. The variables that are subject to dummy coding were labelled 'D' (dummy) and the variables that cannot be recoded as dummy variables labelled 'N' (numerical) and 'Nom' (categorical nominal; see Appendix II). Dummy coding variables means using only the values '1' and '0' to convey all the necessary information to transform categorical data into a number of dichotomies. With this coding, the researcher enters the value '1' to indicate the

[71] The notion of possession and its protection has been the subject of many legal debates for a very long time (Vermond, 1895). For instance, a doctrine supports the notion of the difference between continental civil law and Anglo-American common law based on radically different conceptions of ownership and possession (Gordley & Mattei, 1996). In this study, we do not go into these juridical subtleties, but it is useful to highlight them. Moreover, our research focuses on evidence that fakes can occur in any country, but because the Foundation is domiciled in Switzerland, we assume that the legal concepts used should follow Swiss law. In particular, in the case of the protection of possession, Article 930, paragraph 1 of the Swiss Civil Code (1907) stipulates that the person in possession of movable property is presumed to be its owner.

[72] We argue for this sampling on the basis of practical realisation, and we follow the AAM Guide recommendation mentioned in Section 3.2.2.2.

[73] It should be noted that the majority of the variables in our database are categorical, but some are numerical.

presence or absence of a particular attribute and the value '0' otherwise (Hutcheson, 2011). For example, in the case of an untitled artwork, if the value 1 refers to a work of art without a title, then 0 refers to a work of art with a title. Alternatively, for a titled artwork, if the value 1 refers to an artwork with a title, then 0 refers to an artwork without a title. In other words, by using the values 0 and 1, we can indicate the presence or absence of a particular attribute. In our data, most categorical variables were recoded as having values of 0 or 1. However, some variables could not be recoded as dummy variables.

As indicated above, the variables are sorted into eleven groups, as follows.

1. *Identification number*. This group includes two variables. The first is the number of the original work assigned in the CR, ranging from 1 to 1,704, and the second is the docket number, which consists of the letters A, B, C, D, E, F, or I with the page number. For example, the number 'A11' is read as 'A' (the file name) '11' (the page number in file 'A').

2. *Title*. This group comprises two variables. The variable 'title' indicates the name of the work found in the CR or the archive files. We classify works of art with no title using the variable 'untitled'.

3. *Years*. This group includes four variables. The variable 'year 1' indicates the date of execution, as noted in the CR or the archive files. If two dates are noted together in the CR[74], we recode them using the variables 'year 1' and 'year 2'. For example, if the CR lists the date of creation as '1882 – 1885', then '1882' is the year *before* which the painting was not made (we indicate this year with the variable 'year 1'), and '1885' is the year *after* which the painting was not made (we indicate this year with the variable 'year 2'). Furthermore, this group includes a third variable called 'no date', which contains information about undated paintings on canvas (or paintings on canvas with no date after the signature or monogram of Vallotton)[75], and a fourth variable called 'illegible date', which indicates that the date on the canvas is undecipherable.

[74] The number of dates listed depends on whether authors of the catalogue have doubts about a painting's creation date.
[75] However, the date could be noted in another place, such as on the back of the painting.

4. *Authenticity*. This group includes four variables: 'original', 'fake', 'doubtful', and 'hypothetical'. The variable 'original' includes only the authentic paintings in the CR. The variable 'fake' includes fakes or forgeries in the archive files confirmed by a Foundation expert (see the examples of Pictures 1 and 2 of Section 2.2). In this experimental section, the variable comprising such types of artwork is entitled 'fake/forgery'. According to the connoisseurship analysis, the variable 'doubtful' includes artworks that comprise most of the 'successful' falsification attempts. The Foundation's experts confirmed these works as suspicious (see the examples of Pictures 3 and 4 in Section 2.2). The variable 'hypothetical' refers to works without a signature or monogram which were hypothetically assigned or presumed by the owners to be Vallotton's paintings but were not ultimately authenticated as such. Therefore, these works of art are grouped into a separate category which may include information on forgeries (see the example of Picture 5, Section 2.2).

Table 1: Distribution of works of art in the Authenticity group

Category	**Number of Works of Art**
Original	1,704
Fake	127
Hypothetical	70
Doubtful	13
Total	1,914

5. *Market.* This group includes seven variables. The first variable, 'country of the first appearance', relates to both fakes and originals. It identifies the country of the owner who made a request to the Foundation (usually to provide an estimation for sale) or the country of the physical location mentioned in the CR. The second variable, 'year of appearance', is linked only to fakes. This variable contains information about the year in which the Foundation was first requested to identify the painting. The third variable, 'sale', refers to works of art presented for sale in an auction house or online, but for which the document does not contain any information that the work in question has finally been sold (for example, the work was listed on the auction website). The fourth variable, 'sold', refers to evidence that the paintings were sold. The last three variables

in this group, 'date', 'selling price', and 'estimated price', indicate the date of sale, the price paid, and the estimated selling price, respectively.

6. *Signature/monogram*. This group includes six variables. The first variable, 'without signature', indicates works of art without the painter's signature. The second and third variables, 'monogrammed' and 'signature or seal'[76], indicate works with the artist's monogram or signature, respectively. The fourth variable, 'trace of signature/monogram', refers to the existence of only a trace of a signature on the canvas or other support. The fifth variable, 'atypical signature/with errors or with other letters', indicates cases in which it was observed that the signature was significantly different from that of the artist or contained mistakes. The sixth variable, 'signature on the inside of the pattern or over shoulders of the person', indicates instances in which the signature is situated in an atypical location or, in several cases, over the shoulders of the figure in a portrait[77].

7. *Place of signature/monogram*. This group includes eight variables. It contains information about the location of the artist's signature or monogram, and the relevant variables are 'lower right', 'lower left', 'upper right', 'upper left', 'bottom centre', 'centre right', 'centre left', and 'top centre'.

8. *Technical characteristics*[78]. This group includes twenty-two variables. The first technical characteristic is the dimensions of the painting. All paintings were categorised into three variables according to their size, using subjective criteria. The relevant measurements of the height (X) and width (Y) should fall within these intervals[79]: 'large [61 cm ≤ × ≤ 250 cm and 61 cm ≤ Y ≤ 250 cm]', 'medium [36 cm ≤ × ≤ 60 cm and 36 cm ≤ Y≤ 60 cm]', and 'small [1 cm ≤ × ≤ 35 cm and 1 cm ≤ Y ≤ 35 cm]'. Next, all works are classified according to their

[76] A seal is a signature stamp that is equivalent to the artist's signature and was made by his family after his death.

[77] For example, the painting 'Self-portrait' of the CR No. 108 is signed and dated on the right, above the shoulder.

[78] 'Technical characteristics' is the rather approximate title. It encompasses different types of features such as dimensions, themes, and types of painting techniques and supports.

[79] This division into three sizes was made by considering the artist's largest painting in relation to the smallest one.

subjects. There relevant variables are: 'portrait', 'portrait with identified person', 'still-life', 'landscapes', 'private and public interiors with or without figures', 'great decorations', 'mythological allegorical or biblical subjects', 'various', 'nudes', and 'copy of another famous painter'. The next technical characteristic includes the variables 'oil', 'pastel', 'tempera', and 'gouache'. Paintings with no indication of the technique but for which it can be assumed (via the CR or archive files) are indicated with the variable 'unknown technique'. Finally, the types of support are indicated by the variables 'canvas', 'cardboard', 'wood', and 'paper'.

9. *Back of the painting*. This group includes eight variables and comprises information about the back of the canvas. The variables are 'number/letter', 'sticker', 'seal', 'signature/initials', 'written on the frame', 'written on the canvas', 'written with errors or with letters other than artist's initials', and 'picture'.

10. *Historical documentation*. This group includes nine variables. The first variable is 'without provenance'. This variable represents paintings without an ownership chronology. For example, the picture of Jasinski. This is a portrait whose historical chain of ownership has remained unknown. Thus, the rubric 'provenance' of the CR is empty:

> CR No. 29 *Jasinski in blue jacket,*
> Provenance: (empty)

The second variable, 'recent provenance', indicates paintings for which the ownership chronology begins long after the painting first appeared on the art market. For example, for the painting 'Landscape in Semur', the historical chain of owners in the 'provenance' section only starts in 1970.

> CR No. 1517 Landscape at Semur, 1923
> 'Provenance:
> Sale Ader & Picard, Paris, Palais Galliéra,
> 10.06.1970, no. 127.
> Galerie Vallotton, Lausanne, no. 10264 (acquired for sale).
> Galerie Istvan Schlégl, Zurich (1974).
> Private collection: Zollikon. – Phillips sale, Zurich, 18.03.2002, no. 84'

Paintings acquired directly from the artist, for which there is no complete chronology of ownership, are classified as 'first provenance'. For example, the painting 'Beach at Etretat' has notification of the first owner, Francis Jourdain, who acquired the painting in 1900 (when the artist was alive).

> CR No. 272 'Beach at Etretat', 1899
> Provenance: Francis Jourdain, Paris
> (acquired in 1900 from Bernheim-Jeune).

The other four variables in this group provide information about missing original works, as follows: 'localisation unknown', 'stolen artwork', 'destroyed artwork', and 'absence in commonplace book'. The variables 'restored artwork' and 'relined[80] artwork were also included in this group. Such a notification is explicitly noted in the CR.

11. *Period*. This group includes eight variables[81]. The variable '1882–1889 youth' indicates paintings from the beginning of Vallotton's career. In 1893, Vallotton joined the Nabis group and painted very little, instead devoting himself to engraving wood. This period is categorised using the variable '1890–1900'. A few paintings were produced during this period, in which Vallotton attempted to apply his method of reducing atmospheric phenomena to highly simplified plane figures. The next variable, '1901–1910', indicates Vallotton's paintings after he reduced his work as an engraver. The Nabis theme coexists with the exploration of new techniques. The variable '1901–1905 landscape' indicates works that include the restitution of natural light effects and atmospheric phenomena, which continued to occupy Vallotton's work; nevertheless, sunsets are absent from his work until 1910. Another variable in this group is '1905–1910 nude', indicating the period in which Vallotton presented the nude and its ramifications in vast compositions of mythological or allegorical characters. The motif of the sunset over the sea, accompanied by

[80] Relining a painting means that the linen applied originally applied to the back of an oil painting on canvas is replaced to reinforce it (Oxford Reference, n.d.).
[81] This division over time by theme is approximate and also subjective; not all paintings can be classified according to the themes applied to the periods.

infinite variations in colour, occurred frequently during the period indicated by the variable '1909–1915 sunsets'. Vallotton attempted to artistically express his perception of war, as indicated by '1915–1917 war'. The last variable, '1916–1925 sunset', indicates paintings of the sun setting over the sea from that period.

Appendix II summarises the entire set of variables extracted from original and fake artworks.

5.3.4 Missing data

As mentioned previously, our database includes information from two sources. The first set of information refers to original works of art, whereas the second contains fakes and forgeries. Consequently, the missing data (MD) mechanisms in the case of original and fake data may differ. In particular, the CR and archive files define the scope of our research. Therefore, any additional data for the corresponding artwork that exist outside these two sources can be seen as missing information. Furthermore, several variables in our database give the impression of incompleteness owing to their low frequency[82]. The reason why some information is not included in the CR can be explained by the lack of direct examination of the artwork (in the case of unknown location or geographical distance) and/or incomplete documentation on which the experts based their assessment. For example, CR indicates that painting No. 199 was signed and dated on the reverse side.

> CR No. 199. The funeral, approx. 1896
> Oil on cardboard, 26 × 24,5 cm
> Signature and date stamp on the back: «F. Vallotton. 95»
> Private collection (1976)

For another painting (e.g. No. 200) from the same period, information concerning the back of the painting was not mentioned (Ducrey & Poletti, 2005, p. 110, vol. 2).

> CR No. 200. Perros in Ploubazlanec, approx. 1896 – 1897

[82] This issue relates to several categories in the *'Market'* group, such as 'restored', 'relining', 'sale', 'sold', 'estimated price', 'sale price', 'date of sale', and all categories in the *'Back of the painting'* group.

Oil on cardboard, 25,5 × 35 cm
Signed and dated lower right: « F. Vallotton. 95 »
Private collection, Switzerland

Similarly, some characteristics of forgeries are not included in the archive because owners may choose to provide only partial data on their work. To set these data limits, we have coded the different characteristics with dummy variables, which are all coded '0' when the corresponding characteristic is not indicated in the sources. For example, if the size of a painting is not noted in the CR or the archive files, the variables 'large', 'medium', and 'small' all take the value 0, even though theoretically one of them could take a value of 1. This approach, called 'zero imputation', corresponds to a particular way of dealing with MD in which unknown information is treated as the absence of a characteristic (McKnight, McKnight, Sidani, & Figueredo, 2007). Furthermore, we coded the missing characteristic as '1' when this characteristic was implicitly identified in the source as not belonging to the painting. For instance, an undated painting was coded as '1' because there is a clear indication in the CR that the date is absent on the painting's medium. Therefore, this technique allows us to distinguish between missing characteristics and exploit them as variables.

Ideally, a dataset should be as complete as possible. Indeed, these sets should include all known original works and forgeries. However, it is possible that some forgeries have not been recorded in our database because experts have not yet detected them. Similarly, some original works could still await identification and inclusion in the CR. Such MD remains outside the scope of our sources and therefore has no direct impact on the study results.

6 GETTING TO GRIPS WITH THE DATABASE: PRELIMINARY INSIGHTS AND TECHNIQUES FOR ANALYSIS

This chapter presents a preliminary analysis made up of descriptive analysis and data mining. We outline these techniques, test their feasibility and determine the most appropriate approach for the operational phase. All numerical examinations will be carried out with the statistical software 'R'.

6.1 Descriptive analysis

The observation phase included a descriptive exploration. This analysis provides a general idea of data distribution, helping to identify associations between variables and prepare for further statistical analyses.

At this point, the distribution of each variable is presented in tables as a function of its degree of authenticity (i.e. originals, fakes/forgeries, or hypothetical). To simplify interpretation, we combined the categories of 'fake' and 'doubtful' into one new variable that can be considered to reflect fakes. The tables show the results with the explained variables in the *columns* and the explanatory variables in the *rows*. These tables include the values of each explanatory variable.

However, some of the variables included in Tables 3, 10, 13, and 16 are not mutually exclusive. This implies that an explanatory variable can simultaneously belong to different categories. For example, a painting can have a number, letter, seal, sticker, handwriting, date, or image on the canvas or frame. Accordingly, Table 16 presents the distribution of all of them. Nevertheless, one painting may have only one inscription on the back, another may have several inscriptions, or there may be no information on the back. Thus, in tables in which the variables are not mutually exclusive, the totals per category differ from the general total of the explanatory variables in brackets.

For the sake of clarity, our commentary focuses on the variables 'original' and 'fake/forgery' and only considers the variable 'hypothetical' in exceptional cases (i.e. in

the case of an extremely high frequency). The data are presented in tables according to the groups described in Section 5.3.3 and in Appendix II.

6.1.1 Title and Years groups

Table 2 and Table 3 show that the fakes have no titles or dates much more frequently than the originals do. Specifically, 50.7% of the fakes have no title and 59.3% have no date on the canvas, whereas only 7.6% of the originals have no date, and none are untitled. We obtained information on the sale price and year of appearance only for fakes. However, information on the current country of location is provided for both originals and fakes.

Table 2: Distribution within the Title group

Category	**Original 1,704**		**Fake 140**		**Hypothetic 70**		**Total 1,914**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
With title	1,704	100.0	69	49.3	17	23.3	1,790	93.5
Without title	0	0.0	71	50.7	53	75.7	124	6.5

Table 3: Distribution within the Years group

Category	**Original (1,704)**		**Fake (140)**		**Hypothetic (70)**		**Total (1,914)**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
No date	130	7.6	83	59.3	68	97.1	281	14.7
Without Year 1	0	0.0	81	57.9	67	95.7	148	7.7
Without Year 2	1,691	99.2	139	99.3	69	98.6	1,899	99.2
Illegible date	0	0.0	3	2.1	0	0.0	3	0.2

***** Table 3 includes variables that are not mutually exclusive; the totals per category differ from the total value of the explanatory variables in brackets.

6.1.2 Market group

Table 4 shows that the majority of both the originals and the fakes are found in four countries: France, Germany, the United States, and Switzerland. However, roughly

the same percentages of originals are located in the United States (2.2%) and Germany (2.4%), whereas Germany has a much higher percentage of the fakes (13.6%) than the United States (5.7%) has. The largest percentage of the originals is found in Switzerland (53.6%), and the percentage of the fakes there is much lower (17.1%). The majority of the fakes are in France (29.3%), whereas France has only 10% of the originals. The frequencies for the other countries are insignificant for both categories.

Table 4: Distribution according to present location

Category	**Original 1,704**		**Fake 140**		**Hypothetic 70**		**Total 1,914**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
France	172	10.1	41	29.3	29	41.4	242	12.6
USA	37	2.2	8	5.7	3	4.3	48	2.5
Italy	4	0.2	4	2.9	3	4.3	11	0.6
Germany	40	2.4	19	13.6	2	2.9	61	3.2
Belgium	0	0.0	1	0.7	1	1.4	2	0.1
Switzerland	913	53.6	24	17.1	10	14.3	947	49.5
Bulgaria	2	0.1	0	0.0	0	0.0	2	0.1
Netherlands	4	0.2	4	2.9	0	0.0	8	0.4
Spain	0	0.0	2	1.4	1	1.4	3	0.2
Finland	1	0.1	0	0.0	0	0.0	1	0.1
Austria	0	0.0	1	0.7	1	1.4	2	0.1
Monaco	2	0.1	1	0.7	0	0.0	3	0.2
England	5	0.3	3	2.1	0	0.0	8	0.4
Brazil	0	0.1	1	0.0	0	0.0	1	0.1
South Africa	1	0.0	0	0.7	0	0.0	1	0.1
Canada	9	0.5	0	0.0	0	00	9	0.5
Japan	2	0.1	0	0.0	0	0.0	2	0.1
Algeria	3	0.2	0	0.0	0	0.0	3	0.2
Russia	8	0.5	0	0.0	0	0.0	8	0.4
No information in sources	502	29.4	31	22.1	19	27.1	552	28.8

Table 5 shows that among the counterfeits for which sales were attempted, only half were sold; among the originals, the number of paintings sold equals the number for which sales were attempted. The variable 'year of appearance' refers to the first year

of application for identification and concerns the categories of false and hypothetical paintings. The CR was published in 2005, meaning that after that date, works that are included in the CR are presumed to be authentic. To compare the distributions of the fake and hypothetical paintings before and after the CR's publication, we reorganise the variable 'year of appearance' into two identical periods, that is, the 13 years following the publication of the CR (2005 to 2018) and the 13 years preceding the creation of the CR (1991 to 2004).

Table 5: Distribution according to selling status

Category	**Original 1,704**		**Fake 140**		**Hypothetical 70**		**Total 1,914**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Sale	291	17.1	38	27.1	4	5.7	333	17.4
Sold	295	17.3	13	9.3	3	4.3	311	16.2
No information in sources	1,118	65.6	89	63.6	63	90.0	1,270	66.4

Table 6 also includes data for a 21-year period (1960–1991) that includes the period from the date of the first registration of a fake in the archives (1960) to the year in which the period becomes informative (1991). The last column in the table refers to false and hypothetical paintings with no indication of the year of appearance in the archives. This distribution comparison does not show much difference between the two comparable periods, before and after 2005. In other words, the distributions were quite similar before and after the publication of the CR (30.7% vs. 32.9% before 2005 and 33.6% vs. 38.6% after 2005).

Table 6: Distribution according to the year of first appearance

Category	**Original 1,704**		**Fake 140**		**Hypothetical 70**		**Total 1,914**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
1960–1991	0	0.0	37	26.4	14	20.0	51	2.7
1991–2004	0	0.0	43	30.7	23	32.9	66	3.4

2005–2018	0	0.0	47	33.6	27	38.6	74	3.9
No information in sources	1,704	100.0	13	9.3	6	8.6	1,723	90.0

Table 7 shows that very few of the originals and fake/forgeries have dates of sale indicated, whereas the data sources for many of the paintings do not contain such information.

Table 7: Distribution according to the date of sale

Category	**Original (1,704)**		**Fake (140)**		**Hypothetical (70)**		**Total (1,914)**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
1920–1939	3	0.2	0	0.0	0	0.0	3	0.2
1940–1959	10	0.6	0	0.0	1	1.4	11	0.6
1960–1979	42	2.5	0	0.0	0	0.0	42	2.2
1980–1999	139	8.2	4	2.9	1	1.4	144	7.5
2000–2019	98	5.9	10	7.1	2	2.9	110	5.7
No information in sources	1,412	82.9	126	90.0	66	94.3	1,604	83.8

The variable ‘date of sale’ was recoded into several periods of comparable length as well as one group containing paintings with no information about their selling status. The variables ‘sold price’ and ‘estimated price’ are recoded into five groups; four of them refer to equally sized intervals, with the same intervals for each variable, and one group includes paintings with no information on their selling or estimated prices.

It can be assumed that all fake forgeries are listed. However, Tables 8 and 9 show that very little data were obtained for sold or estimated prices.

Table 8: Distribution according to the sold price (CHF)

Category	**Original 1,704**		**Fake 140**		**Hypothetical 70**		**Total 1,914**	
	al	**%**	***n***	**%**	***n***	**%**	***n***	**%**
1–5,000	0	0.0	6	4.3	2	2.9	8	0.4
5,001–10,000	0	0.0	1	0.7	0	0.0	1	0.1
10,001–15,000	0	0.0	0	0.0	0	0.0	0	0.0
15,001–20,000	0	0.0	1	0.7	0	0.0	1	0.1
No information in sources	1,704	100	132	94.3	68	91.7	1,904	99.4

Table 9: Distribution according to the estimated price (CHF)

Category	**Original 1,704**		**Fake 140**		**Hypothetical 70**		**Total 1,914**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
1–5,000	0	0.0	9	6.7	0	0.0	9	0.5
5,001–10,000	0	0.0	3	2.2	0	0.0	3	0.2
10,001–15,000	0	0.0	1	0.7	0	0.0	1	0.1
15,001–20,000	0	0.0	1	0.7	0	0.0	1	0.1
No information in sources	1,704	100	121	89.6	70	100	1,895	99.0

6.1.3 Signature and Place of signature groups

Table 10 shows that fakes and hypotheticals frequently have no signature and are more often monogrammed than originals are. In particular, 13.6% of the fakes (88.6% of the hypotheticals) are not signed, and 14.3% of the fakes are monogrammed, whereas only 1.5% of the originals are not signed and 0.6% of the originals are monogrammed. The fakes have atypical signatures four times as often (5.5% vs. 1.2%)

as the originals do and have signatures located in an atypical place almost twice as often as the originals do (2.1% vs. 1.6%).

Table 10: Distribution according to the signature or monogram

Category	**Original (1,704)**		**Fake (140)**		**Hypothetical (70)**		**Total (1,914)**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Without signature	25	1.5	19	13.6	62	88.6	106	5.5
Monogram	10	0.6	20	14.3	0	0.0	30	1.6
With signature	1,566	91.9	92	65.7	1	1.4	1,659	86.7
Trace	1	0.1	5	3.6	0	0.0	6	0.3
Atypical signature	21	1.2	8	5.7	3	4.3	32	1.7
Atypical location signature	28	1.6	3	2.1	0.0	0.0	31	1.6

* Table 10 shows the variables that are not mutually exclusive.

Table 11: Distribution according to signature location

Category	**Original 1,704**		**Fake 140**		**Hypothetical 70**		**Total 1,914**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Lower right	947	55.6	57	40.7	1	1.4	1,005	52.5
Lowe left	354	20.8	42	30.0	2	2.9	398	20.8
Upper right	174	10.2	9	6.4	0	0.0	183	9.6
Upper left	66	3.9	2	1.4	0	0.0	68	3.6
Bottom centre	0	0.0	2	1.5	0	0.0	2	0.1
Centre right	2	0.1	0	0.0	0	0.0	2	0.1
Centre left	2	0.1	2	1.4	0	0.0	4	0.2
Top Centre	0	0.0	0	0.0	0	0.0	0	0.0

No information in sources	159	9.3	26	18.6	67	95.7	252	13.2

Table 11 shows that most of the originals and fakes have signatures or monograms in the lower right corner. The signatures or monograms of the originals are located in the lower right corner slightly more often (55.6% vs. 40.7%), whereas the signatures or monograms of the fakes are found in the lower left corner more frequently than for the originals (30% vs. 20.8%). The signature is not located at the top centre for any of the three categories.

6.1.4 Technical characteristics group

Table 12 shows that the fakes are usually smaller than the originals. In particular, the fakes are twice as likely to be small (22.9% vs. 11%) and half as likely to be large than are the originals (61.9% vs. 30.7%).

Table 12: Distribution according to dimensions

Category	**Original 1,704**		**Fake 140**		**Hypothetical 70**		**Total 1,914**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Large	1,055	61.9	43	30.7	14	20.0	112	58.1
Medium	359	21.1	52	37.1	24	34.3	435	22.7
Small	203	11.9	32	22.9	10	14.3	245	12.8
No information in sources	87	5.1	13	9.3	22	31.4	122	6.4

Table 13 shows that landscapes are the most frequent subjects of both the originals and the fakes (40.6% vs. 31.4%, respectively). The fakes are more likely to be still lifes than the originals are (21.4% vs. 14.9%).

Table 14 shows that the most common support for both the fakes and the originals is canvas. However, the fakes are created on paper nine times as often as the originals are. Specifically, 9.3% of the fakes and only 0.4% of the originals are made on paper.

Table 13: Distribution according to subject

Category	**Original (1,704)**		**Fake (140)**		**Hypothetical (70)**		**Total (1,914)**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Portrait	251	14.7	25	17.9	14	20	290	15.2
Portrait with identified person	164	9.6	4	2.9	5	7.1	173	9
Still-life	254	14.9	30	21.4	2	2.9	286	14.9
Landscape	692	40.6	44	31.4	17	24.3	753	39.3
Interiors w	71	4.2	5	3.6	10	14.3	87	4.5
Decorations	2	0.1	0	0.0	0	0.0	2	0.1
Nudes	324	19.0	24	17.1	24	34.3	372	19.4
Copy	7	0.4	1	0.7	1	1.4	9	0.5
Mythological	13	0.8	0	0.0	0	0.0	13	0.7
Various	58	3.4	11	8.9	3	4.3	72	3.8

* Table 13 shows variables that are not mutually exclusive.

Table 14: Distribution according to the type of support

Category	**Original 1,704**		**Fake 140**		**Hypothetical 70**		**Total 1,914**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Canvas	1,267	74.4	67	47.9	33	47.1	1,367	71.4
Cardboard	279	16.4	23	16.4	8	11.4	310	16.2
Wood	81	4.8	10	7.1	5	7.1	96	5.0
Paper	6	0.4	13	9.3	3	4.3	22	1.2
No information in sources	71	4.2	27	19.3	21	30	119	6.2

Table 15 shows that both fakes and originals are most often made in oil. In particular, 91.6% of the originals and 75% of the fakes are made in oil. However, only 5.3% of the originals are made with other techniques (i.e. pastel and tempera), relative to 25% of the fakes. Regarding the fakes, 14.3% are made using techniques for which experts have no information, and about 11% are made using other techniques. Pastel and gouache are more common among the fakes, but Vallotton used tempera (3.4%) more often than the fakers did (0.7%)

Table 15: Distribution according to technique

Category	**Original 1,704**		**Fake 140**		**Hypothetical 70**		**Total 1,914**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Oil	1,561	91.6	105	75.0	49	70	1,715	89.6
Pastel	13	0.8	7	5.0	2	2.9	22	1.1
Tempera	58	3.4	1	0.7	1	1.4	60	3.1
Gouache	8	0.5	8	5.7	1	1.4	17	0.9
Unknown technique	52	3.1	19	13.5	15	21.4	86	4.5
No information in sources	12	0.7	0	0.0	2	2.8	14	0.7

6.1.5 Back of the painting group

Little information is available in the CR of the canvases. For this reason, comparisons were made between the fakes and originals. Consequently, Table 16 shows that a signature or initials is more common; numbers, letters, and stickers appear more rarely; and a seal appears only once on the back of an original. Writing on a frame or canvas was more common than writing in other places among the fakes. None of the three categories has writing on the back that contains errors.

Table 16: Distribution according to the painting back

Category	Original 1,704		Fake 140		Hypothetical 70		Total 1,914	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Number/Letter	7	0.4	5	3.6	2	2.9	14	0.7
Sticker	12	0.7	10	7.1	2	2.9	24	1.3
Signature/initial	20	1.2	9	6.4	0	0.0	29	1.5
Seal	1	0.0	2	1.4	4	5.7	7	0.4
Written on the canvas	3	0.2	12	8.6	6	8.6	21	1.1
With errors	0	0.0	0	0.0	0	0.0	0	0.0
Date	12	0.7	1	0.7	0	0.0	13	0.7
Written on the frame	14	0.8	4	2.9	2	2.7	20	1.0
Under the other painting	10	0.6	1	0.7	0	0.0	11	0.6

* Table 16 includes variables that are not mutually exclusive.

6.1.6 Historical documentation group

For most of the fakes, no information is provided about their provenance (88%). In contrast, most originals had a complete history of the property or location (Table 17). Only 3.5% of the originals have no provenance, and 3.2% of them have provenance only by the last owner. Approximately 3% of the fakes were restored, and 1.4% were relined, whereas these frequencies are much lower for the originals (0.3% and 0.1%, respectively). Information on the actual locations of the originals, whether they were stolen or destroyed, and whether they were absent from the commonplace book is found only in the first data source, which describes the originals.

Few originals are missing, mainly because their locations were unknown (13.3%). Vallotton's commonplace book failed to note 4.3% of the paintings, and only a small percentage were stolen or destroyed (0.4% and 0.5%, respectively).

Table 17: Distribution according to provenance

Category	**Original 1,704**		**Fake 140**		**Hypothetic 70**		**Total 1,914**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Unknown location	226	13.3	0	0.0	0	0.0	226	11.8
Stolen	7	0.4	0	0.0	0	0.0	7	0.4
Destroyed	9	0.5	0	0.0	0	0.0	9	0.5
Absence in commonplace	83	4.9	0	0.0	0	0.0	83	4.3
Without provenance	59	3.5	123	87.9	63	90	245	12.8
Recent provenance	55	3.2	14	10	5	7.1	74	3.9
First provenance	92	5.4	0	0.0	0	0.0	92	4.8
With provenance	1,498	87.9	3	2.1	2	2.9	1,503	78.5

6.1.7 Period group

Table 18 shows that the fakes represent the period '1916–1925 landscapes and sunset' (12.9%) more often than they represent other periods. The originals were frequently made during the periods '1901–1910' (14.4%) and '1890–1900 wood' (15%). There are no fakes or hypotheticals in the category of '1917 war'.

Table 18: Distribution according to period

Category	**Original 1,704**		**Fake 140**		**Hypothetic 70**		**Total 1,914**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Before 1882	2	0.1	1	0.7	1	1.4	4	0.2
1882–1889 youth	96	5.6	6	4.3	0	0.0	102	5.3
1890–1900 wood	245	14.4	9	6.4	0	0.0	254	13.3
1901–1910	256	15	7	5	0	0.0	263	13.7
1901–1905 landscape	126	7.4	7	5	1	1.4	134	7.0
1906–1909 nude	98	5.8	2	1.4	0	0.0	100	5.2
1910–1914 sunset	32	1.9	6	4.3	0	0.0	38	2.0
1915–1916	1	0.1	3	2.1	0	0.0	4	0.2
1917 war	8	0.5	0	0.0	0	0.0	8	0.4
1916–1925 sunset, landscape	2	0.1	18	12.9	1	1.4	21	1.1
No information in sources	838	49.2	81	57.9	67	95.7	986	51.5

6.1.8 Concluding remarks

A summary of the descriptive analysis provides guidance for understanding what lies behind the statistical differences as follows:

∞ Nearly half of the fakes have no titles, and slightly more than half of them are undated on the canvas. In contrast, very few originals are undated and all of them have titles.

∞ Most of the fakes are in France, and the majority of the originals are in Switzerland.

∞ The fakes are much more often unsigned and much more often monogrammed than the originals.

∞ The fakes have atypical signatures four times more frequently than the originals do.

∞ The signatures/monograms are most frequently located in the lower right corner for both the originals and the fakes, but the signatures/monograms of the fakes are more often located in the lower left corner than among the originals. No works in any of the three categories have signatures at the top centre.

∞ The fakes are usually smaller than the originals. The most common size of the fakes is 'medium', whereas that of the originals is 'large'.

∞ Landscapes are the most frequent subjects of both originals and fakes, although they are slightly more common among the originals. The fakes are about one and a half times more likely to be still-lifes than the originals.

∞ The most common form of support for both originals and fakes is canvas. However, fakes were created nine times more often on paper than the originals.

∞ Both originals and fakes are most often created in oil, but far fewer originals than fakes are made using other techniques (pastel and tempera). The same percentages of fakes and originals were made using techniques for which the experts had no information. Pastels and gouaches were more common among the fakes, whereas the artist used tempera more often than the fakers did.

∞ The backs of originals more commonly have the signature or initials of the artist and more rarely have numbers, letters, stickers, and seals. Writing on the frame is more common than writing on the backs of the fakes. No paintings in either category have writing on the back that contains errors.

∞ Most fakes have no information about their provenance, whereas most originals do.

∞ The fakes represent the period '1916–1925 landscapes and sunset' more often than they represent the other periods. Originals were frequently made during the periods '1901–1910' and '1890–1900 wood'. There are no fakes or hypotheticals in category '1917 war'.

Overall, the observation phase revealed a definite difference between the characteristics of the original and fake works of art. The next stage of our analysis was to determine a way to quantify these distinctions.

6.2 Creating an experimental set for testing statistical methods

A limited set is necessary when a database contains many variables (Devijver & Kittler, 1982). Guyon and Elisseeff (2003) noted that reducing the number of variables can help improve the indicators' prediction performance and offer a better understanding of the underlying process generating the data. Accordingly, the goal is to provide a set that includes the most representative variables for each group of characteristics specified in the tables in Section 6.1.

This example of selecting the most representative variable from the group 'Historical documentation' may illustrate how the experimental set is formed in detail. Table 17 presents the percentages corresponding to the characteristics of the 'Historical documentation' group. The fragment of Table 17 for the distribution of the category fake/forgery with corresponding rates to each variable of the group:

Without provenance	123	87.9%
Recent provenance	14	10%
First provenance	0	0.0%
With provenance	3	2.1%

By comparing these rates, we choose the most frequent characteristic – 'without provenance' which has the highest percentage (87.9%). Hence, at the end of the selection process, all the distribution tables were processed using the same technique.

Overall, the dataset of 80 features was reduced to a set of 15 variables:

Without provenance
Untitled
Without date on painting
Without signature or monogram
Appearance in Switzerland
Low right
Oil
Canvas
Still-life
Medium
Signature, monogram (on back)
1916–1925 landscape
Restored/recast
Sold
Assigned without signature

The next section is designed to test the Principal Component Analysis (PCA), which can help identify the underlying relationships across a set of variables and replace the raw data with an easier-to-use dataset by removing repetitive information. The results are presented as coefficients and plots.

6.2.1 Principal component analysis (PCA) of the experimental set

a) PCA with all significant components

The experimental set was initially visualised using PCA to obtain a global view of the multivariate data. As previously shown, this technique aims to reduce the number of variables (or characteristics) to PCs. The interpretation mechanism was based on determining the variables having the highest correlations with each component (Clausen, 1998). When only one variable is strongly associated with a component, this variable represents very different information from all other variables. First, several useful components must be selected. There are different rules, but the most commonly used method is that of Kaiser (1960), who retains factors with proper values (eigenvalues) greater than 1 (Braeken & van Assen, 2017). The associations between

the retained components and variables are then examined. Statisticians have proposed that the correlation size should be greater than or equal to 0.3 (Larson-Hall, 2013; Kent, 2015).

Table 19 presents the coefficients that indicate a strong association between the six retained components and original variables.

Table 19: Correlation matrix (Experimental set)

Variables (Experimental set)	**Components**					
	1	**2**	**3**	**4**	**5**	**6**
Without provenance	0.8					
Untitled	0.8					
Without date on painting	0.7					
Without signature or monogram	0.7					
Appearance in Switzerland	−0.4					
Low right	−0.4					
Oil		0.7				
Canvas		0.6				
Still-life			0.7			
Medium			0.5			
1916–1925 landscape				−0.4		
Restored/recast					0.6	
Sold						0.7
Signature, monogram (back)						0.5
Assigned without signature						0.3
Percentage of variance reproduced	21%	10%	8%	8%	8%	7%

The PCs can be interpreted by answering the question of how the variables that are highly correlated with the same component share information. Every PC was interpreted according to its correlations with the following variables:

– PC1 is strongly correlated with six variables: 'without provenance', 'untitled', 'without date on painting', 'without signature or monogram', 'appearance in Switzerland', and 'low right'. Therefore, these variables contain common information. When *r* is positive, the variables and the PC vary in the same direction. Thus, the value of the PC1 increases as 'without provenance', 'untitled', 'without date on painting', and 'without signature or monogram'

switch from 0 to 1. 'Appearance in Switzerland' and 'low right' are negatively associated with PC1 (i.e. *r* takes a negative value). Consequently, the likelihood of the variables 'without provenance', 'untitled', 'without date on painting', and 'without signature or monogram' tend to increase simultaneously, when the likelihood of the variables 'appearance in Switzerland' and 'low right' decreases. In other words, it follows that a painting without provenance is more often untitled and lacks a date/signature but is very rarely signed in the lower-right corner or found in Switzerland.

– PC2 is positively correlated with two variables, 'canvas' and 'oil'. Thus, canvas paintings are often made in oil.

– PC3 is positively correlated with two variables, 'still-life' and 'medium'. It follows that paintings with still-life subjects are often medium-sized.

– PC4 mainly represents 'period 1916–1925' and no other variables.

– PC5 mainly represents 'restored/recast'.

– PC6 is positively correlated with 'sold', 'signature initials on back', and 'assigned without signature'.

b) PCA with two components

This approach was used to summarise the data in a PC score plot. The plot can reveal patterns in the data, such as clusters that may not be apparent in the raw data (Koch, 2014). Often, the first two PCs exhibit the main structure of the data; therefore, it is advisable to consider a two-dimensional score plot. Nevertheless, it should be noted that an interesting structure (i.e. a split into clusters) may not always appear (Koch, 2014).

We designed a score plot that showed the projection of the variable set into the span of the first two components. Figure 2 plots PC1 on the *x*-axis and PC2 on the *y*-axis.

The convergence of the variables into clusters is shown in Figure 2. To determine the clusters, we must locate the position of each variable (i.e. blue dots) in the graph. Variables within clusters vary together (Jolliffe, 2002).

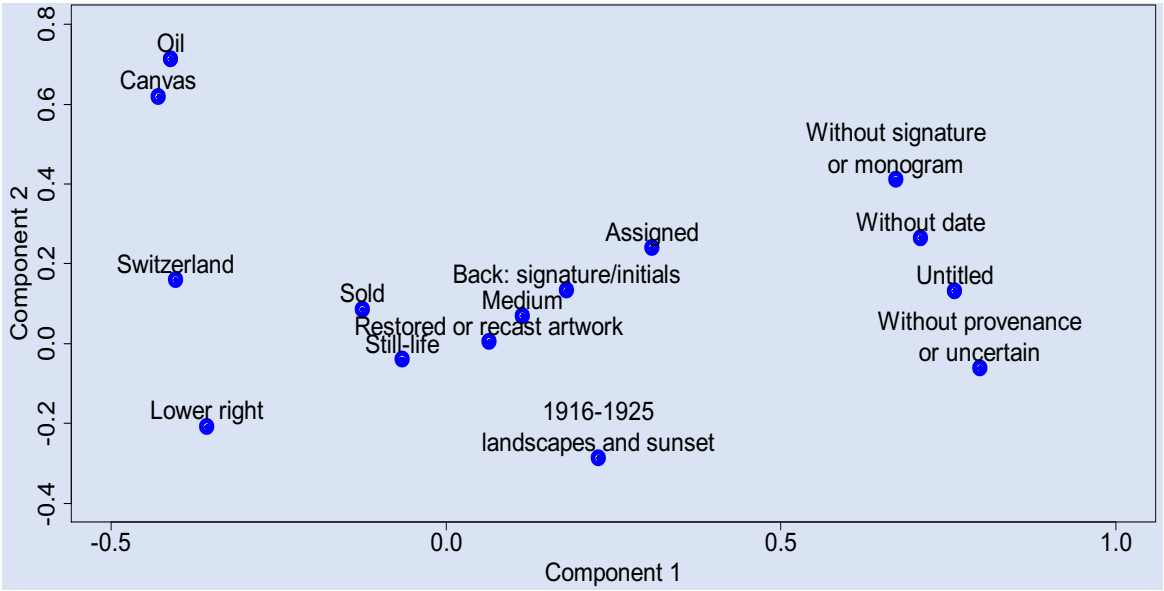


Figure 2: Two-dimensional score plot
(Experimental set)

Interpretation

The first component reproduces 21% of the original information, and the second 10%. All variables in the experimental set are represented in the space of the first two components. The main findings in Table 19 are shown in Figure 2. Regarding the first component (PC1), variables 'without signature and monogram', 'without date on painting', 'untitled', and 'without provenance' on the right are grouped together and are opposed to 'appearance in Switzerland', 'low right', 'oil', and 'canvas' on the left. Regarding the second component (PC2), the variables 'oil' and 'canvas' are grouped together in the top left of the figure and 'Switzerland' and 'low right' in the lower left. First, we observe that variables in the right group share a common meaning, that is, the absence of certain information (i.e. unknown date, unknown title, no signature, or unknown provenance). This signifies that paintings without provenance are more likely to be undated, untitled, or unsigned. The variables in the lower-left group reveal that a painting signed in the lower-right corner is more likely to be found in Switzerland. The upper-left group indicates that a painting on canvas is more likely to be in oil. Second, the three above-defined groups occupy three antagonist positions in Figure 2, meaning that characteristics defined by one of the three groups of variables are very unlikely to be found in the same artwork as the characteristics defined by both other groups. For example, works of art without title and provenance

are not likely to be signed in the lower right-hand corner, and works of art in oil on canvas are not likely to be undated. Finally, variables closer to the origin of the figure, that is, the position of the coordinates (0,0), are more difficult to analyse because their positions can only indicate that the first two components are not sufficient to represent them correctly.

