

# A Novel Dataset for Non-Destructive Inspection of Handwritten Documents

Eleonora Brecci, Luca Guarnera, Sebastiano Battiato; Department of Mathematics and Computer Science, University of Catania, Italy

## Abstract

Forensic handwriting examination is a branch of Forensic Science that aims to examine handwritten documents in order to properly define or hypothesize the manuscript's author. These analysis involves comparing two or more (digitized) documents through a comprehensive comparison of intrinsic local and global features. If a correlation exists and specific best practices are satisfied, then it will be possible to affirm that the documents under analysis were written by the same individual. The need to create sophisticated tools capable of extracting and comparing significant features has led to the development of cutting-edge software with almost entirely automated processes, improving the forensic examination of handwriting and achieving increasingly objective evaluations. This is made possible by algorithmic solutions based on purely mathematical concepts. Machine Learning and Deep Learning models trained with specific datasets could turn out to be the key elements to best solve the task at hand. In this paper, we proposed a new and challenging dataset consisting of two subsets: the first consists of 21 documents written either by the classic "pen and paper" approach (and later digitized) and directly acquired on common devices such as tablets; the second consists of 362 handwritten manuscripts by 124 different people, acquired following a specific pipeline. Our study pioneered a comparison between traditionally handwritten documents and those produced with digital tools (e.g., tablets). Preliminary results on the proposed datasets show that 90% classification accuracy can be achieved on the first subset (documents written on both paper and pen and later digitized and on tablets) and 96% on the second portion of the data. The datasets are available at <https://iplab.dmi.unict.it/mfs/forensic-handwriting-analysis/novel-dataset-2023/>.

## Introduction

Handwritten document analysis is a specialized field within forensic science, dedicated to the meticulous examination of intrinsic features [11, 12, 20, 23] in order to ascertain the authorship of documents. Its primary objective is to unveil the unique traits and patterns within handwritten materials, enabling the identification and attribution of authorship through a comprehensive investigative process. There are two methods of manuscript analysis: *destructive inspection* and *non-destructive inspection*. *Destructive inspection* involves all techniques that require the removal of a part of the writing. This investigation heavily depends on the type of document preservation, as exposure to sources of heat or humidity; it can degrade the document much more rapidly than normal. *Non-destructive inspection* is used when it is necessary to preserve the original manuscript and can be carried out through the use of imaging techniques.

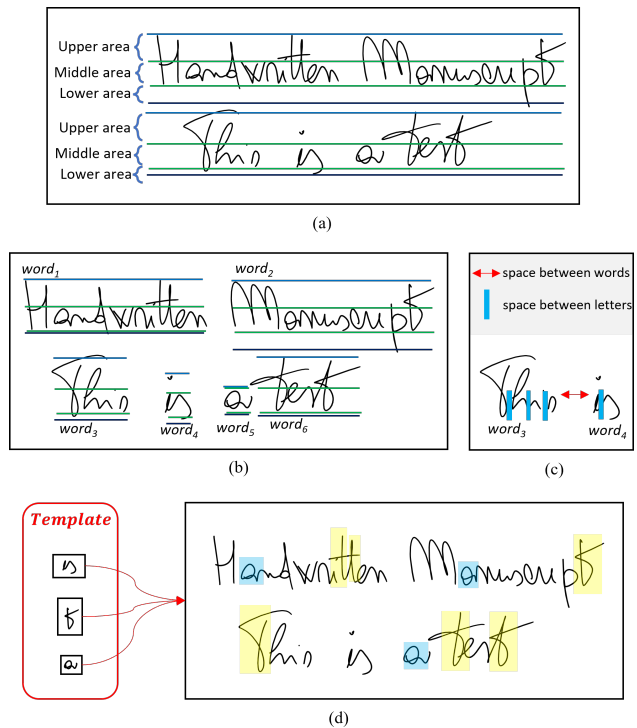


Figure 1. (a) Upper, middle and lower area detection; (b) Word detection; (c) Word and letter spacing; (d) Character recognition.

