



Exploring Medieval Manuscripts Writer Predictability: A Study on Scribe and Letter Identification

Francimaria Rayanne Dos Santos Nascimento, Universidade Federal de Pernambuco, BR, francimariasantos123@gmail.com

Stephen Smith, University of York, UK, stephen.smith@york.ac.uk

Marjory Da Costa Abreu, Sheffield Hallam University, UK, m.da-costa-abreu@shu.ac.uk

Handwriting communication is a long-established human activity that has survived into the 21st century. Accordingly, research interest in handwritten documents, both historical and modern, is significant. The way we write has changed significantly over the past few centuries. For example, texts of the Middle Ages were often written and copied by anonymous scribes. The writing of each scribe, known as his/her "scribal hand" is unique. It can be differentiated using a variety of consciously and unconsciously produced features. Distinguishing between these different scribal hands is a central focus of the humanities research field known as "paleography." Character recognition within each scribal hand has also posed an interesting challenge. Some issues make these digital processes difficult, such as paper degradation and the soiling of the manuscript page. Thus, in this paper, we propose an investigation in both perspectives, character recognition and writer identification, in medieval manuscripts to better understand the specific behaviour of two 800-year-old scribes based on their manuscripts in comparison with a modern calligrapher. The experiments demonstrated that degradation and tremor can influence the analysis of medieval handwriting documents. However, the results presented an efficient accuracy with a better accuracy rate in letter classification than in writer identification.

La communication manuscrite est une longue tradition humaine qui a persisté jusqu'à nos jours, au 21^e siècle. Par conséquent, l'intérêt de la recherche concernant des documents manuscrits historiques et modernes est grand. La façon dont nous écrivons a changé au cours des derniers siècles. Par exemple, des textes du Moyen Âge ont souvent été écrits et copiés par des scribes. L'écriture de chaque scribe, appelée son « écriture scribale » (anglais scribal hand), est unique. Nous pouvons la différencier en observant une gamme de caractéristiques produites consciemment et inconsciemment. Faire la distinction entre ces écritures sribales différentes est au centre des préoccupations du domaine de recherche de paléographie. La reconnaissance de caractères dans chacune des écritures sribales pose des défis intéressants. Certains problèmes, tels que la dégradation de papier et l'encrassement de la page manuscrite, rendent difficiles ces processus numériques. Dans cet article, nous proposons ainsi



une enquête sur les deux perspectives, la reconnaissance de caractères et l'identification de scribes, dans les manuscrits médiévaux dans le but de mieux comprendre le comportement spécifique de deux scribes vivant il y a 800 ans, en se basant sur leurs manuscrits en comparaison avec un calligraphe moderne. Les expériences démontrent que la dégradation et le tremblement peuvent influencer sur l'analyse des documents manuscrits médiévaux. Cependant, les résultats ont présenté une précision efficace, avec un meilleur taux de précision dans la reconnaissance de caractères que dans celui de l'identification de scribe.

1 Introduction

Analysis of handwritten documents is a relevant problem in the fields of machine learning, image processing, and pattern recognition research. For such applications, it is essential to consider the different types of handwritten documents that can be analyzed, such as pre-modern and contemporary manuscripts (Toledo et al. 2017; Zamora-Martinez et al. 2014).

One of the most relevant tasks in the handwritten analysis is writer identification, which aims to determine the author of a specific text, given a training dataset (Christlein et al. 2017). The data extraction process can be classified into two types: online and offline (Wang et al. 2018; Kamble and Hegadi 2015). In the online process, the data are collected based on temporal-space data. On the other hand, the offline process obtains information by processing only document image data without using any dynamic features.

Another relevant (and related) problem is Handwritten Text Recognition (HTR), in which handwriting is extracted from images and converted into a similar symbolic value as ASCII text, with a computer (Romero, Toselli, and Vidal 2012). However, distinct languages, writing styles, and input contents may increase HTR system complexity.

An important consequence (or contribution if found) of analyzing medieval documents is to understand the characteristics of each scribe and try to identify any relevant issues that the scribe might have, such as high alcohol consumption (Haubenberger et al. 2011) or neurological conditions (Thorpe 2015). By finding written features in each of these conditions, with only off-line handwriting analysis of the manuscripts, we can better understand modern conditions as well (Schiegg and Thorpe 2017; Zhi et al. 2017).

However, analyzed medieval manuscripts may present particular problems, such as ink degradation, background noise, dye stain, and writing failures. Thus, the handwriting analysis systems may be less accurate than machine-printed documents, especially when dealing with degraded historical documents (Saleem et al. 2014).

These problems have motivated us to offer a study of medieval documents in delicate conditions. We proposed an investigation using techniques described in the literature in particular documents written in Middle English by three different scribes. Two of them were thirteenth-century scribes, the third being a modern expert calligrapher writing using medieval tools and techniques. In one scribe, there are no apparent pathological writing features, and the other is an anonymous scribe known as “The Tremulous Hand of Worcester,” or only “Tremulous Hand” due to his distinctive writing tremor.

A previous study was made by Thorpe and Alty (Thorpe and Alty 2015) based on the scribe's writing features, indicating that Tremulous Hand had essential tremor. This common neurological condition typically presents a "fine" or "fine-moderate" amplitude of tremor, with a relatively fast frequency. The authors found that handwriting showed regular amplitude tremors, and individual letters in words presented the same degree of lateral derivation. The scribe Tremulous Hand is of great importance for investigating pathological evidence in writing because he is the only medieval scribe widely known with a tremor. Moreover, he had a particular interest in translating documents written centuries earlier.

Identifying the impact of tremor in handwriting (when present) on recognizing letters and scribing him/herself is of great relevance since pathological evidence visible in the document can influence handwriting analysis. Although different contributions have been dedicated to investigating character recognition and writer identification from degraded historical documents (Bukhari et al. 2012; Diem and Sablatnig 2010a; Gilliam, Wilson, and Clark 2010; Saleem, Hollaus, and Sablatnig 2014), scribe's pathological evidence visible in writing did not receive attention in these researches. This contribution aims to investigate the impact of tremors in handwritten character recognition, writer identification, and pixel density analysis in the context of degraded historical handwriting documents.

We used an approach called diagonal feature extraction for the feature extraction of these documents. This approach uses image segmentation in zones and obtains the features based on three different ways of feature extraction: horizontal, vertical, and diagonal. We use the classifiers Naive Bayes, Support Vector Machines (SVM), and the Multi-Layer Perceptron (MLP). Furthermore, we analyze the pixel density of characters using global and local metrics to investigate changes in character characteristics.

This paper is organized as follows: Section 2 presents the related work. Section 3 explains the methodology adopted for character recognition, writer identification, and metrics for character analysis. Section 4 discusses the results obtained. Section 5 shows the work conclusions and our future work.

2 Related work

The literature investigates some approaches to the recognition and treatment of degraded medieval manuscript documents, based on the technological approach to automating the analysis of medieval documents. Paleography, on the other hand, also investigates the characteristics of the neurological condition of medieval scribes of these documents. Accordingly, some related work is described in this section.

2.1 Automatic character and writer recognition in degraded medieval handwritten documents

The automatic process of character and writer recognition in medieval documents is not trivial, mainly due to degradation and noises. There are several approaches based on image processing, pattern recognition, and machine learning research areas proposed to address some problems, such as degradation treatment, binarization, character recognition, text segmentation, and scribe identification.

The International Conference on Document Analysis and Recognition (ICDAR) and the International Conference on Frontiers in Handwriting Recognition (ICFHR) have proposed several competitions to explore medieval handwriting documents. For instance, the competitions ICDAR2017 (Cloppet et al. 2017) and ICFHR2016 (Cloppet et al. 2016) presented tasks that focused on the classification of medieval handwritings in Latin Script.

In Stokes (2007), image-processing and data-mining techniques were discussed for scribal detection. The author considered the automatic feature-extraction methods proposed in the literature and an automatic clustering approach using the AutoClass package. Although the results were obtained in the literature with these techniques, the researcher highlights that they were obtained under controlled conditions and points out that these applications all require careful thought.

In Gatos, Pratikakis, and Perantonis (Gatos, Pratikakis, and Perantonis 2006), the researchers proposed an adaptive binarization technique to enhance degraded historical handwritten documents, old newspapers, and poor-quality modern documents using Wiener's filter and Sauvola's approach. Su, Lu, and Tan (Su, Lu, and Tan 2010) used an approach based on local maximum and minimum for binarization of handwritten historical documents. According to the authors, the method was efficient in treating different types of degradation in documents such as smear and uneven illumination.

The use of local features and SVM to select an optimal global threshold for binarization of degraded historical documents were present in Xiong et al. (Xiong et al. 2018). The authors evaluated the performance of the method on 21 images of degraded manuscripts. The results showed superior performance compared to other state-of-the-art techniques.

An adaptive method based on character models for segmentation of highly degraded historical documents was proposed in Bar-Yosef et al. (Bar-Yosef et al. 2009) applied to an antique Jewish prayer book, written between the 11th and 13th centuries. They used 500 degenerate characters from 16 different Hebrew letters for identification. The idea was to construct a small set of shapes, based on the training set, with variable shapes, so the system could classify different forms of writing a determined letter. Normalized

cross-correlation (NCC) was applied to identify matching characters, yielding 98% of correct recognition.

A layout analysis of Arabic historical manuscript documents with machine learning was proposed in Bukhari et al. (Bukhari et al. 2012). The authors performed the extraction with discriminative and straightforward features at the connected component level, sequentially generated in a robust feature vector. The neural network MLP was used for text component classification and then a voting system as the final classifier.

Diem and Sablatnig (Diem and Sablatnig 2010b) presented a methodology for character recognition in degraded ancient Slavonic manuscripts from the 11th century. The authors proposed a system for OCR divided into two steps: character classification and character location. In this approach, they explored local descriptors directly extracted from grey-scale images. Multiple SVMs with Radial Basis Function (RBF) kernels were then used to classify descriptors. The k-means algorithm was employed for localization of characters based on interest points using local descriptors previously computed. The authors used a voting system for the final character classification. Diem and Sablatnig (Diem and Sablatnig 2010a) described an approach based on local descriptors where characters were localized using clustering techniques. The authors attempted to identify and classify Glagolitic characters in degraded manuscript documents like those of St. Catherine's Monastery. They used SVM to classify local descriptors and then a voting system for character recognition.

An approach to the recognition of Glagolitic characters in degraded historical documents was proposed in Saleem et al. (Saleem et al. 2014). Character recognition with Dense Scale Invariant Feature Transform (SIFT) and Nearest Neighbour Distance Maps (NNDM) was used to localize and classify Glagolitic characters. A new pre-processing method proposed by the authors and Total Variation (TV) were used for image restoration and noise reduction. The experiments performed showed that the best results were obtained using image restoration as a pre-processing step for Dense SIFT for analyzed documents.

Saleem, Hollaus, and Sablatnig (Saleem, Hollaus, and Sablatnig 2014) used Dense SIFT and NNDM to recognize Glagolitic characters in degraded documents written in the 11th century. In this approach, local minima were used for localization and character recognition in documents according to the NNDM output algorithm.

An approach of scribe identification in medieval English manuscripts of the 14th–15th century was presented in Gilliam, Wilson, and Clark (Gilliam, Wilson, and Clark 2010). The authors compared the performance of Sparse Multinomial Logistic

Regression (SMLR) with the k-NN classifier, achieving a classification accuracy of 78% using the grapheme code-book method.

2.2 Analyze documents in the context of disease and disorders

In the context of paleographic studies, diseases and disorders of ancient scribes are investigated based on their handwriting features. In this way, the researchers try to analyze and understand the clinical condition of old writers using some techniques and tools of modern medicine. Some of the most recent works will be discussed below.

An analysis of historical handwriting distortions from patients in an early 20th-century psychiatric hospital in southern Germany was presented in Schiegg and Thorpe (Schiegg and Thorpe 2017). The study demonstrated that combining historical handwriting analysis with modern medicine can help re-contextualize individual writing disorders and offer information about medical conditions involving writing disorders.

Thorpe, Alty, and Kempster (Thorpe, Alty, and Kempster 2019) proposed a retrospective diagnosis of John Ruskin (1819–1900). Based on the paleographical study, the authors investigated the relationship between features of writing through his letters and diary entries and Ruskin’s clinical condition. Thus, they concluded that he had an organic neurological disorder.

3 Understanding how to analyze user and letter identification from medieval data

From what we have observed in Section 1, the analysis of medieval handwritten texts is a relevant research question, and for such related applications, we investigate different tasks. Nevertheless, we need first to understand the problems of analyzed medieval documents that could influence the experiment’s accuracy rate.

Medieval handwritten documents present problems such as degradation, ink stains, failure to write, background noise, so on. Besides, particular characteristics caused by diseases and disorders can increase the difficulty of the handwriting analysis process.

Considering these problems, we performed a series of steps to extract and understand the data of the medieval handwritten document written by scribe “Tremulous Hand of Worcester” in contrast to texts written in Middle English and by a modern calligrapher.