Overall, the principle of PCA is to reduce the number of variables by creating composite variables (the components) that group together information that is included in a similar way in several of the original variables. In this way, the relationships between the variables become clearer and the user can work with a data set that retains most of the original information but with a reduced number of variables. However, the components are often less interpretable than the original variables. Indeed, the PCA results show that the variables extracted from the different subgroups of the full dataset have some similarities. These results also imply that an even smaller set of variables may be sufficient to design an explanatory model for identifying fake artwork. We expect these findings to contribute to our understanding of the subsequent analyses. However, as can be seen from the percentages of variance reproduced by each component shown in Table 19, the PCA analysis indicates that it is not really possible to concentrate most of the information from the original variables on just 2 or 3 components. As a result, an interpretation based on the first two components alone, as shown in Figure 2, is insufficient. On the other hand, and this is the purpose of the PCA analysis, Table 19 clearly shows that not all the selected variables are perfectly independent, and that close relationships exist between some of them, notably on the first component. Thus, PCA analysis as used in this thesis should be seen as an aid to understanding the data, rather than an analysis tool in its own right.

Despite the precise results of the PCA, their interpretation did not specify whether such clusters belonged to the original artworks or forgeries. To further understand relationships within the data and achieve our objective of identifying the most explanatory variables and the probabilities of authenticity, we will apply the CART model. The CART approach is a flexible and widely used algorithm that constructs a decision tree by iteratively partitioning the data into subsets that are as homogeneous

as possible in relation to the dependent variable. The purpose of the CART model is to predict the value of a dependent variable and classify data instances into distinct classes based on a set of input features (Breiman et al., 1984). In our case, we use CART to predict whether an artwork is authentic or fake, gaining insights into how various characteristics contribute to these predictions and classifications. Moreover, the CART model's ability to handle collinearity, as explained in Chapter 4.2, is crucial, given that the results of the PCA analysis indicate close relationships between some variables.

The following section presents the results of testing this approach. Similar to the approach taken for PCA, each step of the CART analysis will be explained using an experimental dataset.

6.2.2 Classification tree analysis (CART) of the experimental set

a) Original vs. Fake/forgery

1) General principle

Figure 3 shows the classification tree for the *Original vs. Fake/forgery* approach. We analysed a sample of 1,831 works (1,704 originals and 127 fakes/forgeries) using the 15 variables included in the experimental set. This model is built to predict which of these works have a high probability of being fake/forgery (or original) works, based on the most significant of the 15 explanatory variables. The classification tree algorithm segregates all of the works based on the values of these 15 variables.

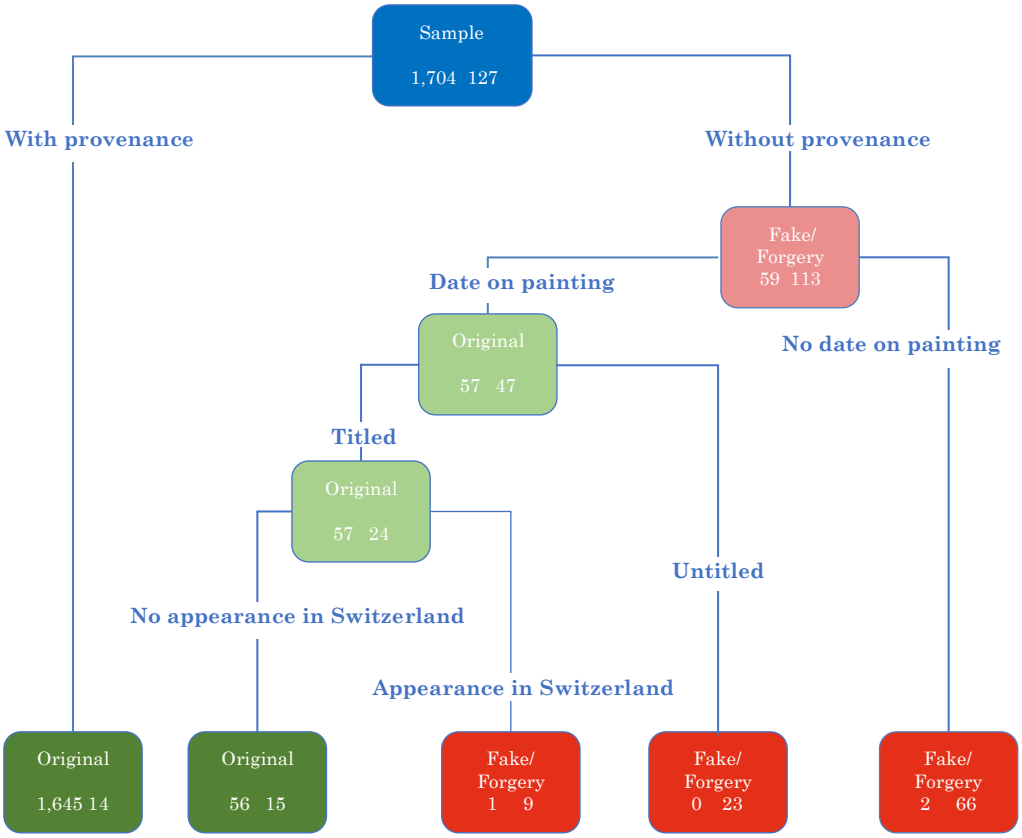


Figure 3: CART model: Original vs. Fake/forgery (Experimental set)

In each step, a parent node is optimally split into two children using one of the 15 available variables. After each split, the homogeneity of the resulting sub-nodes is increased. In other words, the purity of each node increases with respect to the target variable. The decision tree divides the nodes according to all the available variables and then chooses the most homogeneous sub-nodes (Analytics Vidhya, 2016).

2) Algorithm operation

To better explain the tree elements, we constructed Figure 4, which shows one of the segments of the classification tree shown in Figure 3.

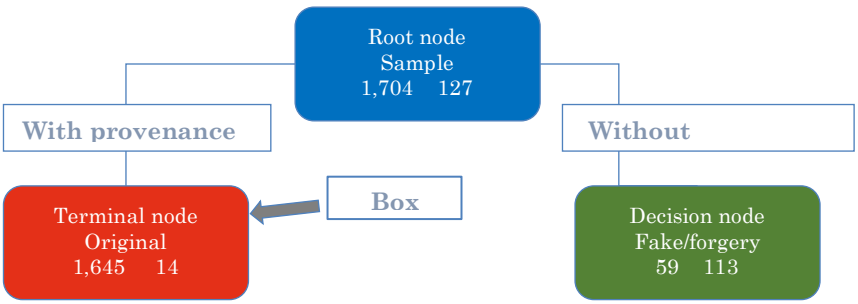


Figure 4: Example tree segment (from Figure 3)

In this example, the reader can observe that the root node is divided into a terminal node and a decision node. The algorithm used by the classification tree selects the variable 'without provenance' as the potential explanatory variable with the greatest power to discriminate between original and fake/forgery works. First, the algorithm checks whether the artwork has a provenance. Then, if the variable 'without provenance' has a value below 0.5 (i.e. 0, as this variable is dichotomous), the work is classified as original; if the variable takes a value greater than or equal to 0.5 (or equal to 1 for dichotomous variables), the work is classified as a fake/forgery.

3) Box reading operation

All terminal and decision nodes have labels. For example, the terminal node reads 'Original'. This label means that the majority of artwork included in this box is 'original', meaning that without additional information, the best guess is that a work of art in this box is original. The decision node is labelled 'Fake/forgery', which means that artwork included in this box belongs mainly to the 'fakes/forgeries' category. The categories of the potential explanatory variables are noted on each line. For example, 'with provenance' is written on the left-hand line going down to the terminal node, and 'without provenance' is written on the right-hand line going to the decision node.

Each box contains two numbers. For example, the root node contains the numbers '1,704' and '127'. These numbers indicate that our sample comprises 1,704 original

works and 127 fakes/forgeries. The numbers '1,645' and '14' are written on the terminal node, meaning that of the 1,704 originals at the root node, 1,645 works with provenance are assigned to the terminal node. Of the 127 fakes/forgeries at the root node, 14 fakes/forgeries with provenance are (wrongly) classified as originals at the terminal node. The decision node box can be read similarly: 59 original works without a provenance were classified as fakes/forgeries, and 113 fakes/forgeries without a provenance were classified as fakes/forgeries. It is important to stress that a classification tree is a probabilistic approach that searches for the most likely situation rather than an exact situation. Thus, not all of the works assigned the label 'original' are in fact originals (and not all of the works assigned the label 'fake/forgery' are fakes/forgeries). Indeed, these works can be considered originals (or fakes/forgeries) with a certain probability, as explained below.

4) Rules and interpretation of the classification tree

In our example with the experimental set (Figure 3), the classification algorithm produces a tree with five final nodes, two of which are labelled 'original' and three of which are labelled 'fake/forgery'. We proceed through the tree step by step until the final nodes are reached, where the final prediction is made. These nodes comprise all the variables with the greatest discriminatory power recognised in each step of the tree. The final nodes represent the niches.

Using a simple calculation, we can determine the proportion of our data classified correctly at each node and the proportion incorrectly classified. In this calculation, *O* denotes the original artwork (the left number in each box) and *F* denotes fakes/forgeries (the right number).

According to this calculation, 99.8% of the original works are classified correctly and 0.17% are classified erroneously.

Artwork	**Originals**

Correct classification	$\frac{1{,}645\ (O\ 1st\ Terminal\ node)^{83}+56\ (O\ 2nd\ Terminal\ node)}{1{,}704\ (O\ Root\ node)}$
Erroneous classification	$\frac{1\ (O\ 3d\ Teminal\ node)+0\ (O\ 4th\ Terminal\ node)+2(O\ 5fth\ Terminal\ node)}{1{,}704\ (O\ Root\ node)}$

According to these calculations, 77.2% of fakes/forgeries are classified correctly and 22.8% are classified erroneously. Hereafter, we use the term 'sensitivity' to indicate the percentages of correctly classified fakes/forgeries (or doubtful works) and the term 'specificity' to indicate the percentage of correctly classified originals.

Artwork	**Fakes/Forgeries**
Correct classification	$\frac{9\ (F\ 3d\ Terminal\ node)+23\ (F\ 4th\ Terminal\ node)+66\ (F\ 5th\ Terminal\ node)}{127\ (F\ Root\ node)}$
Erroneous classification	$\frac{14\ (F\ 1st\ Terminal\ node)+15\ (F\ 2nd\ Terminal\ node)}{127\ (F\ Root\ node)}$

In addition, to better understand the rules of each node, we need to know the probabilities that correctly classified works are originals or fakes/forgeries. 'Probability is commonly estimated by the ratio of the number of successful cases to the total number of possible cases, derived mathematically using known properties of the distribution of events, or estimated logically by inferential or inductive reasoning' (Oxford English Dictionary, n.d.).

Using the logic of the box configuration, we can estimate the probabilities of each terminal node, which are presented as the following ratios[84]:

Originals	$\boldsymbol{\frac{O}{O+F}}$

[83] The text in parentheses is meant to help to explain the meanings of the numbers. In particular, 1,645 is the number shown on the left side of the box corresponding to the first terminal node of originals. The abbreviations 'O' and 'F' are defined above.

[84] The percentages for each node are explained in a table of rules and interpretations below.

First node	1,645/(1,645+14) × 100% = 99%
Second node	56/(56+15) × 100% ≈ 79%

Fakes/Forgeries	$\frac{F}{O+F}$
Third node	9/(1+9) × 100% = 90%
Fourth node	23/(0+23) × 100% = 100%
Fifth node	66/(2+66) × 100% ≈ 97%

The classification tree algorithm is easily interpretable because it allows access to all the splits for each variable. We can clearly see how the rules are defined from the root node to the terminal node (Moisen, 2008). We simply need to define follow-up rules based on the explanatory variables and list them to explain each prediction. Indeed, according to our tree configuration (Figure 3), the most important variable for discriminating between originals and fakes/forgeries is 'without provenance', the second is 'untitled', the third is 'without date', and the fourth is 'appearance in Switzerland'; the other input variables are not included in the model by the algorithm. Thus, each node (niche) corresponds to a set of rules identifying the artwork assigned to each terminal node, as listed in Table 20.

Table 20: Rules and interpretations for Originals vs. Fakes/Forgeries (Experimental set)

Node	Rule	Interpretation
1	Having a provenance	A work of art with a provenance has a high probability (99%) of being original.
2	Not having a provenance + Having a date on the painting + Not being untitled + Not having appeared in Switzerland	When a work of art does not have a provenance and has not appeared in Switzerland (defined as either the country from which a request was made to the Vallotton Foundation or the country of the painting's physical location) but does have a date on the painting and a title, the probability of being original is 79%.
3	Not having a provenance + Having a date on the painting + Not being untitled + Having appeared in Switzerland	According to the interpretation of the algorithm, the absence of a provenance combined with having appeared in Switzerland places the artwork in a risky zone (90% probability of being a fake/forgery).
4	Not having a provenance + Having a date on the painting + Being untitled	The probability of being a fake/forgery is 100% if a work of art has neither a provenance nor a title but does have a date.
5	Not having a provenance + Not having a date on the painting	According to the interpretation of the algorithm, the absence of a provenance and a date on the painting means that it has a 97% probability of being a fake/forgery.

b) Original vs. Fake/forgery/Doubtful

This analysis is conducted in the same manner as the *Original vs. Fake/forgery* approach. The theoretical explanations mentioned above remain valid for this model. Thus, the following trees should be interpreted according to the explanations provided in the previous section.

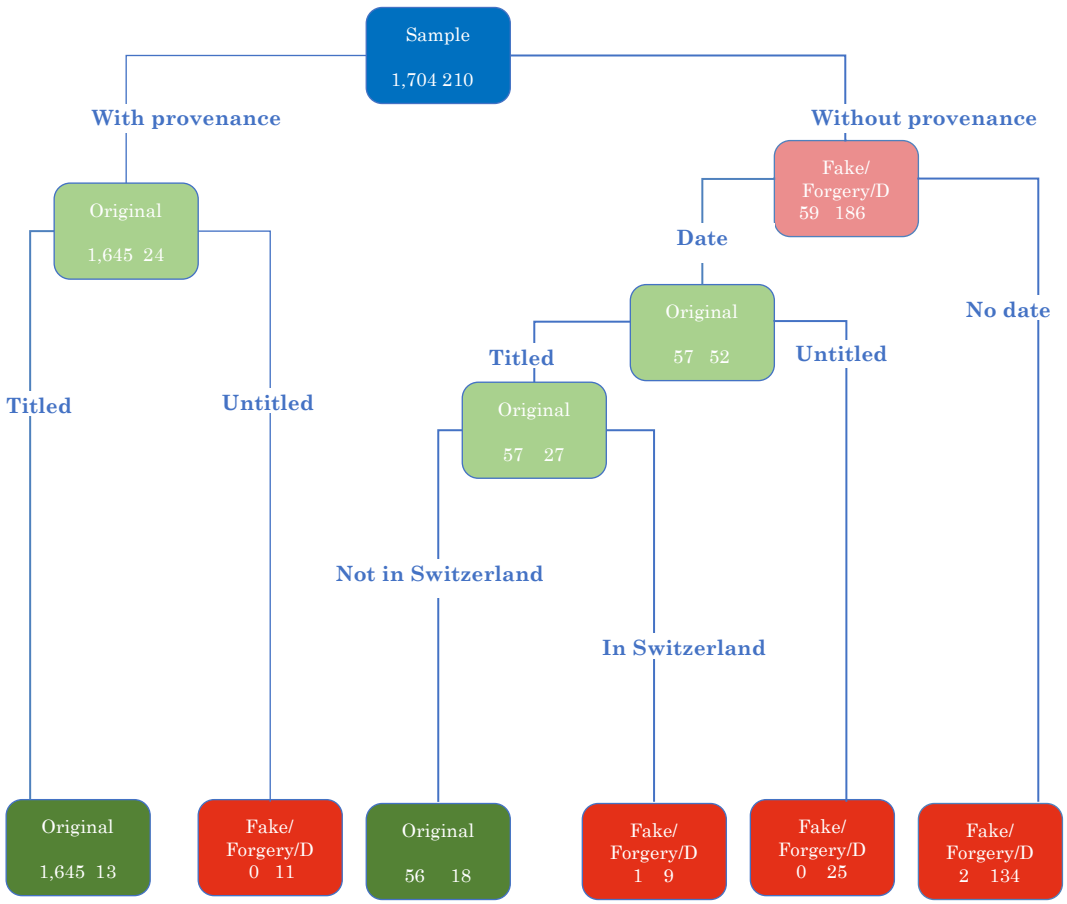


Figure 5: CART model: Original vs. Fake/forgery/Doubtful (Experimental set)

According to the tree configuration in Figure 5, the most important variables for discriminating between originals and the fake/forgery/doubtful category are 'without provenance', 'untitled', 'without date', and 'in Switzerland'; the algorithm does not include the other input variables in the model.

Table 21: Rules and interpretations for the Original vs. Fake/forgery/ Doubtful (Experimental set)

Node	Rule	Interpretation
1	Having a provenance + Not being untitled	Artwork with a provenance and title has a high probability (92%) of being original.
2	Having a provenance + Being untitled	Artwork has a 100% probability of being a fake/forgery or doubtful if it has a provenance and no title.
3	Not having a provenance + Having a date on the painting + Not being untitled + Not having appeared in Switzerland	This interpretation is similar to that of the 2nd node from Table 20: when artwork does not have a provenance and did not appear in Switzerland but has a date on the painting and a title, the probability of being original is 76%.
4	Not having a provenance + Having a date on the painting + Not being untitled + Having appeared in Switzerland	This interpretation is similar to that of the third node in Table 20: according to the interpretation of the algorithm, the absence of a provenance combined with having appeared in Switzerland places the artwork in a risky zone (i.e. a 90% probability of being a fake/forgery or doubtful).
5	Not having a provenance + Having a date on the painting + Being untitled	The probability of being a fake/forgery or doubtful is 100% provided that the artwork has neither a provenance nor a title but has a date.
6	Not having a provenance + Not having a date on the painting	Works of art that fit these rules have a 98.5% probability of being a fake/forgery or doubtful.

The sensitivity of the model is 85.2%, and the specificity is 99.8%. The classification algorithm produces a tree with six terminal nodes. Each terminal node corresponds to a set of rules that identify the works of art assigned to it as listed in Table 21.

c) *Original vs. Fake based on PCs*

Figure 6 shows the classification tree obtained when the six PCA components are included as potential explanatory variables. The algorithm selects PC1, PC3, and PC5 as the potential explanatory components with the highest power to discriminate between original and fake artworks.

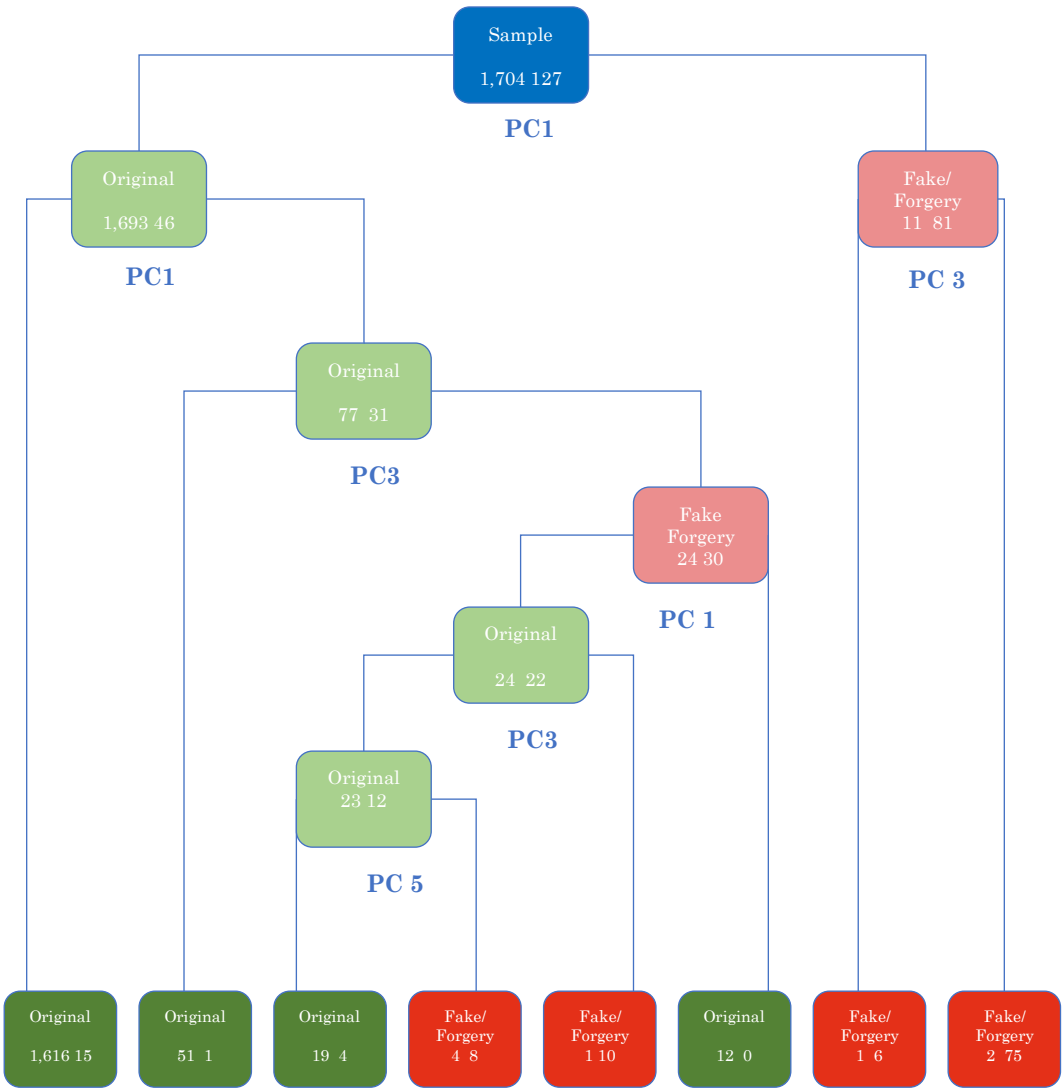


Figure 6: CART model: Original vs. Fake based on PCs (Experimental set)

Globally, the sensitivity was 101/127 = 79.53% because 101 fakes were assigned correctly, and the specificity was 1,697/1,704 = 99.59% because 1,697 original works were correctly classified. In this model, both the sensitivity and specificity are high. The advantage of using PCs over the 15 original variables is that each of the 6 PCs contains more information than any of the 15 original variables. However, the results of this approach are difficult to conceptualise. Certain nodes contain information about many variables. For example, the third node in Fig. 6 comprises PC1, which correlates with six variables; PC3, which correlates with two variables; and PC5, which correlates with one variable. Thus, to better understand the rules leading to this specific node, we must understand the relationship between the original variables and the PCs. Overall, the added complexity of this approach does not seem to be counterbalanced by better results. This model does not perform better in terms of specificity and sensitivity than the model based on 15 variables.

d) *Original vs. Fake/forgery/Doubtful based on PCs*

Figure 7 shows the classification tree when the six PCs are included as potential inputs. The classification algorithm produces a tree with eight terminal nodes using PC1 and PC3 (see the PCA model).

Globally, the sensitivity of this model is 183/210 = 87.14% because 183 fake/forgery/doubtful works are assigned correctly, and the specificity is 1,702/1,704 = 99.59% because 1,697 of the 1,704 original works are classified correctly.

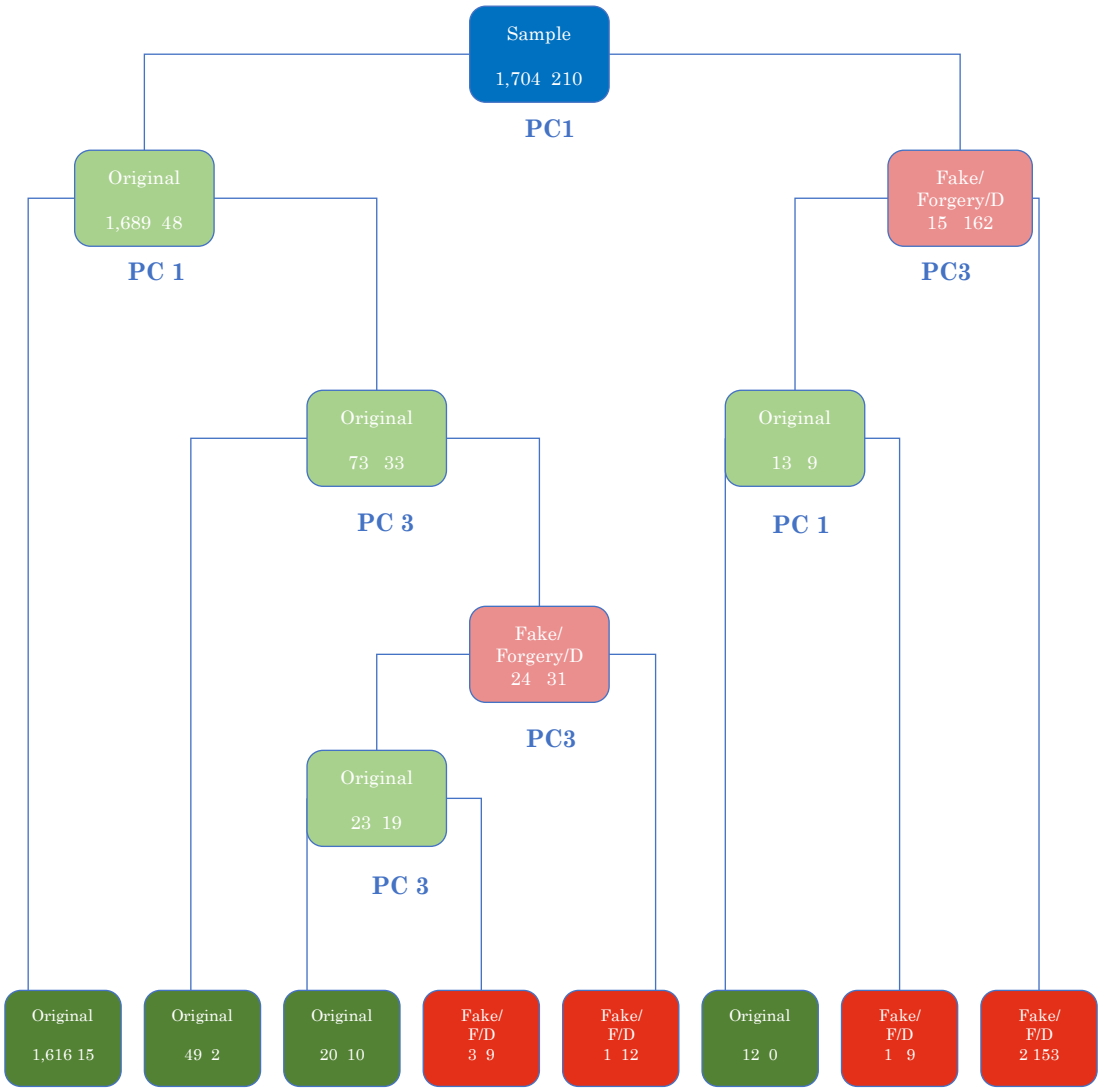


Figure 7: CART model: Original vs. Fake/forgery/Doubtful based on PCs (Experimental set)

6.2.3 CART Validation of the analyses of the experimental set

Because predictive statistical models are estimated based on the available sample dataset, a fundamental challenge in statistics is that of obtaining an accurate estimate of a model's prediction error, that is, the expected loss of the estimated model on future observations (Borra & Di Ciaccio, 2010). However, the classification tree method remains a statistical model; therefore, its output contains errors. In particular, some paintings from the catalogue were classified as fakes, and vice versa. Therefore, as explained earlier, to measure the model accuracy (or performance), we used the proportion of correct predictions made by the model, which was close to 100%. Nevertheless, we cannot rule out the possibility that the good performance of our model is due to certain particularities of our data rather than to the overall quality of the chosen method. In the relevant literature, studies of model evaluation refer to different approaches for validating such errors, including the cross-validation model (Borra & Di Ciaccio, 2010). Cross-validation is a validation technique for assessing how the results of a statistical analysis can be generalised to an independent data set (Picard & Cook, 1984). This technique involves splitting a data sample into complementary subsets, performing an analysis on one subset, called the *training set*, and validating the analysis using the other subset, called the *validation set* (Blockeel & Struyf, 2002). To assess the quality of the results, we used a cross-validation procedure. First, the complete sample of artwork was randomly divided into training and validation sets of approximately equal size. The classification-tree method was then applied to the training set. Finally, the model was applied to the validation set to determine its classification power. The following tables show the results of 10,000 cross-validation iterations for each dataset combination (Table 23).

We focused on the means of the last two distributions (75.5% and 99.5%), which apply to the validation set for the CART 'Original vs. Fake' model approach (Table 22). The sensitivity for the total sample is 77.2%[85], whereas that for the set of 10,000 cross-validations is 75.5%. Thus, the sensitivity of the validation test is slightly lower, but still higher than that of the total sample.

[85] See the explanation of the CART '*Original vs. Fake'* model approach.

The specificity of 10,000 cross-validations is 99.5%, almost equal to the specificity of the total sample (99.8%), and this result is considered very good.

Table 22: Sensitivity and specificity for the CART 'Original vs. Fake' model approach (Experimental set): summary of the cross-validation study (%)

	Minimum	**Q1**	**Median**	**Mean**	**Q3**	**Maximum**
Training Sensitivity	54.7	76.9	80.0	79.9	83.1	94.4
Training Specificity	98.6	99.4	99.5	99.6	99.8	100.0
Validation Sensitivity	49.1	71.7	76.5	**75.5**	80.0	93.6
Validation Specificity	97.4	99.3	99.5	**99.5**	99.8	100.0

Table 23: Sensitivity and specificity for the CART 'Original vs. Fake/forgery/Doubtful' model approach, (Experimental set): summary of the cross-validation study (%)

Set 5	**Minimum**	**Q1**	**Median**	**Mean**	**Q3**	**Maximum**
Training Sensitivity	71.8	81.8	84.3	84.3	87.0	96.4
Training Specificity	98.4	99.4	99.5	99.6	99.8	100.0
Validation Sensitivity	57.7	77.6	80.6	**80.6**	83.8	95.6
Validation Specificity	97.0	99.3	99.5	**99.5**	99.8	100.0

We obtain a sensitivity of 80.6% and a specificity of 99.5% for the CART 'Original vs. Fake' model approach (Table 23). The sensitivity (80.6%) is slightly lower but still higher than that of the total sample (85.2%), and the specificity is almost equal to that of the total sample (99.8%). These results validate the fact that fake/forgery/doubtful works have many characteristics that differentiate them from the original paintings. Moreover, we note that over 10,000 replications of the consolidation, the specificity is never below 97.4% and the sensitivity is never below 49.1%. Thus, even in the worst cases, the model correctly identified the vast majority of the original and approximately half of the fakes/forgeries and doubtful works.

Performing accurate cross-validation for *Original vs. Fake based on PCs* and for *Original vs. Fake/forgery/Doubtful based on PCs* is not practically feasible. When the same calculation is repeated for different subsamples, the structure of the PCA model may change radically, hampering accurate comparisons between models. Moreover, since models based on PCs have already been shown to be not helpful, these cross-validations would also yield limited utility.

6.2.4 Results of the observation phase

In this research stage, different characteristics were analysed using two methods: PCA and CART. The PCA results imply that a relatively small set of variables may be effective for developing an explanatory model to identify counterfeit artwork. The PCA results were precise, but their interpretation does not specify whether the identified clusters belonged to the original art or forgeries.

To overcome this obstacle, we tested the CART. The results show that CART allows us to determine when the grouped variables are indicative of originals and fakes. The CART performs very well. The error level of the classification based on all the variables was very low. However, classification tree approaches based on PCs have three main limitations.

- ∞ They are very difficult to interpret because the components themselves are complicated.
- ∞ In terms of specificity and sensitivity, they are not necessarily better than models that use variables.
- ∞ Given the variability of the PCA and the number of components selected, it is difficult to perform a meaningful cross-validation.

Considering these difficulties, the approaches *Original vs. Fake based on PCs* and *Original vs. Fake/forgery/Doubtful based on PCs* were not retained in the next step of the analysis. The algorithmic interpretations of other CART approaches are straightforward and explicit; therefore, we use them as examples in the main phase of the analysis.

6.3 Creating working sets based on vulnerable painting categories: destroyed, unknown location, stolen and absent from the common place book

In the previous step, we used a set of randomly selected variables to test and explain the statistical tools. In this section, we examine the features of Vallotton's work that may indirectly reflect the historical background of his artistic legacy, particularly the uncertainties that might have attracted the interest of fakers.

Based on assumptions concerning the lost works of other artists (see Section 3.2.2.1), we proposed a hypothesis about Vallotton's lost works. Theoretically, a faker may pass off a fake or forgery as a lost painting in the hope that the provenance of the truly lost art object may become the provenance for the fake or forgery, thereby decreasing the likelihood of fraud detection. Therefore, a faker can use the available information (e.g. size, technique, support, title, or date) about lost artwork as a source for fabricating a fake artwork. We expect that this approach will contribute to the discovery of new niches and achieve an excellent model performance. The CR points out four possible reasons why Vallatton's paintings might be lost: 'unknown location', 'stolen work', 'destroyed work', and 'absence from the commonplace book'.

Consequently, we divide all the characteristics into four working sets corresponding to each reason and provide the distributions for the lost categories in Appendix III. We then selected the most representative characteristics by comparing their percentage rates.[86] Finally, four sets were formed as described below:

[86] For example, the most destroyed works have a title and large dimensions. They are often a portrait, landscape, private interior, nude, or painting with a mythological subject made in oil or using an unknown technique and made on canvas. They often do not have provenance or only the first provenance, and they were most often made in 1882–1889 or 1890–1910.

Destroyed	**Unknown location**	**Stolen**	**Absent from the commonplace book**
Working Set 1	Working Set 2	Working Set 3	Working Set 4
With title	With title	With title	With title
Large	With date	With date	Without date
Portraits	With signature	Large	Sale/online
Portrait with identified/ hypothetical person	Sale/online	Portrait	Sold
Landscapes	Sold	Portrait with identified/ hypothetical person	Monogrammed
Private and public interiors	With error/ atypical signature	Landscapes	Lower left
Mythological subjects	Lower right	Private and public interiors	Lower right
Nudes	Large	Mythological subjects	Small
Oil	Unknown technique	Nudes	Portrait
Unknown technique	Landscape	Oil	Landscape
Canvas	Oil	Pastel	Nudes
Without provenance	Canvas	Unknown technique	Copy
First provenance	Signature/ initials (on the back)	Without provenance	Oil
1882–1889	First provenance	First provenance	Tempera
1890–1900	Stolen	1882–1889	Cardboard
1901–1910	Missing from the commonplace book	1890–1900	Wood
1906–1909	1890–1900	1901–1910	Signature/ initials (on the back)
	1901–1910	1901–1905	Date (on the back)
	France	1906–1909	Under another painting (on the back)
	Switzerland		First provenance
	England		Unknown location
			1901–1910

PART 3 – FINDINGS

7 DISCOVERING PATTERNS: RESULTS OF PCA AND CART ANALYSES ACROSS WORKING SETS

In this Chapter, we examine the most representative characteristics of the lost categories. Four working sets were examined using the methods described in the previous section. Their interpretation is similar to that of the experimental set.

7.1 Set 1: Destroyed

7.1.1 PCA model for the Destroyed set

a) PCA with all significant components

Table 24 shows that there are correlations between the variables in this set.

Table 24: Correlation matrix (Working Set 1)

Variables (Working Set 1)	**Components**					
	1	**2**	**3**	**4**	**5**	**6**
Unknown technique	0.7					
Oil	−0.7					
Canvas	−0.6					
Large	−0.6					
Without provenance	0.5					
Portraits with identified (or hypothetical person)		0.7				
Portraits		0.7				
1882-1889		0.5				
Nudes			0.7			
Landscape			−0.7			
1906–1909 nude			0.6			
1890–1900 wood				−0.6		
Untitled				0.5		
1901–1910					0.6	
Private and public interiors					0.5	
First provenance					−0.4	
Mythological /allegorical or biblical						0.5
Percentage of variance reproduced	18%	13%	11%	8%	8%	6%

Indeed, the PC1 is strongly correlated with five variables: the likelihood of the variables 'without provenance' and 'unknown technique' increase simultaneously when the likelihood of the variables 'oil', 'canvas', and 'large' decrease, meaning that a

painting without provenance more often has a technique that is not indicated in the files or the CR and very rarely is a painting in oil on canvas with large dimensions. With the help of the PC2, we apprehend that portraits occur more frequently if they show an identified or hypothetical person and are made during 1882–1889. The PC3 indicates that a painting created during 1906–1909 is more likely to be nude and rarely a landscape work of art. Through interpretation of the PC4, it emerges that untitled paintings are very rarely found during 1890–1900. Furthermore, the correlations of PC5 mean that a painting created during 1901–1910 is more likely to be a painting with the theme of private/public interiors and very rarely has a first provenance (i.e. no complete provenance when only the first owner is known). PC6 mainly correlates with works of art with mythological/allegorical/biblical themes.

The next step was to examine whether the observed links were visible in the plot.

b) PCA with two components

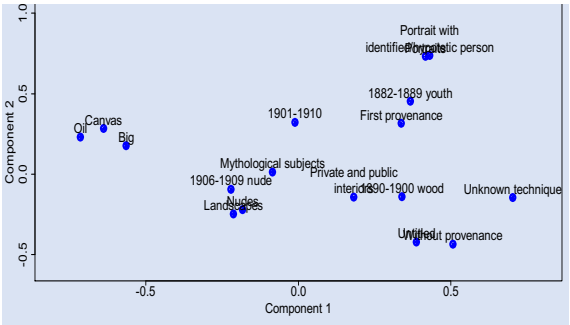


Figure 8: Two-dimensional score plot (Working Set 1)[87]

Figure 8 shows the projections of the 17 variables onto the space spanned by the two principal components. The issue is to analyse whether the main results in Table 24 also appear in Figure 8. The first component reproduces 18% of the original information, and the second 13%.

Considering the PC1 plot configuration, we can see that 'oil', 'canvas', and 'large' on the left are grouped together and are opposed to 'unknown technique' and 'without provenance' on the right. Regarding the PC2, variables 'identified or hypothetical

[87] See Appendix VIII.

person', 'portraits', and '1882–1889' are grouped together at the top of the figure. Furthermore, Figure 8 provides additional information to that of the correlation matrix (Table 24). First, the variable 'untitled' is clearly associated with the group already composed of 'unknown technique' and 'without provenance', since the location of the 'untitled' dot is very close to the one of the 'unknown provenance'. Similarly, the variable 'first provenance' is associated with 'identified or hypothetical person', 'portraits', and '1882–1889'. Variables in the lower-right group share a common meaning: the absence of certain information (i.e. unknown provenance, unknown technique, or unknown title). This implies that a painting without a provenance is more likely to have an unknown technique and/or title. The left group reveals that a larger painting is more likely to be oil on canvas. The upper group indicates that a painting with a first provenance is more likely to be a portrait with/without an identified or hypothetical person and created during the period 1882–1889. Second, the three abovementioned groups occupy three antagonist positions, meaning that the characteristics defined by one of the three groups of variables are very unlikely to be found in the same artwork as the characteristics defined by the other two groups. For instance, large oil artworks on canvas are not likely to be portraits, or portraits are not likely to be untitled or without provenance. Finally, variables closer to the origin of the figure, that is, the position of the coordinates (0,0), are more difficult to analyse because their positions can only indicate that the first two components do not suffice to represent them correctly.