Law enforcement agencies use standard protocols based on manual processing of handwritten documents. This methodology is time-consuming, is often subjective in its evaluation, and is not fully replicable. To overcome these limitations, it is important to design new sophisticated software, based mainly on Deep Learning paradigms and techniques [9, 16, 28]. The absence of specific and sophisticated datasets could severely affect the performance of algorithms based on artificial intelligence solutions. Therefore, in this paper, we present a new and challenging dataset consisting of 362 handwritten manuscripts by 124 different people, compiled following a specific scheme and rules. Each involved person wrote 3 documents: the first is a copy of a text projected on the screen; the second is a copy of a document written by an anonymous person, with the goal of "copying/cloning" as much as possible calligraphy; and the last is a document containing a few sentences of a piece of music chosen at will. This approach to dataset creation is different from all other existing ones (such as [6, 19]), which instead include multiple texts written by the same author but all differing from each other. In addition, 21 more images

written by different people were included, using not only the classic “pen and paper” approach and later digitized but also written directly on digital devices such as tablets.

A preliminary analysis of the proposed dataset was carried out by using the method [4]. In detail, for each involved document, measures related to text line heights, space between words, and character sizes using image processing and deep learning techniques were extracted and processed. Each algorithm deals with a different aspect of text handling and retrieval, collectively contributing to a versatile set of tools for text analysis. The text line detection algorithm is designed to efficiently identify entire sentences from a given text document. In this phase all the discriminative measurements are extracted: upper, middle and lower area (Figure 1 (a)). Subsequently, the word detection algorithm was used to identify all the words (Figure 1 (b)). We obtain also other discriminative measurements, such as: space between two words and the space between letters (Figure 1 (c)). The third method, based on deep learning, requires the active involvement of a forensic expert who chooses a specific character, called a “template”, that will be searched in the entire document (Figure 1 (d)). Measurements such as its height and width are extracted as discriminative data points.

For each involved document, the final feature vector consists of the mean ( $\eta$ ) and standard deviation ( $\sigma$ ) for every type of measure collected. In case the Euclidean distance between these features exceeds a predefined threshold, it is possible to affirm that the documents are written by the same person, otherwise by different authors. Experimental results on the proposed dataset show that with a classification accuracy of about 90% on the first subset (documents written both on pen and paper and later digitized and on tablets) and 96% on the second portion of the data, we can well define the authorship of a manuscript. The obtained results could be improved by creating a more explainable and robust approach, considering the features computed through the Discrete Cosine Transform (DCT), which has been shown to achieve high results in resolution and explainability in several domains of image forensics, such as Deepfake detection. [10, 13].

The main contributions of this paper are summarized below:

- A new and challenging dataset consisting of 362 manuscripts written by 124 different people and another 21 manuscripts also written with digital devices was proposed.
- For the first time, forensic analysis of a manuscript incorporates digital documents, including those written on tablets.
- Our dataset includes not only the original calligraphy of the authors but also how they attempt to mimic another calligraphy.