The analyzed datasets consist of characters cropped manually from handwritten documents (see **Figure 1**). The characters were extracted from:

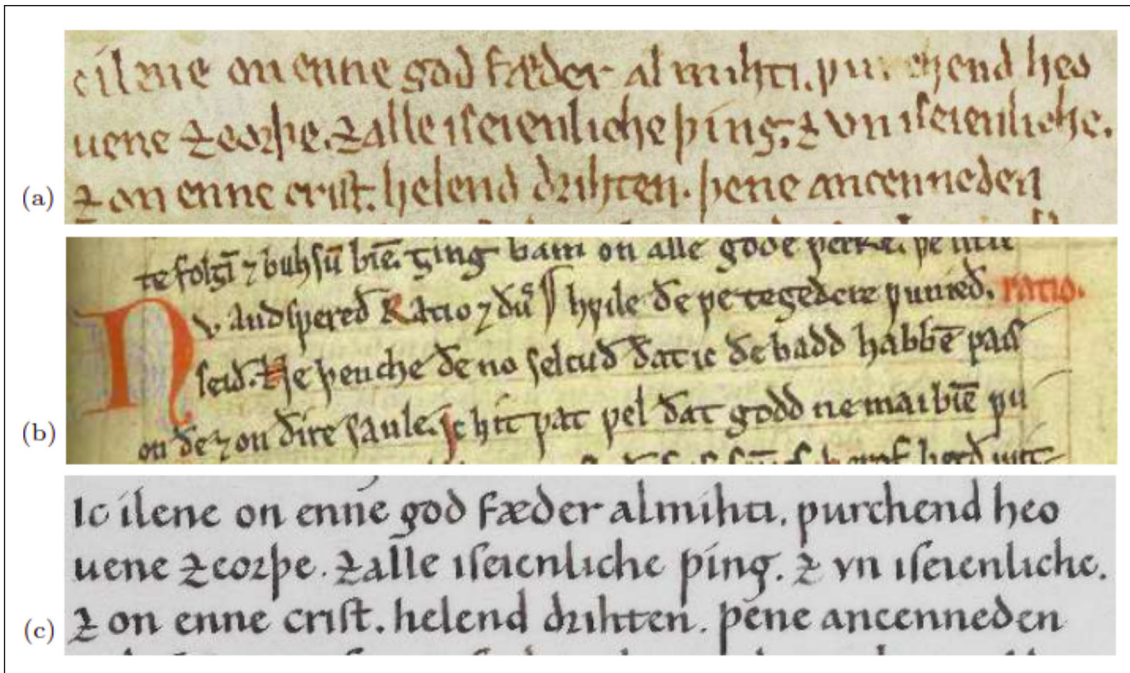


Figure 1: Examples of handwritten documents: (a) Page extract from the Tremulous Hand of Worcester. Source: Detail of Oxford, Bodleian Library, Manuscript Junius 121, Folio vi recto. (b) Page extract from the non-tremulous scribe. Source: London, British Library, Stowe MS 34 (formerly 240), f. 32r, Vices and virtues. (c) Page extract from the modern expert calligrapher.

- a) TREMULOUS – a sample from the 13th century written by an anonymous scribe known as “Tremulous Hand of Worcester” (171 characters samples of TREMULOUS),
- b) NON-TREMULOUS – a sample from the 13th century written by a non-tremulous scribe (171 character samples of NON-TREMULOUS) and
- c) CALLIGRAPHER – a reproduction of the text copied by the tremulous writer by a modern expert calligrapher (80 character samples of CALLIGRAPHER).

The dataset used for the present study contains a total of 422 samples of letter images (of the characters, “a,” “e,” “h,” and “l”). We have chosen the letters “a” and “e” because they have a similar shape and are more frequent in the text to explore a possible OCR system and, on the other hand, the letters “h” and “l” because they have a long vertical stroke; thus, the tremor in writing is more evident in them. An evident issue of this study is the fact that we have a limited dataset size, which is due to the availability of the data collected. This issue has been addressed in other instances which can be seen in Cilia et al. (Cilia et al. 2020) and Liang, Guest, and Fairhurst (Liang, Guest, and

Fairhurst 2012), and it highlights that it is possible to reliably analyze a small dataset from handwriting documents medieval to modern times.

The TREMULOUS and NON-TREMULOUS datasets have 30 examples of the letter “a,” 90 of the letter “e,” 20 of the letter “h,” and 31 of the letter “l,” resulting in 171 examples of NON-TREMULOUS and 171 of TREMULOUS, with a total of 342 samples. The CALLIGRAPHER dataset is formed from 20 examples of each letter (“a,” “e,” “h,” “l”), resulting in 80 samples. For the task of character recognition, the dataset used is classified in the “a,” “e,” “h,” and “l” classes. In the task of writer identification, on the other hand, the data are classified in the TREMULOUS, NON-TREMULOUS and CALLIGRAPHER classes.

In this paper, the process of handwriting analysis consists of pre-processing, segmentation and feature extraction, classification and recognition, and post-processing, which will be presented in the next sections.

3.1 Pre-processing

Pre-processing is the initial step where a series of essential operations are performed for the treatment of input images, mainly in our context of degraded images (Gatos, Pratikakis, and Perantonis 2006).

The global threshold was used to convert from a grey-scale image to a binary image, where pixels with a value of 0 represent the background and with a value of 1 represent the object. This method allows defining the data with intensity below the threshold as belonging to one phase and the remaining to the other, performing the separation between the background and foreground (Sezgin and Sankur 2004; Gatos, Pratikakis, and Perantonis 2006). Thus, a threshold T from the grey-scale image I that separates these models using Otsu’s method (Otsu 1979) is selected to extract the object from the background. The thresholded image $g(x, y)$ is defined in Equation 1.

$$g(x, y) = \begin{cases} 1 & \text{if } g(x, y) > T \\ 0 & \text{if } g(x, y) \leq T \end{cases} \quad (1)$$

Subsequently, the Sobel technique was used for edge detection in the binarized image. Compared to other edge detection techniques, Sobel has the advantages of performing some smoothing effect in the image noise, detection, and differentiation of two rows or two columns. Thus, the edges of both sides are highlighted (Gupta and Mazumdar 2013). Each image was segmented into individual characters, uniformly re-sized into 90×60 pixels.

3.2 Feature extraction

At this stage, we used a diagonal feature extraction scheme proposed in Pradeep, Srinivasan, and Himavathi (Pradeep, Srinivasan, and Himavathi 2011) (see Figure 2). The method is based on three ways of feature extraction: vertical direction, horizontal direction, and diagonal direction. An extensive study performed in handwritten English alphabet characters in Pradeep, Srinivasan, and Himavathi (Pradeep, Srinivasan, and Himavathi 2011) showed that this approach provides good recognition accuracy.

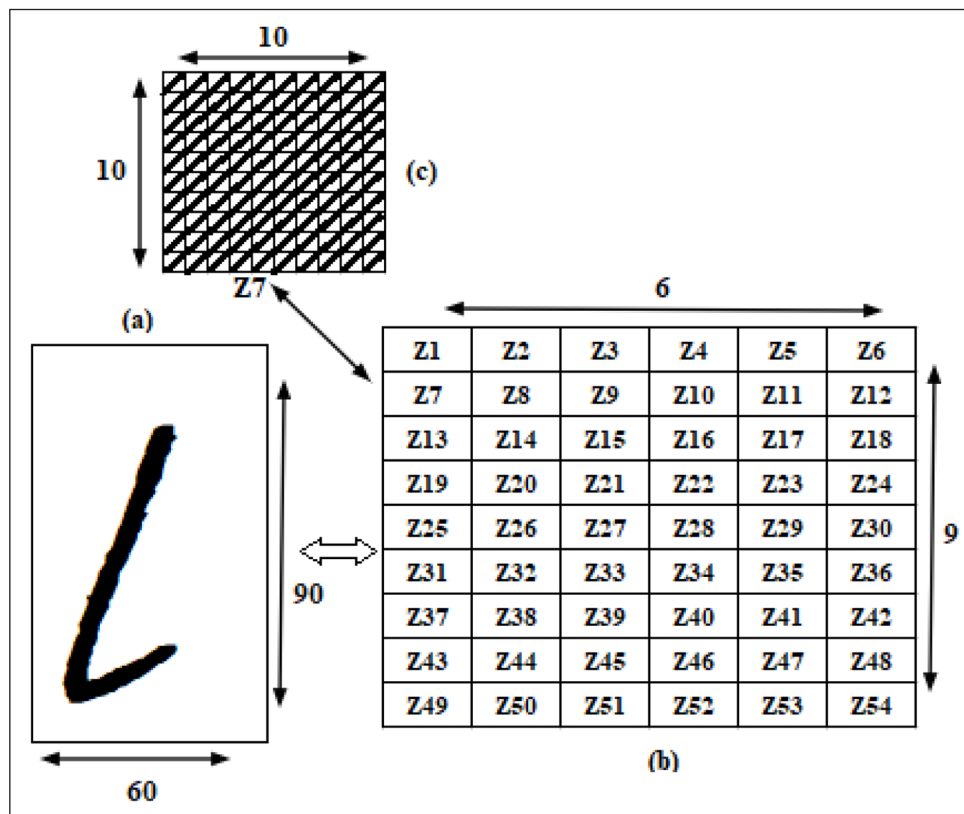


Figure 2: Classification and Characters Recognition (a) Image re-sized 90×60 . (b) Partitioning image in 54 zones with size 10×10 . (c) Diagonal features extraction. Taken and adapted from Pradeep, Srinivasan, and Himavathi (2011).

Images of size 90×60 were partitioned in 54 zones with equal dimensions, with sizes 10×10 (an example can be seen in Figure 2c). Features are extracted from zone pixels. Each zone has 19 diagonals. The values along each diagonal line are summed, resulting in 19 sub-features of each zone. Thus, the average of 19 sub-features forms the single feature value of the corresponding zone (see Figure 2b). The process repeats for all blocks, resulting in 54 features.

In addition to the diagonal features, we also measured the average of rows and columns in the zones, resulting in nine characteristics corresponding to rows and six to columns. Sixty-nine features thus represented each character.

3.3 Classification and character recognition

After the feature extraction stage, character recognition and writer identification were performed to analyze the potential of the database. Thus, we used SVM and MLP classifiers because they had efficient performance in the related work (Bukhari et al. 2012; Diem and Sablatnig 2010a, 2010b; Xiong et al. 2018). The Naive Bayes was applied to data classification from a probabilistic perspective. We did not use the Deep Learning approach because, to achieve prediction accuracy efficiently, we would need a vast amount of training data (Chilimbi et al. 2014). Since we only have a small amount of data, this has become an impossibility. The used classifiers are available in the Weka toolbox (Weka 2021a) that can be described as:

- Naive Bayes (Mitchell 2017) is a probabilistic classifier based on the application of Bayes' theorem. The method uses the idea that probabilities amongst features are strongly independent, rather than calculating the value for each attribute x_i given a hypothesis $h P(x_1, x_2, x_3|h)$, simplifying the probability calculation for each value as $P(x_1|h) * P(x_2|h) * P(x_3|h)$. All experiments performed used WEKA's implementation of the Naive Bayes (WEKA 2021b) with the following configuration: `useKernelEstimator false, useSupervisedDiscretization false`.
- SVM is a classifier based on Statistical Learning Theory, proposed in Vapnik (Vapnik 1998). The algorithm finds the optimal separating hyperplane for the greatest number of points belonging to the same class stays on the same side while maximizing the distances of each class to that hyperplane. The subsets of the training points of the two classes, closest to the optimal separating hyperplane, are called support vectors (Surinta et al. 2015). WEKA's implementation of the SVM (WEKA 2021c) algorithm is based on the implementation from Keerthi et al. (Keerthi et al. 2001) and Platt (Platt 1998). In our experiments, the following configuration was used: `C 1.0, Kernel Function PolyKernel`.
- Neural network MLP (Haykin 2007) is a generalization of the simple algorithm Perceptron. MLP is formed by an input layer, one or more hidden layers, and one output layer. The training of MLP is supervised, and it uses the error back-propagation algorithm. In WEKA's implementation of MLP (WEKA 2021d),

all nodes are sigmoid. The classifier allows parameters to be modified as the number of hidden layers, learning rate, training time, and momentum rate. All experiments were performed using the following configurations: hidden layers $a \left(\frac{\text{number of attributes} + \text{number of classes}}{2} \right)$, learning rate 0.3, momentum 0.2, training time 500, threshold 20, validation set 0.

Different simulations were performed to select configurations of classifiers. Thus, the settings that achieved the best results in most of the cases investigated were selected.

The k -fold cross-validation, with k , equals ten, was used to estimate the accuracy of classification algorithms. Extensive experiments on different datasets with various machine learning techniques have shown that ten-fold cross-validation presents a good compromise to obtain the best estimate of error (McLachlan, Do, and Ambroise 2005; Witten et al. 2016). The method consists of partitioning the dataset randomly into ten mutually exclusive subsets, where one subset is used for model validation and the others for model estimation (training), where each part is held out in turn (Kohavi 1995). The method is applied ten times on a different training dataset. The averaged error estimate is then used to produce an overall error estimate. Thus, all our presented confusion matrices will have an overall error estimate, provided by ten iterations.

3.4 Character analysis

For the character analysis, we have investigated a small subset consisting of five samples of the letter “a,” “e,” “h,” and “l” of scribes TREMULOUS, NON-TREMULOUS, and CALLIGRAPHER; due to this task, it requires a lot of time for pre-processing. This analysis intended to detect changes in the characteristics of documents such as the amount of ink deposited, size-reduction, and pixel density within a writing sample.

As previously mentioned, we have chosen the letters “a” and “e” because they have a similar shape. On the other hand, we have chosen the letters “h” and “l” because they have a longer vertical stroke, where the tremor is more visible. We have made some image pre-processing before we calculate the metrics. The grey-scale image background was removed, but only the CALLIGRAPHER dataset image background was automatically removed. We have made manual adjustments in the medieval document due to its degradation, writing failures, background noise, and dye stain, which were considered visual noise. This step was developed by the authors and performed in MATLAB (MathWorks 2021) to remove marks not relevant to the characters.

Posterior, two types of metrics were used for character analysis based on pixel values: *Global* and *Local* metrics. The metrics will be described in the next section.

3.4.1 Global metric

Global metrics are based on the function $f(x, y)$ which is defined on the plane (x, y) . Where, at the point (x_0, y_0) $\rho(x, y)$, whenever a pixel belongs to the background of the letter $f(x, y)$, is zero; otherwise, f is the pixel value of the digitized image at the grey-scale at point (x, y) because as we used historical documents the pixel value is needed in each sample. The Horizontal Projection Profile (HPP) is calculated by Equation 2, and the Vertical Projection Profile (VPP) by Equation 3 (Zhi et al. 2017).

$$HPP(x) = \sum_y f(x, y) \quad (2)$$

$$VPP(y) = \sum_x f(x, y) \quad (3)$$

The HPP denotes the sum of all pixels along a row in the image. Similarly, VPP is the sum of columns (Mamatha and Srikantamurthy 2012). The HPP or VPP curves area indicates the number of pixels accumulated in samples, which corresponds to the amount of ink used in the manuscripts. We, therefore, expect these distortions in writing to influence the shape of HPP or VPP curves.

3.4.2 Local metric – Pixel Density Variation (PDV)

The PDV metric evaluates compression/distortion resources in samples because some effects may arise progressively on writing. Such effects can be detected based on characteristics changing locally from left to right within a sample (Zhi et al. 2017).