The PCA results provide strong evidence of the relationships between the 17 variables and allow us to proceed to the next step of the analysis by using them as potential explanatory variables for the CART.

7.1.2 CART model for the Destroyed set

a) Original vs. Fake/forgery

Figure 9 shows the classification tree when all the variables from Working Set 1 are included as potential inputs.

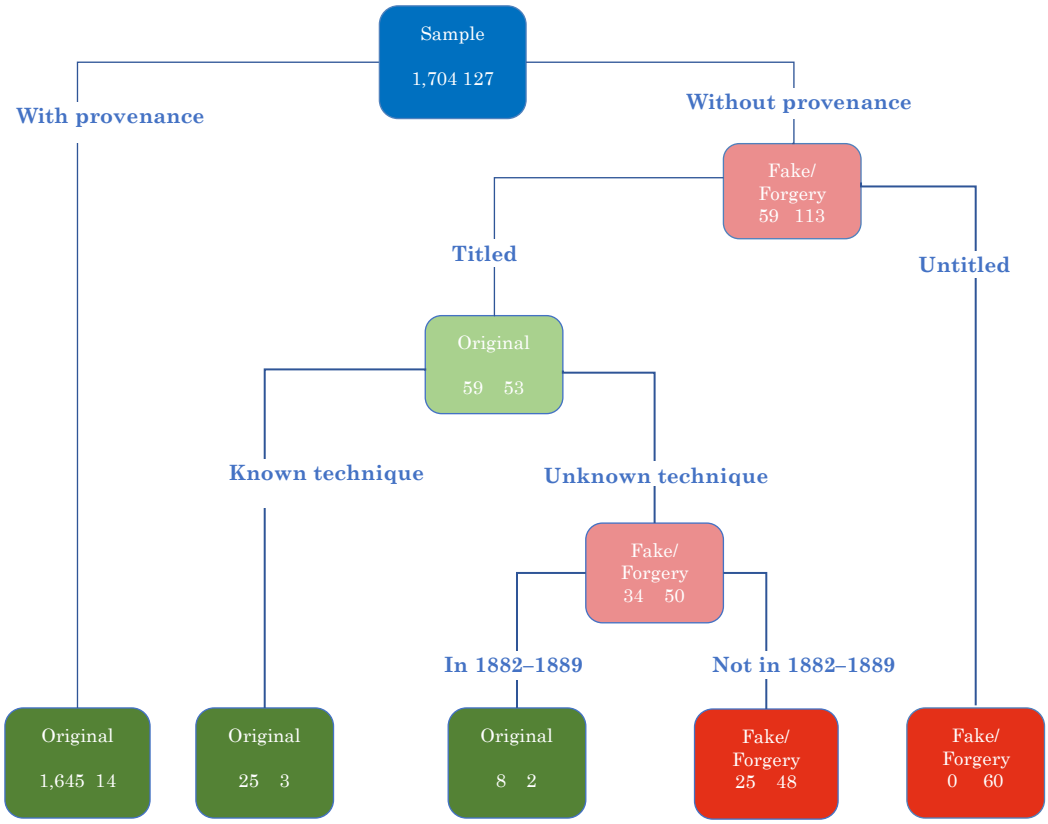


Figure 9: CART model: Original vs. Fake/forgery (Working Set 1)

The classification algorithm produces a tree with five terminal nodes. Each node corresponds to a set of rules that identifies the works of art assigned to each terminal node, as listed in Table 25.

According to the tree configuration, the most important variables for discriminating between originals and fakes/forgeries are 'without provenance', 'untitled', 'unknown technique', and 'in 1882–1889'. The algorithm did not include other input variables in the model.

Table 25: Rules and interpretations for Original vs. Fake/forgery (Working Set 1)

Node	Rule	Interpretation
1	Having a provenance	Artwork with a provenance has a high probability (99%) of being original.
2	Not having a provenance + Not being untitled + Not having an unknown technique	Artwork without a provenance but with a title and a known technique has an 89% probability of being original.
3	Not having a provenance + Not being untitled + Having an unknown technique + Made during 1882–1889	A painting without a provenance and with a title that was made with an unknown technique during the years 1882–1889 has a good probability of being original (90%).
4	Not having a provenance + Not being untitled + Having an unknown technique + Not made during 1882–1889	Artwork with similar characteristics to those of node 3 in this table but that was not created in the period 1882–1889 (i.e. before 1910) has a 66% probability of being a fake/forgery.
5	Not having a provenance + Being untitled	According to the algorithm's interpretation, the absence of a provenance and a title on painting means that the work has a 100% probability of being a fake/forgery.

b) Original vs. Fake/forgery/Doubtful

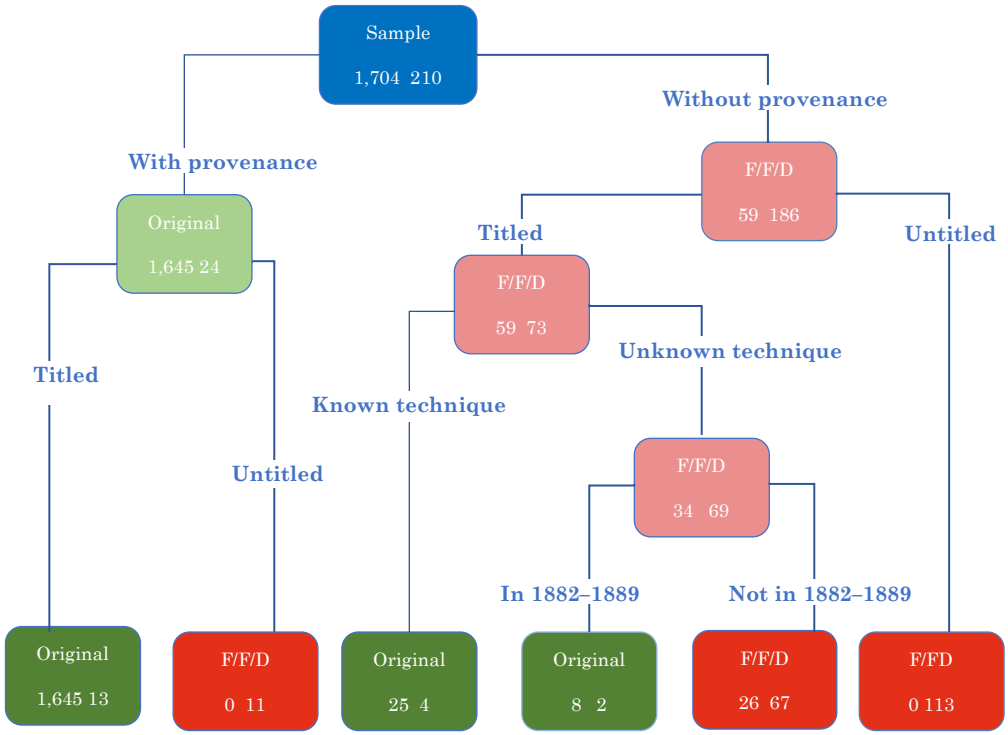


Figure 10: CART model: Original vs. Fake/forgery/Doubtful (Working Set 1)

Each node in Figure 10 corresponds to a set of rules identifying the works of art assigned to each terminal node, as listed in Table 26.

Table 26: Rules and interpretations for Original vs. Fake/forgery/ Doubtful (Working Set 1)

Node	Rule	Interpretation
1	Having a provenance + Not being untitled	Artwork with a provenance and a title has a high probability (99%) of being original.
2	Having a provenance + Being untitled	Artwork with a provenance but no title has a 100% probability of being a fake/forgery or doubtful. We suppose that a painting with a provenance such that we can identify the chronology of its existence from its creation to today must have a title; otherwise, it may be suspicious.
3	Having a provenance + Not being untitled + Not having an unknown technique	The probability of being original is 86% if the work of art has a provenance, a title, and a known technique.
4	Not having a provenance + Not being untitled + Having an unknown technique + Made during 1882–1889	The probability of being original is 80%. The explanation for the third node of Table 25 applies.
5	Not having a provenance + Not being untitled + Having an unknown technique + Not made during 1882–1889	The probability of being a fake/forgery or doubtful is 72%. The explanation for the fourth node of Table 25 applies.
6	Not having a provenance + Being untitled	The probability of being a fake/forgery or doubtful is 100%. The explanation for the fifth node of Table 25 applies.

7.2 Set 2: Unknown location

7.2.1 PCA model for the Unknown location set

a) PCA with all significant components

Table 27 lists the nine optimal components of the PCA structure for the variables in Working Set 2.

Table 27: Correlation matrix (Working Set 2)

Variables (Woking set 2)	**Components**								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
With date	0.6								
With title	0.6								
With signature	0.5		0.6						
Sale/online		0.9							
Sold		0.9							
With error/ atypical signature									0.5
Lower right				−0.5					
Large	0.6								
Unknown technique	−0.6								
Landscape				−0.5					
Oil	0.6								
Canvas	0.6								
Signature/initials (on the back)							0.4		
First provenance				0.4					
1890–1900				−0.5					
1901–1910							−0.5		
France						0.6			
Switzerland						−0.6			
England								−0.7	
Stolen					0.4				
Missing from the commonplace book					0.5				
Percentage of variance reproduced	15%	9%	9%	7%	6%	6%	5%	%	5%

These results indicated the presence of new clusters. Indeed, according to the PC1 interpretation, a large-dimension painting in oil on canvas often has a title and date,

and very rarely has an unknown technique. There is a relation between 'sale/online' and 'sold' (PC2), a negative correlation between 'in France' and 'in Switzerland' (PC6) and between 'signature/initials (on the back)' and '1901–1910' (PC6). We discover that a painting created during 1890–1900 is more likely to be a landscape with a signature on canvas located in the lower right corner and rarely has a first provenance (PC4).

b) PCA with two components

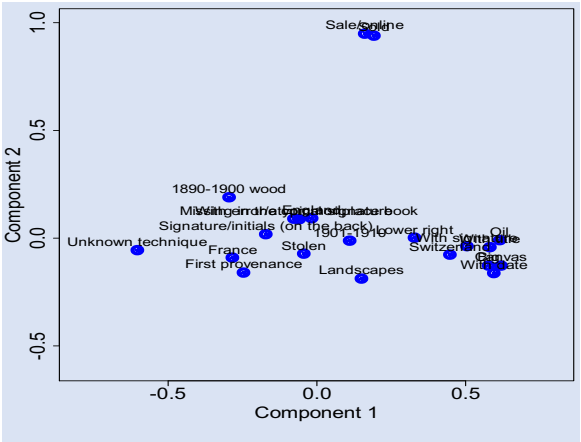


Figure 11: Two-dimensional score plot (Working Set 2)[88]

The plot also shows the findings of the correlation matrix (Table 27) also appear in the plot (Figure 11). The first component reproduces 15% of the original information, and the second 9%. Similar to the plot for Working Set 1 (Figure 8), this plot provides additional clusters. Indeed, we can see that the variable 'in Switzerland' is associated with the group already composed of 'with date', 'with signature', 'large', 'oil', 'canvas', and 'with title'. The variables in the lower-right group share a common meaning: the presence of certain information (i.e. date, signature, title, known technique, and support). This means that a painting with a date and signature is more likely to be a large-dimension painting in oil whose owner is situated in Switzerland.

[88] See Appendix VIII.

7.2.2 CART model for the Unknown location set

a) Original vs. Fake/forgery

Figure 12 shows the classification tree when all the variables from Working Set 2 are included as potential inputs.

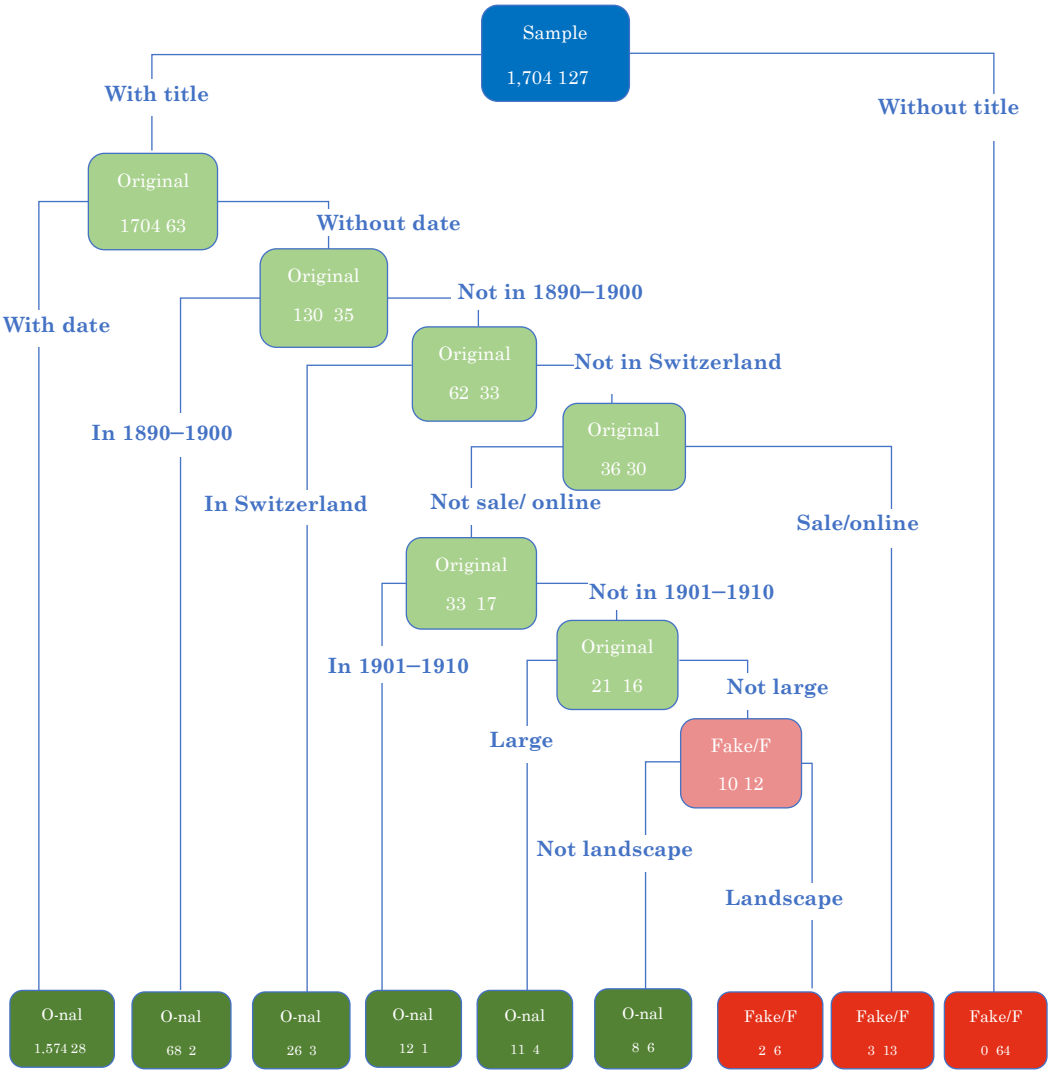


Figure 12: CART model: Original vs. Fake/forgery (Working Set 2)

The classification algorithm produces a tree with nine terminal nodes. Each node corresponds to a set of rules identifying the works of art assigned to the node, as listed in Table 28.

Table 28: Rules and interpretations for Original vs. Fake/forgery (Working Set 2)

Node	Rule	Interpretation
1	Having a title + Having a date	Artwork with a title and a date has a high probability (98%) of being original.
2	Having a title + Not having a date + Made during 1890–1900	Artwork with a title, without a date, made during the period 1890–1900 has an 87% probability of being original.
3	Having a title + Not having a date + Not made during 1890–1900 + Being in Switzerland	Artwork with a title, without a date, not made during the period 1890–1900, and being in Switzerland has a 90% probability of being original.
4	Having a title + Not having a date + Not made during 1890–1900 + Not being in Switzerland + Not being sale/sale online + Made during 1901–1910	Artwork with a title, without a date, not being in Switzerland, not being on sale, and not being made during 1890–1900 but made during 1901–1910 has a 92% probability of being original.
5	Having a title + Not having a date + Not made during 1890–1900 + Not being in Switzerland + Not being sale + Not made during 1901–1910 + Large	A large artwork with a title, without a date, not being in Switzerland, not being on sale/sale online and not made during 1890–1910 has a 73% probability of being original.
6	Having a title + Not having a date + Not made during 1890–1900 + Not being in Switzerland + Not being sale/sale online + Not made during 1901–1910 + Not a large + Not being a landscape	A medium or small size artwork with a title, without a date, not being in Switzerland, not being on sale, not made during 1890–1910 and not being a landscape has a 57% probability of being original.

7	Having a title + Not having a date + Not made during 1890–1900 + Not being in Switzerland + Not being sale/sale online + Not made during 1901–1910 + Not a large + Being a landscape	Being a landscape with a combination of medium or small size, with a title, without date, not being in Switzerland, not being for sale, not having been made between 1890–1910 has a 75% probability of being fake/forgery.
8	Having a title + Not having a date + Not made during 1890 – 1900 + Not being in Switzerland + Being sale/sale online	Artwork with a title, without a date, not being in Switzerland, made between 1890–1900 and being on sale has a 81% probability of being a fake/forgery.
9	Not having a title	According to the algorithm's interpretation, the absence of a title on painting means that the work has a 100% probability of being a fake/forgery.

b) Original vs. Fake/forgery/Doubtful

The classification algorithm produces a tree with seven terminal nodes.

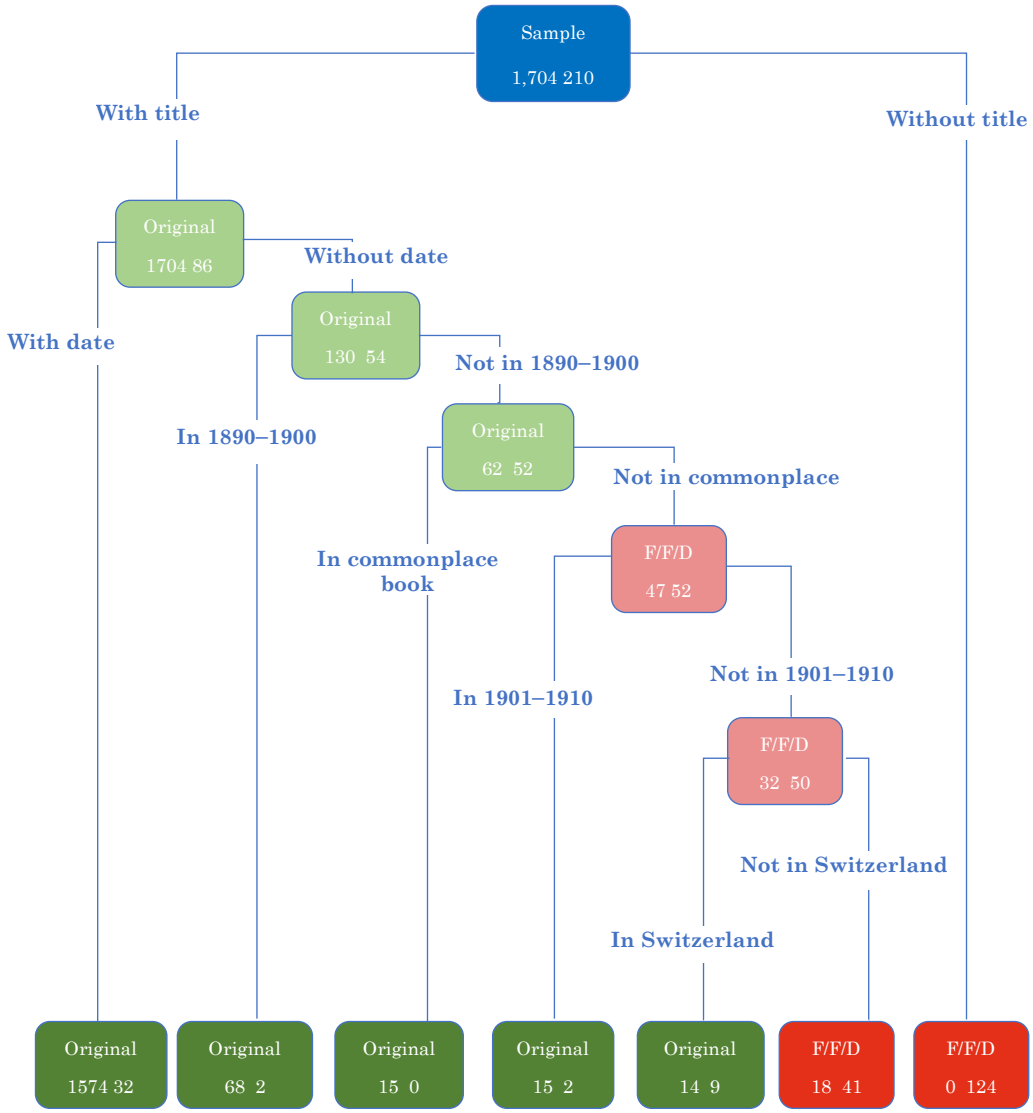


Figure 13: CART model: Original vs. Fake/forgery/Doubtful (Working Set 2)

Each node corresponds to a set of rules identifying the works of art assigned to each terminal node, as listed in Table 29.

Table 29: Rules and interpretations for Original vs. Fake/forgery/ Doubtful (Working Set 2)

Node	Rule	Interpretation
1	Having a title + Having a date	Artwork with a title and a date has a high probability (98%) of being original.
2	Having a title + Not having a date + Made during 1890–1900	Artwork with a title without a date, made during the period 1890–1900 has an 87% probability of being original.
3	Having a title + Not having a date + Not made during 1890–1900 + Being in commonplace book	The commonplace book is Vallotton's primary source of information. After 1900, the artist began to use more precise descriptions in his commonplace book to protect his artistic heritage. Thus, a work of art, even without a date noted on the support but realized after the period 1890–1900 and noted in his notebook, is considered by the model as 100% original.
4	Having a title + Not having a date + Not made during 1890–1900 + Not being in commonplace book + Made between 1901–1910	Artwork with a title, without a date, not being in the commonplace book, and made during 1901–1910 has an 88% probability of being original. The probability of being original decreases when Vallotton did not note the artwork in the commonplace book.
5	Having a title + Not having a date + Not made during 1890–1900 + Not being in commonplace book + Not made during 1901–1910 + Being in Switzerland	Artwork with a title, without a date, not being in commonplace book, not made during 1890–1910, and being in Switzerland has a 61% probability of being original. These characteristics indicate the relatively low probability of being original.
6	Having a title + Not having a date +	The artwork whose owner is not located in Switzerland, with a title,

	Not made during1890–1900 + Not being in commonplace book + Not made during1901–1910 + Not being in Switzerland	without a date, not noted in commonplace book, and not made during 1890–1910 has a 69% probability of being fake/forgery or doubtful
7	Not having a title	According to the algorithm's interpretation, the absence of a title on painting means that the work has a 100% probability of being a fake/forgery or doubtful.

7.3 Set 3: Stolen

7.3.1 PCA model for the Stolen set

a) PCA with all significant components

Table 30 lists the eight optimal components of the PCA structure for the variables in Working Set 3.

Table 30: Correlation matrix (Working Set 3)

Variables (Woking set 3)	**Components**							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
With title		0.5						
With date	−0.5							
Large	−0.5							
Portraits		0.7						
Portraits with identified/ hypothetical person		0.7						
Landscapes			−0.7					
Private and public interiors					0.6			
Mythological subjects								0.9
Nudes								
Oil	−0.7		0.8					
Pastel							0.8	
Unknown technique	0.7							
Without provenance	0.6							
First provenance				0.4				
1882–1889		0.4						
1890–1900					0.7			
1901–1910				−0.5				
1901–1905			−0.5					
1906–1909			0.6					
Percentage of variance reproduced	16%	12%	10%	8%	7%	6%	6%	5%

The PC1 was negatively correlated with ‘with date’, ‘oil’, and ‘large’ and positively correlated with ’unknown technique’ and ‘without provenance’. The same components were observed in the experimental set (Table 19). The other components of Set 3

indicate novel clusters. Indeed, artworks are likely to be portraits of identified/hypothetical persons when they have a title and were created in 1882–1889 (PC2). The PC3 correlation means that a painting created during the period 1906–1909 was likely to be in oil and not a landscape. Moreover, artwork in oil was rarely created during 1901–1905. PC4 was positively correlated with 'first provenance' and negatively with '1901–1910'. This means that a painting with a first provenance was unlikely to have been created during 1901–1910. With the help of PC5, we can see that the paintings made during 1890–1900 were likely to have a theme of private and public interiors. Components 7 and 8 strongly correlate with 'pastel' and 'mythological subjects'.

b) PCA with two components

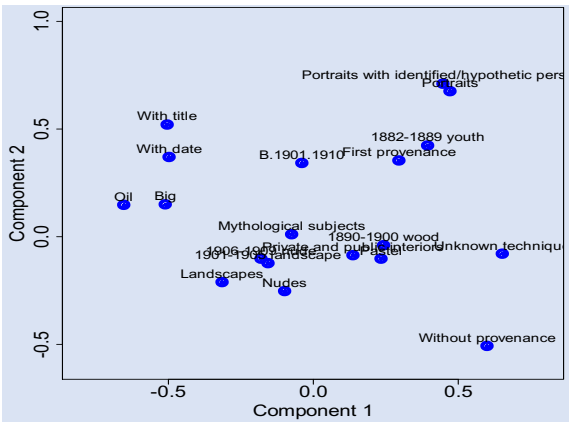


Figure 14: Two-dimensional score plot (Working Set 3)[89]

Similar to the previous sets, the main findings in Table 30 appear in a two-dimensional score plot (the first and second components reproduce 28% of original information).

The plot configuration shows that the variable 'with title' is correlated with variables of the PC1 and not PC2 like in the correlation matrix (Figure 14). In addition, the left group reveals that larger paintings are more likely to be in oil, titled, and dated. The upper-right group indicates that a painting with a first provenance is more likely to be a portrait with or without an identified or hypothetical person created during 1882–1889. The left group shows that a larger painting is more likely to be an oil, titled, or

[89] See Appendix VIII.

dated painting. The upper-right group indicates that a painting with a first provenance is more likely to be a portrait with or without an identified or hypothetical person created during 1882–1889.

7.3.2 CART model for the Stolen set

a) Original vs. Fake/forgery

Figure 15 shows the classification tree when all the variables from Working Set 3 are included as potential inputs.

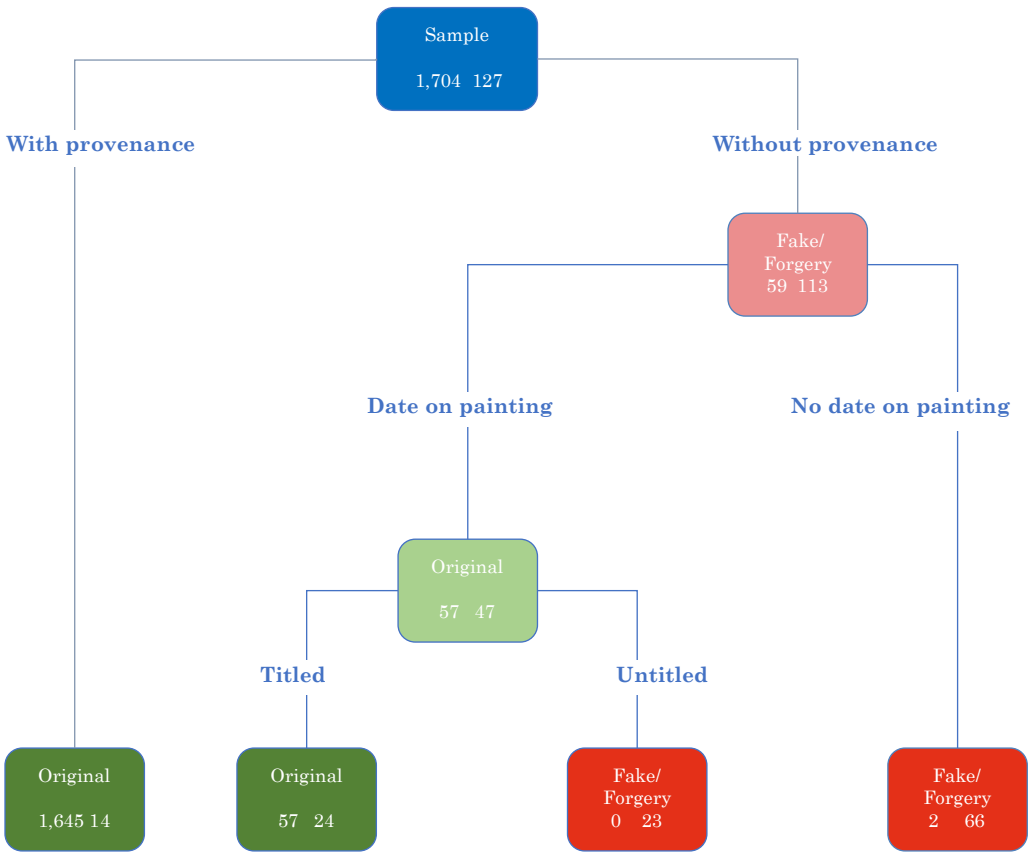


Figure 15: CART model: Original vs. Fake/forgery (Working Set 3)

According to the tree configuration, the most important variables for discriminating between originals and fakes/forgeries are 'without provenance', 'with date', and 'with title'. The algorithm did not include other input variables in the model. Each node corresponds to a set of rules identifying the works of art assigned to each terminal node, as listed in Table 31.

Table 31: Rules and interpretations for Original vs. Fake/forgery (Working Set 3)

Node	Rule	Interpretation
1	Having a provenance	A work of art with a provenance has a high probability (99%) of being original.
2	Not having a provenance + Having a date + Having a title	When a work of art does not have a provenance but has a title and a date, the probability of its being original is 70%.
3	Not having a provenance + Having a date + Being untitled	The probability of being a fake/forgery is 100% when a work of art has neither a provenance nor a title but does have a date.
4	Not having a provenance + Not having a date	According to the interpretation of the algorithm, the absence of a provenance and a date on the painting means that it has a 97% probability of being a fake/forgery.

b) Original vs. Fake/forgery/Doubtful

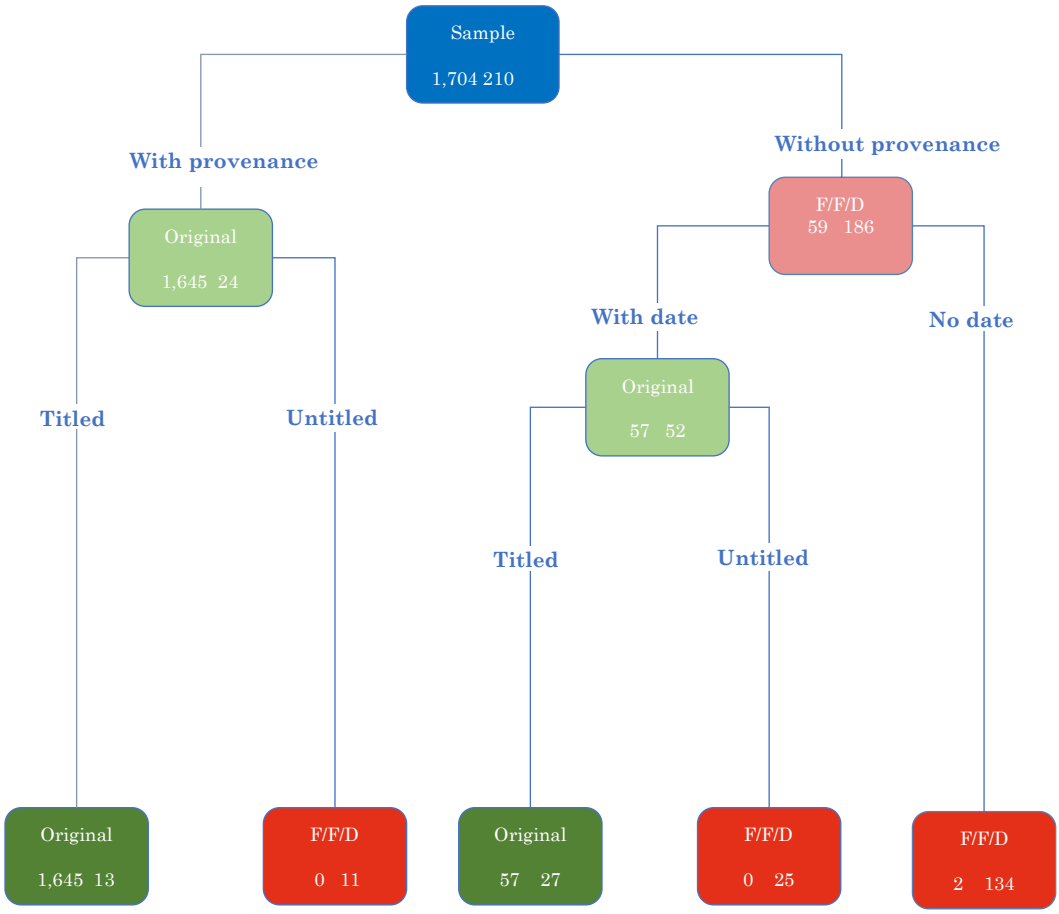


Figure 16: CART model: Original vs. Fake/forgery/Doubtful (Working Set 3)

According to the tree configuration, the most important variables for discriminating between originals and fakes/forgeries/doubtful are 'provenance', 'date', and 'title'. The algorithm did not include other input variables in the model. The classification algorithm produces a tree with five terminal nodes. Each node corresponds to a set of rules identifying the works of art assigned to each terminal node, as listed in Table 32.

Table 32: Rules and interpretations for the Original vs. Fake/forgery/ Doubtful (Working Set 3)

Node	Rule	Interpretation
1	Having a provenance + Titled	Artwork with a provenance and title has a high probability (92%) of being original.
2	Having a provenance + Being untitled	Artwork has a 100% probability of being a fake/forgery or doubtful if it has a provenance and no title.
3	Not having a provenance + Having a date + Having a titled	When artwork does not have a provenance but has a date and a title, the probability of being original is 68%.
4	Not having a provenance + Having a date + Being untitled	The probability of being a fake/forgery or doubtful is 100% provided that the artwork has neither a provenance nor a title but has a date.
5	Not having a provenance + Not having a date	Works of art that fit these rules has a 99% probability of being a fake/forgery or doubtful.

7.4 Set 4: Absent from the commonplace book

7.4.1 PCA model the Absent from the commonplace book set

a) PCA with all significant components

Table 33: Correlation matrix (Working Set 4)

Variables (Woking set 4)	**Components**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
With title					0.5					
Without date	0.5									
Sale/online		0.9								
Sold		0.9								
Monogrammed										−0.3
Lower left							−0.7			
Lower right				0.6						
Small	0.5									
Portrait			−0.5							
Landscape			0.5							
Nudes						−0.7				
Copy										0.7
Oil	−0.6									
Distemper				0.7						
Cardboard				0.5						
Wood			0.5							
Signature/initials (on the back)					0.6					
Date (on the back)					0.6					
Another painting (on the back)										0.6
First provenance			−0.5							
Unknown location			−0.5							
1901–1910								0.7		
Percentage of variance reproduced	10%	9%	9%	8%	8%	7%	6%	6%	5%	5%

The results of the PCA for Set 4, with the most frequent variables of category 'absent from the commonplace book', resulted in novel clusters (Table 33). In particular, with the help of the interpretation of the PCs, we understand that a small painting without a date is rarely a painting in oil (PC1), that a painting that has been sold is likely to be on sale or for sale online (PC2), and that a landscape is more likely to be on a wooden

support and to have a known rather than unknown location; a picture on wood support is rarely a portrait and rarely has the first provenance (PC3). The interpretation of PC4 reveals that a painting made on a cardboard support is more likely to be made using the tempera technique and has a signature in the lower-right corner. Since PC5 is positively correlated with 'date on back', 'signature or initials on the back', and 'with title', the explanation implies that the painting with a title is more likely to have a date, signature, or initials on the back. PC10 also provides additional awareness: a painting that is a copy of another painting is more likely to have a picture on its back and is less likely to have a monogram. Although the PCA of this set provided ten components, which was more than that of the other sets, only six of the components are fully interpretable.

b) PCA with two components

The first component reproduces only 10% of the original information, and the second 9%. However, the two-dimensional plot provides additional information compared to the correlation matrix (Figure 17).

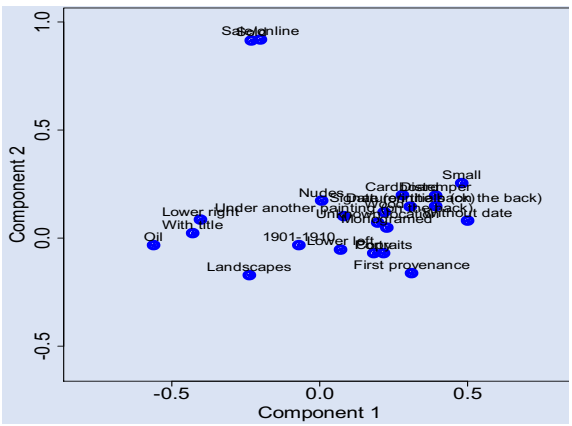


Figure 17: Two-dimensional score plot (Working Set 4)[90]

Indeed, a group of two variables is distant from the other variables in Set 4. This means that a painting whose records retain information that it has been sold and evidence that a work of art has been sold online is unlikely to be a large-dimensional signed oil painting with a title and signature in the lower right-hand corner.

[90] See Appendix VIII.

Furthermore, such a work of art is unlikely to be a tempera painting of a small size which does not have a date on the canvas and has a date on the back.

7.4.2 CART model for the Absent from the commonplace book set

a) Original vs. Fake/forgery

Figure 18 shows the classification tree when all the variables from Working Set 4 are included as potential inputs.

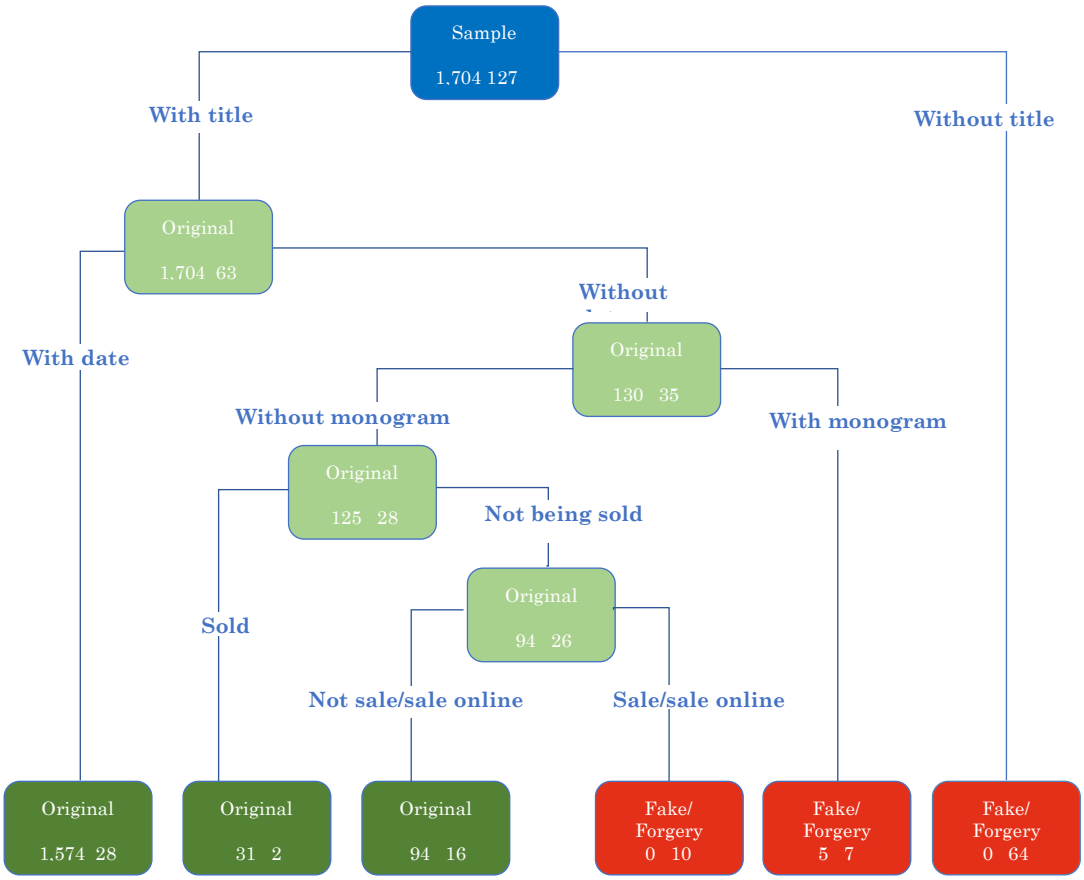


Figure 18: CART model: Original vs. Fake/forgery (Working Set 4)

The most important variables for discriminating between originals and fakes/forgeries are 'without provenance', 'with date', and 'with title'. Each node corresponds to a set of rules identifying the works of art assigned to each terminal node, as listed in Table 34.