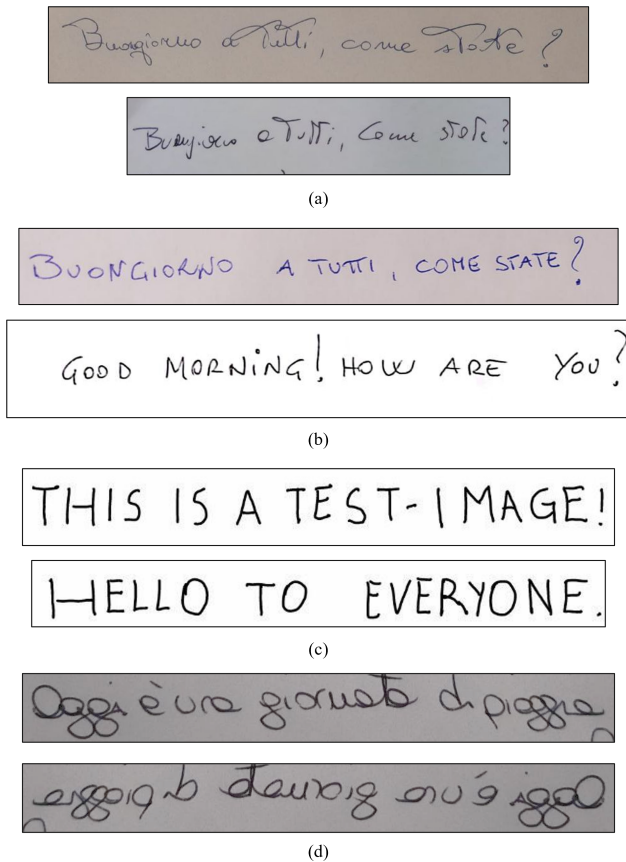
## Related Works

Several studies [8, 20, 23] have analyzed the problem of writing identification with the aim of defining the most discriminating features to assess the identity of the writer. Morris [23] emphasized the importance of evaluating dimensional parameters by comparing absolute and relative quantities in order to analyze speed, slant, and style and detect possible attempts at forgery. According to Koppenhaver [20], the analysis of character heights can provide discriminative information about an individual’s handwriting variability. The importance of absolute dimensions in graphology has been studied by Hayes [15]. These dimensions

reflect hand and finger movements, which are influenced by individual characteristics. For example, some people tend to produce small handwriting, while others have larger handwriting. Kelly and Lindblom [18] have shown that the relationship between the heights of upper and lower case letters can be used to identify the person who wrote a text. Guarnera et al. [11, 12] showed that the relationship between certain intrinsic measures (e.g., distance between words and between letters, height and width of characters, etc.) can be used to identify the person who wrote a text.

Some groups of researchers have suggested the use of computational methods to perform text author recognition and verification. Instead, Saad [25] designed a system using fuzzy logic and genetic algorithms to identify authors of Arabic handwritten texts offline by managing the ambiguity of handwriting similarities. Chahi et al. [5] proposed the “blockwise local binary count” operator to characterize each author’s writing style by computing a set of histograms from the connected components detected in handwriting. This method allows authors to be identified in handwritten documents. Shivram et al. [27] introduced a theoretical framework based on latent Dirichlet allocation to model writing styles as a shared component of an individual’s handwriting. To identify the writer, Hannad et al. [14] proposed a method based on splitting the writing into small fragments. Each fragment is represented using texture-based descriptors such as histograms of local binary patterns, local ternary patterns and local phase quantization. The extracted descriptors are used to compare two documents by calculating the distance between them. Several handwriting analysis tasks beyond author identification have also been explored. Bhardwaj et al. [3] defined a method that can estimate similar handwriting styles corresponding to a given query image. In contrast, Ramaiah et al. [24] estimated the approximate age of historical handwritten documents by learning a distribution of different styles across centuries.

Recent state-of-the-art approaches have addressed the task of identifying authors using machine learning and deep learning techniques on different datasets. Crawford et al. [7] presented a statistical approach to model various handwriting styles, aiming to calculate the likelihood of a questioned document’s authorship from a known set of writers. Semma et al. [26] extracted key points from handwriting and using Convolutional Neural Networks (CNNs) for classification. FAST key points and Harris corner detectors identify points of interest, and a CNN is trained on patches around these points. Bennour et al. [2] explored writer characterization in handwriting recognition using an implicit shape codebook technique. Key points in handwriting are identified and used to create a codebook, leading to promising results in various experimental scenarios. Kumar et al. [21] introduced a writer identification model (SEG-WI) using a CNN and weakly supervised region selection. It achieves segmentation-free writer identification on various datasets, outperforming state-of-the-art methods. Lai et al. [22] proposed novel techniques using pathlet and unidirectional SIFT features for fine-grained handwriting description. He et al. [17] presented FragNet, a deep neural network with two pathways to extract powerful features for writer identification from word and page images. Bahram [1] used co-occurrence features extracted from preprocessed regions of interest and contour texture.