For the calculation of the PDV metric, each image is split into cells of the same width. The amount of ink in the upper and lower borders determines the height of the cells. The area A_{ijk} is measured by the product of height and width of the cell, with i and j being the subject and sample number respectively, and k being the cell number. The pixel density ρ_{ijk} is calculated by the number of ink pixels P_{ijk} in each cell divided by area A_{ijk} by Equation 4.

$$\rho_{ijk} = \frac{P_{ijk}}{A_{ijk}} \quad (4)$$

The ρ_{ijk} is plotted along with the cell localization $k = 1, \dots, K$ in the sample counting from left to right, with K being the total cell number.

3.4.3 Statistical analysis of metrics

We perform a statistical analysis of global metrics (HPP and VPP) using Pearson's correlation coefficient, which measures the linear relationship between two variables (Taylor 1997). The metric is defined by Equation 5, in terms of covariance, given a pair of random variables (A, B).

$$\rho(A,B) = \frac{\text{cov}(A,B)}{\sigma_A \sigma_B} \quad (5)$$

where cov is the covariance, and σ_A and σ_B are the standard deviations of A and B , respectively. The correlation coefficient matrix is defined by Equation 6, which is the matrix of correlation coefficients for each pairwise variable combination.

$$R = \begin{pmatrix} \rho(A,A) & \rho(A,B) \\ \rho(B,A) & \rho(B,B) \end{pmatrix} \quad (6)$$

The value $\rho(A,B)$ is in the range -1 and 1 . Values near -1 indicate a direct, negative linear correlation; values near 0 indicate little or no linear correlation; and values near 1 indicate a direct, positive linear correlation (Boddy and Smith 2009). Given that the samples are always directly correlated to themselves, the diagonal of the entries is 1 . We have hidden the diagonal value (1) for a better view of the variation of the other values.

4 Experiments and results

The techniques we have presented for character recognition have the aim of evaluating degraded medieval documents written in Middle English and a manuscript written by a modern calligrapher. Experiments with different perspectives were performed on datasets and will be presented in the following sections.

4.1 Letter classification

Figure 3 shows the accuracy of Naive Bayes, SVM, and MLP classifiers obtained in the letter classification task ("a," "e," "h," and "l") for each individual scribe. The experiments showed good classification accuracy for each of the three writers. The classifier accuracy for each scribe is greater than 92.9%.

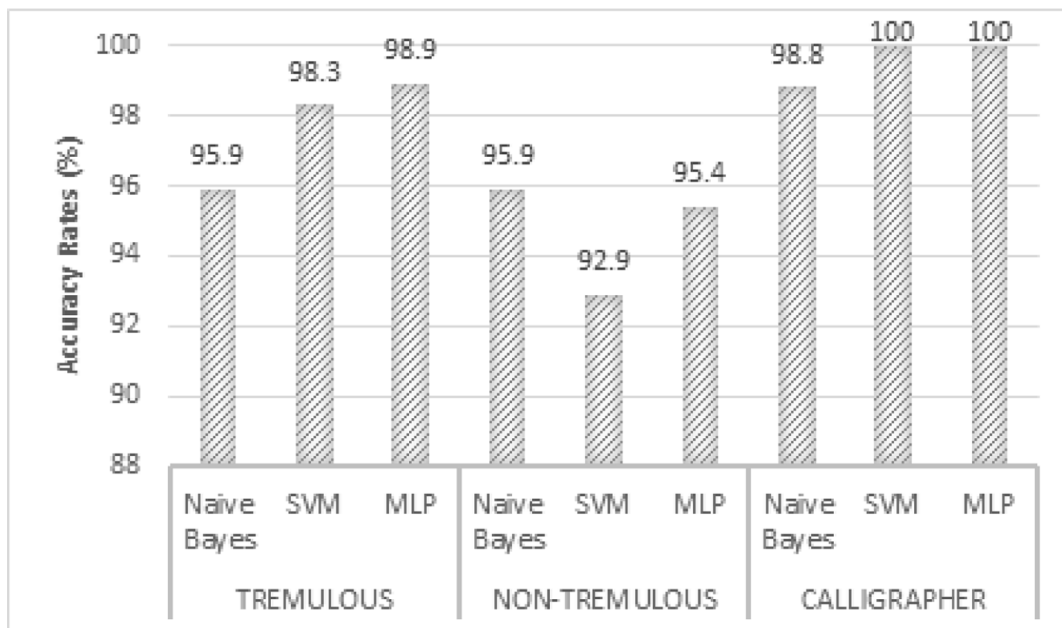


Figure 3: Comparison between the Naive Bayes, SVM, and MLP classifiers based on letters (“a,” “e,” “h,” and “l”) for each scribe (TREMULOUS, NON-TREMULOUS, and CALLIGRAPHER).

The confusion matrices for the TREMULOUS dataset (Table 1) show accuracy rates above 96%. However, all classifiers confounded “l” by “e” and “e” by “l” (see the samples incorrectly classified by all classifiers in Figure 4a, only MLP and SVM in Figure 4b and Naive Bayes in Figure 4c). The letter “l” has irregularities present in its shape with the top with a rounded stroke in some cases, noise background, and tremor present in write. Also, the letter “e” has degraded ink. These problems can have influenced the erroneous classification. For the NON-TREMULOUS dataset, confusion matrices of classifiers (Table 2) show that the letter “l” presented the most significant error rate. Typically, all classifiers confounded “l” with “e” (see samples in Figure 5). The letter “h,” on the other hand, was better classified by SVM and MLP. The classifiers obtained better performance on the TREMULOUS dataset than on the NON-TREMULOUS dataset. These problems may have been caused by dense noise in the test dataset because some data samples have parts of adjoining letters.

The dataset CALLIGRAPHER (Table 3) presented better results than the two other analyzed datasets. The absence of background noise may have improved the results obtained, with positive results achieved for all three classifiers investigated.

In general, in this experiment, we performed the letter classification to investigate character recognition of specific scribes. The results obtained in this experiment showed that the letters “l” and “e” of TREMULOUS and NON-TREMULOUS, which usually have

different shapes, are confounded by classifiers, due to irregularities present in the letter as background noise and tremor in the case of scribe TREMULOUS.

| Naive Bayes | | | | | SVM | | | | |
|-------------|----|----|----|----|-----|----|----|----|----|
| | a | e | h | l | | a | e | h | l |
| a | 29 | 1 | 0 | 0 | a | 30 | 0 | 0 | 0 |
| e | 0 | 88 | 0 | 2 | e | 0 | 89 | 0 | 1 |
| h | 1 | 0 | 19 | 0 | h | 0 | 0 | 20 | 0 |
| l | 0 | 3 | 0 | 28 | l | 0 | 2 | 0 | 29 |

| MLP | | | | |
|-----|----|----|----|----|
| | a | e | h | l |
| a | 30 | 0 | 0 | 0 |
| e | 0 | 89 | 0 | 1 |
| h | 0 | 0 | 20 | 0 |
| l | 0 | 1 | 0 | 30 |

Table 1: Confusion matrices for TREMULOUS.

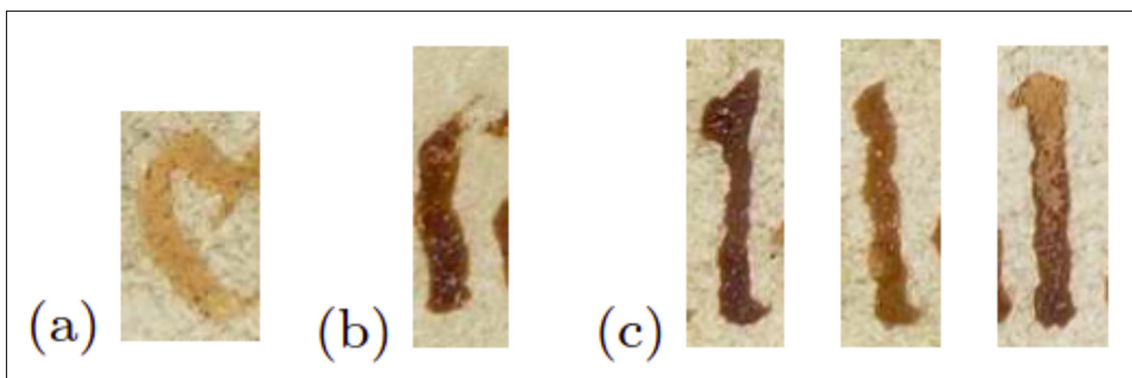


Figure 4: Letters samples from TREMULOUS scribe erroneously classified: (a) letter “e” confounded with “l” by all classifiers (Ref.: s1-e-56); (b) Letter “l” confounded with “e” by SVM and MLP classifier (Ref.: s1-l-21); (c) Letter “l” confounded with “e” by Naive Bayes (Ref.:s1-l-10, s1-l-12, s1-l-19).

| Naive Bayes | | | | | SVM | | | | |
|-------------|----|----|----|----|-----|----|----|----|----|
| | a | e | h | l | | a | e | h | l |
| a | 26 | 1 | 1 | 2 | a | 26 | 2 | 1 | 1 |
| e | 0 | 89 | 0 | 1 | e | 0 | 88 | 0 | 2 |
| h | 2 | 0 | 18 | 0 | h | 0 | 0 | 20 | 0 |
| l | 0 | 0 | 0 | 31 | l | 1 | 5 | 0 | 25 |

| MLP | | | | |
|-----|----|----|----|----|
| | a | e | h | l |
| a | 27 | 1 | 1 | 1 |
| e | 1 | 89 | 0 | 0 |
| h | 0 | 0 | 20 | 0 |
| l | 0 | 4 | 0 | 27 |

Table 2: Confusion matrices for NON-TREMULOUS.



Figure 5: Samples of "l" confounded with "e" by the Naive Bayes, SVM, and MLP classifiers from the NON-TREMULOUS scribe (Ref.: s3-l-06, s3-l-13, s3-l-22, s3-l-30).

| Naive Bayes | | | | | SVM | | | | |
|-------------|----|----|----|----|-----|----|----|----|----|
| | a | e | h | l | | a | e | h | l |
| a | 20 | 0 | 0 | 0 | a | 20 | 0 | 0 | 0 |
| e | 0 | 20 | 0 | 0 | e | 0 | 20 | 0 | 0 |
| h | 0 | 0 | 20 | 0 | h | 0 | 0 | 20 | 0 |
| l | 0 | 0 | 1 | 19 | l | 0 | 0 | 0 | 20 |

| MLP | | | | |
|-----|----|----|----|----|
| | a | e | h | l |
| a | 20 | 0 | 0 | 0 |
| e | 0 | 20 | 0 | 0 |
| h | 0 | 0 | 20 | 0 |
| l | 0 | 0 | 0 | 20 |

Table 3: Confusion matrices for CALLIGRAPHER.

4.2 Letter classification by writer

For letter classification by writer, we used subsets of the letters “a,” “e,” “h,” and “l,” each composed by samples of the scribes TREMULOUS, NON-TREMULOUS, and CALLIGRAPHER. **Figure 6** shows the accuracy rates obtained with the classifiers analyzed. This analysis presents encouraging results, where the subset “a,” which has lower accuracy rates, resulted in 93.8%, 88.8%, and 88.8% using Naive Bayes, SVM, and MLP classifiers, respectively.

For the letter “a,” the confusion matrices present in **Table 4** show that all classifiers confound CALLIGRAPHER with NON-TREMULOUS (see samples incorrectly classified in **Figure 7a**, **Figure 7b**). On the other hand, only the SVM and Naive classifiers assigned the NON-TREMULOUS with CALLIGRAPHER (see **Figure 7c**). We believe this is due to the similarity between the letter shapes produced by the two scribes.

The results of the letter “e” (see confusion matrices in **Table 5**), on the other hand, show that usually TREMULOUS is incorrectly classified as the NON-TREMULOUS and vice-versa (see **Figure 8**). Problems in the texts of both scribes may be responsible for incorrect letter classification, such as background noise. Besides, NON-TREMULOUS has dye stains, while the TREMULOUS dataset has notable ink degradation.

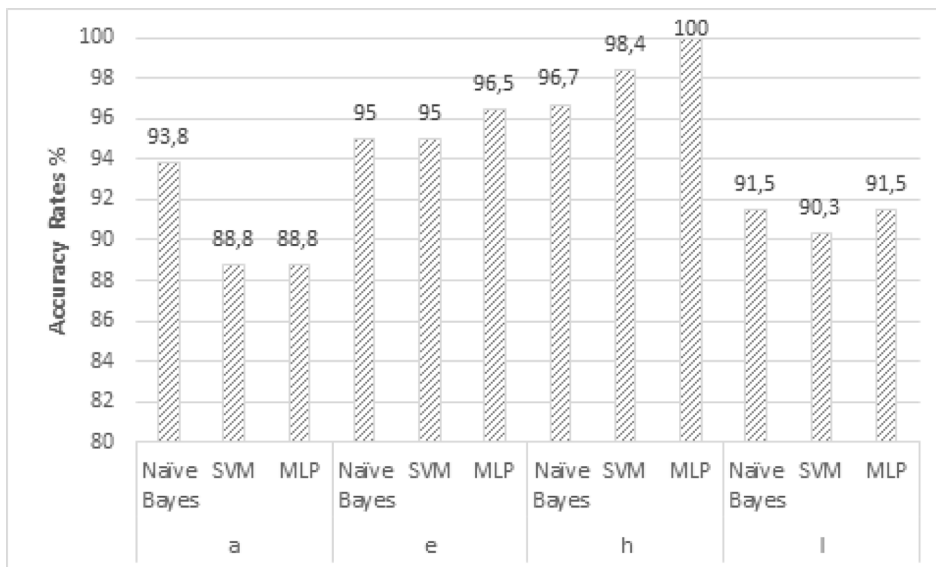


Figure 6: Comparison between the Naive Bayes, SVM, and MLP classifiers based on the scribes (TREMULOUS, NON-TREMULOUS and CALLIGRAPHER) for each letter (“a,” “e,” “h,” and “l”).

| Naive Bayes | | | | SVM | | | |
|-------------|-----|-----|-----|-----|-----|-----|-----|
| | TRE | NON | CAL | | TRE | NON | CAL |
| TRE | 29 | 1 | 0 | TRE | 29 | 1 | 0 |
| NON | 1 | 29 | 0 | NON | 1 | 25 | 4 |
| CAL | 0 | 3 | 17 | CAL | 0 | 3 | 17 |

| MLP | | | |
|-----|-----|-----|-----|
| | TRE | NON | CAL |
| TRE | 29 | 1 | 0 |
| NON | 2 | 25 | 3 |
| CAL | 0 | 3 | 17 |

Table 4: Confusion matrices for “a.”