Table 34: Rules and interpretations for Original vs. Fake/forgery (Working Set 4)

Node	Rule	Interpretation
1	Having a title + Having a date	Artwork with a title and a date has a high probability (98%) of being an original.
2	Having a title + Not having a date + Not having a monogram + Being sold	Artwork that has been sold recently which has a title but no date on the painting and is not monogrammed has a 94% probability of being an original.
3	Having a title + Not having a date + Not having a monogram + Not being sold + Not being sale/online	Artwork that has not been sold, was not offered for online sale, which has a title but no date on the painting, and is not monogrammed has an 85.5% probability of being original.
4	Having a title + Not having a date + Not having a monogram + Not being sold + Being sale/online	Artwork that has not been sold, was on sale, or online sale, which has a title but no date on the painting, and is not monogrammed has a 100% probability of being a fake/forgery.
5	Having a title + Not having a date + Not having a monogram	Artwork which has a title, is monogrammed, and does not have a date on painting has a 58% probability of being a fake/forgery.
6	Not having a title	Artwork not having a title has a 100% probability of being a fake/forgery.

b) Original vs. Fake/forgery/Doubtful

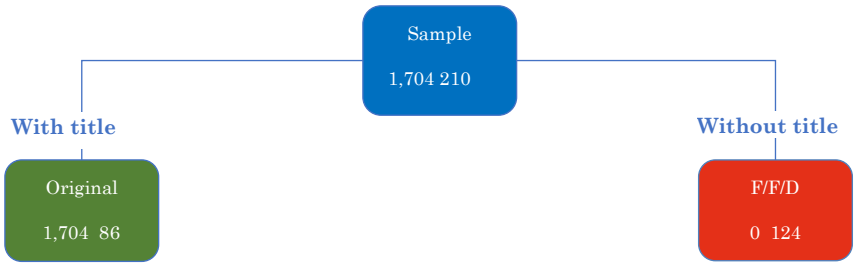


Figure 19: CART model: Original vs. Fake/forgery/Doubtful (Working Set 4)

The classification algorithm produces a tree with two terminal nodes. Each node corresponds to a set of rules identifying the works of art assigned to each terminal node, as listed in Table 35.

Table 35: Rules and interpretations for Original vs. Fake/forgery/ Doubtful (Working Set 4): Summary of the cross-validation study

Node	Rule	Interpretation
1	Having a title	Artwork with a title and a date has a high probability (95%) of being original.
2	Not having a title	Artwork with a title and a date has a high probability (100%) of being fake/forgery or doubtful.

7.5 CART validation of the analyses of the four working sets

An ideal model should have a 100% correct classification rate. In practice, this statement signifies that the closer the proportion of correct classifications of the model is to 100%, the less influence the errors have on the results. To check for such an impact, we calculated for each working set the percentage of correct classification – *sensitivity and specificity* –of original and fake works (see Section 6.2.2). These rates are listed in Table 36.

Table 36: Sensitivity and specificity for the CART's approaches (Working sets)

Working sets	**Original vs Fake/forgery (%)**	**Original vs Fake/forgery/Doubtful (%)**
Set 1		
Specificity	98.5	98.0
Sensitivity	85.0	91.0
Set 2		
Specificity	99.7	98.9
Sensitivity	65.4	78.6
Set 3		
Specificity	99.9	99.9
Sensitivity	70.0	80.9
Set 4		
Specificity	99.7	100.0
Sensitivity	63.8	59.0

Overall, the performance of CART in classifying original works of art was robust across all the work sets. The specificity rates ranged from 98.0% to 100.0%. The correct classification of fake works varied, and their rates were lower than those of the original works. This result is not surprising. Indeed, there are much more data for the originals (1,704 works) than for the counterfeits (210 works). This may explain the differences between the accuracies of the two categories. Concerning sensitivity rates, CART's performance is excellent with Sets 1 (85.0% and 91.0%) and 3 (70.0% and 80.9%). The performance of fake classification was also robust for Set 2 (65.4% and 78.6%). The percentage of correct classifications of fakes was the lowest with Set 4. Nevertheless,

both ratios (63.8% and 59.0%) were greater than 50%, indicating statistically robust results.

Taking precautions to ensure that the heightened performance is not due to the peculiarity of our data but to the quality of the chosen method, we tested how the results can be generalised to a limited dataset. Accordingly, 10,000 cross-validation tests were conducted for each working set.[91] Summaries of the working sets are presented in Tables 37–44.

For *Working Set 1* (Tables 37 and 38), the CART for the fake/forgery approach demonstrated a sensitivity of the validation set of 77.0% (Table 37), which was approximately equal to that of the sample (85.0%; Table 36). The specificities were similar (98.4% and 98.0%, respectively). The test validation results were good for the fake/forgery/doubtful approaches.

Table 37: Sensitivity and specificity for the CART 'Original vs. Fake/forgery' model approach (Working Set 1): summary of the cross-validation study (%)

Set 1	**Minimum**	**Q1**	**Median**	**Mean**	**Q3**	**Maximum**
Training Sensitivity	51.0	78.3	82.35	81.7	85.9	98.2
Training Specificity	97.4	98.5	98.8	98.8	99.0	100.0
Validation Sensitivity	37.7	72.1	78.3	**77.0**	83.1	95.6
Validation Specificity	96.0	98.0	98.4	**98.4**	98.8	100.0

The sensitivity of the validation set (84.6%; Table 38) was approximately equal to that of the sample set (91.0%; Table 37), and its specificity (98.1%; Table 38) was similar to that of the sample (98.5%; Table 36).

[91] Section 6.2.3 provides a detailed explanation of the interpretations of cross-validation processes.

Table 38: Sensitivity and specificity for the CART 'Original vs. Fake/forgery/Doubtful' model approach (Working Set 1): summary of the cross-validation study (%)

Set 1	**Minimum**	**Q1**	**Median**	**Mean**	**Q3**	**Maximum**
Training Sensitivity	72.5	85.0	87.5	87.5	90.0	98.0
Training Specificity	96.9	98.1	98.4	98.4	98.7	99.8
Validation Sensitivity	51.4	81.7	85.0	**84.6**	87.9	98.9
Validation Specificity	95.7	97.8	98.1	**98.1**	98.5	99.8

For *Working Set 2* (Tables 39 and 40), a comparison of the rates shows that the specificity (Table 39) of the validation set (99.2%) for the fake/forgery approach was almost equal to that of the sample (99.7%; Table 36). Its sensitivity was approximately (60.5%; Table 39) that of the sample (65.4%; Table 36).

Table 39: Sensitivity and specificity for the CART 'Original vs. Fake/forgery' model approach (Working Set 2): summary of the cross-validation study (%)

Set 2	**Minimum**	**Q1**	**Median**	**Mean**	**Q3**	**Maximum**
Training Sensitivity	39.6	60.9	65.5	65.3	70.3	88.2
Training Specificity	98.4	99.3	99.5	99.5	99.8	100.0
Validation Sensitivity	32.7	55.7	60.7	**60.5**	65.8	84.5
Validation Specificity	96.4	98.8	99.3	**99.2**	98.8	100.0

The specificity of the validation set for the fake/forgery/doubtful approach was 98.6% (Table 40). This proportion was close to the specificity of the sample (98.9%). The sensitivity (73.2%; Table 40) was similar to that of the sample (78.6%; Table 36).

Table 40: Sensitivity and specificity for the CART 'Original vs. Fake/forgery/Doubtful' model approach (Working Set 2): summary of the cross-validation study (%)

Set 2	**Minimum**	**Q1**	**Median**	**Mean**	**Q3**	**Maximum**
Training Sensitivity	54.0	73.5	77.3	76.6	80.2	93.1
Training Specificity	97.1	98.8	99.2	99.1	99.4	100.0
Validation Sensitivity	45.5	68.8	74.4	**73.2**	78.5	92.2
Validation Specificity	95.0	98.2	98.7	**98.6**	99.1	100.0

For *Working Set 3* (Tables 41 and 42), the results of the cross-validation set show that the specificity of the validation set is almost equal to that of the total sample for both approaches. The sensitivity of the validation set (73.2%; Table 41) was slightly higher than that of the total sample (70.0%; Table 36) for fake or forgery.

Table 41: Sensitivity and specificity for the CART 'Original vs. Fake/forgery' model approach (Working Set 3): summary of the cross-validation study (%)

Set 3	**Minimum**	**Q1**	**Median**	**Mean**	**Q3**	**Maximum**
Training Sensitivity	53.7	72.4	76.8	76.8	81.0	96.7
Training Specificity	98.2	99.4	99.7	99.6	99.9	100.0
Validation Sensitivity	38.0	69.0	73.0	**73.2**	77.3	92.9
Validation Specificity	96.2	99.1	99.5	**99.4**	99.9	100.0

For the category 'fake/forgery/doubtful', the validation set sensitivity (79.0%; Table 42) is approximately that of the total sample (80.9%; Table 36).

Table 42: Sensitivity and specificity for the CART 'Original vs. Fake/forgery/Doubtful' model approach (Working Set 3): summary of the cross-validation study (%)

Set 3	**Minimum**	**Q1**	**Median**	**Mean**	**Q3**	**Maximum**
Training Sensitivity	66.7	78.4	81.8	82.0	85.3	97.4
Training Specificity	97.6	99.4	99.8	99.6	99.9	100.0
Validation Sensitivity	54.9	75.3	78.6	**79.0**	82.3	97.9
Validation Specificity	95.6	98.2	98.8	**99.4**	99.9	100.0

For *Working Set 4* (Tables 43 and 44), the specificity of the validation set (99.9%; Table 43) was almost equal to that of the sample set (99.7%; Table 37). The sensitivity (52.5%; Table 43) was similar to that of the sample (63.8%; Table 36), and similarly for the fake/forgery/doubtful approach (Table 44).

Table 43: Sensitivity and specificity for the CART 'Original vs. Fake/forgery' model approach (Working Set 4): summary of the cross-validation study (%)

Set 3	**Minimum**	**Q1**	**Median**	**Mean**	**Q3**	**Maximum**
Training Sensitivity	33.3	50.8	55.0	55.5	60.3	80.6
Training Specificity	98.5	100.0	100.0	99.9	100.0	100.0
Validation Sensitivity	32.7	48.4	52.2	**52.5**	56.5	76.5
Validation Specificity	96.3	100.0	100.0	**99.9**	100.0	100.05

Table 44: Sensitivity and specificity for the CART 'Original vs. Fake/forgery/Doubtful' model approach (Working Set 4): summary of the cross-validation study (%)

Set 3	**Minimum**	**Q1**	**Median**	**Mean**	**Q3**	**Maximum**
Training Sensitivity	48.5	59.2	62.6	64.0	68.2	88.5
Training Specificity	97.4	99.7	100.0	99.8	100.0	100.0
Validation Sensitivity	45.5	57.4	60.5	**61.5**	64.6	84.0
Validation Specificity	94.8	99.5	100.0	**99.5**	100.0	100.0

In summary, the results showed that the performance of the validation sets approached that of the working sets (sample), confirming that the CART model was valid.

8 MAXIMIZING DATA MINING POTENTIAL BY EXPLORING ADDITIONAL SETS

This Chapter provides an additional assessment to maximise the potential for data mining. Given that we conducted the statistical analysis with the most representative variables, only half of the recorded variables were used. The remaining variables were not analysed. Therefore, it is necessary to check whether the remaining variables have discriminatory power to ensure that all possible niches have been revealed in the preliminary and primary phases. Because there are multiple variables, we must select sets of variables to optimise the results. Accordingly, additional sets were formed to analyse this issue. The following sections describe the selection procedure and assessment results.

8.1 Analyses of sets based on variables available in both sources of the database

Some variables in our database are available from both sources, while others are available only in the archive or the CR (see Appendix II)[92]. Consequently, we have marked those that are available in both sources by the label 'Both'. Therefore, it would be interesting to examine this category separately to determine whether new niches appear. This assessment is based on the CART model[93].

[92] For instance, the information that the painting has been restored or stolen is only available for the originals.

[93] PCA cannot be applied because multiple variables of this set have the value 0, making it mathematically impossible to perform the calculations.

a) Original vs. Fake/forgery

Figure 20 shows the classification tree when all the variables from the set with variables referred to in both sources are included as potential inputs.

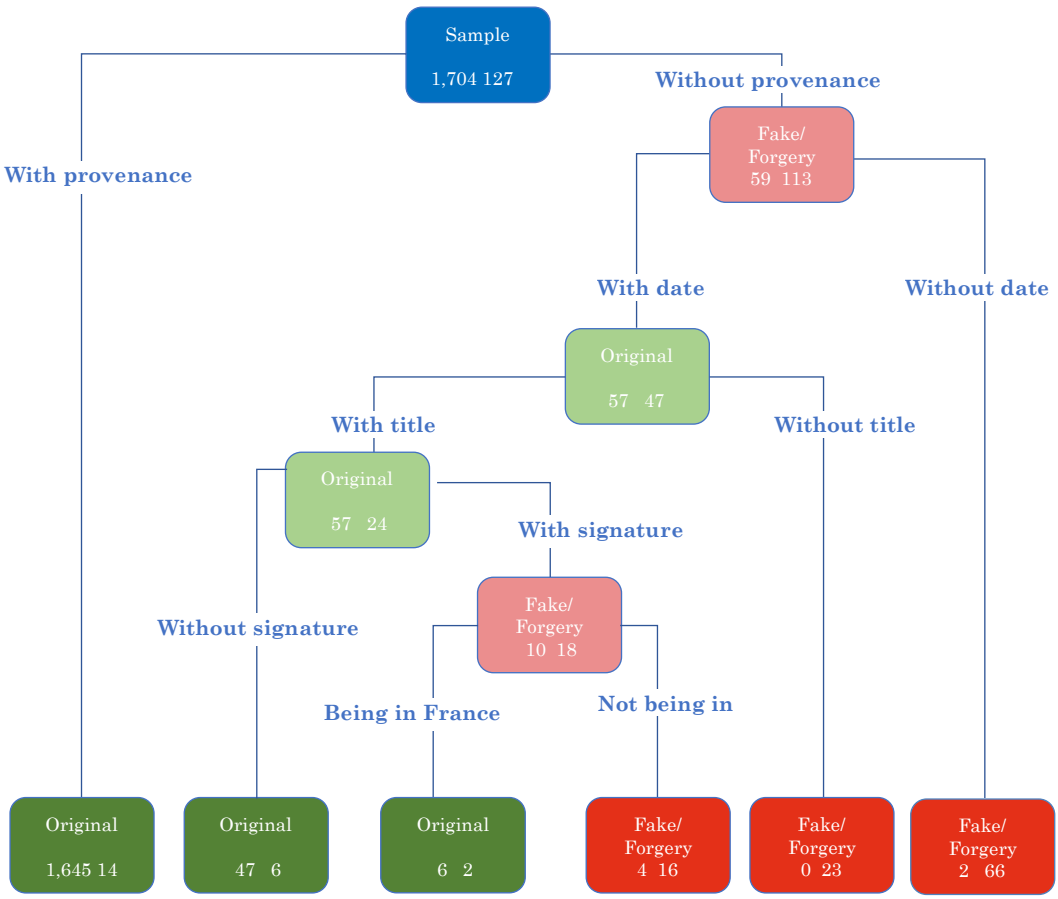


Figure 20: CART model: Original vs. Fake/forgery
(Set of variables referring to both sources)

Each node corresponds to a set of rules, as listed in Table 45.

Table 45: Rules and interpretations for the Original vs. Fake/forgery approach (Set of variables from both sources)

Node	Rule	Interpretation
1	Having a provenance	Artwork with a provenance has a 99% probability of being original.
2	Not having a provenance + Having a date + Having a titled + Having a signature/seal	When artwork does not have a provenance but has a date, a title, and having a signature/seal, the probability of being original is 89%.
3	Not having a provenance + Having a date + Having a titled + Not having a signature/seal + Being in France	Artwork without a provenance and signature, with a date and a title, and whose owner is in France has a 75% probability of being original.
4	Not having a provenance + Having a date + Having a titled + Not having a signature/seal + Being in France	Artwork without a provenance and signature, with a date and a title, whose owner is not in France has an 80% probability of being a fake or forgery.
5	Not having a provenance + Having a date + Not having a titled	Artwork without a provenance, without a title, and with a date has a 100% probability of being a fake or forgery.
6	Not having a provenance + Not having a date	Artwork without a provenance and without a date has a 97% probability of being a fake or forgery.

b) *Original vs. Fake/forgery/Doubtful*

Figure 21 shows the classification tree when all the variables from the set with variables referred to in both sources are included as potential inputs.

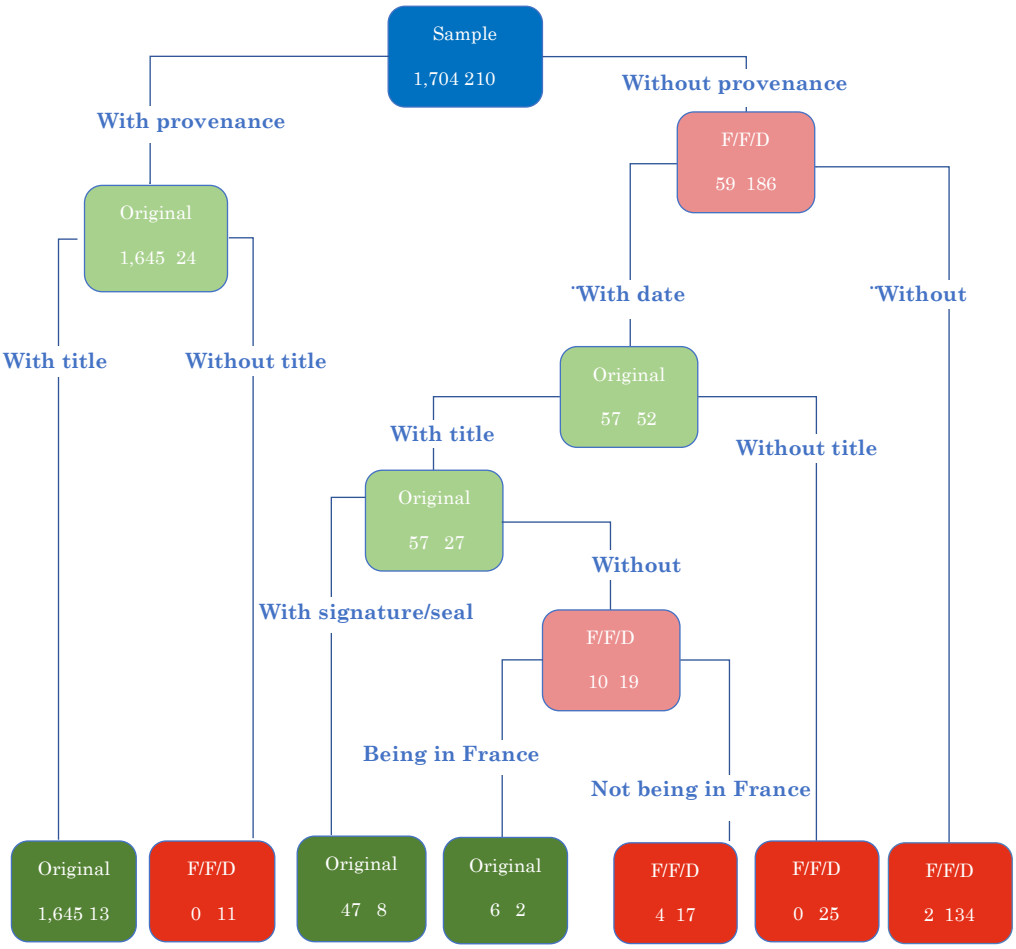


Figure 21: Original vs. Fake/forgery/Doubtful
(Set of variables refer to both sources)

Each node corresponds to a set of rules identifying the works of art assigned to each terminal node, as listed in Table 46.

Table 46: Rules and interpretations for the Original vs. Fake/forgery/ Doubtful approach (Set of variables referring to both sources)

Node	Rule	Interpretation
1	Having a provenance+ Having a title	Artwork with a provenance and a title has a 99% probability of being original.
2	Having a provenance + Not having a title	Artwork with a provenance but without a title has a 100% probability of being a fake/forgery/doubtful.
3	Not having a provenance + Having a date + Having a title + Having a signature/seal	When artwork does not have a provenance but has a date, a title, and a signature/seal, the probability of being original is 85%.
4	Not having a provenance + Having a date + Having a title + Not having a signature/seal + Being in France	Artwork without a provenance and signature, with a date and title, whose owner is in France has a 75% probability of being original.
5	Not having a provenance + Having a date + Having a title + Not having a signature/seal+ Not being in France	Artwork without a provenance and signature, with a date and title, whose owner is not in France has an 81% probability of being a fake/forgery/doubtful.
6	Not having a provenance + Having a date + Not having a title +	Artwork without a provenance, without a title, and with a date has a 100% probability of being a fake/forgery/doubtful.
7	Not having a provenance + Not having a date +	Artwork without a provenance and without a date has a 99% probability of being a fake/forgery/doubtful.

8.2 Analyses of a set including the relabelled variable 'information on back'

As explained in Section 3.2.2.2, visible signs of ownership can be represented by various inscriptions such as marks and labels on the front or back of a painting. These signs provide evidence of the authenticity of an art object. Each sign requires detailed interpretation to clarify the story of the art object. Therefore, we assume that the information on the back of the paintings may have explanatory power to differentiate between the original and fake works. In our database, this information corresponds to the variable group 'Back of the painting', which specifies the different signs on the back. However, this information is scarce in both sources (see Table 16). Consequently, to effectively exploit few information, all variables of this group – 'number/letter', 'sticker', 'seal', 'signature/initials', 'written on the frame', 'written on the canvas', 'written with errors or with letters other than artist's initials', and 'picture' – has been re-labelled in the new variable 'information on the back'.

Thus, the re-labelled variable constitutes a sign of ownership without specifying the type of inscription. Next, we add the variable 'information on the back' to the set with variables common to both sources[94] used in Section 8.1. The set was then analysed using the CART model.

[94] The PCA model cannot be used for the reason explained in the previous section.

a) *Original vs. Fake/forgery*

Figure 22 shows the classification tree when all variables from the set of dichotomous variables including the re-labelled variable 'information on the back' are considered potential input.

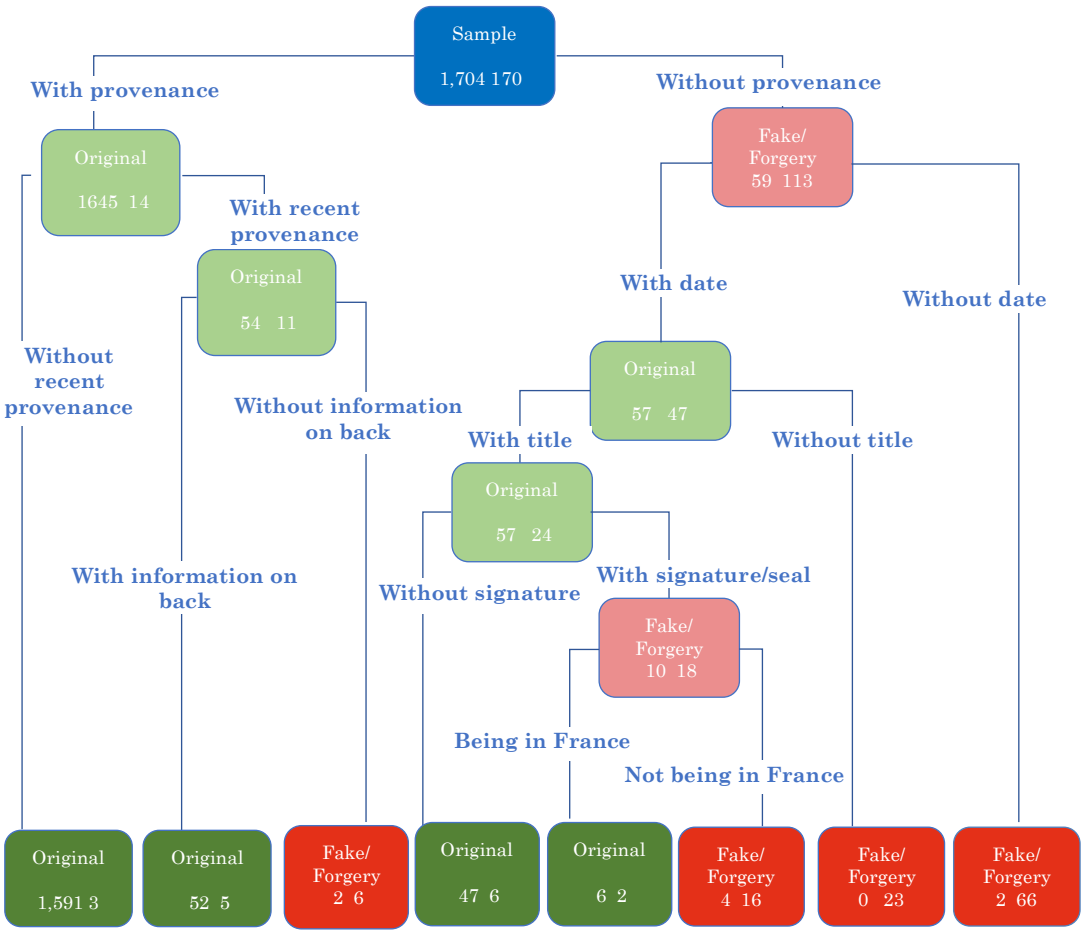


Figure 22: CART model: Original vs. Fake/forgery
(Set of variables refer to both sources including the re-labelled variable 'information on the back')

Each terminal node corresponds to a set of rules identifying the works of art as listed in Table 47.

Table 47: Rules and interpretations for the Original vs. Fake/forgery approach
(Set of variables refer to both sources including the re-labelled variable 'information on the back')

Node	Rule	Interpretation
1	Having a provenance + Not having recent provenance	Artwork with a provenance and without recent provenance has a 99% probability of being original.
2	Having a provenance + Having a recent provenance + Having information on back	Artwork with a provenance, with information on the back, and with recent provenance has a 91% probability of being original.
3	Having a provenance + Having a recent provenance + Not having information on back	When an artwork has a provenance, without information on the back, and with recent provenance, the probability of being fake/forgery is 75%.
4	Not having a provenance + Having a date + Having a title + Not having a signature/seal +	Artwork without a provenance and signature/seal, but with a date and title has a 89% probability of being original.
5	Not having a provenance + Having a date + Having a titled + Having a signature/seal + Being in France	Artwork without a provenance and signature, with a date and title, and whose owner is in France has a 75% probability of being original.
6	Not having a provenance + Having a date + Having a titled + Having a signature/seal + Not being in France	Artwork without a provenance and signature/seal, with a date and title, and whose owner is not in France has an 80% probability of being
7	Not having a provenance + Having a date + Not having a titled	Artwork without a provenance, without title, and with a date has a 100% probability of being a fake or forgery.
8	Not having a provenance + Not having a date +	Artwork without a provenance and with a date has a 97% probability of being a fake or forgery.

b) *Original vs. Fake/forgery/Doubtful*

Figure 23 shows the classification tree when all variables from the set of dichotomous variables including the re-labelled variable 'information on back'.

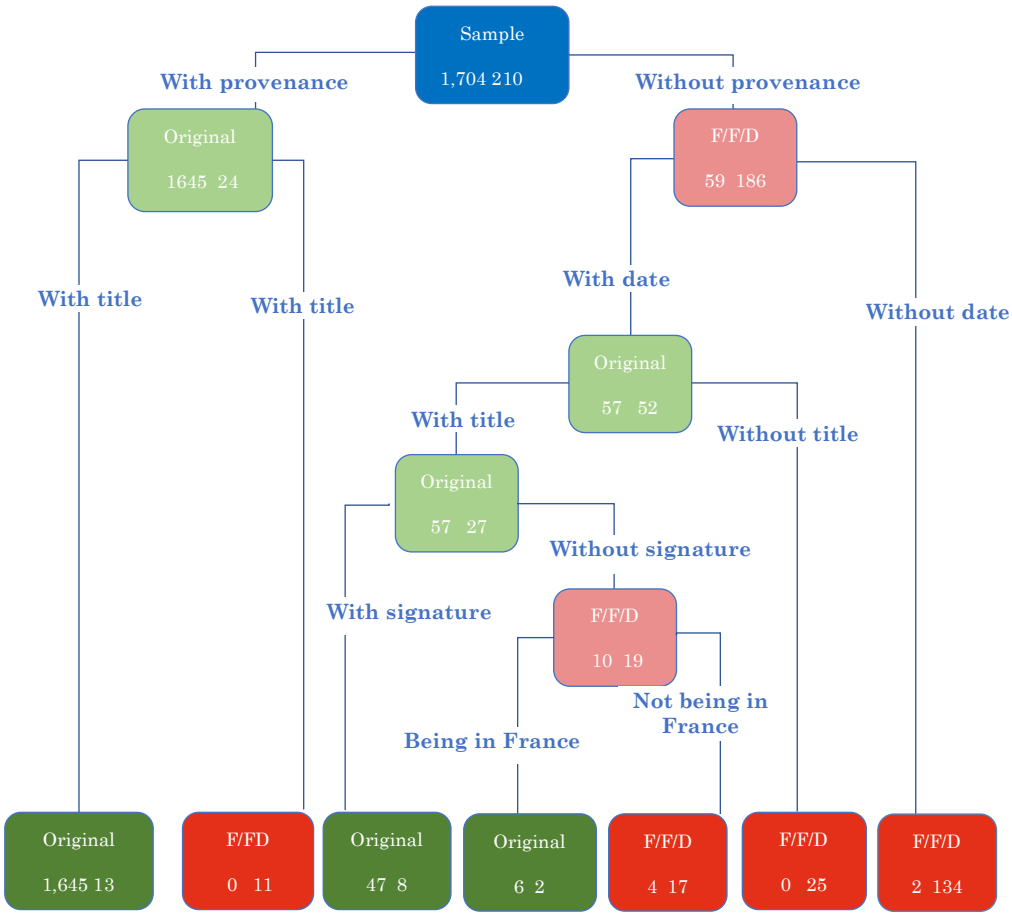


Figure 23: CART model: Original vs. Fake/forgery/Doubtful
(Set of variables refer to both sources including the re-labelled variable 'information on the back')

The interpretation of this tree is identical to that of the working tree of the set referring to both sources (see Figure 21).

8.3 CART validation of the analyses of the additional sets

The CART validation for the additional sets is made with the help of the calculation of the correct classification. The sensitivity and specificity rates for the samples are listed in Table 48.

Table 48: Sensitivity and specificity for the CART's approaches: summary of the sample study

Supplementary set	**Original vs Fake/forgery (%)**	**Original vs Fake/forgery/Doubtful (%)**
Set*		
Specificity	99.6	99.6
Sensitivity	82.7	89.0
Set**		
Specificity	99.5	99.6
Sensitivity	82.7	89.0

* Set of all variables present in both sources.

** Set of all variables present in both sources including the re-labelled 'information on back' variable

Table 48 shows that the specificities and the sensitivities are close to 100%. Such proportions of correct classification mean that the CART's performance with both sets is robust.

The summaries of the 10,000 cross-validation tests are displayed in Tables 49 to 52.

Table 49: Sensitivity and specificity for the CART 'Original vs. Fake/forgery' model approach (Set of variables refer to both sources): summary of the cross-validation study (%)

Set*	**Minimum**	**Q1**	**Median**	**Mean**	**Q3**	**Maximum**
Training Sensitivity	62.5	79.3	82.4	85.2	85.5	96.8
Training Specificity	98.3	99.3	99.4	99.5	99.6	100.0
Validation Sensitivity	46.7	73.3	77.9	**77.5**	82.5	96.2
Validation Specificity	96.0	99.2	99.3	**99.3**	99.5	100.0

Table 50: Sensitivity and specificity for the CART 'Original vs. Fake/forgery/Doubtful' model approach (Set of variables refer to both sources): summary of the cross-validation study (%)

Set*	**Minimum**	**Q1**	**Median**	**Mean**	**Q3**	**Maximum**
Training Sensitivity	73.5	84.2	86.8	86.7	89.4	98.0
Training Specificity	97.9	99.2	99.4	99.4	99.6	100.0
Validation Sensitivity	60.8	79.4	82.9	**82.8**	86.3	96.9
Validation Specificity	95.4	99.1	99.3	**99.3**	99.5	100.0

Table 51: Sensitivity and specificity for the CART 'Original vs. Fake/forgery' model approach (Set of variables referring to both sources including the variable 'information on the back'): summary of the cross-validation study (%)

Set**	**Minimum**	**Q1**	**Median**	**Mean**	**Q3**	**Maximum**
Training Sensitivity	62.5	79.4	82.6	82.5	85.7	96.8
Training Specificity	98.3	99.3	99.4	99.5	99.6	100.0
Validation Sensitivity	45.7	73.2	77.9	**77.5**	82.4	98.1
Validation Specificity	96.0	99.2	99.3	**99.3**	99.5	100.0

Table 52: Sensitivity and specificity for the CART 'Original vs. Fake/forgery/Doubtful' model approach (Set of variables referring to both sources including the variable 'information on the back'): summary of the cross-validation study (%)

Set**	**Minimum**	**Q1**	**Median**	**Mean**	**Q3**	**Maximum**
Training Sensitivity	73.5	84.2	86.8	86.7	89.4	98.0
Training Specificity	97.9	99.2	99.4	99.4	99.6	100.0
Validation Sensitivity	60.8	79.4	82.9	**82.8**	86.3	96.9
Validation Specificity	95.4	99.1	99.3	**99.3**	99.5	100.0

Overall, the specificities of the validation sets are almost equal to those of the samples, and the sensitivities are approximately equal to those of the samples. Therefore, the CART model is valid for both sets.

8.4 Summary of the analyses of the additional sets

The results of the extended statistical analyses provide evidence of the completeness of the preliminary and main analyses. Indeed, according to the tree configuration of both sets, the variable 'France' has power to discriminate original from fake works. The rules that include this variable are interpreted for niches N4 and N5 in Table 47. Furthermore, the classification tree of the set with re-labelled variable indicates the variable 'information on the back', which reveals new rules of probability (niches N2 and N3 in Table 47). Therefore, only two variables that were not included in the previous analysis provided new information for the model.

Additionally, we experimented with other combinations of variables. Notably, we examined whether there were correlations between the most representative variables of the experimental set and the 'information on the back'. In parallel, we tested whether novel niches emerged when CART analysed all discriminatory variables from the working and experimental sets by grouping them into a new set.[95] No new niches were found in either of the experiments.

[95] This set includes the variables: 'without provenance', 'untitled', 'unknown technique', '1882–1889', 'without the date on the painting', 'Switzerland', '1890–1900', 'sale online', '1901–1910', 'large', 'landscapes', 'commonplace book', 'cold', and 'monogrammed'.

9 CONSTRUCTING A PREVENTION TOOL AND TESTING THE MODEL'S UTILITY

In this section, we develop an explanatory scheme with guidelines for practical applications and test the model using two case studies. The first case study concerns a painting that the Foundation's experts classified as a forgery. The second case focuses on a painting that the experts attributed to Vallotton. We provide a step-by-step explanation of each element in the manual application of the model. Finally, we examine the question of the model's standardisation using the CART approach.

9.1 Developing guidelines for building the model

This step provides an optimal framework for presenting niches. We need to determine which experimental results should be retained for further use. For this purpose, the niches that contain the most information are compared. They include more than 50% originals and more than 50% fake works. We grouped the rules into three tables, indicating the corresponding probability rates and repetition times in the experimental analysis (Table 53, Table 54, Table 55).

Table 53: Rules for the likelihood of being an original

Rule	**Probability %**	**Repeated times**
Having a *provenance*	99	3
Having a *provenance* + Having a *title*	92 ⇔ 99	3
Having a *title* + Having a *date*	98	3
Having a *title*	95	1

Table 54: Rules for the likelihood of being a fake/forgery

Rule	**Probability %**	**Repeated times**
Not having a *provenance* + Not having a *date*	97	2
Not having a *title*	100	2

Table 55: Rules for the likelihood of being a fake/forgery/doubtful

Rule	**Probability %**	**Repeated times**
Not having a *provenance* + Not having a *date*	99	2
Not having a *provenance* + Not having a *title*	100	1
Not having a *title*	100	2

A comparison of the rules shows that the most frequent characteristics with discriminatory power for both categories are *provenance* and *title*. In other words, a work of art is likely to be original when its provenance and title are noted in a document which is the source of this information. By contrast, the absence of these characteristics should considerably increase vigilance.

The variable 'provenance' implies that the chain of owners (or possessors) of the artwork is noted in the source. It is crucial to understand that this variable should not be interpreted in the sense of provenance confirmation; it is merely a statement that the owners' names are noted in the source. Consequently, when the source does not contain any information about the owners, this information is interpreted as 'without a provenance'. When several owners are noted, it should be interpreted as 'with a provenance'. A chain of ownership implies that the artwork has changed ownership several times. However, in some cases, the source may provide information about a single owner. Such information also corresponds to the variable 'with a provenance'. To be precise, we created two supplementary variables. One is 'recent provenance', which indicates paintings for which the ownership chronology begins long after the

painting first appeared on the art market. The other is 'first provenance', corresponding to paintings acquired directly from the artist and for which there is no complete chronology of ownership. Their interpretation is better explained using examples (see Section 5.3.3). Accordingly, the variables 'with a first provenance' and 'with a recent provenance' are figuratively similar to the variable 'with provenance', and vice versa: 'without a first provenance' and 'without a recent provenance' are equivalent to 'without a provenance'. This interpretation could be complicated for external users of the model. Thus, this interpretation can be simplified. Indeed, since these two supplementary variables are absorbed by the variable 'provenance', we suggest facilitating their interpretation as follows: if information about one or several owners is noted in the source, it should be interpreted as 'with provenance'; if the document does not contain any information about owners, it should be interpreted as 'without provenance'.

Likewise, the variable 'monogrammed' is absorbed by the variable 'signature'. It means the variant 'with a signature' implies that the artwork is signed by the signature 'Vallotton' or the monogram 'FV'. Consequently, the variable 'without a signature' means that either abbreviation of the name of the artist or his signature (full name) is observable in the source.

To better understand how to interpret the 17 variables for the manual approach evaluation, we propose a series of questions with options for possible answers (see Appendix V) to guide the user in transforming raw data into variables. Sections 9.2 and 9.3 will explain how to use questions to evaluate artworks in practice.

Furthermore, for easier exploitation, all niches are synthesised in the seven-block worksheet in Table 56 from the following tables:

1. Experimental set: Table 20, Table 21
2. Working Set 1: Table 25, Table 26
3. Working Set 2: Table 28, Table 29
4. Working Set 3: Table 31, Table 32
5. Working Set 4: Table 34, Table 35

6. Set of variables referring to both sources: Table 45, Table 46
7. Set of variables with re-labelled 'information on the back': Table 47

Table 56: Grouped CART tables

1.

Niche No.	**1** 1,645 (O)	**2** 56 (O)	**3** 9 (F)	**4** 23 (F)	**5** 66 (F)	
Rules: experimental set: O (original) vs F (fake/forgery)	With a provenance = 99% (O)	Without provenance + With a date + With title + Not in Switzerland = 79% (O)	Without a provenance + With a date + With title + In Switzerland = 90% (F)	Without a provenance + With a date + Without a title = 100% (F)	Without a provenance + Without a date = 97% (F)	
Niche No.	**1** 1,645 (O)	**2** 11 (F/D)	**3** 56 (O)	**4** 9 (F/D)	**5** 25 (F/D)	**6** 134 (F/D)
Rules: experimental set: O (original) vs F/D (fake/forgery/doubtful)	With a provenance + With a title = 92% (O)	With a provenance + Without a title = 100% (F/D)	Without a provenance + With a date + Without a title + Not in Switzerland = 76% (O)	Without a provenance + With a date + With a title + In Switzerland = 90% (F/D)	Without a provenance + With a date + Without a title = 100% (F/D)	Without a provenance + Without a date = 99% (F/D)

2.

Niche No.	**1** 1,645 (O)	**2** 25 (O)	**3** 8 (O)	**4** 48 (F)	**5** 60 (F)	
Rules: **Working Set 1: O vs F**	With a provenance = 99% (O)	Without a provenance + With a title + With a known technique = 89% (O)	Without a provenance + With a title + Unknown technique + In 1882–1889 = 90% (O)	Without a provenance + With a titled + With an unknown technique + Not in 1882–1889 = 66% (F)	Without a provenance + Without a title = 100% (F)	
Niche No.	**1** 1,645 (O)	**2** 11 (F/D)	**3** 25 (O)	**4** 8 (O)	**5** 67 (F/D)	**6** 113 (F/D)
Rules: **Working Set 1: O vs F/D**	With a provenance + With a title = 99% (O)	With a provenance + Without a title = 100% (F/D)	With a provenance + With a title + With a known technique = 86% (O)	Without a provenance + With a title + Unknown technique + In 1882–1889 = 80% (O)	Without a provenance + With a title + Unknown technique + Not in 1882–1889 = 72% (F/D)	Without a provenance + Untitled = 100% (F/D)

3.