**Figure 2.** Examples of four different comparisons. (a) Two handwritten manuscripts written by different persons. (b) A handwritten manuscript is compared to a digitized text. (c) Two digitized texts written by the same person are compared. (d) Two identical manuscripts are compared, but one is mirrored.

## Proposed Dataset

The proposed dataset comprises two distinct subsets. The initial subset comprises 21 documents created through the traditional “pen and paper” method, subsequently digitized, or directly acquired using common devices like tablets. The second subset comprises 362 handwritten manuscripts contributed by 124 individuals, obtained through a defined pipeline. The subsequent sections provide additional information on the acquisition, organization, and compilation of the data.

The first subset involved 21 manuscripts written by 13 different individuals. The peculiarity of this subset is its non-uniformity; each author has written their texts in their original handwriting style, language, using the medium they most frequently prefer, whether it be pen (without any distinction of colour) or digital devices (e.g. tablet). As a result, there are 21 distinct sentences with varying spontaneous styles (including bold or italics), all of which the proposed method can equally compare. A preliminary analysis of this first portion of the proposed dataset was performed in order to define the authorship of the manuscripts, using the algorithms proposed in [4]. In detail, all documents were compared with each other, resulting in a classification accuracy value (in the task of author identification by analyzing manuscript documents) of about 90%. Table 1 shows

Experiment ID	Euclidean Distance	Expected Result	Actual Result
1	5,86	D	D
2	9,72	D	D
3	1,28	S	S
4	11,88	D	D
5	7,04	D	D
6	5,44	S	D
7	0,046	S	S
8	7,05	D	D
9	10,97	S	S
10	1,42	S	S

**Table 1.** Examples of handwritten document comparisons. The first column describes the experiment ID number carried out, the second column shows the Euclidean distance resulting from the comparison of two feature vectors, the third column shows the result obtained (represented as “D” for “Different writer” and “S” for “Same writer”), and the fourth column shows the expected result.

some examples of the obtained results, by plotting 10 of the 169 examples of comparisons.

The second portion of the proposed dataset includes manuscripts by various authors, organized and collected according to very specific rules and methods. The dataset is composed by 362 images written by 124 students from the University of Catania, Italy. Each student wrote 3 manuscripts with the following rules: (i) to copy a computer-written projected text; (ii) to copy a text written in a sheet of paper with the main objective of reproducing as much as possible the same calligraphy; (iii) to write a few lines of a lyric of any song chosen by the user. Point (ii) turns out to be extremely important in the field of forensic investigations, mainly in cases where a malicious individual would attempt to simulate the calligraphy of any person for illicit purposes. This new dataset can then be extensively analyzed to create increasingly sophisticated algorithmic solutions capable of detecting forged documents of all kinds. In summary, the 3 documents written by each student consist of:

1. The same text of the projected document with their original calligraphy.
2. The same text of the projected document trying to copy the calligraphy of the writer.
3. An arbitrary text with their original calligraphy.

Figure 3 shows the projected documents to be copied (for tasks 1 and 2) and some examples of the results obtained by 4 different students. Each image was scanned and saved to test the [4] method in defining authorship. The experimental results show a classification accuracy of 96%.

## Datasets in Comparison

Several datasets used by researchers in this area are available in the literature such as:

- (i) Computer Vision Lab (CVL) Database [19], includes handwritten text from 311 unique writers, comprising both English and German samples. We selected a total of 1600 handwritten documents.
- (ii) Center for Statistics and Applications in Forensic Evidence (CSAFE) Handwriting Database [6], that consist of a total of 2430 handwriting sample images collected from surveys of 90 writers.

Handwriting analysis is a crucial aspect of forensic science, contributing to the investigation of various criminal and civil cases.

Forensic experts meticulously examine and compare handwriting samples to determine authenticity, identify potential forgeries, or establish authorship.