Figure 7: Samples of the letter “a” erroneously classified by more than one classifier: (a) samples of the letter “a” from CALLIGRAPHER incorrectly classified as NON-TREMULOUS by all classifiers (Ref.: s4-a-01,s4-a-09); (b) sample of the letter “a” from CALLIGRAPHER incorrectly classified as the NON-TREMULOUS, by SVM and MLP classifiers (Ref.: s4-a-20); (c) samples of the letter “a” from NON-TREMULOUS incorrectly classified as CALLIGRAPHER by SVM and MLP classifiers (Ref.:s3-a-19, s3-a-22, s3-a-26).

| Naive Bayes | | | | SVM | | | |
|-------------|-----|-----|-----|-----|-----|-----|-----|
| | TRE | NON | CAL | | TRE | NON | CAL |
| TRE | 84 | 6 | 0 | TRE | 85 | 6 | 0 |
| NON | 3 | 87 | 0 | NON | 5 | 85 | 1 |
| CAL | 0 | 1 | 19 | CAL | 0 | 0 | 20 |

| MLP | | | |
|-----|-----|-----|-----|
| | TRE | NON | CAL |
| TRE | 86 | 5 | 0 |
| NON | 4 | 87 | 0 |
| CAL | 0 | 0 | 20 |

Table 5: Confusion matrices for “e.”

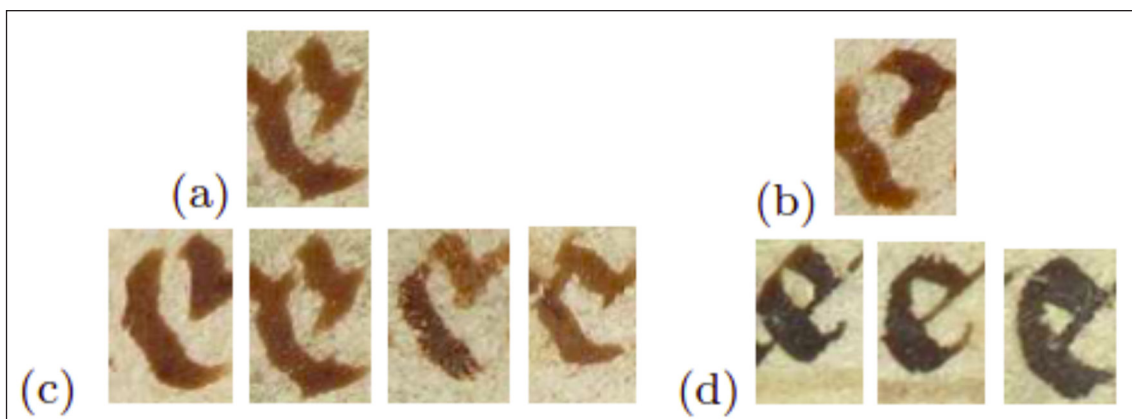


Figure 8: Samples of the letter “e” erroneously classified by more than one classifier: (a) sample of the letter “e” from TREMULOUS incorrectly classified as the NON-TREMULOUS by all classifiers (Ref.: s1-e-16); (b) sample of the letter “e” from TREMULOUS incorrectly classified as the NON-TREMULOUS by the Naive Bayes and MLP classifiers (Ref.: s1-e-04); (c) samples of the letter “e” from TREMULOUS incorrectly classified as the NON-TREMULOUS by SVM and MLP classifiers (Ref.: s1-e-10, s1-e-16, s1-e-17, s1-e-74); (b); (d) Sample of the letter “e” from NON-TREMULOUS incorrectly classified as the TREMULOUS by all classifiers (Ref.: s3-e-02, s3-e-04, s3-e-69).

The confusion matrices of the letter “h” (Table 6) show similar results with encouraging accuracy levels. The distinctions in the shape of the letter “h” may have motivated the results obtained (see Figure 9), although both documents of scribes TREMULOUS and NON-TREMULOUS have been severely degraded. Furthermore, TREMULOUS presents a visible tremor in writing, and NON-TREMULOUS has a slight inclination in writing, whereas CALLIGRAPHER has no background noise.

| Naive Bayes | | | | SVM | | | |
|-------------|-----|-----|-----|-----|-----|-----|-----|
| | TRE | NON | CAL | | TRE | NON | CAL |
| TRE | 20 | 0 | 0 | TRE | 19 | 1 | 0 |
| NON | 1 | 19 | 0 | NON | 0 | 20 | 0 |
| CAL | 0 | 1 | 19 | CAL | 0 | 0 | 20 |

| MLP | | | |
|-----|-----|-----|-----|
| | TRE | NON | CAL |
| TRE | 20 | 0 | 0 |
| NON | 0 | 20 | 0 |
| CAL | 0 | 0 | 20 |

Table 6: Confusion matrices for “h.”

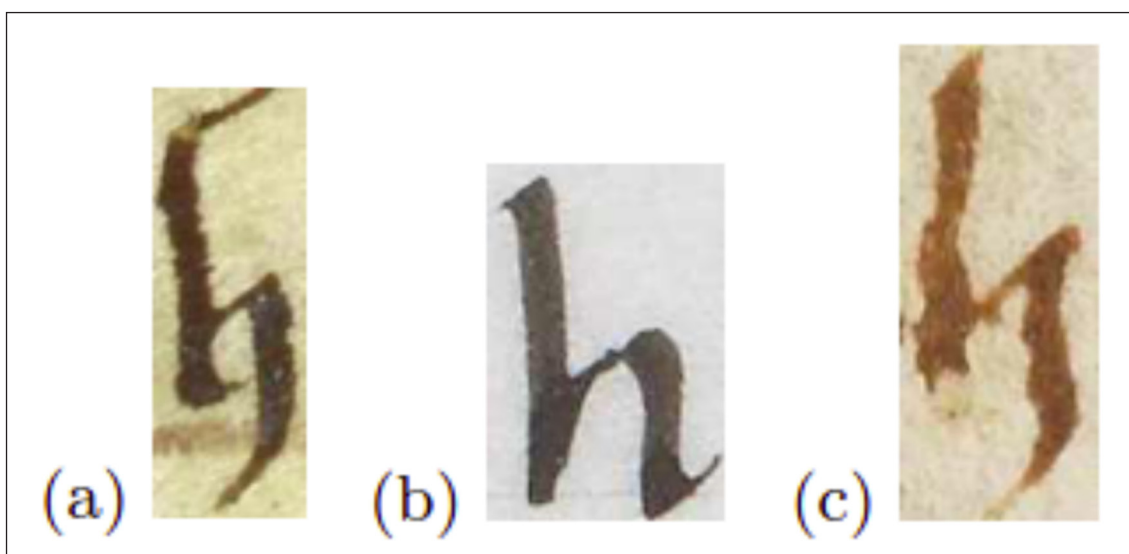


Figure 9: Samples of the letter “h” erroneously classified: (a) sample of the letter “h” from NON-TREMULOUS incorrectly classified as the TREMULOUS by Naïve Bayes classifier (Ref.: s3-h-09); (b) sample of the letter “h” from CALLIGRAPHER incorrectly classified as the NON-TREMULOUS by Naïve Bayes classifier (Ref.: s4-h-05); (c) sample of the letter “h” from TREMULOUS incorrectly classified as the NON-TREMULOUS by SVM classifier (Ref.: s1-h-04).

The confusion matrices for the letter “l” (Table 7) present small variations between the results of classifiers. All classifiers confound NON-TREMULOUS with TREMULOUS (see Figures 10a, 10b, and 10c), and CALLIGRAPHER with NON-TREMULOUS (see Figures 10d and 10e) and vice versa. The letter “l” has a single stroke with a similar

shape in the writing of scribes. Furthermore, the dense noise presented in NON-TREMULOUS samples may have contributed to the results obtained.

| Naive Bayes | | | | SVM | | | |
|-------------|-----|-----|-----|-----|-----|-----|-----|
| | TRE | NON | CAL | | TRE | NON | CAL |
| TRE | 28 | 3 | 0 | TRE | 29 | 2 | 0 |
| NON | 2 | 28 | 1 | NON | 3 | 26 | 2 |
| CAL | 0 | 1 | 19 | CAL | 0 | 1 | 19 |

| MLP | | | |
|-----|-----|-----|-----|
| | TRE | NON | CAL |
| TRE | 27 | 3 | 1 |
| NON | 0 | 29 | 2 |
| CAL | 0 | 1 | 19 |

Table 7: Confusion matrices for “l.”

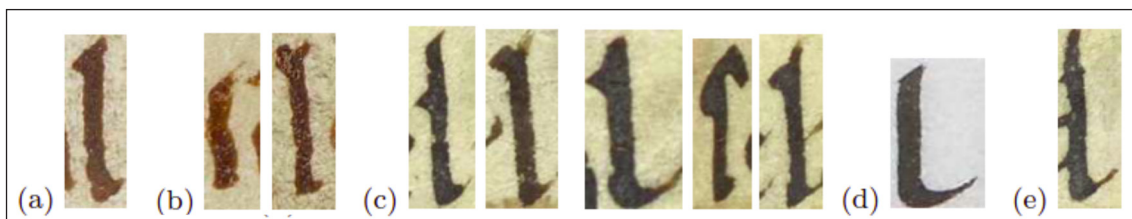


Figure 10: Samples of the letter “l” erroneously classified by more than one classifier: (a) Sample of the letter “l” from TREMULOUS incorrectly classified as the NON-TREMULOUS by all classifiers (Ref.: s1-l-06); (b) sample of the letter “l” from TREMULOUS incorrectly classified as the NON-TREMULOUS by Naive Bayes and MLP classifiers (Ref.: s1-l-21) and by Naive Bayes and SVM classifiers (Ref.: s1-l-22); (c) Sample of the letter “l” from CALLIGRAPHER incorrectly classified as the NON-TREMULOUS by Naive classifier (Ref.:s3-l-17, s3-l-19) and by SVM classifier (Ref.:s3-l-05,s3-l-15,s3-l-26); (d) sample of the letter “l” from CALLIGRAPHER incorrectly classified as the NON-TREMULOUS by all classifiers (Ref.:s4-l-15); (e) sample of the letter “l” from NON-TREMULOUS incorrectly classified as the CALLIGRAPHER by all classifiers (Ref.:s3-l-22).

From these results, we observed that the letters present quite different behaviours, where in some cases scribes are more mistakenly classified than in others, mainly NON-TREMULOUS by TREMULOUS and TREMULOUS by NON-TREMULOUS. The similarity in the writing period and the noises presented in the authors’ handwritten documents may have caused the wrong classification between scribes.

In this experiment, we verified the relationship between each letter and the writers. The results showed that the letters “a” and “l” are incorrectly classified as the NON-TREMULOUS by CALLIGRAPHER and vice versa, and the letters “e” and “l” as TREMULOUS by the NON-TREMULOUS. Furthermore, we observed that the letter “h” presented better accuracy in the results. These results can be related to particularities in the writing of each scribe.

4.3 Classification of writer

Writer identification is a related task in the analysis of medieval handwritten documents. Since English texts of the Middle Ages were usually written and copied by anonymous scribes (Gillespie and Wakelin 2011), the automation of the writer identification process of medieval manuscripts can be an essential contribution.

The results obtained for writer identification are present in **Figure 11**. The Naive Bayes classifier performed worse when compared to other classifiers analyzed. On the other hand, the MLP classifier had the best accuracy in writer identification, with an average of 89% accuracy.

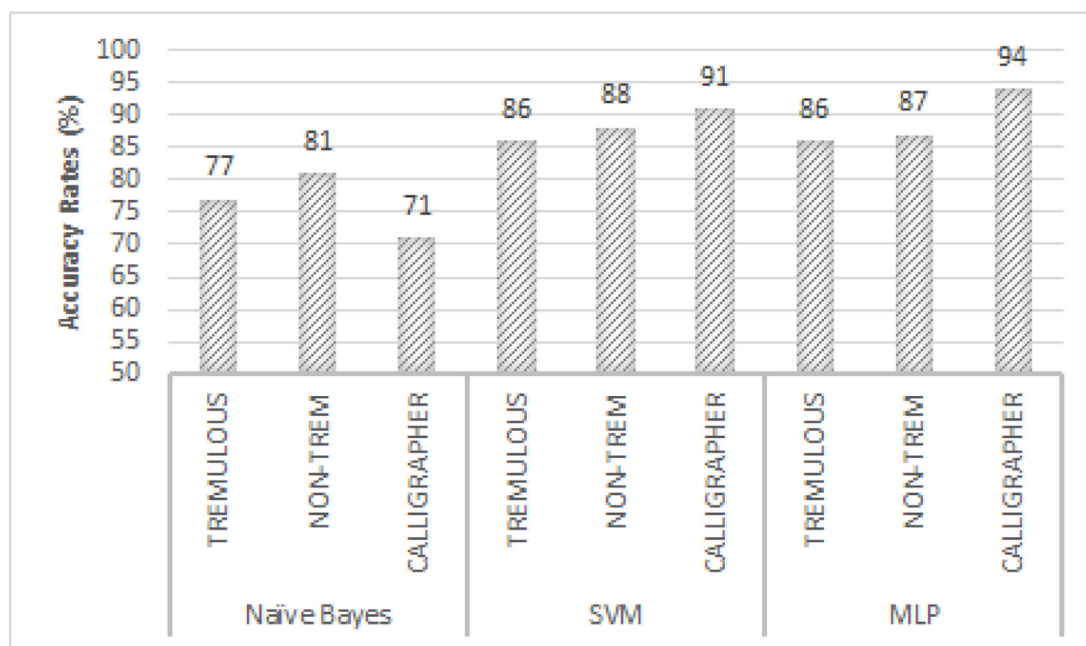


Figure 11: Comparison between the Naive Bayes, SVM, and MLP classifiers based on scribe (TREMULOUS, NON-TREMULOUS, and CALLIGRAPHER).