Niche No.	**1** 1,574 (O)	**2** 68 (O)	**3** 26 (O)	**4** 12 (O)	**5** 11 (O)	**6** 8 (O)
Rules: **Working Set 2:** **O vs F**	With a title + With a date = 98% (O)	With a title + Without a date + In 1890–1900 = 87% (O)	With a title + Without a date + Not in 1890–1900 + In Switzerland = 90% (O)	With a title + Without a date + Not in 1890–1900 + Not in Switzerland + Not being sale/sale online + In 1901–1910 = 92% (O)	With a title + Without a date + Not in 1890–1900 + Not in Switzerland + Not being sale + Not in 1901–1910 + Large = 73% (O)	With a title + Without a date + Not in 1890–1900 + Not in Switzerland + Not being sale/sale online + Not in 1901–1910 + Not a large + Not being a landscape = 57% (O)
Niche No.	**7** 6 (F)	**8** 13 (F)	**9** 64 (F)			
Rules: **Working Set 2:** **O vs F**	With a title + Without a date + Not in 1890–1900 + Not in Switzerland+ Not being sale/sale online + Not in 1901–1910 + Not a large + Landscape = 75%(F)	With a title + Without a date + Not in 1890–1900 + Not in in Switzerland+ Being sale/sale online = 81%(F)	Without a title = 100%(F)			
Niche No.	**1** 1,574 (O)	**2** 68 (O)	**3** 15 (O)	**4** 15 (O)	**5** 14 (O)	**6** 41 (F/D)
Rules: **Working Set 2:** **O vs F/D**	With a title + With a date = 98% (O)	With a title + Without a date + In 1890–1900 = 87% (O)	With a title + Without a date + Not in 1890–1900 + In commonplace book = 100% (O)	With a title + Without a date + Not in 1890–1900 + Not in commonplace book +	With a title + Without a date + Not in 1890–1900 + Not in commonplace book +	With a title + Without a date + Not in 1890–1900 + Not in commonplace book + Not in 1901–1910 +

				In 1901–1910 = 88% (O)	Not in 1901–1910 + In Switzerland = 61% (O)	Not in Switzerland = 69% (F/D)
Niche No.	**7** 124 (F/D)					
Rules: Working Set 2: O vs F/D	Without a title = 100 (F/D)					

4.

Niche No.	**1** 1,645 (O)	**2** 57 (O)	**3** 23 (F)	**4** 66 (F)	
Rules: Working Set 3: O vs F	With a provenance = 99%(O)	Without a provenance + With a date + With a title = 70%(O)	Without a provenance + With a date + Without a title = 100%(F)	Without a provenance + Without a date = 92%(F)	

Niche No.	**1** 1,645 (O)	**2** 11 (O)	**3** 57 (O)	4 25 (F)	**5** 134 (F)
Rules: Working Set 3: O vs F/D	With a provenance + With a title = 92% (O)	With a provenance + Without a title = 100% (F/D)	Without a provenance + With a date + With a titled = 68% (O)	Without a provenance + With a date + Without a title = 100% (F/D)	Without a provenance + Without a date = 99% (F)

5.

Niche No.	**1** 1,574 (O)	**2** 31 (O)	**3** 94 (O)	**4** 10 (F)	**5** 7 (F)	**6** 64 (F)
Rules: **Working Set 4:** **O vs F**	With a title + With a date = 98% (O)	With a title + Without a date + Without a monogram + Being sold = 94% (O)	With a title + Without a date + Without a monogram + Not being sold + Not being sale/online = 86% (O)	With a title + Without a date + Without a monogram + Not being sold + Being sale/online = 100% (F)	With a title + Without a date + Without a monogram = 58% (F)	Not having a title = 100% (F)

Niche No.	**1** 1,704 (O)	**2** 124 (F/D)
Rules: **Working Set 4:** **O vs F/D**	With a title = 95% (O)	Without a title = 100% (F/D)

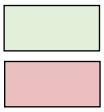
6.

Niche No.	**1** 1,645 (O)	**2** 47 (O)	**3** 6 (O)	**4** 16 (F)	**5** 23 (F)	**6** 66 (F)
Rules: Working set of variables refers to both sources: O vs F	With a provenance = 99% (O)	Without a provenance + With a date + With a title + With signature/seal = 89% (O)	Without a provenance + With a date + With a title + Without a signature/seal + In France = 75% (O)	Without a provenance+ With a date + With a title + Without a signature/seal + In France = 80% (F)	Without a provenance + With a date + Without a title = 100% (F)	Without a provenance + Without a date = 97% (F)
Niche No.	**1** 1,645 (O)	**2** 11 (F/D)	**3** 47 (O)	**4** 6 (O)	**5** 17 (F/D)	**6** 25 (F/D)
Rules: Working set of variables refers to both sources: O vs F/D	With a provenance + With a title = 99% (O)	With a provenance + Without a title = 100% (F/D)	Without a provenance With a date + With a title + With a signature/seal = 85% (O)	Without a provenance With a date + With a title + Without a signature/seal + In France = 75% (O)	Without a provenance With a date + With a title + Without a signature/seal + Not in France = 81% (F)	Without a provenance With a date + Without a title = 100% (F)
Niche No.	**7** 134 (F/D)					
Rules: Working set of variables refers to both sources: O vs F/D	Without a provenance + Without a date = 99% (F)					

7.

Niche No.	**1** 1,591 (O)	**2** 52 (O)	**3** 6 (F)	**4** 47 (O)	**5** 6 (O)	**6** 16 (F)
Rules: Working set with dichotomous variables including 'back': O vs F Figure 22	With a provenance + Without recent provenance = 99% (O)	With a provenance + With a recent provenance + Having information on back = 91% (O)	With a provenance + With a recent provenance + Not having information on back = 75% (F)	Without a provenance + With a date + With a title + Without a signature/ seal = 89% (O)	Without a provenance + With a date + With a titled + With a signature/seal + In France = 75% (O)	Without a provenance + With a date + With a titled + With a signature/seal + In France = 80% (F)
	7 23 (F)	**8** 66 (F)				
Rules: Working set with dichotomous variables including 'back': O vs F Figure 22	Without a provenance + With a date + Without a titled = 100 (F)	Without a provenance + Without a date = 97% (F)				

The worksheet contains the row 'Niche' corresponding to the table's number (No.). The "rules" row indicates rules conforming to the number of niches. Green specifies a niche with a predominant probability of being the original, and red indicates a predominant probability of being fake or a forgery/doubtful.[96] The letters O, F, and F/D specify the category.



Colour of niches with original works

Colour of niches with fake works

In addition, a schematic illustrating the logic behind using this model is presented. The diagram contains the following abbreviations.

A – characteristic of the questionable artwork
B – characteristic of the model that user can find in the worksheet
Y – niche
E – likelihood
X – questionable artwork

To illustrate how to read this diagram, we propose that the reader imagine an art player, such as a collector, who would like to evaluate a work of art that is not in the catalogue raisonné. He asks for the probability that the artwork is not original. To answer this, the collector must compare the characteristics (A_n) of the painting in question with those specified by our model (B_n). If they correspond to the characteristics of one or more niches (Y), then the niche indicates the probability (E) that the work is original or fake.

[96] While we will use the worksheet in the next sections to evaluate two case studies, it contains additional colours: grey and blue. Blue corresponds to the niches that match the characteristics of case study 1, and grey to the characteristics of case study 2.

Schematically, the user of this worksheet needs to (1) check how the artwork's characteristics fit into the 17 variables included and (2) search for the niches to which these characteristics belong to find the corresponding probability.

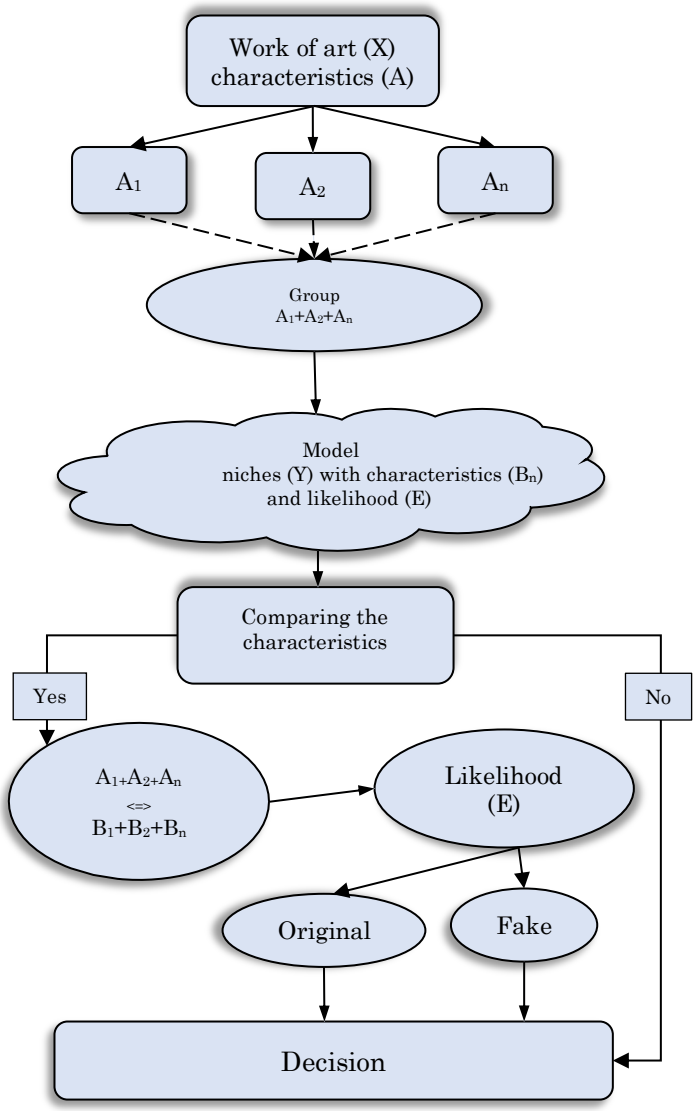


Figure 24: Scheme of the reasoning of the theoretical model

To better illustrate the assessment using the worksheet and diagram, we evaluated two studies that were not included in the database.[97] The first case study concerns a painting that the Foundation's experts classified as a fake work, the second a painting

[97] Since the Vallotton experts authenticated them after this doctoral research had begun, they were not included in the database.

that the experts attributed to Vallotton. The focus is on the model's operation and comparing the results of this evaluation with those of the experts.

9.2 Test of the tool with Case Study 1: picture without title

a) Artwork's record[98]

The Foundation's experts detected a painting supposedly made by Vallotton on an online sales platform[99].

Photo: Front of the painting

The experts alerted the sales management because of substantial doubts about the authenticity of this object. A few days later, the painting was withdrawn from sale. At the same time, the presumed owner of the painting in question wrote to the Foundation, asking them to confirm its authenticity. He emailed photographs of the painting, front and back.

A brief explanation stated that the painting had been purchased from a private collection (without an indication of the collector's name). The oil painting on canvas is

[98] These data are derived from the communication recorded in the files and the explanations provided by the experts.

[99] The announced price was EUR 5,000. Such a price is quite low compared to the average market price for Vallotton's works (see Section 2.2)

untitled and measured 65 × 82.5 cm – a signature dated 'F. Vallotton 99' is in the lower right corner. The theme of the picture is *the interior of a house*. There are several inscriptions on the back: '1865–1925', the number '344', a stock label of Germany, and the number '62550' with the inscription 'Bourse France'. The owner lives in France.

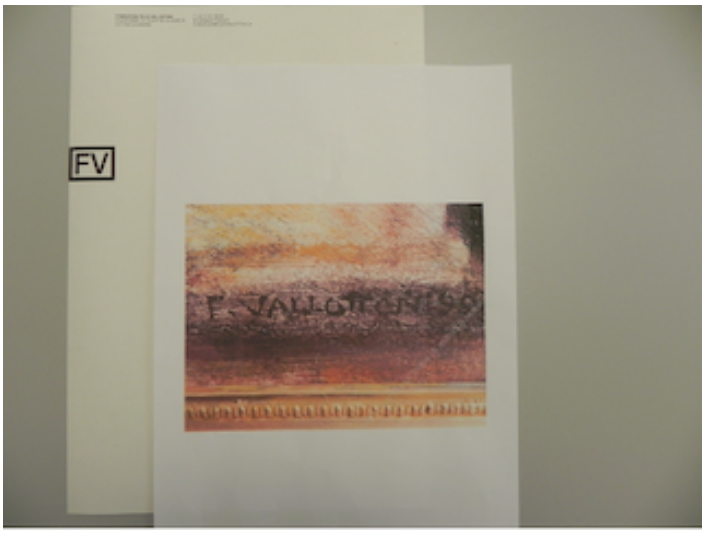


Photo: Fragment with the signature (front of the painting)

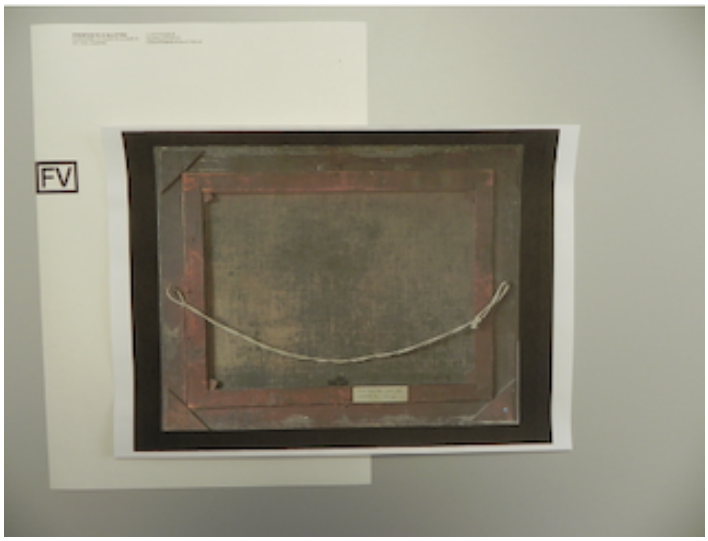


Photo: Back of the painting

No further details regarding the origin of this work are available. Additionally, there was no information on what happened once the sales platform received the message from the experts. We can only assume that the painting owner was urged to provide an authentication certificate and perhaps advised to contact the Foundation. In

addition, the owner ceased all contact with the Foundation after the Foundation sent him an email stating their refusal to certify the painting. Similarly, the owner did not request on-site inspection of the painting. The experts provided their opinion [100] without direct observation.

b) *Sampling of variables*

Based on the information mentioned in point *a*, we define the characteristics the list of questions (see Appendix V). These characteristics must be transformed into variables. This information may be noted in the document (a source of information) or observed in a photo of the artwork. To process the sampling of the variables, the model user is asked to determine the presence or absence of information. For example, to interpret provenance data, the user of the model should question whether one or more owners are noted in the source. If yes, such data is transformed to the variable 'with provenance'; if not, it should be interpreted as 'without provenance'. If the current owner is known, this data should be interpreted in two variables 'with provenance' and 'with recent provenance'. In contrast, it should be interpreted as 'without provenance' and 'without recent provenance' (see Section 9.1).

The question that corresponds to a variable can be obtain with help of the questions like: 'Is a date noted in the source or observable in the photograph of the work?' or 'Is a signature or monogram observable on the artwork's photo (or mentioned in the source)?' We explain the logic of the variable interpretation in detail in Appendix VI. Consequently, the variables used for evaluation were as follows:

Without a title
With a date
Not in Switzerland
In France
Sale online
With a signature
Without a monogram

[100] The Foundation's experts evaluate works of art through the connoisseur's stylistic analysis and the historical documentation detailed in Section 3.2.1.

Large
Having information on back
Not a landscape
Not in 1882–1889
Not in 1901–1910
Having a known technique
Not in commonplace book
Without a provenance
Without recent provenance

c) *Evaluation*

To obtain essential awareness through the likelihood, the user must check whether a combination of the variables of the final list matches one or several niches grouped in Table 56. For example, a niche comprises the following variables.

Without a provenance +
With a date +
With a title +
In Switzerland = 90% (F)

The variables in the list were compared with those in this niche (see Figure 24). In this example, only one variable – 'without a provenance' – matches our list. Indeed, the other three variables – 'with a date', 'with a title', 'in Switzerland' – do not match. Thus, this niche is not useable for prediction. However, the other niche may comprise the fitted combination of variables. For example, the niche below contains three variables identical to those in our list.

Without a provenance +
With a date +
Without a title = 100% (F)

Thus, prediction of this niche was valuable in our case: the group of three variables indicated a 100% probability of being fake. Because several niches may match our list

of variables, we marked every related niche in *blue* in the worksheet. At the end of this comparison, we found 10 niches[101], all of which indicated a high probability (92%–100%) of being fake.

As mentioned under point *a*, the experts who examined this painting practiced the connoisseurship technique (see Section 3.2.1). Trained visual memory allows them to make comparisons, orient their perception, and classify the artwork as fake. According to our evaluation, the work of art in question has characteristics that distinguish a high likelihood of being a doubtful or fake work. Hence, the evaluation of the model was compatible with expert opinions.

9.3 Test of the tool with Case Study 2: picture with title

a) *Authentication survey*

In 2016, the Cantonal Museum of Fine Arts in Lausanne announced its entry into its collection of an unpublished and newly identified painting by Félix Vallotton. The work in question, titled *Mer haute, Villerville*, was authenticated by the Félix Vallotton Foundation (Fibicher & Poletti, 2016). The authentication survey was published as a press release for a conference organised by the Cantonal Museum of Fine Arts of Lausanne [102] (Fibicher & Poletti, 2016). A brief explanation is provided below.

An art dealer contacted the Foundation requesting an appraisal of a painting by Vallotton signed 'Vallotton. 02', which is not included in the catalogue raisonné and has always been in the same private collection. The Foundation asked him for a photograph of the work. The first impression of Vallotton's previous attribution seemed correct. However, first impressions alone are not sufficient to authenticate it as an original work.

[101] They are marked in blue colour in Table 54.

[102] The Cantonal Museum of Fine Arts of Lausanne is considered the leading international institution in conserving artists' works (Fibicher & Poletti, 2016).

Photo: Front of the painting[103]

Félix Vallotton (Lausanne, 1865–Paris, 1925),
'High sea, Villerville' (*Mer haute, Villerville*), 1902,
Oil on board, 34 × 59.5 cm,
Museum Cantonal of Fine Arts, Lausanne.
Long-term deposit from the Collection
of Dr Marcel Bahro, 2015. Inv. 2015-030
© Cantonal Museum of Fine Arts of Lausanne

The fact remains that Vallotton often did not mention small-format landscapes individually in his commonplace book. Notably, the commonplace book mentions nine landscapes, but only seven have been found. Given the date of the work, its format, and its formal proximity to a landscape in this series, the experts hypothesised that this painting could be part of the same series as the painting of Vallotton titled *Mer basse, Villerville*, which is preserved at the Kunsthalle in Bremen and could correspond to one of the missing landscapes.

To test this hypothesis, the Foundation asked for a photo of the back of the painting, as it could contain essential evidence (see the photo below). The reverse side shows a title and date, 'Mer haute Villerville 1902'; a number 28; a number 53; a label 'Kunstsalon Emil Richter Dresden'. Within the framework of a crosscheck with archival documents (notebooks, lists, or exhibition catalogues), the indications on the back make it possible to identify this work of art. While the artist's hand inscriptions give the work's title and confirm its originality, the other three indications refer to exhibitions.

[103] In French : *Félix Vallotton (Lausanne, 1865–Paris, 1925), 'Mer haute, Villerville', 1902, Huile sur carton, 34 × 59.5 cm, Musée cantonal des Beaux-Arts de Lausanne. Dépôt à long terme de la Collection du Dr Marcel Bahro, 2015. Inv. 2015-030 © Musée cantonal des Beaux-Arts de Lausanne* (in French).

Photo: Back of the painting
Félix Vallotton (1865–1925), 'High sea, Villerville', 1902,
Foundation Vallotton[104]

After some research, the following cross-references were made:

- Number 53 corresponds to the *Vallotton and Vuillard Exhibition* held in 1903 at the Galerie Bernheim-Jeune in Paris, whose catalogue mentions the hitherto unidentified painting '*Marée haute (Villerville)*' under number 53.
- The label 'Kunstsalon Emil Richter Dresden' refers to a personal exhibition of Vallotton's work in 1910 without a catalogue. The paintings in the exhibition were identified based on a list of works that Vallotton sent to Munich on 28 February 1910 found in a notebook kept at the Vallotton Foundation. This list includes a '*Villerville, mer haute'* priced at 600 Fr. This statement is preceded by the number 28, which the artist wrote on the back of the work.

This cross-checking of information, facilitated by the artist's methodical mind, made it possible to confirm an initial impression and to conclude, after examining the painting, that it was an original work of Vallotton's.

b) *Sampling*

We define the variables using the same questions and reasoning on answers as explained for Case Study 1 (Appendix VII). Finally, based on these questions, we

[104] '*Félix Vallotton (1865–1925), Verso du tableau 'Mer haute, Villerville', 1902 Fondation Vallotton, photo'* (in French).

transformed the data into variables. The variables marked with markers are indicators of the ratio of the probability of being original or fake. They are as follows:

With a title
With a date
In Switzerland
Not in France
Not sale online
With a signature
Without a monogram
Not large
Having information on back
Landscape
Not in 1882–1889
Not in 1890–1900
Not in 1901–1910
Having a known technique
Not in commonplace book
With a provenance
With recent provenance

c) *Evaluation*

The evaluation was performed with the help of Table 56 using the same reasoning as in the previous case study. After comparing every niche contained in the worksheet with the list of variables, we found 10 niches[105] with combinations that matched the variables belonging to the artwork under examination. The comparison shows that different combinations of 'with provenance', 'with recent provenance', 'with title', 'with date', 'not in commonplace book', and 'information at back' are repeated in several sets. Remarkably, all related niches with these combinations indicated a high probability (between 88 and 99%) of it being an original work. The prediction of the model was thus consistent with the experts' evaluations.

[105] They are marked in *grey* in Table 56.

9.4 Alternative test of both cases using data mining

The case studies illustrate how users can manually evaluate a work of art. Alternatively, the artwork in the case studies can be evaluated using data mining analysis. It includes all the variables identified and listed in Studies 1 and 2 as potential inputs and prospectively runs the CART model. We tested this alternative, and the results are presented in Appendix IV. We conclude that the machine-learning outcomes are similar to those of the manual evaluations. Indeed, the artwork of the first case study has a 100% probability of being fake (niche no. 4, Figure 25, Appendix IV) and that of the second study has a probability of 99.8% of being original (niche no. 1, Figure 26, Appendix IV). Such an evaluation can only be performed by a computer specialist, which reduces the ease of use for another person without statistical training. However, this option is a way to standardise the methodology introduced in this thesis so that it can be generalised to other artists. From this perspective, generic software can be developed to allow any database to be processed in a standard form.

PART 4 – DISCUSSION AND CONCLUSIONS

10 INTRODUCTORY REMARKS TO PART 4

The following two chapters include discussions and conclusions stemming from the doctoral research results, and constitute the final part of the thesis. We evaluate the relevance of our research approach to state-of-the-art methods. The question we address is how the roles of the authentication and detection functions are allocated according to the examiner's objectives and the technical means. In this discussion, we outline the differences and affinities of our model and highlight the role of fake detection compared to the identification function.

The other aim of this section is to direct the reader's attention to the limitations and particularities of the database, together with recommendations for their refinement in future research. The final objective is to explain how the data-mining model functions as a tool for the early detection of fakes and forgeries. We present perspectives on developing statistical models for other artists and consider their implementation in practice. Overall, this section aspires to affirm the need for a new response, which includes building the capacity to understand the real risks of counterfeit art.

11 DISCUSSION

11.1 Unveiling the distinction between forgery detection and authentication functions

A literature review has shown that counterfeit detection plays a secondary role when these methods are applied in practice. To explain this finding, the discussion examines how the roles of authentication and detection functions are allocated according to examiners' objectives and the technical means at their disposal.

The stylistic connoisseurship method is a visual examination by a *trained eye* which experts most often apply in practice. In principle, this method has two functions. One is to identify features that correspond to an artist's originality or unconscious personal characteristics. The other is to detect various errors committed by authors, such as negligence, omissions, and incoherence. These two functions are interdependent. Indeed, when the examiner concludes that certain specific features of the work of art under examination are identical to those of the original work and that there are no obvious errors, such a positive assessment may provide evidence of the work's authenticity. Otherwise, it turns into a negative evaluation, implying a fake work. This is an exclusive method in which only a person with extensive knowledge and training in the original works of a given artist and counterfeit works is suited to practice.

Moreover, compared to other authentication methods, it is the least objective method in terms of scientific objectivity. Scientific objectivity refers to freedom from personal bias. This concept does not exclude the fact that scientists may have a certain degree of subjectivity, for example, in interpreting results or selecting samples. However, an important premise of this doctrine is that the results should be objective and independent of the personal skills or experiences of the test-takers (Reiss, & Sprenger, 2020). In the case of stylistic connoisseurship, the method of providing knowledge that relies on the examiner's trained eye depends on the intellectual skills, learning, and experience of the person who conducted the analysis. Therefore, this method is

scientifically subjective[106]. One can argue that this method is not, in absolute terms, a scientific approach, because its results cannot be tested using a similar tool. Indeed, different and even contradictory opinions are possible from experts who have used the *trained eye* method on the same issue (see Section 3.2.1). Nevertheless, this method is significant because of its exclusiveness. In such a delicate situation, art market participants should pay great attention to an expert's professional reputation to ensure the reliability of the assessment results.

Historical documentation analysis is more scientifically objective than stylistic analysis. This method involves establishing the history of an artwork from its creation to the present day. Despite the uncertainties in the field of provenance research, this approach plays a significant role in art authentication (Section 3.2.2). Provenance information can be found in primary sources, such as receipts and bills of sale that document ownership transfers, and in secondary sources, such as catalogues and artist monographs. Similar to stylistic analysis, this method operates in two ways. It can help prove the authenticity of a work of art by reconstructing its chain of ownership. By contrast, incoherence in a chronological sequence may reinforce doubts of an artwork's authenticity. In practice, historical analysis is usually applied to complement the stylistic connoisseurship analysis. Furthermore, it is often performed by the same experts.

Experts who doubt the authenticity of an art object may solicit scientific expertise. For example, the incoherence between the composition of the materials and pigments used in the production of the work and the materials available during the artist's lifetime may confirm suspicions of authenticity. In other words, such a test looks for a feature that can distinguish the original from a fake. For this purpose, there is a spectrum of scientific approaches (laboratory tests) called technical or scientific methods. For example, Section 3.2.3.1 explains the cases in which laboratory analyses were effective in corroborating the doubts of authenticity.

[106] Subjectivity is usually contrasted with objectivity, where knowledge is seen as independent of who produces it (Sharp, 2020).

Overall, when the technical analysis does not identify suspicious elements, the evidence is insufficient to confirm authenticity without the help of other methods. For instance, Section 3.3 illustrated a case in which, although the scientists did not find any suspicious elements, they decided not to confirm the authenticity. In this regard, Reeves (2011), Kallier (2012), and Neuhaus (2014) pointed out that art experts usually view technical/scientific analysis as impractical because they are convinced that it can help detect fakes but cannot establish authenticity (Section 3.2.3). Conversely, when suspicious evidence is uncovered, conclusions drawn using technical means become much less dependent on other methods. Nevertheless, this method cannot be considered a universal tool for detecting counterfeit artworks. Indeed, practice shows that sometimes the results of technical expertise are not convincing enough to corroborate or refute such doubts (Section 3.3).

Although technical/scientific methods are scientifically objective, they do not function in two ways, except for computer-assisted art authentication and scientific classification methods that exploit machine learning. This method may identify specific characteristics associated with authentic or counterfeit artworks. However, most studies using computer technology have proposed methodologies with a one-way identification function. Their methods can classify the similarity of the characteristics of original works (like the studies of Sablatnig et al., 1998; Lombardi, 2005; Saleh et al., 2014; Trochim et al., 2016; and Chen et al., 2017). Indeed, according to Elgammal et al. (2018), by employing global features that mainly capture the composition of a painting, the algorithms can classify a painting made in an artist's style but cannot identify the work. Only a few studies open to the public[107] propose methods capable of indicating originals and detecting forgeries (e.g. Montani, 2015 and Elgammal et al., 2018).

Overall, computer-assisted art authentication studies use different techniques, but they all rely on comparative analysis as the research principle. Therefore, the low use of the detection function is probably due to the fact that comparative analysis needs data on counterfeits, but access to this data is difficult due to confidentiality issues

[107] Some platforms use their machine learning methods in both ways, but as they are commercial, their methods are confidential (Section 3.2.3.2).

(see Section 5.2). For example, Neuhaus (2014) explained that technical and scientific methods are challenging to implement because data on other original works by the same artists are often unavailable (Section 3.2.3). Similarly, obtaining data on fake works referenced by the same artist is even more challenging.[108]

In conclusion, stylistic analysis is the most appropriate method for identifying fake works. However, this method relies entirely on the individual skills of the examiner, which reduces its objectivity. On the one hand, provenance research and technical methods may affect the scientific accuracy of this method. On the other hand, technological progress has made it possible that machine learning methods may achieve stylistic recognition. Can they replace the *trained eye* method?

In theory, this is possible, but not in practice. The primary reason for this is that machine learning algorithms can only provide intelligence based on quantifiable data. In fact, in the field of art, where works of art are not standard merchandise (Section 3.1), achieving this condition for all data would only be possible by losing part of the information. The other reason is that to be an authentication tool, such an analysis must be based on valid data from original works and counterfeits. Data on counterfeit works would need to be accessible, which is not currently the case. Moreover, because this type of scientific method is relevant for detecting counterfeit works of art, it may be used as an independent alternative approach for prevention purposes, thereby detaching the role of authentication. In such cases, the indication of the risk of falsification should be seen as a recommendation for early warning but not as evidence of art fraud.

Within such a framework, the purpose of this doctoral study was to develop an alternative method to provide additional arguments regarding counterfeit risks. Similar to other studies, our method uses comparative analysis. Similar to some commercial platforms, our data included information from actual cases. In contrast to methods with detection purposes employed by art market professionals in daily practice, we sought to make our method accessible to the general public.

[108] This statement can be reinforced by our problematic experience accessing data containing the description of fake works (Section 5.2).

11.2 Key considerations in developing a statistical mining model for classifying fakes and forgeries

As the method was conceived as an alternative assessment method that does not require special skills and can be easily interpreted by the user, we started the second part of the study with the aim of finding parameters that are easily accessible to the general public.

A review of provenance research issues (Section 3.2.2.2) helped us find an appropriate solution. According to the AAM Guide, latent information about an art object is included in the details of the object, such as its style, subject, signature, materials, dimensions, or frame. The AAM Guide recommends sources and guidelines on how and where the characteristics that help to identify the art object can be collected. Their recommendation led us to consider that such characteristics can be exploited as variables in our research. Indeed, they contain latent information on art object identity, do not require any specific skills to be recognised, and are recorded in papers – the sales documents of paintings, such as the artist's bill of sale, certificates, auctions, or exhibition catalogues – which are accessible to the public.

Furthermore, to develop a model, the creation of a comprehensive database is crucial. This database should encompass two sets: one with the original works of the artist, and the other with known fakes and forgeries. In order to ensure the efficacy of the model, both sets must strive for maximum completeness. Ideally, they should include all known original works, as well as all known fakes and forgeries. With the permission of the Foundation, the database was meticulously created based on the detailed descriptions provided in the artist's catalogue raisonné and archival documents pertaining to the forgeries archived by the Foundation. Each element captured in the catalogue or the archival documents, such as the presence or absence of a date, title, signature, creation period, or subject matter, constituted a variable in the database, and each painting was treated as a distinct case. The resulting database comprises a total of 1,914 cases.

The creation of a database is a lengthy process, but at the same time, is a key element of the whole model. To provide a comprehensive overview of the raw data included in

the database, we divided the variables into several groups according to their nature (e.g. characteristics related to authenticity, market, and techniques) and presented them in tables (see Section 6.1). A simple comparison of the percentages presented in these tables shows that there are major differences between the original and fake artworks. For example, forgeries are more often monogrammed or unsigned than originals.

Subsequently, we tested whether the variables included in the database were inter-correlated, independent of the character of the original or fake/forgery status of the cases. We used Principal Component Analysis (PCA), which allowed us to detect several statistically significant correlations among variables (see Section 4.2). For example, the variable 'unknown technique' is positively correlated with the variable 'lack of provenance' and negatively correlated with the variable 'in oil'. This means that a painting performed using an unknown technique often has no information about its previous owner and is rarely painted in oil. Despite the accurate PCA results, their interpretation did not clarify whether the correlations were influenced more by the original or fake artworks. In addition, given the percentages of variance reproduced by each component (similar results in all sets examined), we realised that it was not really possible to concentrate most of the information from the original variables into just 2 or 3 components. Consequently, an interpretation based on the first two components alone proved insufficient. At the same time, the correlation matrices showed that not all the selected variables were perfectly independent and that there were close relationships between some of them, especially in the first component. Therefore, the PCA analysis was used as an aid to understand the data, rather than a tool for reaching definitive conclusions.

The next step was to establish whether artwork characteristics presented themselves similarly or differently in originals and fakes/forgeries. Using the CART method[109], we found that it was possible to distinguish fakes and forgeries from originals based on

[109] This method groups explanatory variables into clusters with similar predicted outcome values through a splitting process. The classification tree method can be schematically described: first, the CART's user choses the explanatory variables relying on the hypothesis; then, the mathematical logic – classification tree algorithm – ranked explanatory variables according to the strength of their correlations and indicated those having discriminatory power.

these characteristics. From this perspective, the algorithmic interpretation is straightforward and explicit. For instance, according to the algorithm's interpretation, the absence of *provenance* and *date* in a painting implies that it has a 97% probability of being a fake/forgery. This means that in Vallotton's works, a painting becomes suspicious when it is undated and when the chronology of its owners from the time of its creation until the most recent owner is unknown.

The developed model was then tested through an experiment (Section 6.2). In the preliminary phase of the experiment, only the most representative variables for each group of characteristics were included. This reduced data set was created to explain the statistical models, test their feasibility, and explore their discriminatory power. In the main phase of the experiment, we used several other sets of variables to perform statistical analysis. Unlike the set used in the preliminary phase, the data-selection process for the main phase involved different assumptions regarding Vallotton's lost works. From this perspective, the catalogue raisonné points out four possible reasons for the loss of Vallotton's paintings. Accordingly, we divided the characteristics into four working sets corresponding to each reason, and then selected the most representative variables for each group. Classification tree analysis showed that the working sets containing the lost characteristics performed well. In particular, the study obtained excellent performance (99.9%) for the original works and robust performance (from 70% to 91%) for the fake works in most working sets[110]. However, the classification accuracies of the working sets were not significantly better than those established using the testing set created without these assumptions. This means that the concept used to create the working sets did not have a significant impact on model performance. Nevertheless, one could object that in our sets we analysed only the most representative variables, leaving aside the rest, which were considered less representative based on our previous analyses. Hence, we conducted a series of additional tests with the less representative variables. Only two of the new variables were discriminatory. This result shows that the maximum discriminatory capacity of the database had been largely exploited.

[110] Taking precautions to ensure that significant performance was not due to a peculiarity of our data but to the appropriately adopted CART analysis method, we explored how the results can be generalized from a limited data set. Consequently, we performed a cross-validation using 10 thousand replications of the analysis.

11.3 Exploring research questions through the methods and findings

The study's methodology addresses various research questions, as outlined in Section 4.1, which can be synthesized into two primary issues. These questions can be summarised into two main issues. The first research issue aims to determine whether statistically significant relationships exist among the variables within our comprehensive database. The second issue focuses on examining whether statistical methods can effectively differentiate between original paintings and counterfeits based on various characteristics mentioned in textual documents, such as the creation date, presence or placement of signatures on the canvas, type of support, or technique. To accomplish this, Principal Component Analysis (PCA) was employed.

Initial tests revealed significant correlations between the variables. The analysis yielded accurate results; however, the interpretation alone did not provide insights into whether the interdependent variables were specific to original artworks or indicative of counterfeits. To overcome this limitation, an additional data mining technique, the Classification Tree method (CART), was applied. The results demonstrated that the algorithm enabled the determination of when grouped variables were indicative of originals or fakes. Furthermore, one of the advantages of CART is the straightforward interpretation of its result. Another is that the CART model has a good ability[111] to account for missing values. However, the CART model is statistical; therefore, its output may contain errors, such that some original paintings were classified as fakes/forgeries, and vice versa. Thus, it was important to determine the success rate of correct classification to understand the model's performance. Although the classification accuracy of the originals was close to 100% in all sets, the accuracy of the false categories varied between 59% and 89%[112]. Hence, the specificities were not homogeneous compared with the sensitivities. Indeed, the correct classifications for false works were robust for the experiment, Working Sets 1 and 3 (around 70–89%). However, for Working Sets 2 and 4, the performance was only 59–65%. A pertinent explanation as to why the sensitivities of these sets were lower

[111] '*If explanatory variables are missing, trees can use surrogate variables in their place to determine the split. Alternatively, an observation can be passed to the next node using a variable that is not missing for that observation*' (Moisen, 2008, p. 586).

[112] According to the cross-validation tests, CART's overall performance was considered robust (Section 7.5).

than those of the other sets appeared when we compared the rules. Indeed, all sets with high sensitivity levels include *provenance* as an explanatory variable, whereas the low-sensitivity sets do not include this variable. Thus, *provenance* can affect accuracy. Indeed, 1,498 of the 1,704 original works compiled in the catalogue raisonné have known provenances (Table 17). Accordingly, the variable 'provenance' is a crucial discriminating feature for a predictive model under the condition that most artworks' ownership histories are known and documented by credible sources.

Therefore, we were able to successfully differentiate between the original paintings and fake paintings in our database based on the aforementioned characteristics. Consequently, our affirmative response to the research question led us to conclude that information regarding authenticity can be revealed through statistical analysis.

11.4 Interpreting variables and rules: essential points for accurate analysis

CART successfully identified 17 variables[113] with various degrees of discriminatory power that were classified into two types of niches: one for those with a probability of being a fake, which contains the features correlated to forgeries, and another for those with a probability of being originals, which contains the variables correlated to Vallotton's originals. Accordingly, the niches encompass a combination of characteristics (rules) and their corresponding probabilities. Based on the probability rates, it was possible to predict whether a work of art was suspicious. For example, one niche indicates that the probability of being a fake/forgery is 100% if a work of art has neither a provenance nor a title, but does have a date. For the user of the model, this information can serve as a starting point for decision making[114]. Indeed, likelihood rates are explicit arguments for understanding whether there is a question of authenticity.

Furthermore, it was interesting (as a research challenge) to define the nature of the

[113] 'Provenance', 'first provenance', 'title', 'unknown technique', 'date on the painting', 'Switzerland', '1882–1889', '1890–1900', '1901–1910', 'sale online', 'large', 'landscapes', 'commonplace book', 'sold', 'monogrammed', 'France', 'information on the back'.

[114] In simplistic terms, the model users can go further, for example, by request the assistance of an expert or even abandoning their previous intent.

rules. However, this interpretation is restricted by the author's limited knowledge of Vallotton's works. For example, to understand why certain characteristics may not be simultaneously present in Vallotton's works, it is necessary to understand the stylistic and historical aspects of his works. Nevertheless, we tried to explain the meaning of some niches based on literature sources and verbal explanations provided by the Foundation's experts. For example, rules about a 100% probability of being a fake/forgery if a work of art has neither a provenance nor a title but does have a date mentioned could be explained by the fact that Félix Vallotton signed all the paintings that came out of his studio for external exhibitions. Accordingly, his unexposed paintings are usually undated (except for the application of a signature stamp by his family; Ducrey & Polleti, 2008). Thus, it may be suspicious if such a work of art has a date on the painting. Some variables of the rules can be explained by simple logical links to the creation process. For example, in the niche with the variable 'not large', the rule with this variable indicates a high probability of being fake. Indeed, making miniature paintings may be less complicated than making large paintings in terms of technique and time. The variable 'landscape' is indicative also because the landscapes of Vallotton have had greater success in the market than other subjects. The discriminative force of the variable 'in Switzerland' may relate to the fact that the Vallotton Foundation is located in Switzerland. Thus, the market for Vallotton's art in Switzerland is better protected than in other countries, because art players can easily ask for expert opinions or submit paintings for expertise (Section 2.2). Accordingly, forgeries are likely to appear more often in another country's art market than in Switzerland[115].

The interpretation of some rules was more explicit than others. We propose interpreting this through the perspective of combining the rules of CART and PCA. The same variables were used for both methods. In particular, they correlated with the principal components[116] and were identified as having discriminating power. Accordingly, this finding can be used to extend the scope of the rules.

[115] These examples are not pretended to be expert opinions because the author of the thesis is not a specialist in the art field.

[116] These variables are 'with date', 'with title', 'sold', 'large', '1890–1900', 'Switzerland', 'without provenance', '1882–1889', '1901–1910', landscape', 'sale/online', 'monogrammed', and 'unknown technique'.

For instance, the PCA model's interpretation – 'a painting with no provenance is more likely to be without date, untitled or signature' – can be put into relationship with the rule of the CART' model rule (Table 21[117]).