By analyzing individual characteristics such as letter formation, slant, pressure, and spacing, valuable insights can be obtained regarding the origin and author of a document.

(a)

Handwriting analysis is a crucial aspect of forensic science, contributing to the investigation of various criminal and civil cases.

Forensic experts meticulously examine and compare handwriting samples to determine authenticity, identify potential forgeries, or establish authorship.

(b)

	Author 1	Author 2
Handwritten Document Copied from (a)	HANDWRITING ANALYSIS IS A CRUCIAL ASPECT OF FORENSIC SCIENCE, CONTRIBUTING TO THE INVESTIGATION OF VARIOUS CRIMINAL AND CIVIL CASES	HANDWRITING ANALYSIS IS A CRUCIAL ASPECT OF FORENSIC SCIENCE, CONTRIBUTING TO THE INVESTIGATION OF VARIOUS CRIMINAL AND CIVIL CASES
Handwritten Document Forged by (b)	Handwriting analysis is a crucial aspect of forensic science, contributing to the investigation of various criminal and civil cases	Handwriting analysis is a crucial aspect of forensic science, contributing to the investigation of various criminal and civil cases.
Generic Handwritten Document	BUT LIFE STILL GOES ON I CAN'T GET USED TO LIVING WITHOUT, LIVING WITHOUT LIVING WITHOUT YOU BY MY SIDE I DON'T WANT TO LIVE ALONE	ANSWER HERO, ANOTHER MINDLESS CRIME BEHIND THE CURTAIN, IN THE PASTORAL HOLD THE LINE DOES ANYBODY WANT TO SHARE IT ANYMORE?
	Author 3	Author 4
Handwritten Document Copied from (a)	Handwriting analysis is a crucial aspect of forensic science, contributing to the investigation of various criminal and civil cases. Forensic experts meticulously examine and compare handwriting	Handwriting analysis is a crucial aspect of forensic science, contributing to the investigation of various criminal and civil cases. Forensic experts meticulously examine and compare handwriting samples to determine authenticity, identify potential forgeries, or establish authorship.
Handwritten Document Forged by (b)	Handwriting analysis is a crucial aspect of forensic science, contributing to the investigation of various criminal and civil cases. Forensic experts meticulously examine and compare handwriting samples to determine	Handwriting analysis is a crucial aspect of forensic science, contributing to the investigation of various criminal and civil cases. Forensic experts meticulously examine and compare handwriting samples to determine authenticity, identify potential forgeries, or establish authorship.
Generic Handwritten Document	Show me a piece of your heart, a piece of your love im calling you up to getting down, down, down the way that we touch is never enough I'm leaving you up to getting down down down	We're no strangers to love You know the rules and so do I A full commitment's what I'm looking for You wouldn't get this from any other guy

(c)

**Figure 3.** Some details of the proposed dataset. (a) Document projected and copied by the participants (with their original calligraphy); (b) The handwritten document to be forged (copied in the same calligraphy). (c) Some examples of documents written by 4 different people. In detail, for each person, the first row represents the handwritten document copied by (a); the second row represents the handwritten document forged by (b); finally, the third row represents a handwritten document with a randomly chosen text of a piece of music.

Figure 4 shows some visual examples.

Our dataset sets itself apart from existing ones in the state of the art by including three documents from each author, capturing them in diverse scenarios. This unique aspect enables us, unlike other datasets, to compare each author's original handwriting on the same text. It also allows us to assess whether the proposed method can discern between authentic handwriting and attempts at imitation. Moreover, we can analyze distinct handwriting styles on various texts for each author. A significant distinction is that

our dataset incorporates a blend of digitized documents (crafted on tablets) and traditional pen-and-paper documents.

### Discussion, Conclusion and Future Works

In conclusion, our study contributes significantly to the evolving landscape of Forensic Handwriting Examination by introducing a novel dataset that encompasses both traditional "pen and paper" manuscripts and documents produced through contemporary digital devices like tablets.