Confusion matrices for writer identification (**Table 8**) show that samples of all scribes are classified incorrectly. However, it was possible to observe a relationship between

TREMULOUS and NON-TREMULOUS, and NON-TREMULOUS with CALLIGRAPHER. This relationship is due to some features of each scribe's writing styles and documents, mainly background problems in medieval manuscripts caused by different factors.

| Naive Bayes | | | SVM | | | | |
|-------------|-----|-----|-----|-----|-----|-----|-----|
| | TRE | NON | CAL | | TRE | NON | CAL |
| TRE | 138 | 25 | 8 | TRE | 151 | 16 | 4 |
| NON | 15 | 132 | 24 | NON | 14 | 147 | 10 |
| CAL | 6 | 17 | 57 | CAL | 3 | 4 | 73 |

| MLP | | | |
|-----|-----|-----|-----|
| | TRE | NON | CAL |
| TRE | 149 | 18 | 4 |
| NON | 16 | 147 | 8 |
| CAL | 1 | 4 | 75 |

Table 8: Confusion matrices for writer identification.

4.4 Analysis of metrics on character samples

Global metrics (HPP and VPP) measure the pixel density of characters. The measurements indicate the amount of ink deposited in the samples, based on the number of pixels accumulated. The HPP denotes the sum of pixels along the rows and VPP of the columns in the image.

For the letter “a,” we observed that the results of HPP and VPP metrics presented a curved shape of different writers, notably distinct (see **Figure 12**). It is also important to note that TREMULOUS has more significant dissimilarity between the curves of each sample in HPP and VPP than the other scribes. The NON-TREMULOUS has little irregularities between rows 60 and 100, around the letter centre, due to changes in stroke thickness and shape of the samples. CALLIGRAPHER has uniform behaviour with irregular peaks in the VPP metric, but with a similar curve shape.

For the letter “e,” the results are shown in **Figure 13**. The TREMULOUS dataset samples have great dissimilarity between them, with a small discrepancy in curves, mainly in VPP results due to writing failures and degradation in documents. NON-TREMULOUS and TREMULOUS have peaks in VPP results at the end of the curve, because the letter “e” of NON-TREMULOUS has a stroke on the upper right side in some samples and CALLIGRAPHER has a long stroke on the down of some letters.

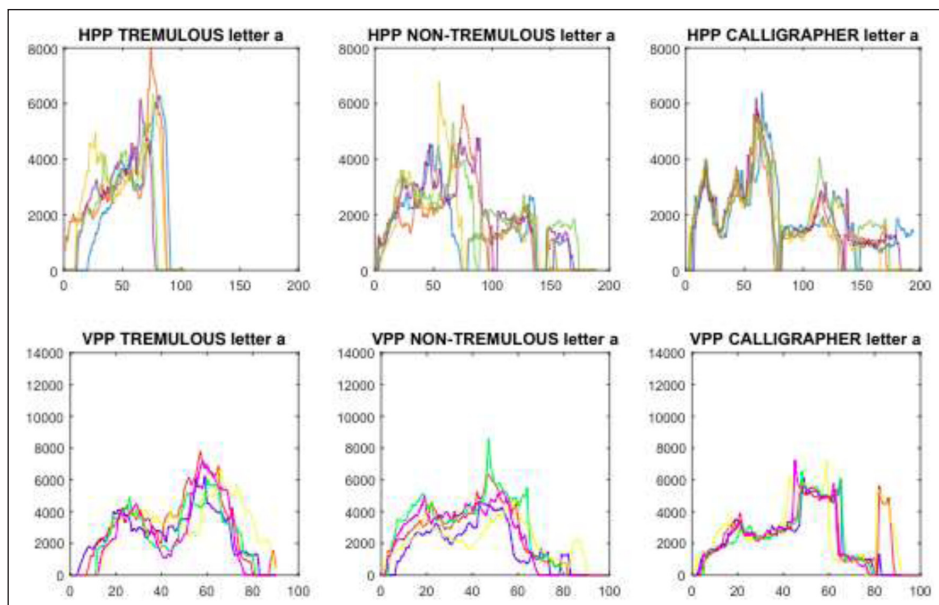


Figure 12: HPP and VPP metrics values for samples of the scribes TREMULOUS, NON-TREMULOUS, and CALLIGRAPHER of the letter “a.” The x-axis represents the height in HPP and width in VPP of the character in pixels and the y-axis is the sum of pixel values belonging to the character.

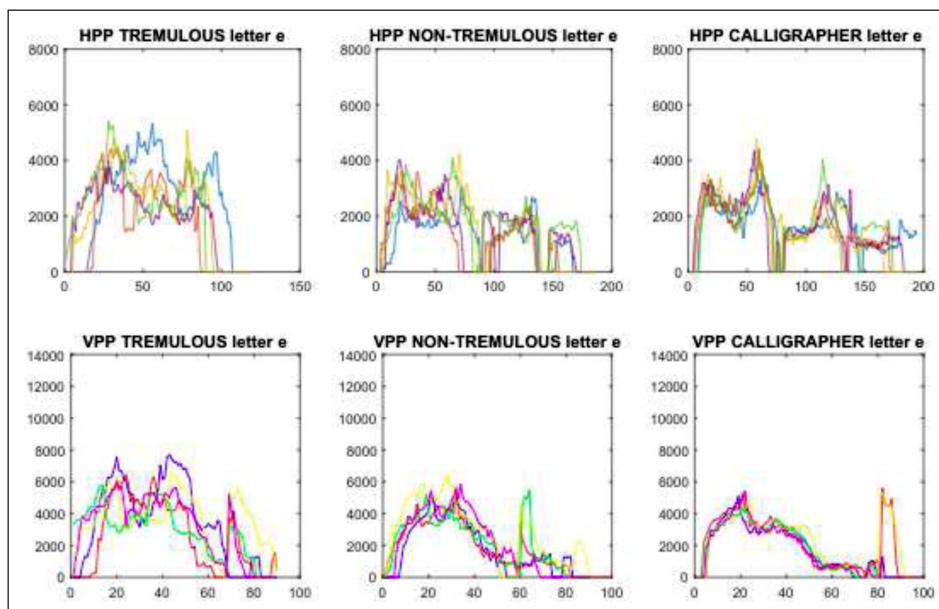


Figure 13: HPP and VPP metrics values for samples of the scribes TREMULOUS, NON-TREMULOUS, and CALLIGRAPHER of the letter “e.” The x-axis represents the height in HPP and width in VPP of the character in pixels and the y-axis is the sum of pixel values belonging to the character.

Compared with the letters “a” and “e,” the “h” and “l” have less dissimilarity between TREMULOUS samples with the HPP metric (see **Figures 14** and **15**). However, the VPP metric has some irregular curves in all writers, mainly TREMULOUS and NON-TREMULOUS scribes.

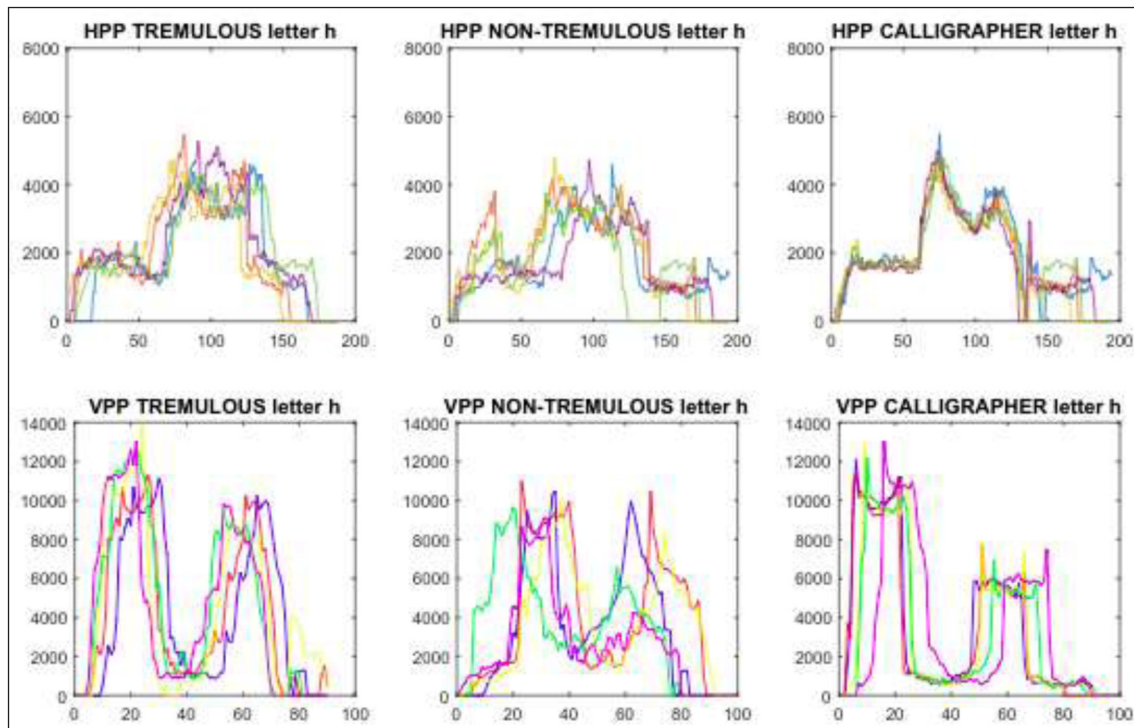


Figure 14: HPP and VPP metrics values for samples of the scribes TREMULOUS, NON-TREMULOUS, and CALLIGRAPHER of the letter “h.” The x-axis represents the height in HPP and width in VPP of the character in pixels and the y-axis is the sum of pixel values belonging to the character.

In general, we observed that the character samples of each writer had similar behaviour between themselves using the HPP metric. On the other hand, the results obtained from different writers are notably distinct. Furthermore, VPP results showed some irregularities in the samples of each scribe, with more regularity in the letters “a” and “e.”

The PDV metric was used to analyze density variation (see **Figure 16**). We split the letters into eight cells ($K = 8$); as the characters “a,” “e,” and “h” have two strokes, we divided the image into four times the number of strokes to explore small parts. Thus, writing effects can be detected based on how characteristics change locally within samples of TREMULOUS, NON-TREMULOUS and CALLIGRAPHER scribes.

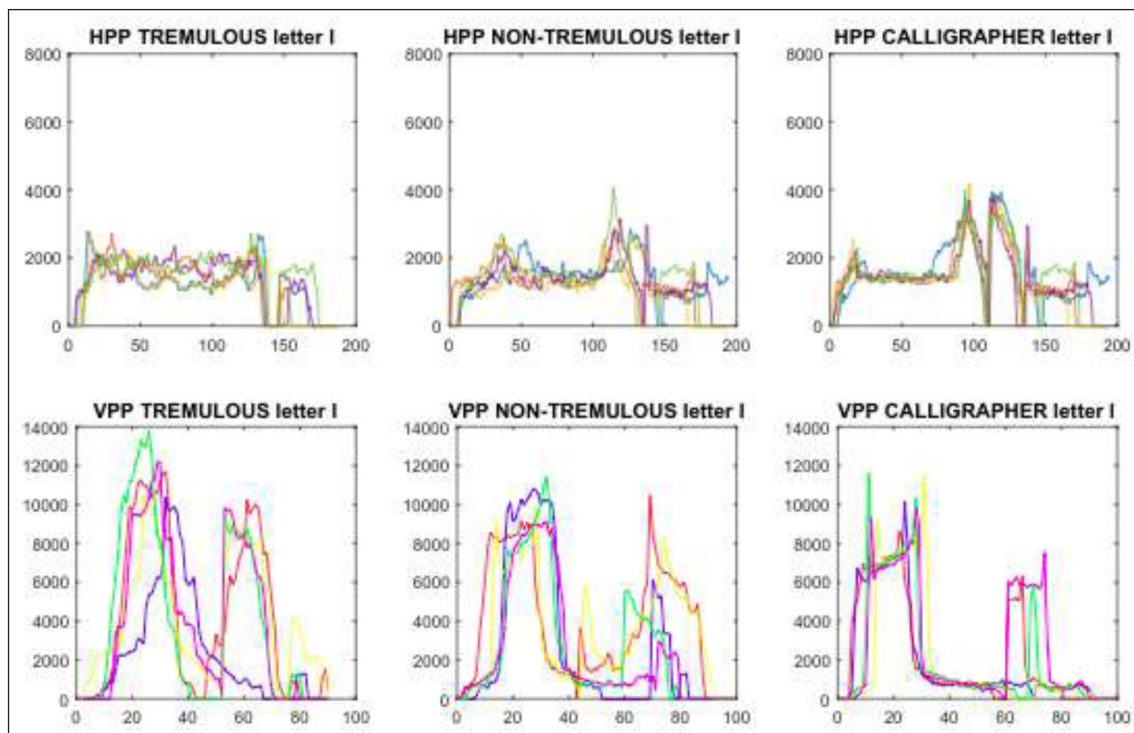


Figure 15: HPP and VPP metrics values for samples of the scribes TREMULOUS, NON-TREMULOUS, and CALLIGRAPHER of the letter "l." The x-axis represents the height in HPP and width in VPP of the character in pixels and the y-axis is the sum of pixel values belonging to the character.

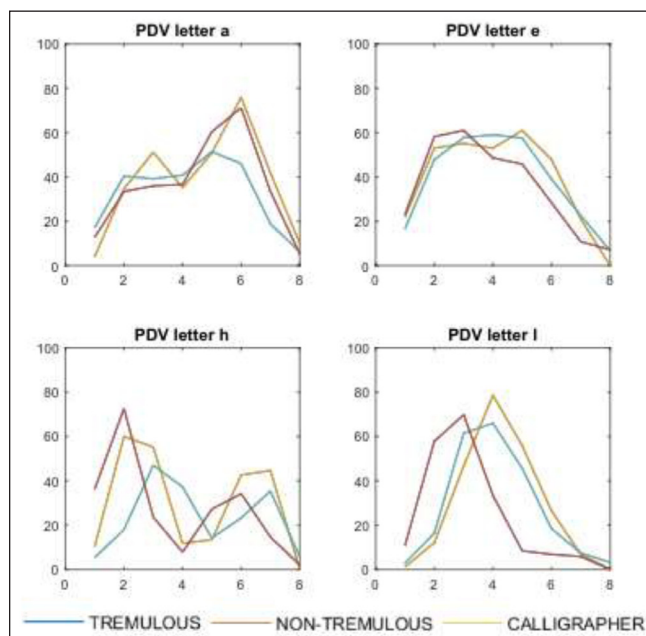


Figure 16: PDV metric values for letters "a," "e," "h," and "l" from TREMULOUS, NON-TREMULOUS, and CALLIGRAPHER scribes. The x-axis represents the cells ($K = 8$) of the characters and y-axis is the pixel density average from five samples of the cells belonging to the character.