Thus, we obtain the following information:

> A painting without provenance is likely to be without a date, title, or signature; when this happens, the picture has a 97–99% chance of being a fake/forgery or doubtful.

A similar example concerns a painting without provenance that is more likely to lack title and date/signature and whose owner is not located in Switzerland. This information can be related to one of the CART rules (Tables 21 and 22) as follows:

> Owners of paintings without a provenance are not often in Switzerland. However, even a painting whose owner is in Switzerland has a high probability (90%) of being a forgery or dubious if the date and title of the painting are unknown.

In addition, the interpretation of some variables was more explicit than others. This difference is related to the logic of the sampling of the variables (Section 9.1). We proposed a series of questions with options for possible answers to guide the user to better understand how interpretative they are (see Appendices V, VI, and VII).

In summary, most niches include variables that are understandable by their direct meaning via their name – such as 'with a title', 'in France', and 'with a signature'. For others, it would be helpful to follow the explanation provided in Appendix II, which specifies the definition of each variable, and Appendix V, which includes a list of supporting questions for sampling.

The hidden meaning of each niche (its nature) can be explained with the help of accumulated knowledge in the literature or with specialists' help. However, such an explanation is optional because awareness can be gained through likelihood. In other

[117] There is the same rule in Tables 22, 36, and 37.

words, when our model is applied in practice, the probability of the rules is a sort of argument. If necessary, the user can later ask an expert why a particular combination of features is suspect.

11.5 Database limitations and recommendations

As explained in Section 3.1, a work of art is a product of individual creativity that reflects the artist's personality, epoch, and culture. This implies that the correlations between the original's characteristics correspond to the artist's individuality and cannot be applied to other artists. In the case of Vallotton's works, these correlations are reflected in the niches with their rules and probabilities. Therefore, the model outcomes summarised in worksheet (Table 56) are exclusively related to Vallotton. The model contains several discriminatory variables that are common to all artists. Features such as *date* or *signature* are standard for the works of any artist. However, their correlations with other variables could have different strengths from one artist to another. Despite this limitation, our methodology is adaptable for detecting forgeries affecting other artists, provided that a new database is developed for them.

For further studies based on our method, researchers should focus on the reliability of the data sources. From a methodological perspective, invalid data sources can affect the quality of research results and conclusions (Kornegay & Segal, 2013; Olabode, Olateju, & Bakare, 2019). Although the accuracy of an unreliable source would not be altered or could even be improved, its validity would be considerably reduced or lost. The main recommendation is to find data containing key variables relevant to the research questions and, more importantly, those from credible sources. In Section 3.2.2.3, we explained some specificities related to the catalogue raisonné. Indeed, there is no formal standardisation or validation of the quality of a source, such as a catalogue raisonné. Under these circumstances, credibility can be guaranteed through public reputation. For example, as we used the catalogue raisonné created by experts from the Vallotton Foundation, its reliability is guaranteed by the internationally renowned specialists who edited it. Furthermore, the Foundation has created archival records of the fakes. In this respect, both sources hold data from previous authentications made by reputable experts that ensure the sustainability of the study results.

Another recommendation for future researchers is to expand the sample beyond the variables examined in this doctoral study. Each available characteristic should be collected to reveal all variables with discriminatory power. Precise guidelines offer the possibility of standardising the sampling process. Since the sampling is similar to the database creation and the variables' interpretation for the evaluation (manual variant), the reader can see a detailed explanation of transforming raw data into variables in the case studies (Sections 9.2 and 9.3) At first sight, this technique could appear relatively complicated, especially for external users adopting the model for the first time. Such a detailed description was necessary to provide an overview of the specific aspects of the interpretation and selection of the variables. In practice, we suggest facilitating this process using a protocol which may contain questions that guide the user to transform data into variables. After the sampling stage, the user is invited to check whether one or more of these variables matches any niches to determine the probability rate. This assessment is simple and requires only a few minutes to complete. The protocol can also include instructions for facilitating this task. Furthermore, with information technology, almost all functions can be automated. Indeed, the sampling and niche-matching procedures can be automated using software.

However, in practice, a combination of variables may not correspond to any niche. In this case, the manual estimation would not lead to a prediction. Therefore, the alternative option provides an opportunity to expand prediction ability (see Section 9.4). All observable characteristics of the evaluated artwork can be integrated into a database and subsequently used as input data in the CART model. In this manner, all features, not only those corresponding to the discriminating variables from the manual variant, would be compared with all features in the database. Although this option performs better, it requires computer skills. Simultaneously, the great advantage of the data-mining evaluation option is that it allows the development of software that can be applied almost automatically to the database for any other artist.

It should also be noted that some numeric variables such as the price of paintings cannot be recorded with dummy codes[118]. Therefore, they cannot be analysed using the CART model. However, we recommend replacing some of them with a nominal variable if it would not be possible to lose much information and then recode it with a dummy code like that for artwork dimensions (Section 5.3.3).

The last proposal concerns the inclusion in the database of characteristics that do not fit the principal criteria of our study because they would be accessible mainly to specialists in the field of scientific analysis. In theory, numerical parameters like brushstrokes, spectral indicators, and technical/scientific pigments (Section 3.2.3.1) can be recorded using a dummy code. From a technical perspective, they can be processed using the CART model. Their advantage is that their integration into the database may improve the model's performance owing to their objectivity.

11.6 Relevance of the authentication authority in implementing the method

A considerable advantage of this method is that data on forgeries were obtained from documents provided by experts. The experts of the Vallotton Foundation actively collaborated with us by providing the necessary explanations during the data collection phase and by analysing the results. Data on forgeries are generally confidential owing to the implications of the owners' names (see Section 5.1). Our approach makes this knowledge available to the public without compromising the owners' confidentiality. Indeed, drawing a parallel with Bandle's (2015) assertion that better access to experience and knowledge is likely to improve the quality of authentication (Section 2.3), it is fair to say that much better access to experience and knowledge can also improve the quality of forgery detection.

[118] We collected the price information from the files but used them only for descriptive analysis (see Tables 8 and 9). Moreover, as noted in Section 2.2, the fluctuation of the market price for Vallotton's works varies depending on their periods and subjects. Exploiting a market price of the artworks as the variable for the comparative analysis would imply comparing the prices of originals and fakes of the same period and subject. On the other hand, codifying such numerical characteristics with dummy code would result in losing some of the information. However, such variables can be collected for exploration in the descriptive analysis of the preliminary phase of statistical research.

Because only the authenticating authority has access to the data and can guarantee the trust that users of the model might expect, it is reasonable for this authority to remain the principal holder of the method. The reader may question why such an authority would implement a counterfeit detection method for nonprofessional art players. Indeed, authenticating authorities such as experts, foundations, research institutes, and museums can benefit from exploiting the detection model. Together with Neuhaus's (2014) suggestion of the utility of technical/scientific expertise in protecting experts from hostile litigation (Section 3.2.3), we argue that our model can help experts for the same purpose. For example, in doubtful cases, art experts can refer to the model's results to maintain previous assertions made through stylistic connoisseurship. Neuhaus (2014, p. 73) stressed that '*although stylistically connoisseurship is to a large extent based on tacit knowledge, art experts should nevertheless make an effort to explain how they arrived at their conclusion*'. Based on this statement, we suppose that the statistical evidence offered by our method may help experts avoid lengthy explanations, especially in 'trivial' cases in which justifying how they arrived at their judgement would take more time than the assessment itself.

As explained in the previous section, applying this method to other artists would require the creation of new databases. However, there are inherent costs. Because the dataset may be exploited indefinitely, a fee charged to external users would progressively amortise these costs. Consequently, this fee could be much lower than the cost of expertise. In addition, the development of software capable of automatically executing the model outputs is relatively inexpensive.

Governance must be established for the utilisation of such models with the aim of securing data accuracy. Ideally, art foundations should oversee the databases related to their respective artists.

From a practical point of view, the question is whether organisations with a public mandate may have the interest and could take the lead in the implementation of our method. In this context, it is worth recalling the ID-Art mobile application for tracing and identifying stolen cultural property developed by Interpol mentioned in Sections 3.2.3.2 and 5.1. Presenting this new app during the event organised by the Art Law

Foundation for its members, Corrado Catesi, *Head of the Works of Art Unit at Interpol*, said that creating the same application for counterfeit art would require the creation of a database of counterfeit items recognised and confirmed by the last instance of judicial authorities (The Art Law Foundation, 2021). Furthermore, he stressed that because collecting such information can take considerable time and effort, Interpol cannot compile it. Nonetheless, Mr. Catesi pointed out that a database of art counterfeits is essential for Interpol (The Art Law Foundation, 2021).

Hence, the development of a counterfeit database by public agencies is currently in limbo (Section 5.1). In such situations, the proposed method can be a compromise. On the one hand, our database contains information on counterfeits not confirmed by the last juridical authorities. However, the findings of the model based on this data set are statistical extrapolations. They stipulate valid arguments regarding the risks of counterfeiting but do not indicate the concrete artwork of the database. Simultaneously, because the model findings are based on expert knowledge and statistically corroborated by validation tests, their reliability is sufficiently strong to constitute a robust preventive measure. Indeed, rates of probability aim to encourage users to contact experts afterwards, and thus avoid making overly impulsive decisions. Of course, using the art market database of one artist would not spread preventive effects to the public at large. Accordingly, Interpol may initiate the creation and consolidation of databases by art foundations or museums with archives of forgeries, and launch applications (similar to those for stolen art) adapted for each artist's database as a preventive measure.[119]

[119]In the same line, the role we suggest for Interpol could be assumed by the research institutions in art or a public (semi-public) organisation at the national level.

12 CONCLUSIONS

12.1 Using the statistical data mining model as a tool for early identification of fakes and forgeries

This study adopts an original position by postulating that art counterfeits[120] can be distinguished from original works of art based on their ordinary characteristics such as subject, size, date, and place of signature. In support of this thesis, this doctoral work proposes a new method of successive and methodical processing. It also leverages experimentation and case studies. The aim of this PhD study is to test whether a statistical data-mining model can help art players identify potential fakes and forgeries in the art market based on sales documents[121].

The test confirmed that comparing an artwork's ordinary characteristics with those identified by the algorithm made it possible to evaluate the probability of dealing with counterfeits. The model provides alertness in the form of likelihoods. Therefore, these probabilities represent additional arguments about risks, regardless of the presence or absence of suspicion.

We believe that the main advantage of the model is that it is based on readily available information and can therefore serve as a prevention tool for persons without statistical training. For example, a collector interested in a painting attributed to Vallotton that is absent from the catalogue raisonné could run the model and obtain a first estimate of the risk of its being a fake.[122]

To achieve the goal of rendering the model readable and easily applicable, we first summarised and presented the rules with their corresponding probabilities in a table. To exploit them, users do not need to learn the methodology for PCA and CART development; they only need to (1) check how the artwork's characteristics fit in the

[120] We use the terms 'counterfeit', 'fake', and 'forgery' as synonyms (a detailed explanation can be found in Chapter 1).
[121] These documents – such as an artist's bills of sale, certificates, or auctions and exhibition catalogues – include a series of these standard features.
[122] Section 11.1 explains why such a model should be considered a distinctly alternative approach for prevention purposes.

17 variables included in this table, and then (2) search for the niches to which these characteristics belong to find the corresponding probability. In addition, we have included a schematic diagram to illustrate the application of the model. We then analysed two pictures outside our dataset using this schema and table to demonstrate each step of making a prediction.

Because the Foundation's experts had formerly provided an examination of the two paintings, we were also able to compare the model's prediction with the expert's conclusion. Remarkably, the statistical extrapolations were consistent with the experts' judgements, which corroborated the effectiveness of the model. In fact, according to our model, the artwork confirmed by experts as fake had a 92%–100% probability of being a fake; conversely, the second work authenticated as original by experts had a 99% probability of being original via the model's rules.

All statistical models have limitations because they are simplified representations of the relationships between observed data. Our basic model thus has certain limitations as well. For example, some combinations of features of a particular work of art may not match the results of the model. Therefore, we attempted to establish whether it was possible to find a way of extending the prediction range of our model. To this end, we propose a different approach as an alternative evaluation technique to extend the scope of the model (Section 9.4). Using this technique, the user can test the all-sampling features as input data to create a new classification tree. The application of this technique relies on CART, and therefore requires the use of a statistical programme to examine the probability for a particular work of art.

Finally, to avoid ambiguity in the authentication issue, it should be clarified whether the extrapolation of the model is oriented towards detection, prevention, authentication, or attribution. The authentication of an artwork is a process that seeks to provide solid evidence to identify the author of a work of art. In addition, it can be used as a filter for forgery through detection. Thus, the proposed model has authentication functionalities. However, an awareness of the probability based on descriptive evidence in sales documents is not sufficient evidence to warrant an expert's certificate, despite its reliability and validity in a statistical sense. For

instance, the presence of the provenance, date, signature, and period indicates an optimistic prediction according to the model. However, in such a case, all prior ownerships, the validity of the signature, and the relevance with its stylistics should be confirmed with provenance research, technical examination, and stylistic analysis to evidence the authenticity of the work. In other words, if the model prediction indicates a high probability of originality, then the work has an excellent chance of having been created by Vallotton. However, even in that case, the work should also be evaluated using traditional authentication approaches. Likewise, the probability of a high risk of forgery is a matter of recommendation but not a guarantee of fraud. The advantage of using our model for guidance is the possibility of alerting art market players to falsification risk. In practice, suspecting that a piece might be a forgery is one thing, but disposing of a tool that establishes that the work in question has, for example, a 90% chance of being a fake can considerably change intentions. Given these points, we conclude that our model is essentially a preventive tool that can act as an early sign or warning of a potential problem. Such a function constitutes a complementary input to the authentication process but can also be used as an alternative preventive tool.

12.2 Implications for future research and the development of statistical models

A large part of this dissertation deals with methodological issues, but the results obtained may influence future research on art fraud detection and prevention. Platforms that use statistical models for computer-assisted art authentication have been made commercially available; consequently, their methods are confidential. Nevertheless, their presence demonstrates the potential of this tool. In parallel, scholarly research applies data mining analysis mainly to facilitate authorship attribution but focuses less on forgery prevention. This is probably because a comparative analysis requires data on forgeries, but accessing such data is challenging because of confidentiality issues (see Section 5.2); consequently, in most of the studies that use comparative analysis, the researchers work with forgeries that they have themselves commissioned from artists who are asked to create a painting following the style of one painter (e.g. Montani, 2015; Elgammal et al., 2018). Researchers often

exploit specific characteristics of works of art such as brushstrokes, measurable parameters, and colour characteristics to develop computer-based authentication methods (Sablatnig et al., 1998; Teegen, 2002; Lombardi, 2005; Montani, 2015; Elgammal et al., 2018). To the best of our knowledge, the typical characteristics of sales documents have not yet been studied.

The only study whose methodology was comparable to ours was Elgammal et al. (2018). These researchers aimed to detect the typical characteristics (strokes) of an artist and discriminate between artists at the stroke level. These scholars hypothesised that some traits reflect the spontaneity of the way in which an original work of art is created, in contrast to the inhibited nature of imitation artworks. In comparison to our study, their research strategy – a comparison of fakes and originals – was similar, but their techniques and underlying assumptions were different. The accuracy of our model for classifying forgeries was 70–91% for most sets, while the classification accuracy of Elgammal et al. (2018) for fake drawings was 100% in most cases. However, their accuracy is based on a test that used imitations of fakes commissioned for the study, while ours reflects an analysis of '*real*' forgeries in the sense that they were found in the market. Another difference is that the development and implementation of Elgammal et al.'s model requires a specialist in computer science, whereas one of the two options for implementing the predictions of our model does not require such specialisation.

From this point of view, the findings of this PhD study refer exclusively to the works of Vallotton, but the methodology developed in this study can be applied to the works of other artists so long as relevant modifications and adaptations are introduced. For example, some characteristics like 'date' or 'signature' are common to any artist's work, but many factors related to an artist's individuality affect the use or omission of a date on the painting, the location of a signature, or the support size. Consequently, the correlations among the characteristics of the works of one artist may vary from those of another. Thus, the probabilities obtained for Vallotton would not be meaningful for other artists, including those who lived during the same period and worked in the same style. Despite these limitations, our study's sampling and analysis techniques are standard. This method can be replicated to assess a database that

includes the characteristics of another artist.[123] Testing our methodology with another artist will increase the validity of this dissertation. However, the difficulties in obtaining counterfeiting data for other artists and the time required for sampling all their works exceeded the limits of what can reasonably be expected for research conducted by a single person.

As a recommendation for future research on other artists, we suggest primarily focusing on dataset quality. In principle, data on original works should come from a catalogue raisonné, monograph, or archival documents. As there is no regulation or official approval for the quality of such documents, their reliability depends on the public reputation of the experts who issue them. Furthermore, developing a model would be impossible without the data on forgeries detected and documented by credible experts. In summary, this tool can only be developed for artists whose original works and forgeries have been carefully preserved. Ideally, the created database should contain the characteristics of all the original works created by the artist as well as all the known fake works. In addition, the database should be updated regularly when new fakes or originals are discovered. This is also true for our model, which can be improved in the future. From this perspective, researchers must keep in mind that the model developed for other artists must not be limited by the selection variables with discriminatory power detected in our study. Other variables may also be relevant to other artists. Another important issue is that the number of artworks included in the database must be sufficient for accurate statistical analysis. Although the question of the minimum amount of data required cannot be answered precisely, analysis of the cross-validation results may show that a dataset is too small owing to high variability across replications.

Specifically, it is necessary to standardise the methodology introduced in this study so that it can be generalised to other artists. In this sense, all observable characteristics of the evaluated artworks can be integrated into a database and used as input data in the CART model. Generic software can then be developed and made available,

[123] Testing the same method with another artist would have set the model in a broader perspective. The obstacles to obtaining data on forgery for other artists and the resources and time demanded by the task of sampling 1,914 works were decisive to this restriction.

allowing the processing of any organised database in a standard form. This software can then be applied in a quasi-automated manner to a database corresponding to any other artist.

Another possibility is to incorporate new types of sampling features into a database, which can lead to more extensive model predictability. It is possible to combine discerning characteristics with other features such as those that are generally exploitable for technical/scientific authentication methods. For example, specialists can easily sample and integrate binary variables representing the absence or presence of brushstrokes, specific pigments, or materials into a database. The sine qua non is that the two sources of the database must contain identical characteristics for comparison. In theory, leading museums that conduct first-rate scientific research as part of their activities, such as the Van Gogh Museum, may supply credible data in both categories to complete such a database.

From a practical point of view, we can envisage implementing the model, preferably by organisations with authenticating authority that keep records of forgeries, such as foundations. A guideline protocol can be developed for use by anyone who wants to assess an artwork. In particular, the protocol would contain policies based on our methodology for evaluating the painting in question by the user himself. In addition, foundations may ask the person requesting the authentication of a work of art to identify key suspect points using this protocol. Applicants may also be asked to submit a summary of their results to the foundation if they want the authentication process to begin. This rule can facilitate the current practices of foundations, particularly those that are overwhelmed by authentication requests. It can also help filter out low-level counterfeit requests. Likewise, art market intermediaries – auction houses, galleries, and lawyers – may ask their clients to conduct the assessment or perform it themselves to draw a conclusion before asking for an expert opinion. Consequently, this self-investigative approach could benefit both experts and market participants.

Some might argue that an open-access statistical model is risky because it could encourage counterfeiters to evaluate and improve their work. Since counterfeiters are

generally well-versed in art[124], they are familiar with the artwork of the artists they forge. Accordingly, we expect that they will not be surprised by the evidence from the model. Of course, minor parts of the niches are not necessarily so obvious, even to specialists, but most of them contain information that is well-known to connoisseurs of art history. Contrariwise, the model's indications can be novel and valuable to art market participants who are not specialists, namely, to the naively prudent potential victims – the amateurs who rely entirely on the opinions of others without taking serious precautions.

Furthermore, some argue that if niches indicate the characteristics of forgeries, this is because even well-informed forgers make mistakes; consequently, having access to the niches would make it possible for them to improve the quality of their forgeries. However, the model results are useless for this purpose because some mistakes were made by negligence, whereas others were introduced deliberately. Hence, the combination of characteristics constituting the error is, in principle, well known to fakers. From this perspective, one relevant question is why forgers make mistakes that lower the quality of their fakes and make them easier to detect. One plausible reason is that this type of error provides a subtle clue that proves that the picture is not original but was inserted by forgers to protect themselves from accusations of fraud once discovered (Section 3.2.2.1). This deliberate error can be invoked to claim that the painting was created without an attempt to cheat because, for example, it contains an obvious anachronism, even if experts do not discover it for a long time. Such a deliberate error can be illustrated in the Vallotton case where forgers use the monogram 'FV' inappropriately.[125]

Another reason why open access to this model is not risky is related to the exceptional protection of Vallotton's art offered by the commonplace book and other detailed

[124] For example, Van Meegen, Zhang Daqian, John Drewe, Wolfgang Beltracchi, and other forgers mentioned earlier were painters who had studied art history at an expert level.

[125] Even knowing that Félix Vallotton mostly inscribed the monogram 'FV' on his prints, forgers have placed it quite often on paintings. Supposedly, this is because the monogram is easier to imitate than the signature, and mainly because it covers up their intent to deceive since the letters of the monogram may represent another name and not that of the artist. Besides, the monogram indicates only the possibility of Félix Vallotton's attribution and leaves ambiguity increasing the interest of the amateur.

documents held by the Foundation and high-level experts. These factors limit the likelihood of new forgeries entering the art market. Therefore, we conclude that the effect of preventing the circulation of counterfeit Vallottons is superior to the minimal risk of abuse of the model for malicious purposes. Therefore, the open model is a valuable tool for art amateurs to help them against informed counterfeiters. However, such access may be questioned regarding other artists. Indeed, the level of accessibility may depend on concrete factors concerning, for example, database specifications or general protections given to the artist's art. Its access may be partially limited by the software mentioned above, insofar as the user would not see the content of the niches and can only obtain a percentage probability without knowing which groups of characteristics are discriminatory. In other words, it is possible to make the 'technology' totally accessible but not the data, particularly the information on fake works of art. Globally, we believe that transparency is a prerequisite for assessing the quality of research findings and encouraging empirical and methodological research in closed, understudied areas, such as art fraud. Otherwise, there is relatively little incentive for art market actors to acquire new skills to prevent their own victimisation, gather better evidence, or conduct more extensive evaluations.

The repetitive and progressive nature of fakes and forgeries circulating in the art market calls for new solutions beyond the traditional authentication approach, whose historical perspective reveals its insufficiency. These new responses involve building the capacity to understand the real risks of confronting fake art objects. This ability is necessary for imagining, evaluating, and deciding the most appropriate solutions and measures. In this context, the originality of our research is linked to the postulate that some fakes and forgeries are distinguishable from authentic artworks by their standard characteristics. The encouraging results, coupled with an increasingly favourable attitude towards new technological methods, allow us to say that it is possible to transfer our statistical model into art market practices and routines at various levels.

REFERENCES

Adam, G. (2018). *La face cachée du marché de l'art: Controverses, intrigues, scandales.* Aubin, France: Beaux Arts Editions.

Aebi, M. F. (2006). *Comment mesurer la délinquance ?* Paris: Armand Colin.

Aljumily, R. (2016). Quantitative criminology: An evaluation of sources of crime data. *Global Journals Inc.*, *16*(4), 1-14.

Amineddoleh, L. (2015). Purchasing art in a market full of forgeries: Risks and legal remedies for buyers. *International Journal of Cultural Property*, *22*(2-3), 419-[iv].

Amore, A. M. (2015). *The art of the con: The most notorious fakes, frauds, and forgeries in the art world* (1st ed.). New York: Palgrave Macmillan.

Analytics Vidhya. (2016, April 12). *Tree based algorithms: a complete tutorial from scratch (in R & Python).* Retrieved August 18, 2020, from https://www.analyticsvidhya.com/blog/2016/04/tree-based-algorithms-complete-tutorial-scratch-in-python/

Art Law Foundation. (2021). *Exclusive members' event: Due diligence in the art market. Interpol's new app to trace and identify stolen cultural property*, June 10 202. Geneva. Retrieved December 29, 2022, from https://artlawfoundation.com/fda-events/exclusive-members-event-due-diligence-in-the-art-market-interpols-new-app-to-trace-and-identify-stolen-cultural-property/

Artnet Price Database. (n.d.). *Artnet*. Retrieved October 26, 2021, from https://www.artnet.com/price-database/

Art Recognition. (n.d.). *Art Recognition*. Retrieved November 11, 2021, from https://art-recognition.com/about-us/

BBC News. (2014, February 2). Fake £100,000 Marc Chagall painting 'to be destroyed'. *BBC News*. Retrieved October 1, 2021, from https://www.bbc.com/news/entertainment-arts-26000331

Balcells, M. (2013). Plundering Boys: A cultural criminology assessment on the power of cultural heritage as a cause for plunder in armed conflicts along history. In J. Kila (Ed.), *Cultural heritage in the crosshairs: Protecting Cultural property during conflict* (pp. 329–349). Leiden-Boston: Brill.

Balcells, M. (2016). Go research an art thief: The importance of empirical research on art theft. In A. Tompkins (Ed.), *Art crime and its prevention: A handbook for collectors and art professionals* (pp. 31–43). London: Lund Humphries.

Bandle, A. L. (2015). Fake or fortune: Art authentication rules in the art market and at court. *International Journal of Cultural Property*, *22*(2–3), 379–400.

Baril, M. (1984). L'envers du crime. *Etude victimologique. Cahier no 2.* Retrieved October 4, 2019, from https://www.cicc-iccc.org/fr/publications/les-cahiers-de-recherches-criminologiques-1984-2005/cahier-no-2-1984

Barnes, H. E., & Teeters, N. K. (1953). *New horizons in criminology* (2nd ed.). New York: Prentice-Hall.

Bazley, T. (2010). *Crimes of the art world*. Santa Barbara, Calif: Praeger.

BBC One. (2011, August 3). *One-Fake or Fortune? Monet* [Video broadcast]. Retrieved November 6, 2019, from https://www.bbc.co.uk/programmes/b0125by8

Ben-Ari, D., Frish, Y., Lazovski, A., Eldan, U., & Greenbaum, D. (2016). Artificial intelligence in the practice of law: An analysis and proof of concept experiment. *Richmond Journal of Law & Technology*, *23*(2), 2–55.

Bergelson, V. (2005). Victims and perpetrators: An argument for comparative liability in criminal law forum on comparative liability in criminal law. *Buffalo Criminal Law Review*, *8*(2), 385–488.

Bloch, G. (2017, February 6). La start-up vaudoise Artmyn signe avec Sotheby's. *Le Temps*. Retrieved November 12, 2021, from https://www.letemps.ch/economie/startup-vaudoise-artmyn-signe-sothebys

Blockeel, H., & Struyf, J. (2002). Efficient algorithms for decision tree cross-validation. *Journal of Machine Learning Research*, *3*(Feb.), 621–650.

Boba, R. (2005). *Crime analysis and crime mapping*. Thousand Oaks, Calif. : Sage Publications.

Borra, S., & Di Ciaccio, A. (2010). Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. *Computational Statistics & Data Analysis*, *54*(12), 2976–2989.

Braeken, J., & van Assen, M. A. L. M. (2017). An empirical Kaiser criterion. *Psychological Methods*, *22*(3), 450.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees* (1st ed.). Boca Raton: Chapman and Hall/CRC Press.

Butler, J. (1740). *The analogy of religion* (3rd ed.). London: John & Paul Knapton.

Burns, C. (2012, January). Knoedler forgery scandal grows. *The Art Newspaper, 231* (after Hufnagel & Chappell, 2016). Hufnagel, S., & Chappell, D. (2016). The Beltracchi affair: A comment and further reflections on the "most" spectacular German art forgery case in recent times. In N. Charney (Ed.), *Art crime: terrorists, tomb raiders, forgers and thieves* (pp. 9–20). London: Palgrave Macmillan.

Button, M., Lewis, C., & Tapley, J. (2009). *Fraud typologies and victims of fraud*. London: National Fraud Authority.

Button, M., McNaughton Nicholls, C., Kerr, J., & Owen, R. (2015). Online fraud victims in England and Wales: Victims' views on sentencing and the opportunity for restorative justice. *Howard Journal of Criminal Justice*, *54*(2), 193–211.

Charney, N. (2009). *Art and crime: Exploring the dark side of the art world*. Santa Barbara: Praeger/ABC-CLIO.

Charney, N. (2015). *The art of forgery: The minds, motives and methods of master forgers*. London: Phaidon Press Ltd.

Chen, L., Chen, J., Zou, Q., Huang, K., & Li, Q. (2017). Multi-view feature combination for ancient paintings chronological classification. *Journal on Computing and Cultural Heritage*, *10*(2), 7–15.

Chowdhury, S., Lin, Y., Liaw, B., & Kerby, L. (2022). *Evaluation of tree based regression over multiple linear regression for non-normally distributed data in battery performance*. Presented at the 2022 International conference on Intelligent Data Science Technologies and Applications (IDSTA), San Antonio, TX, USA, 2022, pp. 17-25, doi: 10.1109/IDSTA55301.2022.9923169.

Clark, M. J. (2004). The perfect fake: Creativity, forgery, art and the law. *DePaul-LCA Journal of Art and Entertainment Law, 15(1),* 1–36.

Clausen, S.-E. (1998). *Applied correspondence analysis: An introduction*. Thousand Oaks: Sage Publications.

Cogeval, G., Cahn, I., Ducrey, M., Poletti, K., & Collectif. (2013). *Félix Vallotton: Le feu sous la glace*. Paris: RMN.

Cohen, P., & Rashbaum, W. K. (2013, August 15). One Queens painter created forgeries that sold for millions, U.S. says. *The New York Times*. Retrieved June 18, 2019, from https://www.nytimes.com/2013/08/16/nyregion/one-queens-painter-created-forgeries-that-sold-for-millions-us-says.html

Conklin, J. E. (1994). *Art crime*. Westport, Conn: Praeger.

Craddock, P. T. (2009). *Scientific investigation of copies, fakes and forgeries* (1st ed.). Oxford: Elsevier/Butterworth-Heinemann.

Christie's. (2017, November 3). *Salvator Mundi — the rediscovery of a masterpiece: Chronology, conservation, and authentication*. Retrieved December 5, 2019, from https://www.christies.com/features/Salvator-Mundi-timeline-8644-3.aspx

Devasia, V. V. (1980). Victimology and the role of victim in crime. *Cochin University Law Review*, *4*, 219–244.

Devijver, P. A., & Kittler, J. (1982). *Pattern recognition: A statistical approach*. Englewood Cliffs: Prentice/Hall International.

Dobovsek, B., Charney, N., & Vucko, S. (2009). Art crime archives. *Journal of Art*

Ducrey, M., & Poletti, K. (2005). *Félix Vallotton (1865-1925)*. Catalogues raisonnés d'artistes suisses. Lausanne: 5 Continents.

Durney, M., & Proulx, B. (2011). Art crime: A brief introduction. *Crime, Law and Social Change*, *56*(2), 115–132.

Dutton, D. (2003). Authenticity in art. In J. Levinson (Ed.), *The Oxford handbook of aesthetics* (pp. 258–274). Oxford: Oxford University Press.

Ebitz, D. (1988). Connoisseurship as practice. *Artibus et historiae*, *9*(18), 207–212.

Elgammal, A., Kang, Y., & Leeuw, M. D. (2018). *Picasso, Matisse, or a fake? Automated analysis of drawings at the stroke level for attribution and authentication*. Presented at the 32nd AAAI conference on Artificial Intelligence, 2-7 February, 2018, New Orleans, USA.

Ellenberger, H. (1954). Relations psychologiques entre le criminel et la victime. *Revue internationale de criminologie et de police technique et scientifique, 8*(2), 103–121.

Encyclopedia of Mathematics. (n.d.). *Probability*. Retrieved February 17, 2020, from https://www.encyclopediaofmath.org/index.php/Probability

Esterow, M. (2005, June 1). The 10 most faked artists. *ARTnews.com*. Retrieved October 5, 2021, from https://www.artnews.com/art-news/news/the-10-most-faked-artists-119/

Fattah, E. A. (1971). *La victime est-elle coupable?: Le rôle de la victime dans le meurtre en vue de vol*. Montreal: Presses de l'Université de Montréal.

Feigenbaum, G. (2013). Manifest of provenance. In G. Feigenbaum & I. Reist (Eds.), *Provenance: An alternate history of art* (pp. 6–29). Los Angeles: Getty Research Institute.

Feigenbaum, G., & Reist, I. (2013). *Provenance: An alternate history of art*. Los Angeles: Getty Research Institute.

Fenoff, R. (2013). Forgery. In L. M. Salinger (Ed.), *Encyclopedia of white-collar and corporate crime* (pp. 360–363). Thousand Oaks: SAGE Publications, Inc.

Fibicher, B., & Poletti, K. (2016). Communique de presse. Un Vallotton inédit pour le Musée cantonal des Beaux-Arts. *mcb-a*. Retrieved August 12, 2021, from http://www.publidoc.vd.ch/guestDownload/direct/Communiqu%C3%A9%20mcb-a-Vallotton_inedit_06%2001%202016.pdf?path=/Company%20Home/VD/CHANC/SIEL/antilope/objet/CE/Communiqu%C3%A9%20de%20presse/2016/01/543342_Communiqu%C3%A9%20mcb-a-Vallotton_inedit_06%2001%202016_20160106_1224485.pdf

Findlay, M. (2004). The Catalogue raisonné. In R. D. Spencer (Ed.), *The expert versus the object: Judging fakes and false attributions in the visual arts* (pp. 55–62). New York: Oxford University Press.

Flynn, T. (2016). The art world: Compliant victim? In A. Tompkins (Ed.), *Art crime and its prevention: A handbook for collectors and art professionals* (pp. 19–30). London: Lund Humphries.

Fong, W. (1962). The problem of forgeries in chinese painting. Part one. *Artibus Asiae*, *25*(2/3), 95–140.

Foradini, F. (2018, November 19). Notorious forger Wolfgang Beltracchi on ethics, the art market and how to make a great fake. *The Art News Paper*. Retrieved April 4, 2019, from http://theartnewspaper.com/interview/beltracchi-his-own-work-in-his-own-words

Foundation Félix Vallotton. (n.d.). *Publications: Félix Vallotton. Critique d'art*. Retrieved April 8, 2020, from https://felixvallotton.ch/publications/

Friedlander, M. J. (1944). *On art and connoisseurship*. London: Cassirer.

Friedrichsen, G. (2011). Eulenspiegen oder Künstler. *Der Spiegel, 43*, 148-149 (after Hufnagel & Chappell, 2016). Hufnagel, S., & Chappell, D. (2016). The Beltracchi affair: A comment and further reflections on the "most spectacular" german art forgery case in recent times. In N. Charney (Ed.), *Art crime: Terrorists, tomb raiders, forgers and thieves* (pp. 9–20). London: Palgrave Macmillan.

Gérard-Varet, L.-A. (1995). On pricing the priceless: Comments on the economics of the visual art market. *European Economic Review*, Papers and proceedings of the ninth annual congress European Economic Association, *39*(3), 509–518.

Gordley, J., & Mattei, U. (1996). Protecting possession. *The American Journal of Comparative Law*, *44*(2), 293–334.

Groom, N. (2007). Romanticism and forgery. *Literature Compass*, *4*(6), 1625–1649.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*(Mar), 1157–1182.

Haigh, J. (2012). *Probability: A very short introduction*. Oxford: Oxford University Press.

Han, B.-C. (2017). *Shanzhai: Deconstruction in Chinese*. Boston: The MIT Press.

Harris, G. (2020, June 30). Waning market for African artefacts? Controversial Benin bronze fails to sell at Christie's. *The Art Newspaper*. Retrieved August 19, 2021, from https://www.theartnewspaper.com/news/christies-african-objects-sale

Hentig, H. V. (1948). *The criminal & his victim; studies in the sociobiology of crime*. Oxford: Yale University Press.

Ho, T.-N. (1998). Art theft in New York City. *Empirical studies of the arts*, *16*(1), 41–46.

Hoving, T. (1997). *False impressions: The hunt for big-time art fakes*. New York: Simon & Schuster.

Hufnagel, S., & Chappell, D. (2016). The Beltracchi affair: A comment and further reflections on the "most spectacular" German art forgery case in recent times. In N. Charney (Ed.), *Art crime: Terrorists, tomb raiders, forgers and thieves* (pp. 9–20). London: Palgrave Macmillan.

Hutcheson, G. D. (2011). Dummy variable coding. In L. Moutinho & G. D. Hutcheson (Eds.), *The SAGE dictionary of quantitative management research* (pp. 99–102). London: SAGE.

Interpol. (n.d.). Stolen works of art database. *Interpol*. Retrieved June 14, 2019, from https://www.interpol.int/en/Crimes/Cultural-heritage-crime/Stolen-Works-of-Art-Database

ID-Art mobile app. (n.d.). *Interpol*. Retrieved December 29, 2022, from https://www.interpol.int/en/Crimes/Cultural-heritage-crime/ID-Art-mobile-app

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer-Verlag.

Johnson, K. C. (1973). Fakes, forgeries, and other deceptions. In S. Sachs (Ed.), *Fakes and forgeries*. Minneapolis: The Minneapolis Institute of Arts.

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, *20*(1), 141–151.

Kahneman, D. (2012). *Thinking, fast and slow*. London: Penguin Books.

Kallier, J. (2012). Galerie St. Etienne. *ArtBanc Intelligence, 1*(Oct.), 14–14 (after Neuhaus, 2014). Neuhaus, N. M. (2014). Art Authentication: Protection of Art Experts from a Swiss Perspective. *Art Antiquity and Law*, *19*(1), 59–74.

Keating, T., Norman, F., & Norman, G. (1977). *The Fake's progress*. London: Hutchinson.

Kent, R. A. (2015). *Analysing quantitative data: Variable-based and case-based approaches to non-experimental datasets*. Thousand Oaks, Californie: SAGE.

Koch, I. (2014). *Analysis of multivariate and high-dimensional data*. New York: Cambridge University Press.

Kornegay, C., & Segal, J. B. (2013). Developing a protocol for observation comparative effectiveness research: A user's guide. In P. Velentgas, N. A. Dreyer, P. Nourjah, S. R. Smith, & M. M. Torchia (Eds.), *Selection of data sources* (pp. 109–125). US: Agency for Healthcare Research and Quality.

Landesman, P. (1999, July 18). A 20th-century master scam. *The New York Times*. Retrieved April 5, 2019, from https://www.nytimes.com/1999/07/18/magazine/a-20th-century-master-scam.html

Larson-Hall, J. (2013). Factor analysis. In P. J. Robinson (Ed.), *The Routledge encyclopedia of second language acquisition* (pp. 237–240). New York: Routledge.

Lee, T. C. (1981). Van Gogh's vision. Digitalis intoxication? *JAMA*, *245*(7), 727–729.

Lemoine, S. (1992). Le role de l'expert dans la pratique. In Q. Byrne-Sutton & M.-A. Renold (Eds.), *L'expertise dans la vente d'objets d'art - aspects juridiques et pratiques* (pp. 67-71) (after Neuhaus, 2014). Neuhaus, N. M. (2014). Art authentication: Protection of art experts from a Swiss perspective. *Art Antiquity and Law*, *19*(1), 59–74.

Lenain, T. (2015). The Narrative structure of forgery tales. In J. Kila & M. Balcells (Eds.), *Cultural property crime: An overview and analysis on contemporary perspectives and trends* (pp. 37–60). Leiden: Brill.

Levi, M. (2008). Organized fraud and organizing frauds: Unpacking research on networks and organization. *Criminology & Criminal Justice*, *8*(4), 389–419.

Lewis, B. (2019). *The last Leonardo: The making of a masterpiece*. London: HarperCollins Publishers.

Loll, C. (2016). Art, anxiety and the online marketplace. In A. Tompkins (Ed.), *Art crime and its prevention: A handbook for collectors and art professionals* (pp. 60–72). London: Lund Humphries.

Lombardi, T. E. (2005, January 1). *The classification of style in fine-art painting* (thesis). Pace University, Ann Arbor, United States. Retrieved from https://search.proquest.com/docview/305390005/abstract/89F25FEDFFCB467EPQ/1

Lopez, J. (2009). *The man who made Vermeers: Unvarnishing the legend of master forger Han van Meegeren*. Boston: Mariner Books.

Mcgivern, H. (2018, April 30). French museum discovers more than half its collection is fake. *The Art Newspaper*. Retrieved October 10, 2019, from http://www.theartnewspaper.com/news/french-museum-discovers-more-than-half-of-its-collection-is-fake

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction* (1st ed.). New York: The Guilford Press.

Mendelsohn, B. (1976). Victimology and contemporary society's trends. *Victimology: An International Journal*, *1*(1), 8–28.

Merriam-Webster, Inc. (n.d.). *Definition of aesthetic*. Retrieved April 8, 2020, from https://www.merriam-webster.com/dictionary/aesthetic

Merryman, J. H. (1992). Counterfeit art. *International Journal of Cultural Property*, *1*(1), 27–78.