You have killed my love. You used to stir my imagination. Now you don't even stir my curiosity. You simply produce no effect, I loved you because you were marvellous, because you had genius and intellect, because you realised the dreams of great poets and gave shape and substance to the shadows of art. You have thrown it all away. You are shallow and stupid.

(a)

Should the Herald publish such advertisements? This is a difficult question. It would obviously be wrong to refuse all political advertisements with which we disagree. When an advertisement contains statements whose factual truth is doubtful, or whose total content would be deeply repugnant to our readers, it is right to exercise editorial discretion. The MRA advertisement falls into neither category though many readers will dislike it.

(b)

**Figure 4.** Examples of (a) CVL [19] and (b) CSAFE [6] digitized handwritten documents.

Our proposed dataset comprises two distinct subsets. The first subset involves 21 documents, initially written using the classic “pen and paper” approach and later digitized, alongside documents directly created on common devices such as tablets. The second subset includes 362 handwritten manuscripts from 124 different individuals, acquired through a specific pipeline, allowing for a comprehensive exploration of both traditional and digital writing mediums.

Through a preliminary analysis on these datasets, we employed cutting-edge automatic algorithms for feature extraction, considering measures such as the heights of upper, middle, and lower areas of text lines and words, as well as the height and width of each character and the spaces between words, among others. To ascertain authorship, we calculated the Euclidean distance between feature vectors of different manuscripts. The classification accuracy achieved is noteworthy, standing at 90% for the first subset (comprising documents written on pen and paper and later digitized, as well as on tablets) and an impressive 96% for the second subset of the dataset.

This study not only showcases the effectiveness of the state of the art methodology but also highlights the importance of embracing both traditional and digital writing mediums in forensic handwriting examination.

Looking ahead, our study opens avenues for future research and development in the field of Forensic Handwriting Examination. One key aspect that warrants attention is the expansion and

enrichment of the proposed dataset, paving the way for more comprehensive and nuanced analyses. In general, the extension could include the following factors.

- **Diverse Writing Styles and Contexts:** documents that reflect a broader spectrum of writing styles, contexts, and genres. This expansion would enhance the dataset’s applicability to a wider range of real-world scenarios, providing a more realistic foundation for forensic examinations.
- **Temporal Evolution:** capture the temporal evolution of an individual’s handwriting. Adding documents from various time periods in an individual’s life would contribute to understanding how handwriting characteristics may change over time, influencing the accuracy and reliability of authorship determination.
- **Incorporation of Additional Digital Devices:** extend the dataset to include documents produced on a broader array of digital devices. The inclusion of handwriting samples from various tablets, stylus technologies, and other emerging digital platforms would further diversify the dataset and account for the evolving landscape of digital writing tools.
- **Increased Sample Size:** enlarge the dataset with a more extensive collection of handwritten manuscripts and documents. A larger sample size enhances statistical robustness and facilitates a more in-depth exploration of the relationships between extracted features, contributing to the overall efficacy of forensic analyses.
- **Annotation for Ground Truth:** annotate the dataset with additional ground truth information. Including annotations that identify specific characteristics or attributes within the documents can provide valuable insights for algorithmic training and evaluation.
- **Integration of Multimodal Data:** Explore the integration of multimodal data, such as incorporating handwriting samples alongside other biometric or contextual information. This holistic approach could lead to more comprehensive forensic examinations that consider a broader range of factors influencing authorship determination.

By addressing these considerations, future research can build upon the foundation laid by this study, advancing the capabilities of forensic handwriting analysis and ensuring that the proposed methodologies remain robust and applicable in evolving forensic contexts.

## Acknowledgments

This research is supported by Azione IV.4 - “Dottorati e contratti di ricerca su tematiche dell’innovazione” del nuovo Asse IV del PON Ricerca e Innovazione 2014-2020 “Istruzione e ricerca per il recupero - REACT-EU”- CUP: E65F21002580005.