From the results obtained with PDV, it was possible to note that the variation in the density of the letter “a” has a more significant similarity between NON-TREMULOUS and CALLIGRAPHER. The results of “e” for the three writers were similar, while the letters “h” and “l” were distinct: A significant result that may indicate a relationship between these results and the classifiers described in Section 4.2, since the letters “h” and “l” had fewer classification errors.

It is interesting to note that HPP, VPP, and PDV metrics are based on characteristics such as the total amount of ink employed and the density variation in samples. Thus, the obtained results show that these metrics can be used to provide support in medieval document analysis.

4.4.1 Analysis of linear correlation between the letter samples

For the letter “a,” in **Figure 17**, it is possible to see the linear correlation of the metric HPP results, and in **Figure 18**, the VPP results. The results obtained by the samples of all scribes have values near 1, indicating a positive linear correlation. TREMULOUS scribe results show that the linear correlation of some samples is below 0.4. On the other hand, the VPP linear correlation between samples is greater than 0.4. The NON-TREMULOUS scribe also has a below 0.4 linear correlation to HPP results, due to irregularities around the letter centre, described in the previous section. However, for the VPP metric, the linear correlation is greater than 0.6 for all samples. The CALLIGRAPHER has greater uniformity in the linear correlation between samples of the HPP metric, with a linear correlation greater than 0.6 for all samples. However, in VPP, some linear correlation results are below 0.6 but not too distant.

For the HPP metric, the linear correlation results for the letter “e” are shown in **Figure 19**. The linear correlation between the samples is below 0.4. Furthermore, on average, TREMULOUS has a smaller correlation between samples for the VPP metric (see **Figure 20**). As noted in the previous section, these results were caused by writing failures and ink deterioration. On the other hand, NON-TREMULOUS and CALLIGRAPHER have results greater than 0.4 in HPP, and for VPP greater than 0.6.

Figure 21 presents results for linear correlation between letter “h” samples of the HPP metric. TREMULOUS samples correlated greater than 0.5, but for the VPP metric, some results were below 0.4 (see **Figure 22**). The NON-TREMULOUS scribe also had a positive correlation to all samples. However, for the VPP metric, some results were negative and near zero, indicating little correlation. In contrast, CALLIGRAPHER for HPP presented a high correlation between samples with all results greater than 0.8, but for the VPP metric, it had variable results with some below 0.4.

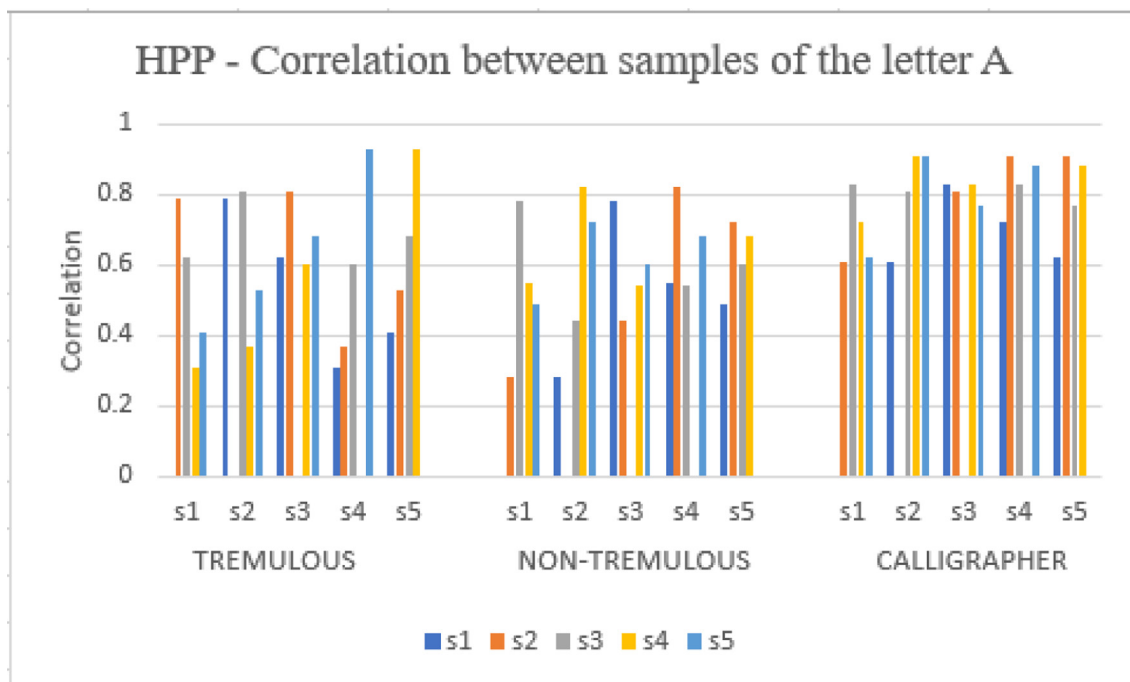


Figure 17: HPP - Correlation between samples of the letter A of each scribe (TREMULOUS, NON-TREMULOUS, CALLIGRAPHER).

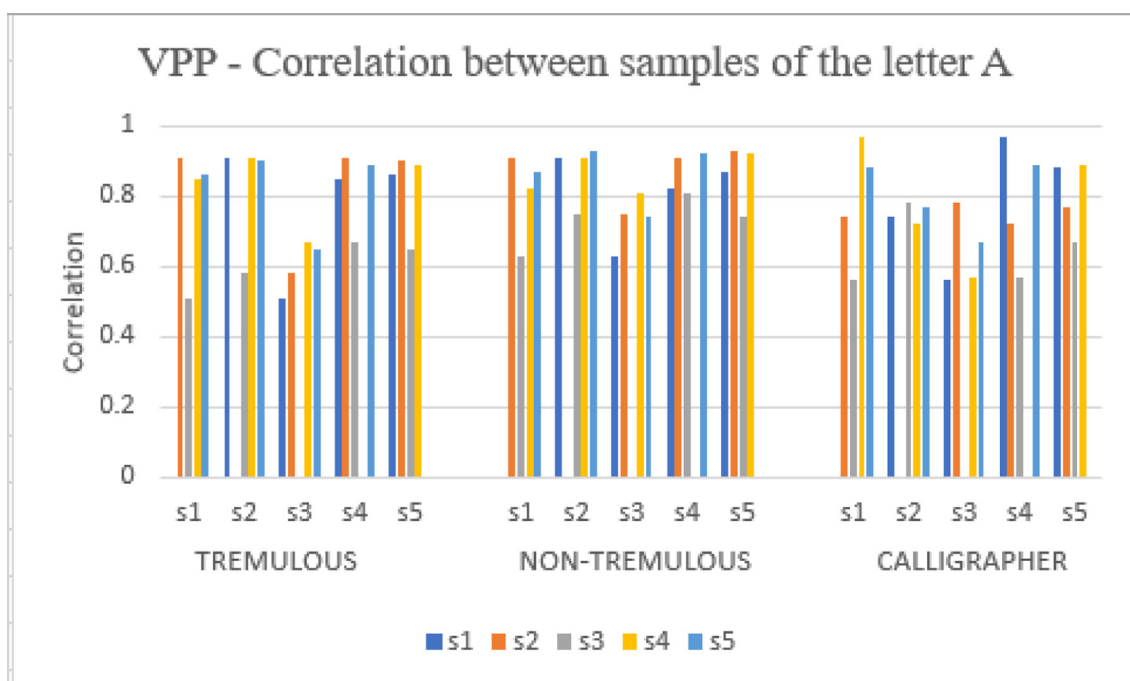


Figure 18: VPP - Linear correlation between samples of the letter A of each scribe (TREMULOUS, NON-TREMULOUS, CALLIGRAPHER).

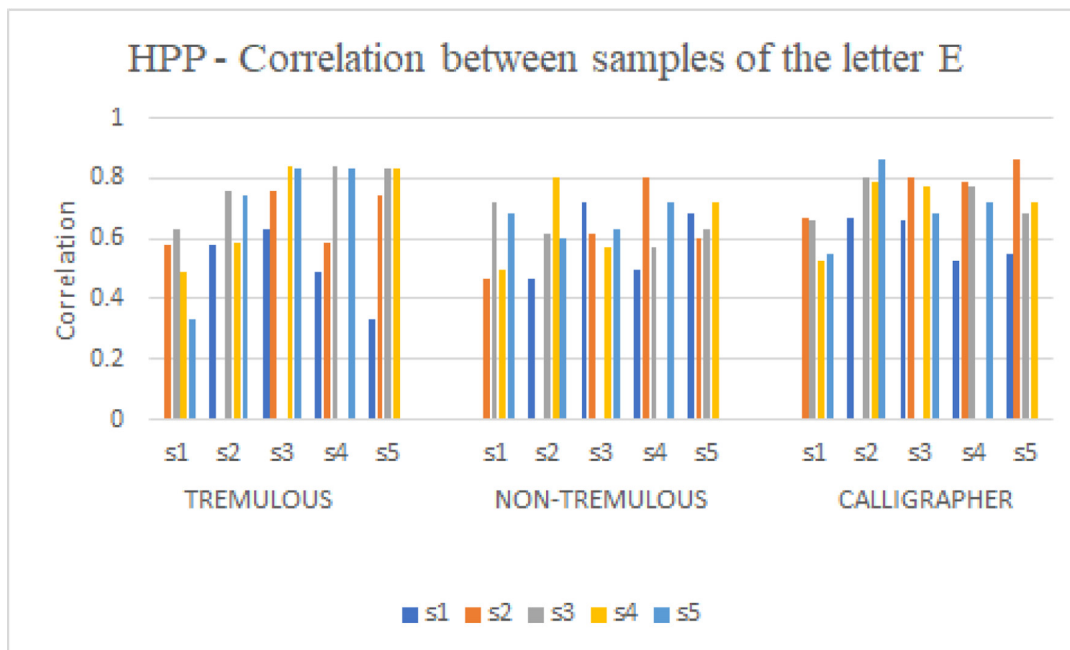


Figure 19: HPP – Correlation between samples of the letter E of each scribe (TREMULOUS, NON-TREMULOUS, CALLIGRAPHER).

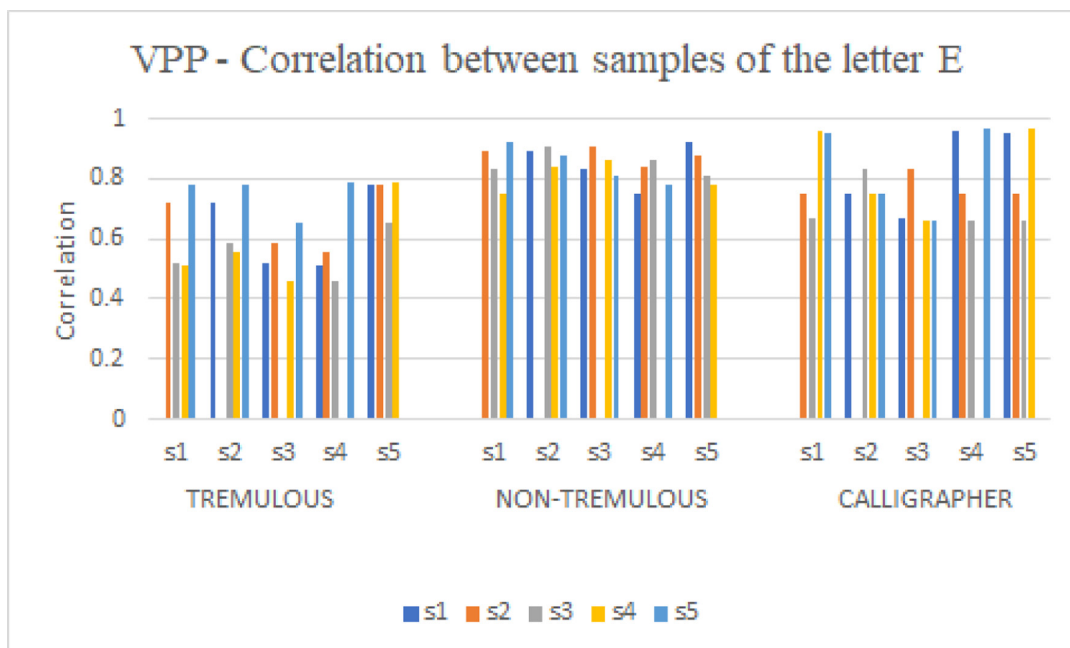


Figure 20: VPP – Correlation between samples of the letter E of each scribe (TREMULOUS, NON-TREMULOUS, CALLIGRAPHER).

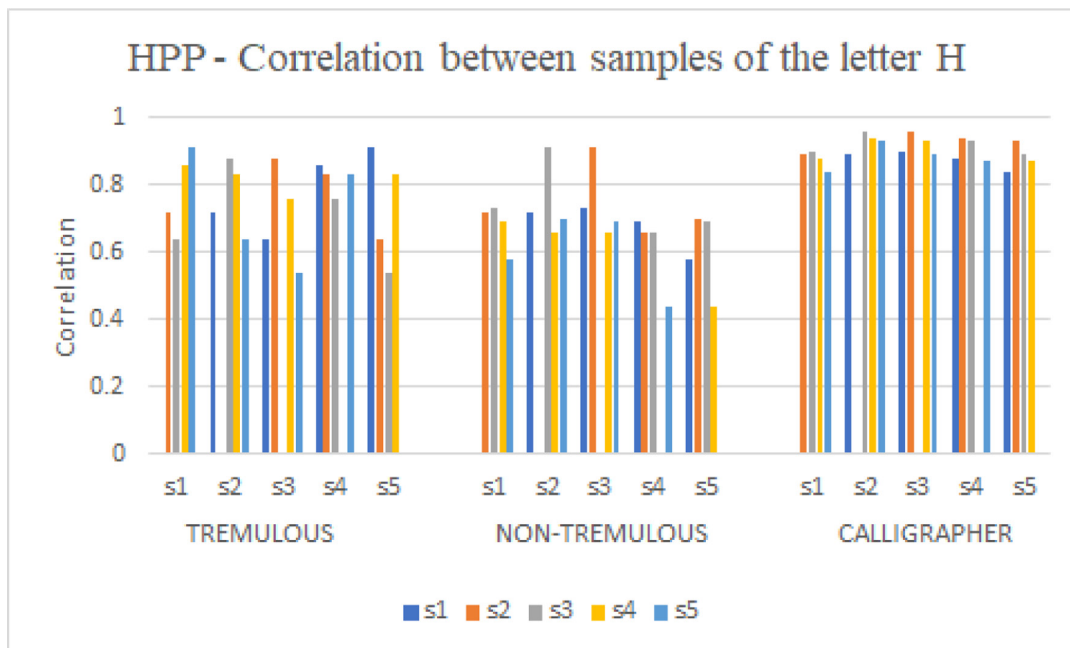


Figure 21: HPP – Correlation between samples of the letter H of each scribe (TREMULOUS, NON-TREMULOUS, CALLIGRAPHER).