Michalska, J. (2011). Charlotte Burns and Ermanno Rivetti. *The Art Newspaper*, *230*(Dec.) (after Hufnagel & Chappell, 2016). Hufnagel, S., & Chappell, D. (2016). The Beltracchi affair: A comment and further reflections on the "most spectacular" German art forgery case in recent times. In N. Charney (Ed.), *Art crime: Terrorists, tomb raiders, forgers and thieves* (pp. 9–20). London: Palgrave Macmillan.

Moisen, G. G. (2008). Classification and regression trees. In S. E. Jorgensen & B. D. Fath (Eds.), *Encyclopedia of ecology* (Vol. 1, pp. 582–588). Oxford, U.K.: Elsevier.

Montani, I. (2015). *Exploring transparent approaches to the authentication of signatures on artwork* (thesis). University of Lausanne, Switzerland.

Moody, O. (2018, June 11). Row over art of spotting forgeries. *The Times*. Retrieved October 4, 2021, from https://www.thetimes.co.uk/article/row-over-art-of-spottingforgeries-zh3rvskht

Morelli, G., & Anderson, J. (1994). *De la peinture italianne: Les fondements de la théorie de l'attribution en peinture à propos de la collection des Galeries Borghèse et Doria-Pamphili*. Paris: Lagune.

Moulin, R. (2003). *Le Marché de l'art: Mondialisation et nouvelles technologies*. Paris: Flammarion.

Moureau, N., & Sagot-Duvauroux, D. (2016). *Le marché de l'art contemporain* (3e édition.). Paris : La Découverte.

Moureau, N., Sagot-Duvauroux, D., Vidal, M. (2016). *Collectionneurs d'art contemporain: Des acteurs méconnus de la vie artistique*. Paris: Ministère de la Culture - DEPS. Retrieved May 27, 2021, From https://www.cairn.info/collectionneurs-d-art-contemporain--9782111281608.htm

Moynihan, C. (2016, February 7). Knoedler Gallery director settles lawsuit over fake Rothko. *The New York Times*. Retrieved June 18, 2019, from

https://www.nytimes.com/2016/02/08/arts/design/knoedler-gallery-director-settles-lawsuit-over-fake-rothko.html

MutualArt. (2018, January 9). The Most Faked Artists in History. Retrieved October 5, 2021, from https://www.mutualart.com/Article/The-Most-Faked-Artists-in-History/ADD1DF7752FEBC5D

Nakache, J.-P., & Confais, J. (2005). *Approche pragmatique de la classification: Arbres hiérarchiques, partitionnements*. Paris: Technip.

Neuhaus, N. M. (2014). Art authentication: Protection of art experts from a Swiss perspective. *Art Antiquity and Law*, *19*(1), 59–74.

Noce, V. (2018, June 12). French dealer awarded rights to Chinese artist T'ang Haywen's work. *The Art Newspaper*. Retrieved October 10, 2019, from http://www.theartnewspaper.com/news/dealer-awarded-t-ang-haywen-rights

Ochberg, F. (1980). An interview with frank Ochberg. *Evaluation and change*, *Special issue*, 21–28 (after Barill, 1984). Baril, M. (1984). *L'envers du crime. Etude victimologique. Cahier no 2.* Retrieved October 4, 2019, from https://www.cicc-iccc.org/fr/publications/les-cahiers-de-recherches-criminologiques-1984-2005/cahier-no-2-1984

O'Connor, F. V. (2004). Authenticating the attribution of art: Connoisseurship and the law in the judging of forgeries, copies, and false attributions. In R. D. Spencer (Ed.), *The expert versus the object: Judging fakes and false attributions in the visual arts* (pp. 3–27). New York: Oxford University Press.

O'Connor, F. V., & Thaw, E. V. (1978). *Jackson Pollock: A catalogue raisonné of paintings, drawings, and other works*. New Haven: Yale University Press.

Olabode, S. O., Olateju, O. I., & Bakare, A. A. (2019). An assessment of the reliability of secondary data in anagement science research. *International Journal of Business and Management Review*, *7*(3), 27–43.

Ouellet, G. (1998). *Statistique et probabilités*. Sainte-Foy: Le Griffon d'argile.

Oxford English Dictionary. (n.d.). *Oxford English Dictionary*. Retrieved March 13, 2019, from http://www.oed.com/

Oxford Reference. (n.d.). *Oxford Reference*. Retrieved September 13, 2019, from https://www.oxfordreference.com/view/10.1093/oi/authority.20110803100413159

Pergam, E. A. (2013). Provenance as pedigree: The marketing of British portraits in gilded age America. In G. Feigenbaum & I. Reist (Eds.), *Provenance: An alternate history of art* (pp. 104–122). Los Angeles: Getty Research Institute.

Picard, R. R., & Cook, R. D. (1984). Cross-Validation of Regression Models. *Journal of the American Statistical Association*, *79*(387), 575–583. Cutruzzolà, O. (2011). Présentaion. *PolCant Info*, *80*(March).

Plattner, S. (1998). A most ingenious paradox: The market for contemporary fine art. *American Anthropologist*, *100*(2), 482–493.

Polk, K., & Chappell, D., (2009). Fakes and deception: Examining fraud in the art market. In N. Charney (Ed.), *Art and crime: Exploring the dark side of the art world* (pp. 72–85). Santa Barbara: Praeger/ABC-CLIO.

Pomfret, J. (1999, January 17). The master forger. *Washington Post*. Retrieved December 11, 2019, from https://www.washingtonpost.com/archive/lifestyle/magazine/1999/01/17/the-master-forger/e2a6cc6a-c60d-4abc-9d6a-c4540c0f1076/

Porter, L. (2007, December 2). Art imitating art. *The age*. Retrieved June 14, 2019, from https://www.theage.com.au/entertainment/art-imitating-art-20071202-ge6fmd.html

Pryor, R. (2016). *Crime and the art market*. London: Lund Humphries.

Rapley, V. (2016). The police investigation of art fraud. In A. Tompkins (Ed.), *Art crime: Terrorists, tomb raiders, forgers and thieves* (pp. 33–40). London: Lund Humphries.

Raux, S. (2013). Frome Mariette to Joulain: Provenance and value in eighteenth—century French auction catalogs. In G. Feigenbaum & I. Reist (Eds.), *Provenance: An alternate history of art* (pp. 88–105). Los Angeles: Getty Research Institute.

Reeves, V. K. (2011). The rights and risks of experts in French and American courts. *IFAR Journal, 12*(4), pp. 18-18 (after Neuhaus, 2014). Neuhaus, N. M. (2014). Art authentication: Protection of art experts from a Swiss perspective. *Art Antiquity and Law*, *19*(1), 59–74.

Reiss, J., & Sprenger, J. (2020). Scientific Objectivity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020.). Metaphysics Research

Lab, Stanford University. Retrieved January 12, 2023, from https://plato.stanford.edu/archives/win2020/entries/scientific-objectivity/

Rousseau, T. (1968). The stylistic detection of forgeries. *The Metropolitan Museum of Art Bulletin*, *26*(6), 247–252.

Russell, J. (1999, August 23). Leo Castelli, Influential Art Dealer, Dies at 91. *The New York Times*. Retrieved May 28, 2021, from https://www.nytimes.com/1999/08/23/arts/leo-castelli-influential-art-dealer-dies-at-91.html

Sablatnig, R., Kammerer, P., & Zolda, E. (1998). *Hierarchical classification of paintings using face- and brush stroke models*. Presented at the 14th International Conference on Pattern Recognition, 20–20 Aug., Brisbane, Queensland.

Saleh, B., Abe, K., Arora, R. S., & Elgammal, A. (2014). Toward automated discovery of artistic influence. *Multimedia Tools and Applications*, *75*(7), 3565–3591.

Salinger, L. M. (2013). *Encyclopedia of White-Collar and Corporate Crime* (2nd edition.). Thousand Oaks, California: SAGE Publications, Inc.

Sanchez, S. (2019). Victims and Victim Typologies. In A. S. Burke, D. Carter, B. Fedorek, T. Morey, L. Rutz-Burri, & S. Sanchez (Eds.), *Introduction to the American Criminal Justice System* (pp. 57–60). Open Oregon Educational Resources. Retrieved September 27, 2019, from https://openoregon.pressbooks.pub/ccj230/chapter/1-14-victims-in-the-cj-system/

Savage, G. (1963). *Forgeries, fakes and reproductions: A handbook for the collector* (1st ed.). New York: Barrie & Rockliff.

Schafer, S. (1977). *Victimology: The victim and his criminal*. Reston: Reston Publishing Company.

Schweizer, M. (2016). The civil standard of proof—What is it, actually? *The International Journal of Evidence & Proof*, *20*(3), 217–234.

Separovic, Z.P. (1974). Victimology: A new approach in the social sciences. In I. Drapkin & E. Viano (Eds.), *Victimology: A new focus* (pp. 15–24). London: Lexington Books (after Baril, 1984). Baril, M. (1984). *L'envers du crime. Etude victimologique. Cahier no 2.* Retrieved October 4, 2019, from https://www.cicc-iccc.org/fr/publications/les-cahiers-de-recherches-criminologiques-1984-

2005/cahier-no-2-1984

Sharp, J. (2020). Subjectivity. In A. Kobayashi (Ed.), *International Encyclopedia of Human Geography (Second Edition)* (pp. 117–120). Oxford: Elsevier. Retrieved January 12, 2023, from https://www.sciencedirect.com/science/article/pii/B9780081022955108698

Shindell, L. M. (2016). Provenance and title risks in the art industry: Mitigating these risks in museum management and curatorship. *Museum Management and Curatorship*, *31*(5), 406–417.

Sotheby's. (2017, December 4). A Year of Scientific Research at Sotheby's. *Sothebys.com*. Retrieved August 19, 2021, from https://www.sothebys.com/en/articles/a-year-of-scientific-research-at-sothebys

Spencer, R. D. (2004a). Authentication in court factors considered and standards proposed. In R. D. Spencer (Ed.), *The expert versus the object: Judging fakes and false attributions in the visual arts* (pp. 189–216). New York: Oxford University Press.

Spencer, R. D. (2004b). The risk of legal liability for attributions of visual art. In R. D. Spencer (Ed.), *The expert versus the object: Judging fakes and false attributions in the visual arts* (pp. 143–188). New York: Oxford University Press.

Stebbins, Jr., T. E. (2004). The art expert, the law, and real life. In R. D. Spencer (Ed.), *The expert versus the object: Judging fakes and false attributions in the visual arts* (pp. 135–142). New York: Oxford University Press.

Strobl, C. (2013). Data mining. *The Oxford Handbook of Quantitative Methods in Psychology*, *2*(Mar.), 1–35.

Surden, H. (2014). Machine learning and law essay. *Washington Law Review*, *89*(1), 87–116.

Swiss Civil Code (1907). Status as of 1 January 2019 [RS 210].

Swiss Criminal Code (1937). Status as of 1 January 2019 [RS 311.0].

Theodorakis, N. (2013). Art fraud. In L. M. Salinger (Ed.), *Encyclopedia of white-collar and corporate crime* (2nd ed., pp. 54–55). Thousand Oaks: SAGE Publications, Inc.

Teegen, W.-R. (2002). Late 19 th–early 20th c. replicas, imitations and falsifications (?) of brooches from the Pyrmont spring: A light and scanning electron microscopic approach. *Journal of Roman Archaeology, Supplementary Series,*

39(2), 287–289 (after Craddock, 2009). Craddock, P. T. (2009). *Scientific investigation of copies, fakes and forgeries* (1st ed.). Oxford: Elsevier/Butterworth-Heinemann.

Tietze, H. (1948). *Genuine and False: Copies, imitations, forgeries* (1st ed.). London: Max Parrish & Co Limited.

Tijhuis, A. J. G. (2009). Trafficking in cultural artifact. In M. H. Tonry (Ed.), *The Oxford handbook of crime and public policy* (pp. 456–475). Oxford: Oxford University Press.

Trochim, W. M. K., Donnelly, J. P., & Arora, K. (2016). *Research methods: The essential knowledge base*. Boston: Cengage Learning.

Valentin, P. (2013, September 12). The Destruction of Fakes. *Art@Law*. Retrieved October 1, 2021, from https://www.artatlaw.com/the-destruction-of-fakes/

Van Dantzig, M. M. (1973). *Pictology: An analytical method for attribution and evaluation of pictures*. Leiden: Brill.

Vermond, E. (1895). Théorie générale de la possession en droit romain. Paris: L. Larose.

Von Stockhausen, T. (2013). The failure of provenance research in Germany. In G. Feigenbaum & I. Reist (Eds.), *Provenance: An alternate history of art* (pp. 124–136). Los Angeles: Getty Research Institute.

Wemmers, J.-A. (2003). *Introduction à la victimologie*. Montreal: Presses de l'Université de Montréal.

Winkleman, E. (2015). *Selling contemporary art: How to navigate the evolving market*. New York: Allworth Press.

Wieseman, M. E. (2010). *A closer look: Deceptions and discoveries*. London: National Gallery Company.

Wilkinson, A. (2013, August 19). The giveaway. *The New Yorker*. Retrieved February 28, 2019, from https://www.newyorker.com/magazine/2013/08/26/the-giveaway

Wurlod, O. (2021, Octobre 12). Marché de l'art – L'intelligence artificielle jette le doute sur un chef-d'œuvre. *24 heures*. Retrieved November 11, 2021, from https://www.24heures.ch/lintelligence-artificielle-jette-le-doute-sur-un-chef-duvre-933789440154

Yeide, N. H., Walsh, A., & Akinsha, K. (2001). *The AAM guide to provenance research*.

Washington: American Association of Museums.

APPENDICES

I. WHO ARE THE VICTIMS OF ART CRIME?

Victims of crime

A large variety of actors involved in art trade and art conservation, including museums, experts, auction houses, galleries, insurance companies, collectors, and archivists, may fall victim to criminal acts (Charney, 2009). Most attempts to analyse victims of art crimes are impeded and obstructed by the lack of appropriate data collection. In this regard, Pryor (2016, p. 51), referring to the research of scholar Truc-Nhu Thi Ho (Ho, 1998) as well as examples of criminal cases, pointed out that '*neither the US nor the UK has means to bring together victim data on art crime*'. Furthermore, the fact that victims are often hesitant to report crimes means that any figures which do exist will, at best, only be a partial representation of the true picture. Moreover, art crime is a relatively new area of study in the field of criminology, and researchers have not yet developed a typology for categorising victims in this area. Thus, the following questions arise: What does the term 'crime victim' mean? What is the mechanism of art fraud, considering the attitudes of the perpetrator and victims? Can a typology of victims based solely on cases of art fraud indicate the most vulnerable category of victims of art crimes?

This Chapter examines the victims of art crimes, focusing on art counterfeiting. First, we discuss several general perspectives on victimology based on previous and current conceptualisations of victim classification. Next, we address issues related to fraud victims and art fraud. By exploiting some elements of the general classification with the help of art counterfeiting cases, we develop a hypothetical typology for art fraud victims.

Crime victim typology

Baril (1984) explained that the term 'crime victim' can be defined in several ways. He extracted definitions from the literature, such as 'the victim is any person, whether natural or legal, who suffers as a result of an accident or harmful, intentional conduct;

victims are those who are killed, injured, or harmed in their right to property'[126], 'victims are people whose rights have been violated'[127], and 'a victim is one who has suffered in a human interaction, has been lowered in a dominance hierarchy as a result of this, and responds with resignation or rage' (Separovic, 1974; Ochberg, 1980 cited by Baril, 1984, pp. 256–257). Baril (1984) pointed out that, in general, these definitions are not very different from one another. Notably, the definitions always include elements of passivity and different forms of damage (i.e. moral, physical, or economic). At the same time, most definitions of the term 'victim' refer only to natural persons (Baril, 1985) even though the term may apply to companies, corporations, or various legal persons (Baril, 1984).

Theorists have developed several victim typologies, mainly focused '[…] on the situational and personal characteristics of victims and the relationship between victims and offenders' (Sanchez, 2019). Benjamin Mendelsohn (1976), the founder of the concept of victimology and creator of the term 'victimology', proposed a victim typology. He formulated several classes according to the degree of victim culpability, ranging from 'innocent victim' to 'guilty victim'[128] (Bergelson, 2005; Sanchez, 2019).

Professor Hans Von Hentig (1948) is another important figure in the history of victimology. In his book *The Criminal and His Victim* that there is a vast range of interactions between criminals and their victims. He postulated that *'[t]he collusion between perpetrator and victim is a fundamental fact of criminology. Of course, there is no understanding or conscious participation, but there is the interaction and an interchange of causative elements*' (Hentig, 1948, p. 436). Hentig (1948) proposed classifying victims according to psychological, sociological, and biological factors,

[126] This quote is a free translation of 'la victime est toute personne, physique ou morale, qui souffre S la suite d'un accident ou d'une conduite préjudiciable, intentionnelle; les victimes sont celles qui sont tuées, blessées ou lésées dans leur droit de propriété'.

[127] This quote is a free translation of 'les victimes seraient ainsi les personnes dont les droits ont été lésés'.

[128] Mendelsohn's classification is as follows: completely innocent victims, victims with minor guilt and ignorant victims, victims that are as guilty as the offender, victims that are more guilty than the offender, most guilty victims and victims who are guilty alone, and simulated or imaginary victims (Mendelsohn, 1976).

ranging from 'society-made' victims to 'born' victims (Hentig, 1948)[129]. First, he described what he called '*the general classes of victims'* (Hentig, 1948, p. 404). This group includes people who are victimised because of their weaknesses, such as the young, elderly, and women. Minorities and immigrants fall into this general category because of their marginalisation within the dominant society.

Mendelsohn's and Hentig's studies were followed by other typologies that used similar criteria by adding new elements to measure the level of susceptibility and participation of a victim in a crime (Wemmers, 2003). Considering socio-biological elements, Schafer (1977) developed a typology based on empirical data. His study provided frequency variations in combinations of age, sex, interpersonal relationships, occupation, and the temporal and spatial aspects of crime.

Barnes and Teeters (1953) proposed the category of the *negligent victim* type, explaining that '*the negligent and careless attitude of a person towards his belongings makes it easy for an offender to commit a crime*' (Barnes & Teeters, 1953, cited by Devasia, 1980, p. 228). In the typology of Ellenberger (1954), which was later adopted by Fattah (1971), Ellenberger[130] introduced the term 'victim-criminal' to indicate that a person who is a victim at one time may become an offender at another. He provided the example of an abused child who later becomes an offender: Ellenberger's second category is 'the latent victim'. Fattah describes persons in this category as '*individuals who reveal an unconscious disposition to become victims*'[131] (Fattah, 1971, p. 26). This category refers to the mutual relationship that exists between criminals and victims in certain cases.

Additionally, Fattah (1971) distinguished between crimes against public property and crimes against individuals by splitting victims into two categories. Notably, his first category is the *specific victim*, that is, natural persons, human beings, legal persons,

---

[129] Hentig's classification is '*female, young, old, mentally defective and mentally deranged, immigrant, minority, dull normal, depressed, acquisitive, wanton, lonesome or heartbroken, tormentor, and blocked, exempted, and fighting victims*' (Hentig, 1948, pp. 404–433).

[130] Ellenberger largely followed Hantig based on his three important concepts of '*the doer-sufferer', 'the potential victim',* and the *'subject-object relation*' (Ellenberger, 1954, p. 104).

[131] This quote is a free translation of '[…] les sugets qui révèlent une disposition inconsciente à devenir victims […]'.

public and private entities (i.e. departments, municipalities, and societies), and international legal persons (i.e. states and international organisations) (Fattah, 1971). Baril's (1984) ideas are a continuation of Fattah's categorisation of specific victims. He questioned whether the definition of victims should be limited to those who are directly harmed by harmful conduct or extended to individuals who suffer secondary or tertiary harm (e.g. families of abused persons or witnesses at the scene of an assault). This hypothesis can also be extended to the business sector because collusion, bankruptcy, fraud, and theft increase prices and interest, and even lead to the collapse of small businesses and layoffs, thus involving subsequently injured victims (Baril, 1984).

The second category identified by Fattah (1971) is *non-specific victims*. Fattah stated that '*any crime, whether or not it involves a specific victim, is committed by one or more social or legal institutions, but certain crimes, by their nature, are not committed against a specific victim but against a non-specific victim. In these cases, the victim of the crime is neither a natural nor a legal person but rather an abstraction. It is not an individual who is harmed who suffers harm; rather, the public is affected by a crime in one of its institutions*' [132] (Fattah, 1971, p. 17). These include public morals, public order, public health, and other institutions.

Later, Wemmers (2003) summarised three main explanations for the causes of becoming a victim. The first is the provocative role of the victim. The second approach is based on the observation of an unhealthy link between the people involved. The third is the victim's lifestyle; for example, people who frequent bars are at a higher risk of victimisation than those who do not.

Fraud and art fraud victims

Victimologists have conducted many surveys and studies, but they have particularly focused on violent crimes, such as aggravated assaults, forced rapes, and robberies. Fraud victims are not the main subject of their research (Croall, 2007; Levi, 2008;

[132] The quote is a free translation of 'Tout infraction, qu'elle comporte une victime spécifique ou non, port attente à une ou plusieurs institutions social ou juridique, mais certaines infractions, de par leur nature, ne se commettent pas contre une victime déterminée mais contre une victime non spécifique. Dans ces cas, la victime de l'infraction n'est ni personne physique, ni une personne morale, mais une abstraction. Ce n'est pas un particulier qui est lésé, qui subit un préjudice, mais le public qui est atteint par l'infraction dans l'une de ses institutions' (Fattah, 1971, p. 17).

Dodge, 2013; Button, McNaughton Nicholls, Kerr, & Owen, 2015). Button, Lewis, and Tapley (2009) explained that fraud in general terms is an infrequently reported crime for many different reasons, including that victims may not know that they are victims of fraud. Button et al. (2009) classified victims as 'unknown' victims and those who realised that they were victims of fraud as 'known' victims. Button et al. (2009) attempted to classify victim of fraud. They noted that the classification may relate to different degrees of cooperation between the victim and perpetrator[133], which may correspond to the type of fraud[134] involving the victim or may be based on losses and repeated victimisation.[135]

Lenain's (2015) criminological research effort sought to understand the mechanisms of art fraud by considering the attitudes of the offender and victim. Lenain (2015) stressed that art forgery should be examined globally, and proposed focusing on chains of action involving different agents. His main criteria were the various functions of the forgery's participants. He also used psychological perceptions to explain agents' behaviours. In particular, Lenain conducted a narratological analysis and employed descriptive expressions to classify offenders and victims. Thus, he replaced the term 'victim' with 'dupe'. According to Lenain, the definition of a dupe (as an actor) involves several sub-actors, either simultaneously or successively. The dupe-actors may be an art expert *who provides the certificate of authenticity, lab scientist, artist, or buyer,* (Lenain, 2015). Lenain determined the behaviour of art lovers by their spiritual commitment and the emotional nature of their investment in an object. Lenain supposed that professional actors in the art world, such as connoisseurs, museum directors, and experts, do not suffer the same kind of deeply personal loss if they realise that they have been deceived unless they are art lovers. This loss does not occur, because the damage is essentially financial or limited to the professional sphere. The key point is that *'[p]art of the forger's operation often consists in triggering an art-lover's kind of response on the part of the certifier who, in principle, should react in a*

[133] This typology includes '*victim making contact with offender; victim providing information about him/herself', 'victim allowing offender to turn business relationship into a personal one', 'victim allows offender to create false perception of situation which can then be exploited'*, and '*victim reveals personal financial information to offender*' (Button et al., 2009, p. 22).

[134] These types may include victims of identity or investment fraud.

[135] One example is a 'chronic victim'.

cold, professional way' (Lenain, 2015, p. 50). It is also important to note that this narratological analysis is based on stories and anecdotes that have flourished in the art literature. Lenain argued that '*[t]hese stories share a common core of recurrent features composing a consistent schema*' (Lenain, 2015, p. 43).

Elaborated victim typology

The assessment was conducted using different criteria, such as the nature of the victim's behaviour, the number of participants in a forgery case, and the location of the offender and victim. Readers must keep in mind that some of the suggested categories are not mutually exclusive and that the typology is strictly hypothetical and should be regarded as indicative. Ultimately, a thorough long-term study of this issue is appropriate.

Indirect and direct victims

Bazley (2010) identified a victim who can be characterised as an 'indirect' victim. He proposed considering museum visitors, '*who view a painting that is not really a work of famous master they so admire, or examine an object from an ancient civilization that was really not created in the era […]*', as victims of forgery even if they are not directly affected by the crime (Bazley, 2010, p. 184). To illustrate this point, we refer to a well-known case in which several museums in the United States held and exhibited fakes/forgeries made by one forger (Mark Landis). Matthew Leininger, the head of the conservation department at the Oklahoma City Museum of Art, discovered during an audit that all the donations received from Marc Landis were fakes/forgeries. Leininger therefore applied to the American Alliance of Museums to make sure that no other museum had received gifts from Landis. Regrettably, several museums (e.g. the Art Institute of Chicago and the Art Museum at the University of Kentucky) held paintings fabricated and gifted as personal donations by Landis that were fakes/forgeries (Wilkinson, 2013). Another example of an indirect victim considering museum forgeries is the Terrus Museum in a village in southern France. The Terrus Museum handed over approximately 80 of the 140 works in the collection of Etienne Terrus to the police after experts declared them fakes/forgeries. The suspicion came from art specialists who noted flagrant incoherencies, such as views of buildings that

did not exist before Terrus's death in 1922 and types of support that the artist did not use. The total amount of damage was estimated to be 160,000 euros (Mcgivern, 2018).

In addition, we suggest considering those who work with historical documents in archives that have been altered by criminals as 'indirect' victims. For example, it is worth mentioning a case in the United States, in which one person was able to modify so many documents in so many different archives by including false information to legitimise his forgeries that even after an investigation and judicial settlement, it was impossible to identify all of the damaged archives[136] and know exactly how many more cases did not come out and how many potential victims were not reported (Landesman, 1999). Theoretically, Fattah's (1971) concept of a *non-specific victim* can be applied to this case. Indeed, it is not a particular individual, but rather an abstraction or, in general terms, the public who is affected by a crime against an institution.

To illustrate a 'direct' victim's behaviour, we select a typical example of art forgery. A physician purchased several Chagall lithographs from a New York gallery via telephone for USD 80,000. When he tried to sell them at a Sotheby's auction house, it was discovered that he had bought forgeries. Furthermore, the owners of the gallery disappeared during the investigation. The physician has accepted the loss of his investment (McGill, 1987, cited by Conklin, 1994).

Multiple victims

Forgery can lead to multiple victimisations, which occur when one forgery successively passes through many art players. One of the most significant cases of art fraud in recent years, the Beltracchi case, illustrates this phenomenon[137]. Hufnagel and Chappell emphasised that Beltracchi's forgeries were sold by Sotheby's and Christie's and passed through the hands of several art dealers, experts, and private collectors (Michalska, 2011, cited by Hufnagel & Chappell, 2016, p. 14). Therefore, many civil law disputes have arisen. Trasteco Limited, a Maltese company, sued the Lempertz Auction House for damages for buying a fake/forgery Heinrich Campendonk for nearly 2.9 million euros. Lempertz requested an investment of 70,000 euros for the purchase

[136] More details on the 'Drewe and Maytt' case are provided in Section 3.2.2.1.
[137] More details on the 'Beltracchi' case are provided in Section 3.2.2.1.

of an X-ray fluorescence analysis machine, but the machine was unable to discover the 'bad' pigments used by Beltracchi (Burns, 2012, cited by Hufnagel & Chappell, 2016, p. 14). Moreover, other civil actions were taken against the art historian Werner Spies (Michalska, 2011, cited by Hufnagel & Chappell, 2016, p. 15). Outside of trial, Christie's and Sotheby's agreed to indemnify several buyers, and the Hilti Art Foundation asked the Dickinson Gallery to indemnify them for the forged Derain that they purchased for 4.5 million euros (Michalska, 2011 cited by Hufnagel & Chappell, 2016, p. 15).

Loll (2016) claimed that artworks are often submitted to qualified experts for authentication only after they have been purchased. Sometimes, when an owner discovers the truth after a purchase, he or she tries to remove the item by selling it again. Thus, the art object continues to circulate in the art market even though the truth is known; Loll (2016) called such objects 'hot potatoes'. A 'hot potato' phenomenon was also identified in the course of research for this dissertation in the archives of the Félix Vallotton Foundation (see the example of picture No. 6 of Section 2.2)

Multinational victims

As noted in the previous section, victims may reside in jurisdictions different from those of offenders. In other words, when a crime is committed, the forger may be located in a different place or on a different continent from the victim. This particularity is shared by art crime and cybercrime. Sometimes, time may pass between the commission of the forgery and its detection by the victim or a third party. These particularities regarding the victim's location and time periods are also described in the literature on organised and online fraud (Levi, 2008; Button et al., 2015).

On the one hand, Conklin (1994) emphasised that victims often live some distance from the location where they bought the counterfeit work and prefer not to spend substantial time and money making claims in the court of another jurisdiction. However, some victims are proactive. For example, a report published by the *Arte* newspaper describes Philippe Koutouzi, a French art dealer based in Hong Kong, and

ownership of the copyright of the works of Chinese artist T'ang Haywen (1927–1991). Koutouzi found many fakes/forgeries of T'ang's on the market and has denounced dozens of fakes/forgeries for sale in Paris, Brussels, and Hong Kong, claiming that his suspicions were confirmed by technical studies by the Swiss Federal Institute of Technology in Zurich (Noce, 2018).

Prudent, naïve, and negligent victims

Referring to art crime victims, Bazley (2010) compared purchasers of artwork to individuals who buy counterfeit luxury items, such as watches or purses, without caring about their authenticity as long as they look like the real thing. This type of potential victim is reminiscent of Barnes and Teeters' (1953) concept of a *negligent victim*, that is, a person with a negligent attitude that facilitates the commission of a crime. From our perspective, the concept of a negligent victim should be viewed with some degree of caution in the context of art forgery victimisation. Certainly, some buyers want to know what they are really buying and have made an effort to determine it, but despite their care, they could nevertheless become victims of counterfeit works of art. In this regard, Conklin (1994) underscored that while '*careful buyers can consult [the] literature and learn to distinguish genuine pieces from fakes, dishonest dealers study [the] literature and, knowing more than many of their customers, are able to dupe them*' (Conklin, 1994, p. 86). His statement can be linked to a case in which an Indian widow paid USD 325,000 for several Russian art pieces said to have been made by Fabergé. The New York dealer sold fake/forged Fabergé pieces to the Indian widow. He knew that the pieces were imitations and not by Fabergé. To seize this opportunity, he invented a false provenance and set prices according to the prices of authentic Fabergé works (Conklin, 1994, p. 86). According to Conklin (1994), naïve buyers often rely on dealer loyalty and do not practice caution or seek awareness of their situation.

The idea of categorising forgery victims according to the attitudes that lead them to be victimised has a place in theory; however, in practice, it is challenging to distinguish between prudent and imprudent victims. As in the example of the Indian widow, the term 'prudent' may have different interpretations depending on the degree of trust in a concrete case. Moreover, negligent victims who do not care about what they buy (as mentioned above) are not common. For this reason, we suggest using the terms

'naively prudent' victim, referring to amateurs who totally rely on the opinions of others involved in the art world without serious precautions[138], and 'prudent' victims, referring to amateurs who do not confide unconditionally in art professionals and take all necessary precautions on their own.

It should be noted that some of the suggested categories are not mutually exclusive. Indeed, one victim's profile can fit into more than one category. For instance, a 'prudent' victim may also be a 'direct', 'multiple', and 'multinational' victim. Notably, with regard to the examples of museums that expose fakes/forgeries[139], it is not fair to describe the museum's visitors as 'naively prudent' because they do not perceive and suspect that a painting is a fake/forgery. Contrariwise, the museum that is the 'direct' victim can be categorised as 'naively prudent' (i.e. the museum accepted a gift without taking the necessary precautions). Museum visitors can be defined as 'indirect' victims. Nevertheless, they can also be categorised as 'multiple'[140] victims, which does not rule out the possibility that some of them may also be 'multinational' victims. Thus, the proposed categorisation of victims is schematic. However, it is necessary to clarify which category of art actors (see Section 2.1) is vulnerable to art fraud.

[138] For example, precautions can include demanding the opinions of several experts, demanding scientific expertise or provenance research, or studying the existing literature.
[139] Examples include the Oklahoma City Museum of Art and the Terrus Museum.
[140] It should be stressed that 'indirect' victims are always 'multiple' victims, as this category applies to the public and therefore involves multiple actors.

II. EXHAUSTIVE LIST OF DATABASE VARIABLES, THEIR TYPES AND CORRESPONDING SOURCES

Appendix II provides a comprehensive summary of the variables extracted for both the original and the fake works. This serves as a valuable reference, offering an overview of the complete set of variables used in the analysis. This summary encompasses a range of characteristics considered during the sampling of the artworks.

The variables that are subject to dummy coding are labelled 'D', indicating that they have been recoded as dummy variables, while variables that cannot be recoded as dummy variables are labelled as 'N' for numerical variables and 'Nom' for categorical nominal variables. This labelling system aids in the organisation and understanding of the variables, allowing researchers and readers to easily identify the type and nature of each variable.

By reviewing the Appendix, readers can gain a deeper understanding of the variables involved in distinguishing between original and fake works, thereby facilitating further research and analysis in the field of art authentication.

Group	Variable	Descriptions	Type	Source
Number	Catalogue number	The numbers of the original works that correspond to the CR's numbers	N	CR
	File number	The numbers of fake/forgery/doubtful works that correspond to the numbers of the archive files	N	Files
Title	Title	The titles originating from the CR and from the archive files, mentioned in the documents (source language)	Nom	Both
	Untitled	Artworks without titles as mentioned in the archive documents	D	Both
Years	Year 1	Date of execution as noted in CR/archive or the first year of the period indicating the year	N	Both

		before which the painting was not made		
	Year 2	The year in which the painting's creation ended or the date after which the painting was not made	N	Both
	No date	Undated artwork	D	Both
	Illegible date	Some elements of numbers	D	Files
Authenticity	Original	Paintings by Vallotton from the CR	D	CR
	Fake	Artwork confirmed as fake by the Foundation	D	Files
	Doubtful	Artwork confirmed as doubtful by the Foundation	D	Files
	Hypothetical (assigned without signature)	Artwork without a signature or monogram which was assigned or presumed by the owners as original but not approved by Foundation	D	Files
Market	File's country	Country of a work's possessor who requested the Foundation's opinion	Nom D	Files
	CR's country	Country of the work's physical location	Nom D	CR
	Year of appearance	Year of application for first-time identification	N	Files
	Sale/online	Whether artwork was presented for sale in an auction house or online sale (last sale)	D	Both
	Sold	Whether there is concrete information that the painting was sold	D	Both
	Date sold	Year of this sale	N	Both
	Sold price	Price that was paid	N	Files
	Estimate price	Estimated price of the online sale	N	Files
Signature/ monogram	Without signature	Lack of a signature or monogram	D	Both

	Monogrammed	Existence of a monogram	D	Both
	Signature or seal (equivalent to a signature made by the family of Vallotton)	Existence of a signature or seal	D	Both
	Trace of signature/ monogram	Existence of only a trace of a signature	D	Both
	Atypical signature	Signature with errors or with other letters	D	Files
	Atypical location of signature	The signature is situated in an atypical location or, at times, over the shoulders of a figure in a portrait	D	Both
Place of the signature/ monogram		Lower right	D	Both
		Lower left	D	Both
		Upper right	D	Both
		Upper left	D	Both
		Bottom centre	D	Both
		Centre right	D	Both
		Centre left	D	Both
		Top centre	D	
Technical features	Size	Large: 61 cm ≤ X, Y ≤ 250 cm	D	Both
		Medium: 36 cm ≤ X, Y ≤ 60 cm	D	Both
		Small: 1 cm ≤ X, Y ≤ 35 cm	D	Both
	Subjects	Portrait	D	Both
		Identified person	D	Both
		Still life		Both
		Landscape	D	Both
		Private and public interior with or without figures	D	Both
		Great decorations	D	Both
		Mythological/allegorical or biblical	D	Both
		Various	D	Both
		Nudes	D	Both
		Copy of another famous painter	D	Both
	Techniques	Oil	D	Both
		Pastel	D	Both
		Tempera	D	Both
		Gouache	D	Both

		Unknown technique	D	Both
	Types of support	Canvas	D	Both
		Cardboard	D	Both
		Wood	D	Both
		Paper	D	Both
Back of the painting	Something on the back	Number/letter/date	D	Both
		Sticker	D	Both
		Seal	D	Both
		Signature/initials	D	Both
		Number/letter/sticker/seal signature/monogram written on the frame	D	Both
		Number/letter/sticker/seal signature/monogram written on the canvas	D	Both
		With errors or with other letters than the initials of Vallotton	D	Both
		Picture on the back		CR
Historical documents	Without provenance or uncertain	Lacking a chronology of the property or location of the painting (the origin is considered as the initial moment of the appearance of a painting).	D	Both
	Recent provenance	Chronology of the property begins much later than the initial moment of the painting's appearance. Owner or possessor of the artwork is known only	D	Both
	First provenance	Acquired from the artist	D	Both
	Localisation unknown	Current location of the work is unknown	D	CR
	Stolen	Artwork was stolen	D	CR
	Destroyed	Artwork was destroyed	D	CR
	Restored	Artwork was restored	D	Both
	Relining	Artwork was relined	D	Both
	Absent in commonplace book	The painting is not cited in the commonplace book	D	Cat

Periods	1882–1889 Youth	The beginning of Vallotton's career	D	Both
	1890–1900 Wood engravings and Nabi period	Vallotton joined the group of Nabis in 1893. He painted very little and devoted himself to engraving on wood. A few paintings were produced during this period, and Vallotton made some attempts to apply his method of reducing atmospheric phenomena to highly simplified plane figures	D	Both
	1901–1910	Vallotton reduced his work as an engraver to focus again on painting. The Nabis theme coexists with the exploration of new techniques	D	Both
	1901–1905 Landscape	The restitution of natural light effects and atmospheric phenomena continues to occupy Vallotton's work, but sunsets remain absent until 1910	D	Both
	1905–1910 Nude	The period in which the nude and its ramifications are presented in vast compositions of mythological or allegorical characters	D	Both
	1909–1915 Sunsets	The motif of the sunset over the sea, accompanied by infinite variations in colour	D	Both
	1915 –1917 War	Vallotton attempts to artistically express his perception of war	D	Both
	1916–1925 Sunsets	Period dominated by paintings of the sun setting over the sea	D	Both

III. TABLES OF VARIABLES DISTRIBUTION USED FOR SELECTION OF WORKING SETS

In the research phase to create the experimental set, we used a relatively randomly selected set of variables to evaluate and illustrate the statistical tools. However, in Section 6.3, we take a closer look at Vallotton's works by creating working sets that delve deeper into his artistic legacy. Our goal is to identify features that may indirectly reflect the historical background of his work and shed light on the uncertainties and intricacies that may have attracted the interest of the forgers.

Appendix III provides the distributions of lost categories and insights into the prevalence of specific characteristics. To identify the most representative features, we compared the percentages across the tables presented in this section. Through this meticulous process, we select the most representative characteristics that capture typical attributes that are no longer available; the reader can find the list of Section 6.3.