## References

- [1] Tayeb Bahram. A Texture-Based Approach for Offline Writer Identification. *Journal of King Saud University-Computer and Information Sciences*, 34(8):5204–5222, 2022.
- [2] Akram Bennour, Chawki Djeddi, Abdeljalil Gattal, Imran Siddiqi, and Tahar Mekhaznia. Handwriting Based Writer Recognition Using Implicit Shape Codebook. *Forensic Science International*, 301:91–100, 2019.

- [3] Anurag Bhardwaj, Achint Oommen Thomas, Yun Fu, and Venu Govindaraju. Retrieving Handwriting Styles: A Content Based Approach to Handwritten Document Retrieval. In *2010 12th International Conference on Frontiers in Handwriting Recognition*, pages 265–270. IEEE, 2010.
- [4] Eleonora Breci, Luca Guarnera, and Sebastiano Battiato. Innovative Methods for Non-Destructive Inspection of Handwritten Documents. *arXiv preprint arXiv:2310.11217*, 2023.
- [5] Abderrazak Chahi, Yassine Ruichek, Raja Touahni, et al. Block Wise Local Binary Count for Off-Line Text-Independent Writer Identification. *Expert Systems with Applications*, 93:1–14, 2018.
- [6] Amy Crawford, Anyesha Ray, and Alicia Carriquiry. A Database of Handwriting Samples for Applications in Forensic Statistics. *Data in Brief*, 28:105059, 2020.
- [7] Amy M Crawford, Danica M Oommen, and Alicia L Carriquiry. A Statistical Approach to Aid Examiners in the Forensic Analysis of Handwriting. *Journal of Forensic Sciences*, 2023.
- [8] Capitaine Marie Deviterne-Lapeyre. Interpol Review of Questioned Documents 2016–2019. *Forensic Science International: Synergy*, 2:429–441, 2020.
- [9] Stefan Fiel and Robert Sablatnig. Writer Identification and Retrieval Using a Convolutional Neural Network. In *Computer Analysis of Images and Patterns: 16th International Conference, CAIP 2015, Valletta, Malta, September 2-4, 2015, Proceedings, Part II 16*, pages 26–37. Springer, 2015.
- [10] Oliver Giudice, Luca Guarnera, and Sebastiano Battiato. Fighting Deepfakes by Detecting GAN DCT Anomalies. *Journal of Imaging*, 7(8):128, 2021.
- [11] Luca Guarnera, Giovanni Maria Farinella, Antonino Furnari, Angelo Salici, Claudio Ciampini, Vito Matranga, and Sebastiano Battiato. GRAPHJ: A Forensics Tool for Handwriting Analysis. In *Image Analysis and Processing-ICIAP 2017: 19th International Conference, Catania, Italy, September 11-15, 2017, Proceedings, Part II 19*, pages 591–601. Springer, 2017.
- [12] Luca Guarnera, Giovanni Maria Farinella, Antonino Furnari, Angelo Salici, Claudio Ciampini, Vito Matranga, and Sebastiano Battiato. Forensic Analysis of Handwritten Documents with GRAPHJ. *Journal of Electronic Imaging*, 27(5):051230–051230, 2018.
- [13] Luca Guarnera, Salvatore Manganello, and Sebastiano Battiato. Improving Video Deepfake Detection: A DCT-Based Approach with Patch-Level Analysis. *arXiv preprint arXiv:2310.11204*, 2023.
- [14] Yaacoub Hannad, Imran Siddiqi, and Mohamed El Youssfi El Kettani. Writer Identification Using Texture Descriptors of Handwritten Fragments. *Expert Systems with Applications*, 47:14–22, 2016.
- [15] Reed C Hayes. *Forensic Handwriting Examination: a Definitive Guide*. ReedWrite Press, 2006.
- [16] Sheng He and Lambert Schomaker. Deep Adaptive Learning for Writer Identification Based on Single Handwritten Word Images. *Pattern Recognition*, 88:64–74, 2019.
- [17] Sheng He and Lambert Schomaker. Fragnet: Writer Identification Using Deep Fragment Networks. *IEEE Transactions on Information Forensics and Security*, 15:3013–3022, 2020.
- [18] J. S. Kelly and B. S. Lindblom. *Scientific Examination of Questioned Documents*. CRC press, 2006.
- [19] Florian Kleber, Stefan Fiel, Markus Diem, and Robert Sablatnig. Cvl-database: An Off-line Database for Writer Retrieval, Writer Identification and Word Spotting. In *2013 12th International Conference on Document Analysis and Recognition*, pages 560–564. IEEE, 2013.
- [20] Katherine M Koppenhaver. *Forensic Document Examination: Principles and Practice*. 2007.
- [21] Parveen Kumar and Ambalika Sharma. Segmentation-free Writer Identification Based on Convolutional Neural Network. *Computers & Electrical Engineering*, 85:106707, 2020.
- [22] Songxuan Lai, Yecheng Zhu, and Lianwen Jin. Encoding Pathlet and SIFT Features with Bagged VLAD for Historical Writer Identification. *IEEE Transactions on Information Forensics and Security*, 15:3553–3566, 2020.
- [23] Ron N Morris. *Forensic Handwriting Identification: Fundamental Concepts and Principles*. In *Academic Press*, 2020.
- [24] Chetan Ramaiah, Gaurav Kumar, and Venu Govindaraju. Handwritten Document Age Classification Based on Handwriting Styles. In *Document Recognition and Retrieval XIX*, volume 8297, pages 195–200. SPIE, 2012.
- [25] Saad M Saad. Application of Fuzzy Logic and Genetic Algorithm in Biometric Text-Independent Writer Identification. *IET Information Security*, 5(1):1–9, 2011.
- [26] Abdelillah Semma, Yaâcoub Hannad, Imran Siddiqi, Chawki Djeddi, and Mohamed El Youssfi El Kettani. Writer Identification Using Deep Learning with Fast Keypoints and Harris Corner Detector. *Expert Systems with Applications*, 184:115473, 2021.
- [27] Arti Shivram, Chetan Ramaiah, and Venu Govindaraju. A Hierarchical Bayesian Approach to Online Writer Identification. *Iet Biometrics*, 2(4):191–198, 2013.
- [28] Youbao Tang and Xiangqian Wu. Text-Independent Writer Identification via CNN Features and Joint Bayesian. In *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 566–571. IEEE, 2016.