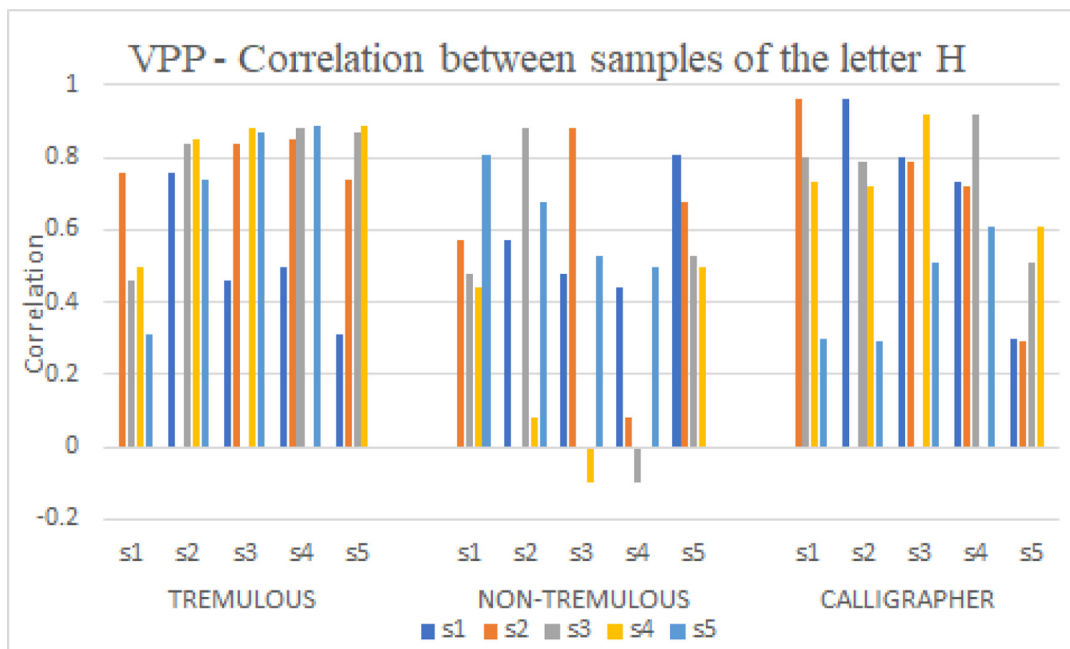


Figure 22: VPP – Correlation between samples of the letter H of each scribe (TREMULOUS, NON-TREMULOUS, CALLIGRAPHER).

For the HPP metric, the linear correlation results for the letter “l” are shown in **Figure 23**. We observed that for TREMULOUS, such as for the letter “h,” the linear correlation between the samples is greater than 0.5. However, for the VPP metric (see **Figure 24**), only the linear correlation with the sample “s1” is below 0.8, indicating a high correlation between the other samples. On the other hand, NON-TREMULOUS, for the HPP metric, had a variation in correlation results with some below 0.2 and for VPP, all results are greater than 0.4, due to ink degradation. CALLIGRAPHER, for the HPP metric, presented near values, where all results are greater than 0.6, but for VPP, the linear correlation of some samples is below 0.6.

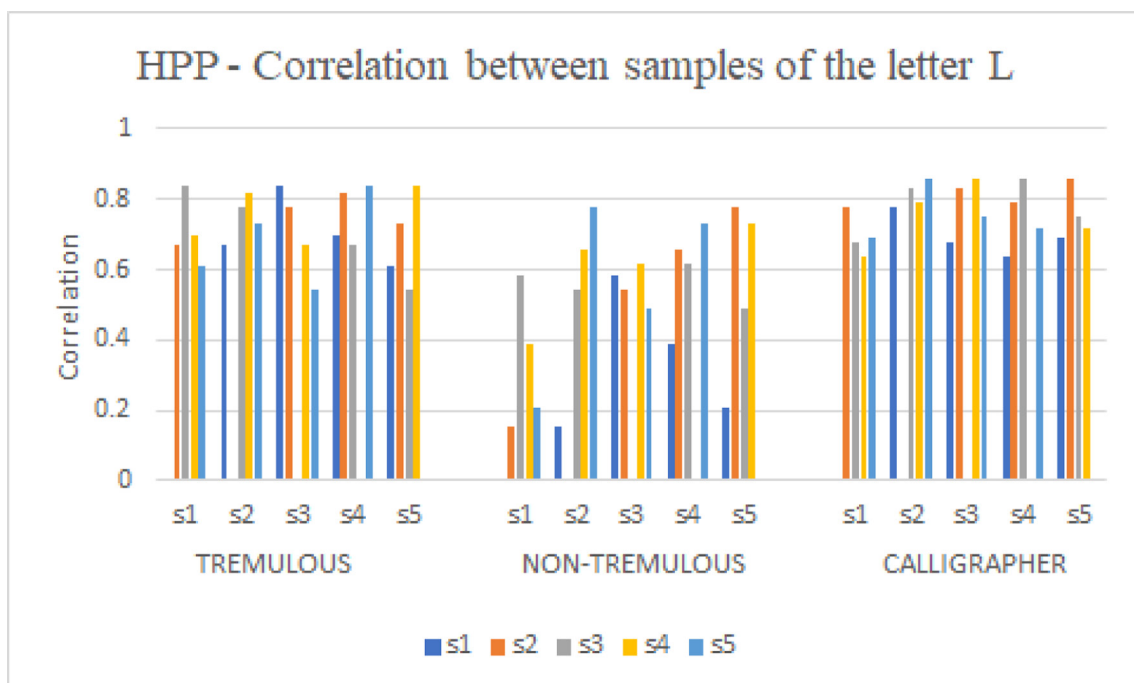


Figure 23: HPP – Correlation between samples of the letter L of each scribe (TREMULOUS, NON-TREMULOUS, CALLIGRAPHER).

Statistical analysis of metrics has shown that the particularities of documents influence the correlation between letter samples from medieval texts. As observed in Section 4.4, TREMULOUS and NON-TREMULOUS scribes have more significant dissimilarity between their samples, therefore, lower correlation, due to document degradation, tremor, writing failures, and changes in letter shape. CALLIGRAPHER,

on the other hand, has uniform conduct. Thus, it had a more significant correlation between the samples.

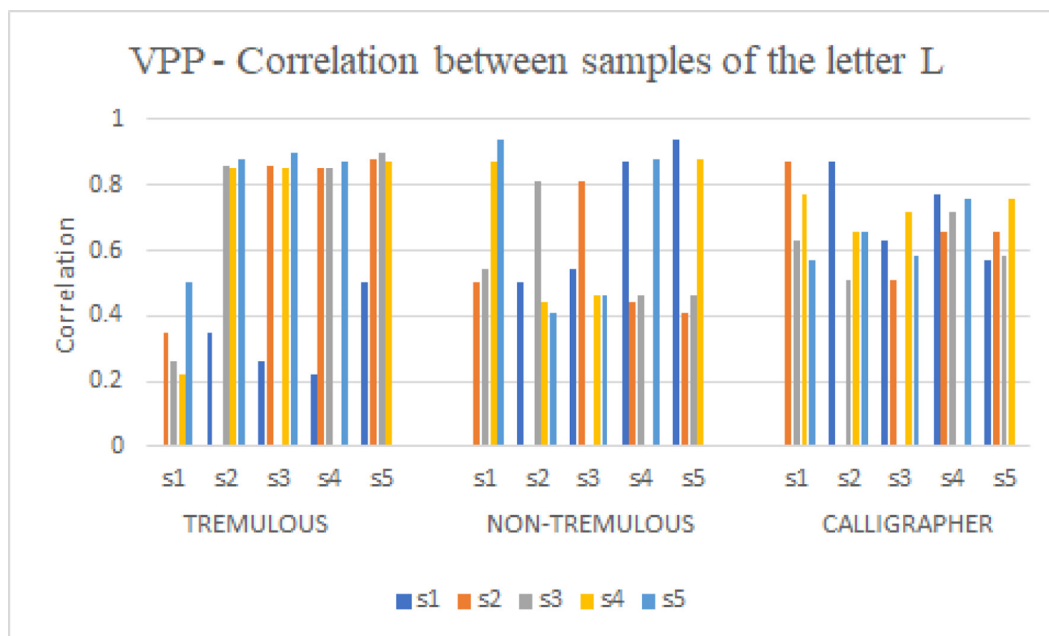


Figure 24: VPP – Correlation between samples of the letter L of each scribe (TREMULOUS, NON-TREMULOUS, CALLIGRAPHER).

4.4.2 Analysis of linear correlation between the scribes

Another aspect to be considered, besides the linear correlation of letter samples, is the linear correlation between scribes. We analyzed the statistical relationship between them to see if there is a correlation between writers.

For the HPP metric (see **Figure 25**), all results had a positive linear correlation with some values below 0.6, but some near 0.8. Based on the results verified, the scribe TREMULOUS had a higher correlation with the scribe NON-TREMULOUS for the letters “a,” “e,” and “l.” In contrast, scribe CALLIGRAPHER had a higher correlation with the NON-TREMULOUS for letters “e,” “h,” and “l.”

For the VPP metric (see **Figure 26**), the linear correlation between scribes was also positive. However, the letter “h” had a smaller correlation than the other letters, with all results below 0.6. As discussed in the previous sections, characteristics such as ink degradation, tremor, slight inclination on writing, may have motivated the results obtained.

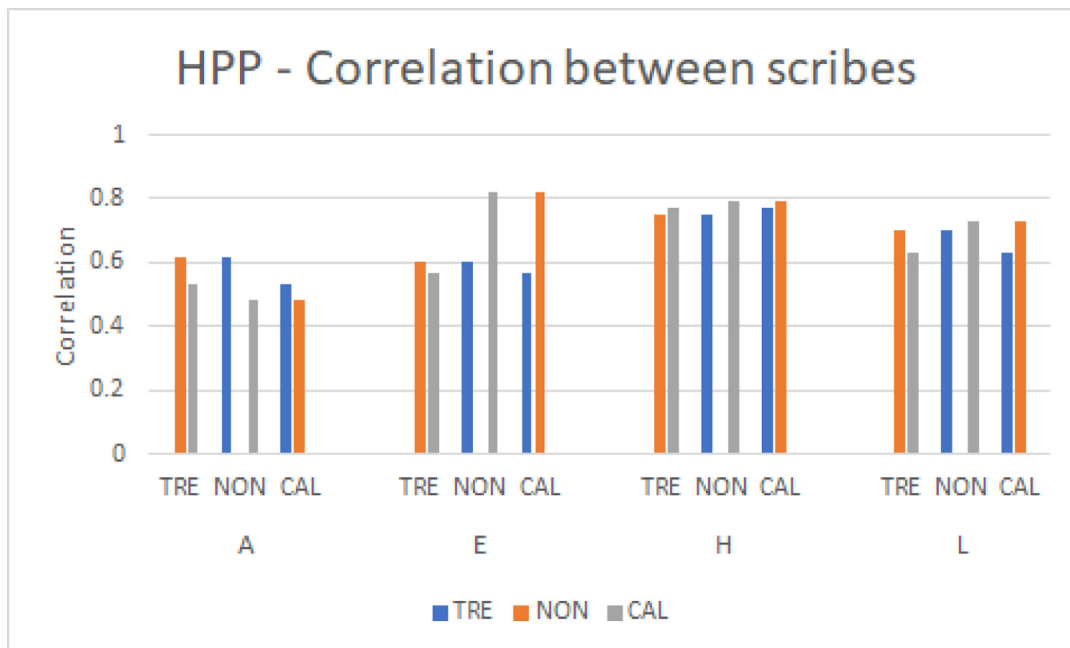


Figure 25: HPP – Correlation between the scribes (TREMULOUS, NON-TREMULOUS, CALLIGRAPHER).

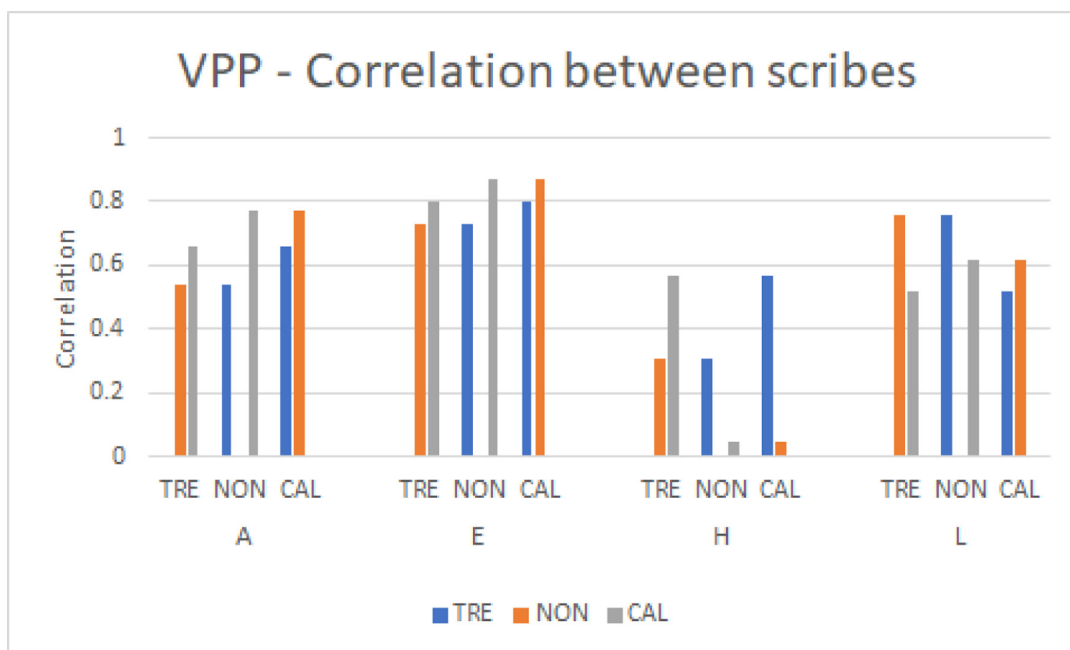


Figure 26: VPP – Correlation between the scribes (TREMULOUS, NON-TREMULOUS, CALLIGRAPHER).