Table: Distribution within the group *Years*

Category	**Stolen (7)**		**Destroyed (9)**		**Common. b. (83)**		**Unknown locale (226)**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Without date	1	14.3	0	0.0	18	21.7	17	7.5
Without year (1)	0	0.0	0	0.0	0	0.0	0	0.0
Without year (2)	7	100	9	100	77	92.8	224	99.1
Illegible date	0	0.0	0	0.0	0	0.0	0	0.0

$N = 1{,}704$

Table: Distribution of the artwork according to the presentation for *Sale* and *Sold* paintings

Category	Stolen (7)		Destroyed (9)		Common. b. (83)		Unknown locale (226)	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Sale	0	0.0	0	0.0	18	21.7	59	26.1
Sold	0	0.0	0	0.0	18	21.7	59	26.1

N = 1,704

Table: Distribution of artwork according to the *Signature* or *Monogram*

Category	Stolen (7)		Destroyed (9)		Common. b. (83)		Unknown locale (226)	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Without signature	0	0.0	0	0.0	5	6.0	2	0.9
Monogrammed	0	0.0	0	0.0	3	3.6	2	0.9
Signature	6	85.7	0	0.0	69	83.1	143	63.3

N = 1,704

Table: Distribution of artwork according to the *Dimension*

Category	Stolen (7)		Destroyed (9)		Common. b. (83)		Unknown locale (226)	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Large	0	0.0	1	11.1	19	22.9	91	40.3
Medium	2	28.6	0	0.0	33	39.8	45	19.9
Small	3	42.9	0	0.0	30	36.1	27	11.9

N = 1,704

Table: Distribution of works according to their present *Countries*

Category	**Stolen (7)**		**Destroyed (9)**		**Common. b. (83)**		**Unknown locale (226)**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
France	3	42.9	0	0.0	9	10.8	4	1.8
USA	0	0.0	0	0.0	3	3.6	0	0.0
Italy	0	0.0	0	0.0	0	0.0	0	0.0
Germany	0	0.0	0	0.0	3	3.6	0	0.0
Belgium	0	0.0	0	0.0	0	0.0	0	0.0
Switzerland	2	28.6	0	0.0	38	45.8	6	2.7
Bulgaria	0	0.0	0	0.0	0	0.0	0	0.0
Netherlands	0	0.0	0	0.0	0	0.0	0	0.0
Spain	0	0.0	0	0.0	0	0.0	0	0.0
Finland	0	0.0	0	0.0	0	0.0	0	0.0
Austria	0	0.0	0	0.0	0	0.0	0	0.0
Monaco	0	0.0	0	0.0	1	1.2	0	0.0
England	0	0.0	0	0.0	1	1.2	1	0.4
Brazil	0	0.0	0	0.0	0	0.0	0	0.0
South Africa	0	0.0	0	0.0	0	0.0	0	0.0
Canada	0	0.0	0	0.0	0	0.0	0	0.0
Japan	0	0.0	0	0.0	0	0.0	0	0.0
Finland	0	0.0	0	0.0	0	0.0	0	0.0
Algeria	0	0.0	0	0.0	0	0.0	0	0.0
Russia	0	0.0	0	0.0	1	1.2	0	0.0

N = 1,704

Table: Distribution of artwork according to an *Atypical Signature* or *Monogram*

Category	Stolen (7)		Destroyed (9)		Common. b. (83)		Unknown locale (226)	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Trace	0	0.0	0	0.0	1	1.2	0	0.0
Atypical signature	0	0.0	0	0.0	0	0.0	4	1.8
Atypical location signature	0	0.0	0	0.0	0	0.0	0	0.0

N = 1,704

Table: Distribution of artwork according to the *Place of the signature*

Category	Stolen (7)		Destroyed (9)		Common. b. (83)		Unknown locale (226)	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Lower right	2	28.6	0	0.0	44	53.0	85	37.6
Lower left	3	42.9	0	0.0	23	27.7	34	15.0
Upper right	1	14.3	0	0.0	4	4.8	17	7.5
Upper left	0	0.0	0	0.0	2	2.4	4	4.8
Bottom centre	0	0.0	0	0.0	0	0.0	0	0.0
Centre right	0	0.0	0	0.0	0	0.0	0	0.0
Centre left	0	0.0	0	0.0	0	0.0	1	0.4
Top centre	0	0.0	0	0.0	0	0.0	0	0.0

N = 1,704

Table: Distribution of artwork according to the *Subject*

Category	Stolen (7)		Destroyed (9)		Commonplace book (83)		Unknown locale (226)	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Portrait	0	0.0	4	44.4	17	20.5	55	22.1
Portrait with p-n	0	0.0	4	44.4	14	16.9	41	18.1
Still-life	2	28.6	0	0.0	9	10.8	40	17.7
Landscape	4	57.1	1	11.1	36	43.4	76	33.6
Interiors w	0	0.0	1	11.1	5	6.0	6	2.7
Decorations	0	0.0	0	0.0	1	1.2	0	0.0
Nudes	1	14.3	2	22.2	13	15.7	43	19
Copy	0	0.0	0	0.0	2	2.4	3	1.3
Mythological	0	0.0	1	11.1	0	0.0	1	0.4
Various	0	0.0	0	0.0	2	2.4	9	4.0

N = 1,704

Table: Distribution of artwork according to the type of *Support*

Category	Stolen (7)		Destroyed (9)		Common. b. (83)		Unknown locale (226)	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Canvas	4	57.1	3	33.3	45	54.2	136	60.2
Cardboard	2	26.6	0	0.0	21	25.3	30	13.3
Wood	1	14.3	0	0.0	14	16.9	10	4.4
Paper	0	0.0	0	0.0	2	2.4	0	0.0

N = 1,704

Table: Distribution of artwork according to the *Technique*

Category	**Stolen (7)**		**Destroyed (9)**		**Common. b. (83)**		**Unknown locale (226)**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Oil	7	100	3	33.3	77	92.8	168	74.3
Pastel	0	0.0	1	11.1	0	0.0	3	1.3
Tempera	0	0.0	0	0.0	2	2.4	9	4.0
Gouache	0	0.0	0	0.0	1	1.2	0	0.0
Unknown technique	0	0.0	4	44.4	1	1.2	44	19.5

N = 1,704

Table: Distribution of artworks according to information about its *Provenance*

Category	**Stolen (7)**		**Destroyed (9)**		**Common. b. (83)**		**Unknown locale (226)**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Without provenance	0	0.0	8	88.9	2	2.4	42	18.6
Recent provenance	0	0.0	0	0.0	4	4.8	10	4.4
First provenance	4	57.1	3	33.3	9	10.8	77	34.1

N = 1,914

Table: Distribution of artworks according to the information on the *Back of the canvas*

Category	**Stolen (7)**		**Destroyed (9)**		**Common. b. (83)**		**Unknown locale (226)**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Number/letter	0	0.0	0	0.0	0	0.0	2	0.9
Sticker	0	0.0	0	0.0	0	0.0	0	0.0
Signature/ initial	1	14.3	0	0.0	5	6.0	5	2.2
Seal	0	0.0	0	0.0	0	0.0	0	0.0
Written on the canvas	0	0.0	0	0.0	0	0.0	0	0.0
With errors	0	0.0	0	0.0	0	0.0	0	0.0
Date	1	14.3	0	0.0	3	3.6	3	1.3
Written on the frame	0	0.0	0	0.0	0	0.0	1	0.4
Under the other painting	0	0.0	0	0.0	1	0.1	0	0.0

N = 1,914

Table: Distribution of artworks according to information about its *Provenance*

Category	**Stolen (7)**		**Destroyed (9)**		**Common. b. (83)**		**Unknown locale (226)**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Without provenance	0	0.0	8	88.9	2	2.4	42	18.6
Recent provenance	0	0.0	0	0.0	4	4.8	10	4.4
First provenance	4	57.1	3	33.3	9	10.8	77	34.1

N = 1,914

Table: Distribution of artwork according to the *Period*

Period	**Stolen (7)**		**Destroyed (9)**		**Commonplace b. (83)**		**Unknown locale (226)**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Before 1882	0	0.0	0	0.0	2	2.4	0	0.0
1882–1889 youth	0	0.0	4	44.4	11	13.3	30	13.3
1890–1900 wood	1	14.3	1	11.1	18	21.7	36	15.9
1901–1910	0	0.0	1	11.1	25	30.1	30	13.3
1901–1905 landscape	1	14.3	1	11.1	09	10.8	17	7.5
1906–1909 nude	0	0.0	1	11.1	3	3.6	13	5.8
1910–1914 sunset	0	0.0	0	0.0	0	0.0	0	0.0
1915–1916	0	0.0	0	0.0	0	0.0	0	0.0
1917 war	1	14.3	0	0.0	0	0.0	0	0.0
1916–1925 sunset and landscape	0	0.0	0	0.0	0	0.0	0	0.0

N = 1,704

Table: Distribution of artworks within the group *Title*

Category	**Stolen (7)**		**Destroyed (9)**		**Common. b. (83)**		**Unknown locale (226)**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Without title	0	0.0	0	0.0	0	0.0	0	0.0
With title	7	100.0	9	100.0	83	100.0	226	100.0

N = 1,704

IV. CART FINDINGS FOR THE ALTERNATIVE ASSESSMENT: FIGURES

The figures in Appendix IV present the results of an alternative approach[141] used to evaluate the artworks discussed in the case studies. This approach involves employing data mining analysis by incorporating all the variables identified and listed in the studies as potential inputs and then running the CART model to generate predictions[142].

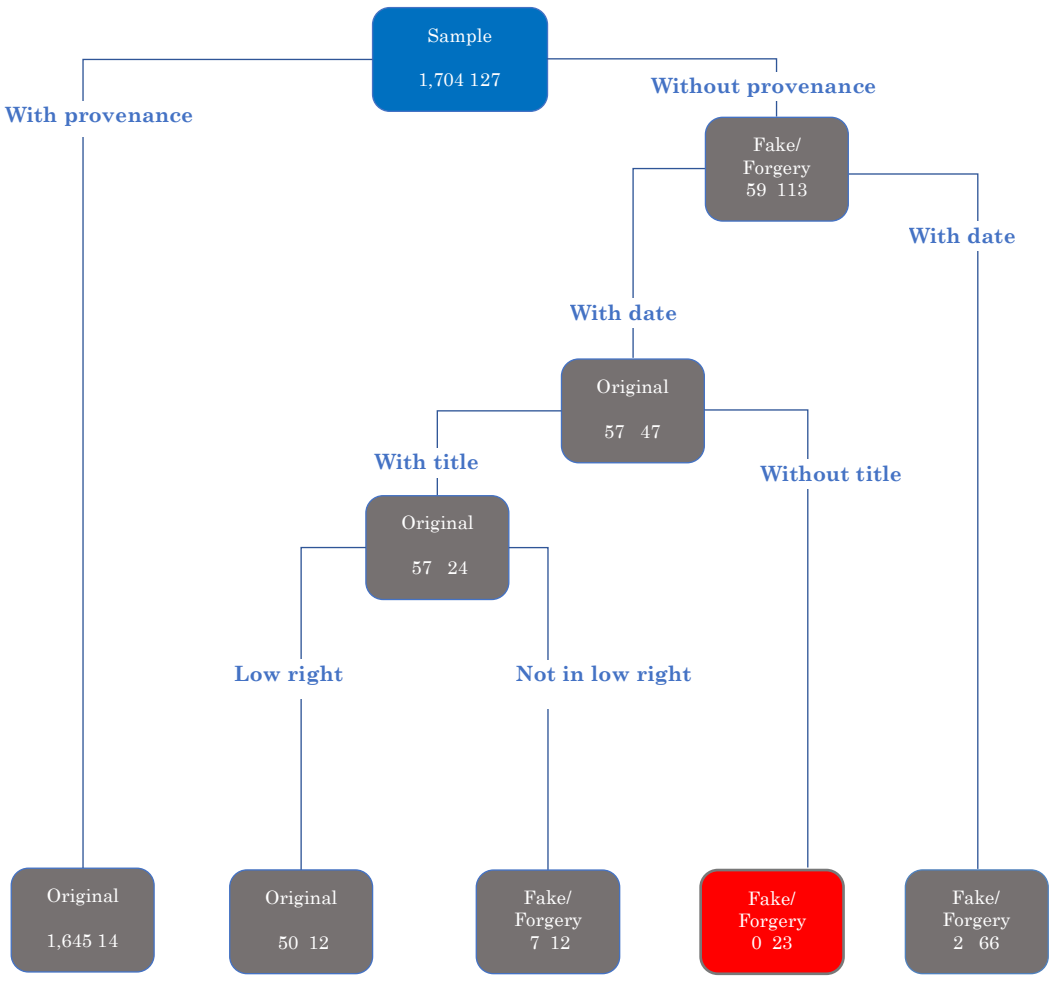


Figure 25: CART model: Original vs. Fake/forgery
(Set of listed variables of Case study 1)

[141] See Section 9.4.

[142] It should be note that these evaluations require individuals with expertise in computer programming and statistical analysis.

The artwork examined in the first case study (Section 9.2) corresponds to the classification tree shown in Figure 25.

Three characteristics of the examined painting match with the niche (in red, as depicted in Figure 25), which includes 'without provenance', 'with date', and 'without title' and has a 100% likelihood of being a counterfeit work.

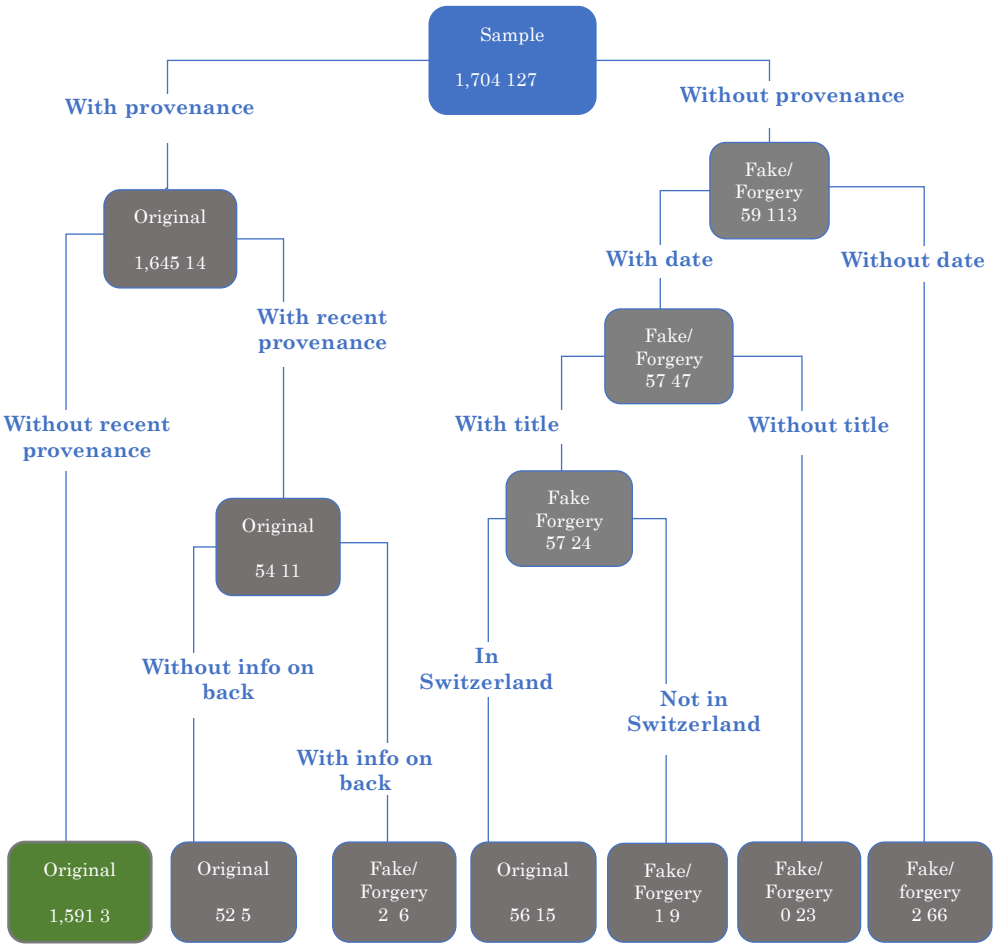


Figure 26: CART model: Original vs. Fake/forgery
(Set of listed variables of Case study 2)

Analysis of the second case study (Section 9.3) corresponds to the tree shown in Figure 26. Its characteristics match those of the green niche, with provenance and recent provenance indicating a 99.8% likelihood of being original.

V. QUESTIONS OF SAMPLING VARIABLES: OPTION OF MANUAL EVALUATION

Appendix V provides a series of questions accompanied by multiple-choice options. These questions were designed to assist users in developing a better understanding of the interpretative aspects of the artworks analysed in this study.

By presenting these questions and their corresponding answer choices, we aim to guide users through the process of critical analysis and interpretation.

Questions and answers

1. Is the title noted in the source (sales documents, such as artist's bills of sale, certificates, auctions, or exhibition catalogues)?

Title	**Answer ⇒**	
Yes		*With a title*
No		*Without a title*

Moreover, users can improve their interpretation skills and gain a deeper understanding of model construction.

2. Is a date noted in the source or observable in the photograph of the work?

Date	**Answer⇒**	
Yes		*With a date*
No		*Without a date*

3. Is a possessor of the work of art located in France? Is a possessor of the work of art located in Switzerland?

Switzerland	**Answer⇒**	
Yes		*In Switzerland*
No		*Not in Switzerland*
France	**Answer⇒**	
Yes		*In France*
No		*Not in France*

4. Does the source include information on the online sale of artworks?

Sale online	**Answer⇒**	
Yes		*Sale online*
No		*Not sale online*

5. Is a signature observable on the artwork's photo or mentioned in the source? Is a signature observable on the artwork's photo or mentioned in the source?

Signature	**Answer⇒**	
Yes		*With a signature*
No		*Without a signature*
Monogram		
Yes		*With a monogram*
No		*Without a monogram*

6. Does the artwork correspond to a size range of 61 cm to 250 cm?

Dimension	**Answer⇒**	
Yes		*Large*
No		*Not large*

7. Is any information in the form of number/letter/date/sticker noted in the source or observable on the photo of this work?

Information on back	**Answer⇒**	
Yes		*Having information on back*
No		*Without information on back*

8. Is the artwork landscape?

Landscape	**Answer⇒**	
Yes		*Landscape*
No		*Not a landscape*

9. In which period the artist created the artwork, and in which not?

1882–1889	**Answer⇒**	
Yes		*In 1882–1889*
No		*Not in 1882–1889*
1890–1900		
Yes		*In 1890–1900*
No		*Not in 1890–1900*
1901–1910		
Yes		*In 1901–1910*
No		*Not in 1901–1910*

10. Are the techniques (oil, gouache, tempera, or pastel) noted in the source?

Unknown technique	**Answer⇒**	
Yes		*Unknown technique*
No		*Having a known technique*

11. Does the source indicate that the artwork is noted in the commonplace book of the artist?

Commonplace book	**Answer⇒**	
Yes		*In commonplace book*
No		*Not in commonplace book*

12. Are one or several owners noted in the source? Are the current owners noted in the source?

Provenance	**Answer⇒**	
Yes		*With a provenance*
No		*Without a provenance*
Recent provenance		
Yes		*With recent provenance*
No		*Without recent provenance*

VI. QUESTIONS OF SAMPLING VARIABLES AND EXPLANATIONS: CASE STUDY 1

Appendix VI comprehensively explains the methodology behind the variable interpretation applied in Case Study 1. It should be noted that the questions and answers presented may vary depending on the artworks examined. The interpretations provided in Appendix VI are specific to this case study and should be understood in its unique context. Nevertheless, by referring to the logic behind the questions, readers can gain an overall understanding of the reasoning behind the various interpretations in the specific case study. This may prove helpful when adopting a specific approach to interpretation by analogy in future studies.

Questions and Answers

1. Is the title noted in the source (sales documents such as the artist's bills of sale, certificates, auctions, or exhibition catalogues)?

Title	**Answer**	
Yes		*With a title*
No	✓	*Without a title*

The record has not provided a title – the variable is *without a title*.

2. Is a date noted in the source or observable in the photograph of the work?

Date	**Answer**	
Yes	✓	*With a date*
No		*Without a date*

The date that is observable on the photo of the artwork on the canvas is '99' – the variable is *with a date*.

3. In which country is a possessor of the work of art located?

Switzerland	**Answer**	
Yes		*In Switzerland*
No	✓	*Not in Switzerland*
France		
Yes	✓	*In France*
No		*Not in France*

The possessor of the artwork is in France – the variable is *in France*. Consequently, by default, the other variable is *not Switzerland*.

4. Does the source include information on the online sale of artworks?

Sale online	**Answer**	
Yes	✓	*Sale online*
No		*Not sale online*

The painting was offered for sale on an online sales platform, and the variable is *sale online*.

5. Is a signature or monogram observable on the artwork's photo (or mentioned in the source)?

Signature	**Answer**	
Yes	✓	*With a signature*
No		*Without a signature*
Monogram		
Yes		*With a monogram*
No	✓	*Without a monogram*

The signature is observable in the photo of the artwork: 'F. Vallotton 99' – the variable is *with a signature*. Simultaneously, the signature is not a monogram; thus, by default, the other variable is *without a monogram*.

6. Does the artwork correspond to a size range of 61 cm to 250 cm?

Dimension	**Answer**	
Yes	✓	*Large*
No		*Not large*

The dimension of the work is 65 × 82.5 cm. Accordingly, this size corresponds to the variable large (see Appendix II): 61 cm ≤ 65, 82.5 ≤ 250 cm – the variable is *large*.

7. Is any information in the form of number/letter/date/sticker noted in the source or observable on the photo of this work?

Information on back	**Answer**	
Yes	✓	*Having information on back*
No		*Without information on back*

The inscriptions on the back are '1865–1925', the number '344', a stock label of Germany and the number '62550' and 'Bourse France' – the variable is *Having information on the back*.

8. Is the artwork a landscape?

Landscape	**Answer**	
Yes		*Landscape*
No	✓	*Not a landscape*

The image in the photo corresponds to the interior of a house. Therefore, it is not a landscape. Consequently, by default, the variable is *not a landscape*.

9. In which period did the artist create the artwork, and in which not?

1882–1889	**Answer**	
Yes		*In 1882–1889*
No	✓	*Not in 1882–1889*
1890–1900		
Yes	✓	*In 1890–1900*
No		*Not in 1890–1900*
1901–1910		
Yes		*In 1901–1910*
No	✓	*Not in 1901–1910*

Félix Vallotton lived from 1865 to 1925. Therefore, '99'could correspond to 1899. The date corresponds to the period of the woodcuts and Nabi from 1890 to 1900 – the variable is *in 1890–1900*. Accordingly, no artwork was created during the other two periods. Thus, the additional variables by default are *not in 1901–1910* and *not in 1882–1889*.

10. Are the techniques (oil, gouache, tempera, or pastel) noted in the source?

Unknown technique	**Answer**	
Yes		*Unknown technique*
No	✓	*Having a known technique*

The painting technique is oil, so the answer is 'no'. The technique is noted in the document source. Therefore, the variable is *having known technique*.

11. Does the source indicate that the artwork is noted in the commonplace book of the artist?

Commonplace book	**Answer**	
Yes		*In commonplace book*
No	✓	*Not in commonplace book*

This painting is not included in the catalogue raisonné. Therefore, there is no indication that this artwork was noted in Vallotton's commonplace book.

12. Are one or several owners noted in the source?

Provenance	**Answer**	
Yes		*With a provenance*
No	✓	*Without a provenance*

13. Is the current owner noted in the source?

Recent provenance		
Yes		*With recent provenance*
No	✓	*Without recent provenance*

The painting came from a private collection with no indication of the owners' names or other relevant information. According to our classification criteria, the absence of ownership traces in the documentation should be interpreted as 'without provenance'. Thus, the painting from an unknown private collection correspond to the variable *without a provenance and without recent provenance.*

VII. QUESTION OF SAMPLING VARIABLES WITH EXPLANATIONS: CASE STUDY 2

Similar to Appendix VI, this section presents a step-by-step breakdown of the logic and reasoning used to interpret each variable in Case Study 2. It outlines the criteria and considerations taken into account when assigning categories to specific variables to ensure transparency and clarity in the interpretation process. This knowledge empowers readers to critically analyse and interpret variables in their own research or when examining similar case studies in the field of art research.

Questions and Answers

1. Is the title noted in the source (sales documents such as the artist's bills of sale, certificates, auctions, or exhibition catalogues)?

Title	**Answer**	
Yes	✓	*With a title*
No		*Without a title*

The title is known – the variable is *with a title*.

2. Is a date noted in the source or observable in the photograph of the work?

Date	**Answer**	
Yes	✓	*With a date*
No		*Without a date*

The date that is observable in the photo of the artwork on the canvas is '02' – the variable is *with a date*.

3. In which country is a possessor of the work of art located?

Switzerland	**Answer**	
Yes	✓	*In Switzerland*
No		*Not in Switzerland*
France		
Yes		*In France*
No	✓	*Not in France*

The possessor of the artwork lives in Switzerland – the variable is *in Switzerland*. Consequently, the other variable is *not in France* by default.

4. Does the source include information on the online sale of artworks?

Sale online	**Answer**	
Yes	✓	*Sale online*
No		*Not sale online*

There is no information in the source indicating that the painting has been offered for sale on an online sales platform – the variable is *not sale online*.

5. Is a signature or monogram observable on the artwork's photo or mentioned in the source?

Signature	**Answer**	
Yes	✓	*With a signature*
No		*Without a signature*
Monogram		
Yes		*With a monogram*
No	✓	*Without a monogram*

The signature is observable on the photo of artwork 'F. Vallotton. 02'– the variable is *with a signature*. At the same time, the signature is not a monogram, so the other variable is *without a monogram*.

6. Does the artwork correspond to a size range of 61 cm to 250 cm?

Dimension	**Answer**	
Yes		*Large*
No	✓	*Not large*

The dimensions are 34 × 59.5 cm. This size corresponds to a medium-size artwork (Medium: 36 cm ≤ 34, 59.5 ≤ 60 cm, see Appendix II). Thus, the variable by default is *not large*.

7. Is any information in the form of number/letter/date/sticker noted in the source or observable on the photo of this work?

Information on back	**Answer**	
Yes	✓	*Having information on back*
No		*Without information on back*

The inscriptions on the back are '*mer haute Villerville 1902*' (in French), the number '28', and the label '*Kunstsalon Emil Richter Dresden'* – the variable is *information on back*.

8. Is the artwork landscape?

Landscape	**Answer**	
Yes	✓	*Landscape*
No		*Not a landscape*

The image shows horizontal parallel lines representing a sea of warm grey and muddy white colours – the variable is *landscape*.

9. In which period the artist created the artwork, and in which not?

1882–1889	**Answer**	
Yes		*In 1882–1889*
No	✓	*Not in 1882–1889*
1890–1900		
Yes		*In 1890–1900*
No	✓	*Not in 1890–1900*
1901–1910		
Yes		*In 1901–1910*
No	✓	*Not in 1901–1910*

The number '02' refers to 1902, as this date is also written on the back of the painting. This date corresponds to the landscape period, 1900–1905, which is characterised by the reproduction of natural light effects and atmospheric phenomena. This period does not correspond to the period of the discriminative variables in the models. Thus, we choose by default *not in 1901–1910, not in 1882–1889*, and *not in 1890–1900*.

10. Are the techniques (oil, gouache, tempera, or pastel) noted in the source?

Unknown technique	**Answer**	
Yes		*Unknown technique*
No	✓	*Having a known technique*

The painting technique is oil, so the answer is 'no'. The technique is noted in the document source. Therefore, the variable is *having known technique.*

11. Does the source indicate that the artwork is noted in the commonplace book of the artist?

Commonplace book	**Answer**	
Yes		*In commonplace book*
No	✓	*Not in commonplace book*

As noted in point *a*, this artwork was not included in the catalogue raisonné. It was not mentioned explicitly in the artist" commonplace book – the variable is *not in the commonplace book*.

12. Are one or several owners noted in the source?

Provenance	**Answer**	
Yes	✓	*With a provenance*
No		*Without a provenance*

13. Is the current owner noted in the source?

Recent provenance		
Yes	✓	*With recent provenance*
No		*Without recent provenance*

The painting was kept in a private collection. The names of the possessors and heirs are noted in the source. Therefore, according to our criteria of interpretation such data should be transformed to two: *with provenance* and *with recent provenance*.

VIII. IMPROVED FIGURES FOR ENHANCED VARIABLE VISIBILITY IN PCA WITH TWO COMPONENTS

In this Appendix, we present improved figures for enhanced variable visibility in Principal Component Analysis (PCA) with two components. It aims to provide readers with a visual representation of the results obtained through PCA, specifically focusing on the enhanced visibility of the variables, as the figures presented in the text of this thesis have a reduced format.

All figures illustrate the distribution and relationship of the variables in a two-dimensional space, allowing for a clearer understanding of their contribution to the overall variance in the dataset. Readers can observe how the variables cluster together, identify potential outliers, and gain insights into the underlying patterns and structures within the dataset.

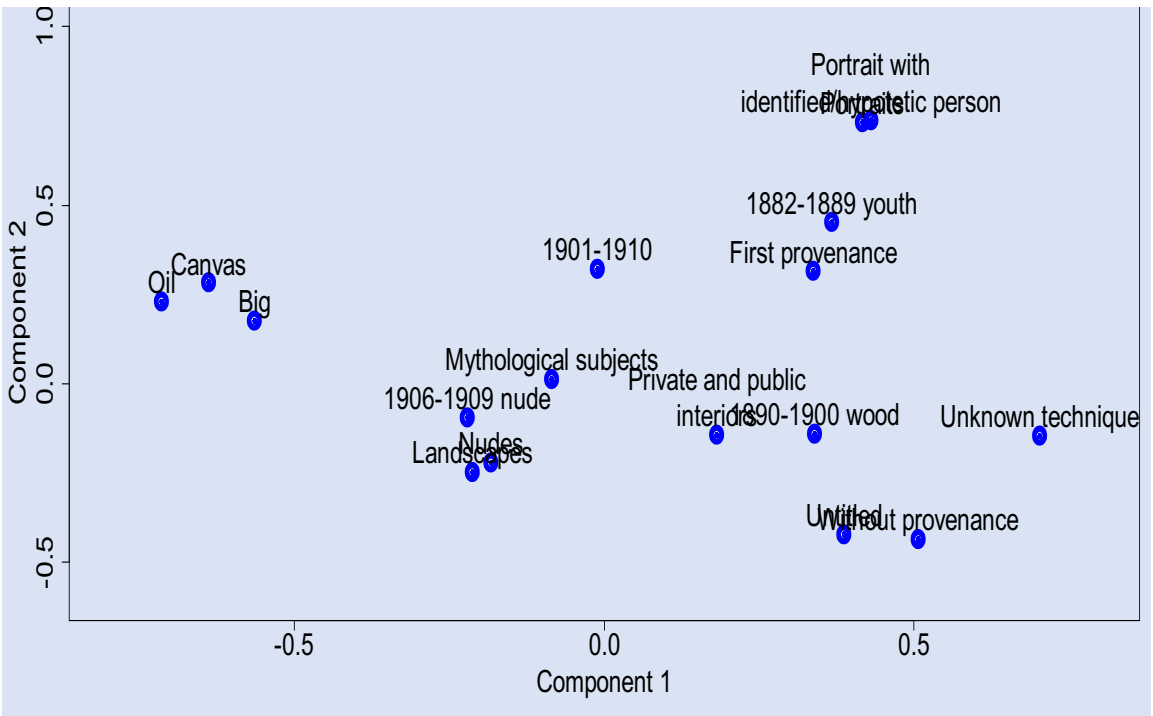


Figure 8: Two-dimensional score plot (Working Set 1)

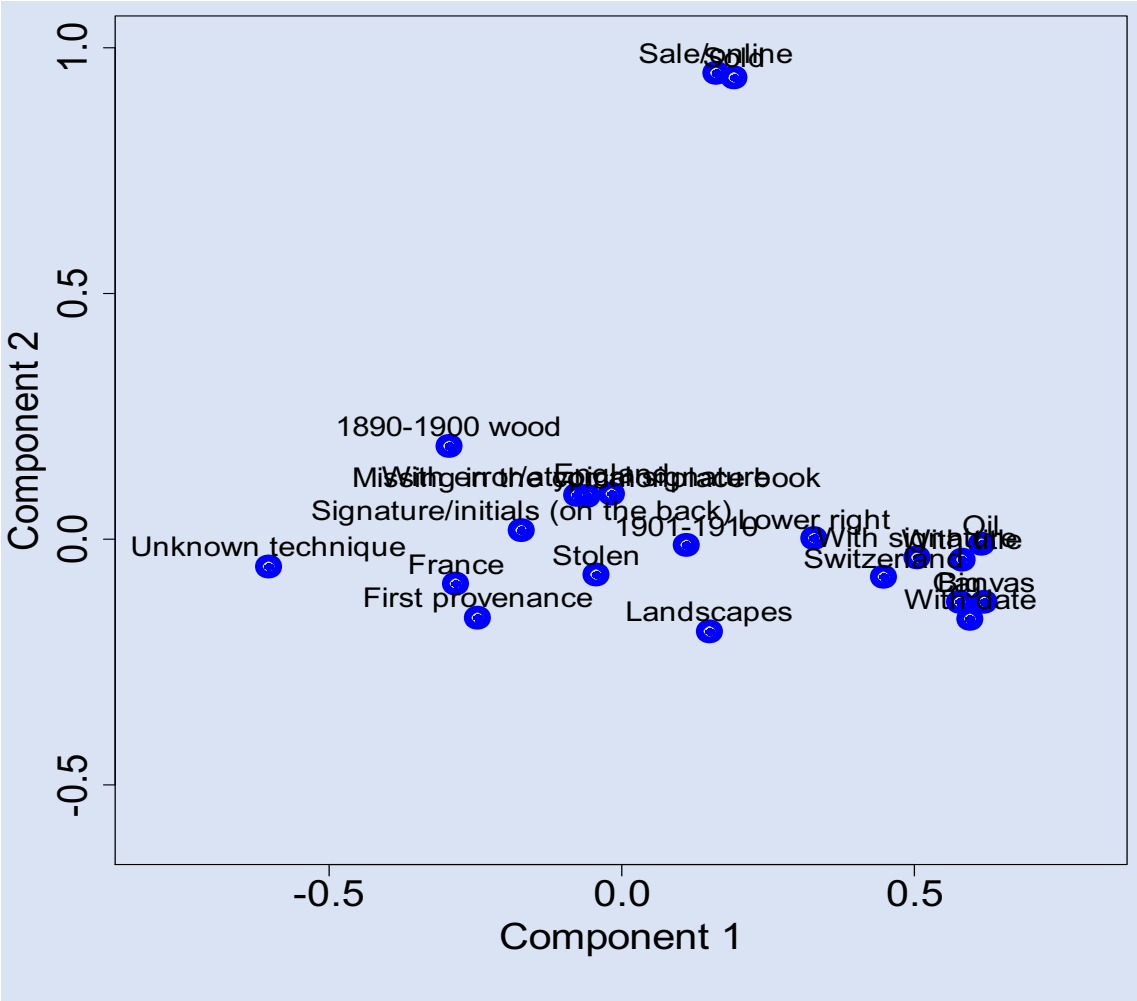


Figure 11: Two-dimensional score plot (Working Set 2)

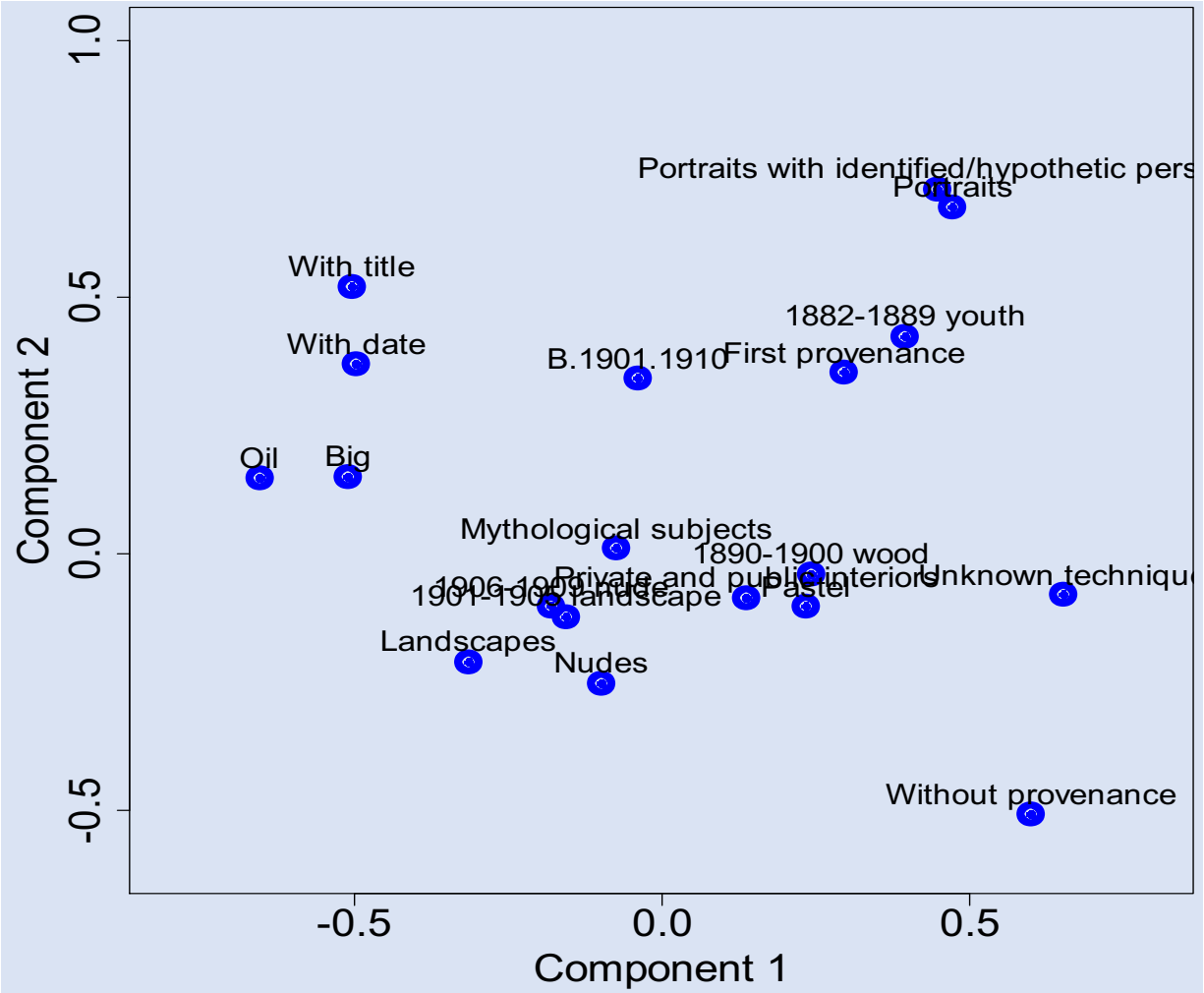


Figure 14: Two-dimensional score plot
(Working Set 3)

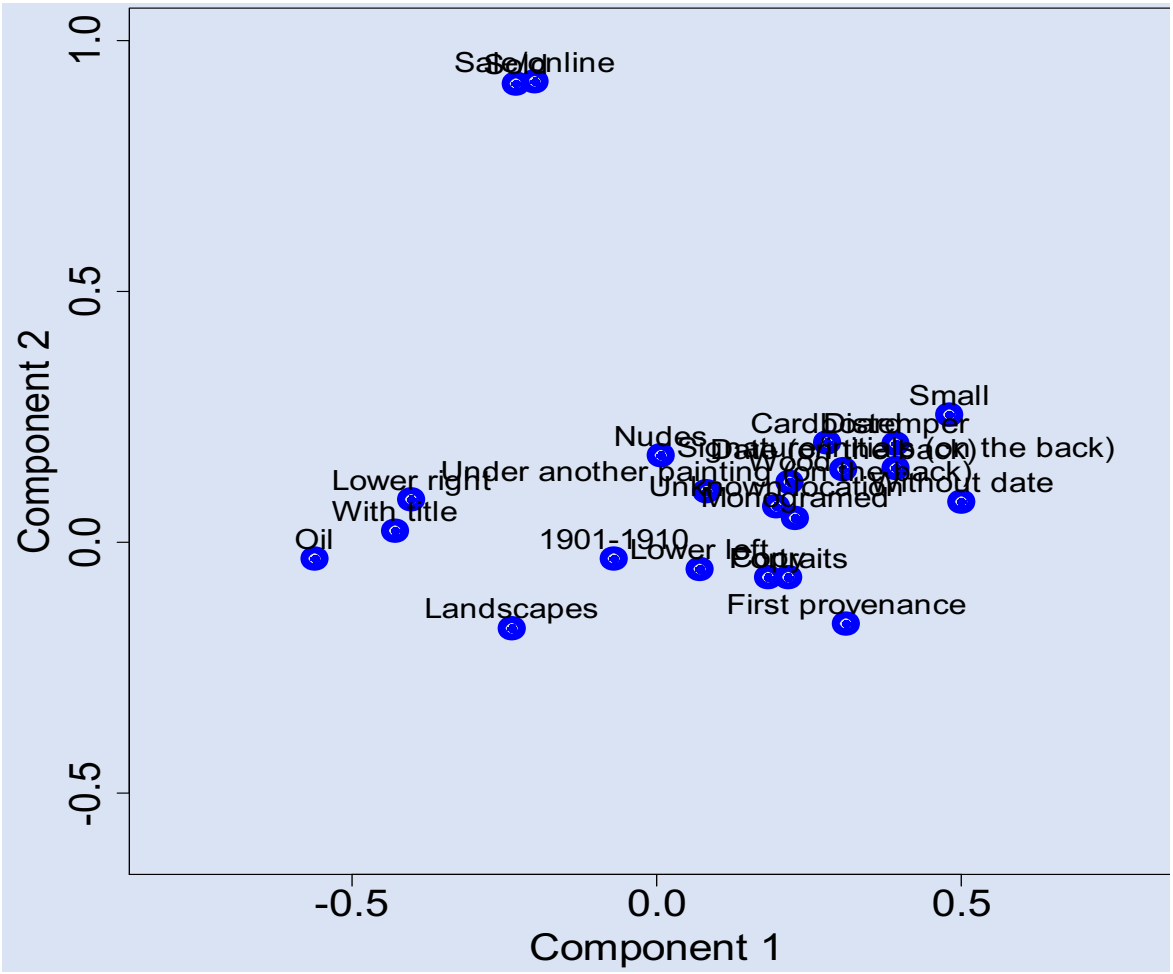


Figure 17: Two-dimensional score plot
(Working Set 4)