## Author Biography

Eleonora Breci was born in Lentini on August 22, 2000. She completed her bachelor's degree in Computer Engineering at the University of Catania, discussing the thesis entitled “GraphJ-3.0: Graphometric Measures for the Forensic Examination of Handwritten Documents” on October 7, 2022. Currently, she is pursuing a Master's program in Cybersecurity at La Sapienza University, Rome. Her primary research areas focus on the forensic study of handwritten documents, with a particular attention on document authentication.

Luca Guarnera was born in Catania on October 26, 1992. Since January 1, 2022, he is a research fellow in Computer Science at the University of Catania. He graduated as Ph.D. in Computer Science (XXXIII cycle, PON number E37H18000330006) on October 14, 2021, discussing the thesis entitled “Discovering Fingerprints for Deepfake Detection and Multimedia-Enhanced Forensic Investigations” at the Department of Mathematics and Computer Science, University of Catania. His main research interests are Computer Vision, Machine Learning, Multimedia Forensics and its related fields with a focus on the Deepfake phenomenon.

Sebastiano Battiato is a full professor of Computer Science at the University of Catania. He received his degree in Computer Science (summa cum laude) in 1995 from the University of Catania and his Ph.D. in Computer Science and Applied Mathematics from the University of Naples in 1999. He has been Chairman of the Undergraduate Program in Computer Science (2012-2017), and Rector's delegate for Education: postgraduates and Phd (2013-2016). He is currently the Scientific Coordinator of the PhD Program in Computer Science (XXXIII-XXXVI cycles) and Deputy Rector for Strategic Planning and Information Systems at the University of Catania. His research interests include Computer Vision, Imaging technology and Multimedia Forensics.