In general, the experiments performed in this paper demonstrate that particular problems in medieval documents and the presence of signs of scribe's neurological condition can affect the performance of the analysis of the handwritten texts. The study of medieval manuscripts, in the context of diseases and disorders, is a relevant task since the characteristics of the scribe's neurological condition can be found in these documents and can help to understand the impact on the medieval context. Given the relevance of this issue, support for the development of technological tools for handwriting analysis is a significant contribution to paleography.

5 Conclusion

This paper presented experimental research on the impact of diseases and disorders in automatic handwritten character and writer recognition using a limited dataset, specifically medieval texts. Moreover, we compared the classification capabilities of medieval handwriting samples with modern samples. We explored the main problems related to medieval samples, such as parchment degradation and ink-related noise. We investigated character classification, writer's classification, and ink distribution.

To have a fair comparison, we decided to select three handwritten documents written in Middle English, two of which are medieval documents (~800 years old) and one written by a modern expert calligrapher. We have based our investigation on character classification, character classification by writer, writer identification, and pixel density analysis. Naive Bayes, SVM, and MLP classifiers were employed to classify character images and writer identification.

The character classification (Section 4.1) and character classification by the writer (Section 4.2) have analyzed the character classification capability and the characteristics that influenced the erroneous classification. On the other hand, the classification of the writer (Section 4.3) investigated the writer identification, and the analyses of metrics on characters samples (Section 4.4) have given us information about characters' pixel density.

Experiments in this work showed us that degradation presented in medieval manuscripts and the tremor (when present) can significantly influence automatic character recognition and scribe identification. However, the results demonstrated an effective accuracy rate in our experiments. These results are significant and may indicate a new line of research for this area of application (historical manuscripts), which considers, in addition to the pure form of the samples, the density of the ink, and the circumstances of that particular scribe.

This paper's main limitation was data availability since our research focused on analyzing medieval documents with pathological evidence from writing (~800 years old). Nevertheless, the anonymous writer, known as "The Tremulous Hand of Worcester," is the only widely known medieval writer with a tremor (Thorpe and Alty 2015). Unfortunately, due to resource constraints for data collection, we have used a limited dataset. Therefore, with this restriction in mind, our efforts were directed to exploring different possibilities.

However, different approaches relevant to the subject still need to be explored better to understand this medieval scribe's handwriting and situation. For future work, we propose investigating characteristics such as the inclination of the ink deposit and specific features, such as the frequency of tremor, and comparing medieval and recent documents, both with pathological evidence visible in the write. Thus, we intend to advance the investigation into the characteristics present in medieval handwritten documents that help to identify the scribe's clinical condition.

Acknowledgements

This work was supported by the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES) (Scholarship no. 88882.375106/2019-01). We would like to acknowledge the great contributions of Dr. Deborah Thorpe in providing the dataset used in this work.

Competing Interests

The authors have no competing interests to declare.

Contributions

Authorial contributions

Authorship is alphabetical after the drafting author and principal technical lead. Author contributions, described using the CASRAI CREDIT typology, are as follows:

Francimaria Rayanne dos Santos Nascimento: FRSN

Stephen Smith: SS

Márjory Da Costa-Abreu: MCA

Authors are listed in descending order by significance of contribution. The corresponding author is Francimaria Rayanne dos Santos Nascimento

Conceptualization: FRSN

Methodology: FRSN, MCA

Software: FRSN

Project Administration: FRSN

Formal Analysis: FRSN

Investigation: FRSN

Visualization: FRSN

Data curation: FRSN, MCA

Writing – Original Draft Preparation: FRSN

Writing – Review and Editing: MCA, SS

Editorial contributions

Recommending Editor

Dr. Barbara Bordalejo, Assistant Professor of Digital Humanities, University of Lethbridge, Canada

Section Editor

Shahina Parvin, The Journal Incubator, University of Lethbridge, Canada

Copy and Layout Editor

Christa Avram, The Journal Incubator, University of Lethbridge, Canada

References

- Bar-Yosef, Itay, Alik Mokeichev, Klara Kedem, Itshak Dinstein, and Uri Ehrlich. 2009. "Adaptive Shape Prior for Recognition and Variational Segmentation of Degraded Historical Characters." *Pattern Recognition* 42(12): 3348–3354. DOI: <https://doi.org/10.1016/j.patcog.2008.10.005>
- Boddy, Richard, and Gordon Smith. 2009. *Statistical Methods in Practice: For Scientists and Technologists*. Chichester, UK: Wiley Online Library. DOI: <https://doi.org/10.1002/9780470749296.ch10>
- Bukhari, Syed Saqib, Thomas M. Breuel, Abedelkadir Asi, and Jihad El-Sana. 2012. "Layout Analysis for Arabic Historical Document Images Using Machine Learning." In *2012 International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Bari, Italy, 639–644. Los Alamitos, CA: IEEE. DOI: <https://doi.org/10.1109/ICFHR.2012.227>
- Chilimbi, Trishul, Yutaka Suzue, Johnson Apacible, and Karthik Kalyanaraman. 2014. "Project Adam: Building an Efficient and Scalable Deep Learning Training System." In *11th USENIX Symposium on Operating Systems Design and Implementation (OSDI '14)*, Broomfield, 571–582. Berkeley: USENIX Association. Accessed May 16, 2021. <https://www.usenix.org/conference/osdi14/technical-sessions/presentation/chilimbi>.
- Christlein, Vincent, David Bernecker, Florian Höning, Andreas Maier, and Elli Angelopoulou. 2017. "Writer Identification Using GMM Supervectors and Exemplar-SVMs." *Pattern Recognition* 63: 258–267. DOI: <https://doi.org/10.1016/j.patcog.2016.10.005>
- Cilia, Nicole D., Claudio De Stefano, Francesco Fontanella, Mario Molinara, Alessandra Scotto di Freca. 2020. "What Is the Minimum Training Data Size to Reliably Identify Writers in Medieval Manuscripts?" *Pattern Recognition Letters* 129: 198–204. DOI: <https://doi.org/10.1016/j.patrec.2019.11.030>
- Cloppet, Florence, Véronique Eglin, Dominique Stutzmann, and Nicole Vincent. 2016. "ICFHR2016 Competition on the Classification of Medieval Handwritings in Latin Script." In *15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Shenzhen, China, 590–595. Los Alamitos, CA: IEEE. DOI: <https://doi.org/10.1109/ICFHR.2016.0113>
- Cloppet, Florence, Veronique Eglin, Marlene Helias-Baron, Cuong Kieu, Nicole Vincent, and Dominique Stutzmann. 2017. "ICDAR2017 Competition on the Classification of Medieval Handwritings in Latin Script." In *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto, Japan, 1371–1376. Los Alamitos, CA: IEEE, 2017. DOI: <https://doi.org/10.1109/ICDAR.2017.224>
- Diem, Markus, and Robert Sablatnig. 2010a. "Are Characters Objects?" In *12th International Conference on Frontiers in Handwriting Recognition*, Kolkata, India, 565–570. Los Alamitos, CA: IEEE. DOI: <https://doi.org/10.1109/ICFHR.2010.93>
- . 2010b. "Recognizing Characters of Ancient Manuscripts." In *Proceedings SPIE 7531: Computer Vision and Image Analysis of Art, IS&T/SPIE Electronic Imaging*, San Jose, California, edited by David G. Stork, Jim Coddington, and Anna Bentkowska-Kafel, 753106. SPIE Digital Library. Accessed March 22, 2022. DOI: <https://doi.org/10.1117/12.843532>
- Gatos, Basilios, Ioannis Pratikakis, and Stavros J. Perantonis. 2006. "Adaptive Degraded Document Image Binarization." *Pattern Recognition* 39(3): 317–327. DOI: <https://doi.org/10.1016/j.patcog.2005.09.010>

- Gillespie, Alexandra, and Daniel Wakelin. 2011. *The Production of Books in England: 1350–1500*. Vol. 14. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511976193>
- Gilliam, Tara, Richard C. Wilson, and John A. Clark. 2010. "Scribe Identification in Medieval English Manuscripts." In *20th International Conference on Pattern Recognition*, Istanbul, Turkey, 1880–1883. Los Alamitos, CA: IEEE. DOI: <https://doi.org/10.1109/ICPR.2010.463>
- Gupta, Samta, and Susmita Ghosh Mazumdar. 2013. "Sobel Edge Detection Algorithm." *International Journal of Computer Science and Management Research* 2(2): 1578–1583.
- Haubenberger, Dietrich, Daniel Kalowitz, Fatta B. Nahab, Camilo Toro, Dominic Ippolito, David A. Luckenbaugh, Loretta Wittevrongel, and Mark Hallett. 2011. "Validation of Digital Spiral Analysis as Outcome Parameter for Clinical Trials in Essential Tremor." *Movement Disorders* 26(11): 2073–2080. DOI: <https://doi.org/10.1002/mds.23808>
- Haykin, Simon. 2007. *Redes Neurais: Princípios e Prática*. Brazil: Bookman Editora.
- Kamble, Parshuram M., and Ravinda S. Hegadi. 2015. "Handwritten Marathi Character Recognition Using R-HOG Feature." *Procedia Computer Science* 45: 266–274. DOI: <https://doi.org/10.1016/j.procs.2015.03.137>
- Keerthi, Sathiya S., Shirish K. Shevade, Chiranjib Bhattacharyya, and Mattur Ramabhadrasastry K. Murthy. 2001. "Improvements to Platt's SMO Algorithm for SVM Classifier Design." *Neural Computation* 13(3): 637–649. DOI: <https://doi.org/10.1162/089976601300014493>
- Kohavi, Ron. 1995. "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." In *IJCAI '95: Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Vol. 2, Montreal, Canada, 1137–1143. San Francisco: Morgan Kaufmann Publishers Inc. <https://dl.acm.org/doi/10.5555/1643031.1643047>.
- Liang, Yiqing, Richard M. Guest, and Michael Fairhurst. 2012. "Implementing Word Retrieval in Handwritten Documents Using a Small Dataset." In *2012 International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Bari, Italy, 728–733. Los Alamitos, CA: IEEE. DOI: <https://doi.org/10.1109/ICFHR.2012.220>
- Mamatha, H.R., and K. Srikantamurthy. 2012. "Morphological Operations and Projection Profiles Based Segmentation of Handwritten Kannada Document." *International Journal of Applied Information Systems (IJ AIS)* 4(5): 13–19. Accessed May 16, 2021. <https://research.ijais.org/volume4/number5/ijais12-450704.pdf>. DOI: <https://doi.org/10.5120/ijais12-450704>
- MathWorks. 2021. "Math. Graphics. Programming." Accessed October 11. <https://www.mathworks.com/products/matlab.html>.
- McLachlan, Geoffrey J., Kim-Anh Do, and Christophe Ambroise. 2005. *Analyzing Microarray Gene Expression Data*. Vol. 422. New York: John Wiley & Sons. DOI: <https://doi.org/10.1002/047172842X>
- Mitchell, Tom M. 2017. "Generative and Discriminative Classifiers: Naive Bayes and Logistic Regression." Accessed May 16, 2021. <http://www.cs.cmu.edu/%7Etom/mlbook/NBayesLogReg.pdf>.
- Otsu, Nobuyuki. 1979. "A Threshold Selection Method from Gray-Level Histograms." *IEEE Transactions on Systems, Man, and Cybernetics* 9(1): 62–66. DOI: <https://doi.org/10.1109/TSMC.1979.4310076>
- Platt, John C. 1998. "Fast Training of Support Vector Machines Using Sequential Minimal Optimization." In *Advances in Kernel Methods: Support Vector Learning*, edited by Bernhard Schölkopf,

- Christopher J. C. Burges, and Alexander J. Smola, 185–208. Cambridge, MA: MIT Press. <https://dl.acm.org/doi/10.5555/299094.299105>.
- Pradeep, Jayabala, E. Srinivasan, and S. Himavathi. 2011. "Diagonal Based Feature Extraction for Handwritten Character Recognition System Using Neural Network." In *2011 3rd International Conference on Electronics Computer Technology (ICECT)*, Kanyakumari, India, 364–368. Los Alamitos, CA: IEEE, 2011. DOI: <https://doi.org/10.1109/ICECTECH.2011.5941921>
- Romero, Verónica, Alejandro Héctor Toselli, and Enrique Vidal. 2012. *Multi-Modal Interactive Handwritten Text Transcription*. Singapore: World Scientific Publishing Company. DOI: <https://doi.org/10.1142/8394>
- Saleem, Sajid, Fabian Hollaus, Markus Diem, and Robert Sablatnig. 2014. "Recognizing Glagolitic Characters in Degraded Historical Documents." In *2014 14th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Heraklion, Greece, 771–776. Los Alamitos, CA: IEEE. DOI: <https://doi.org/10.1109/ICFHR.2014.135>
- Saleem, Sajid, Fabian Hollaus, and Robert Sablatnig. 2014. "Recognition of Degraded Ancient Characters Based on Dense SIFT." In *Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage*, Madrid, Spain, edited by Apostolos Antonacopoulos and Klaus U. Schulz, 15–20. ACM Digital Library. DOI: <https://doi.org/10.1145/2595188.2595201>
- Schiegg, Markus, and Deborah E. Thorpe. 2017. "Historical Analyses of Disordered Handwriting: Perspectives on Early 20th-Century Material from a German Psychiatric Hospital." *Written Communication* 34(1): 30–53. DOI: <https://doi.org/10.1177/0741088316681988>
- Sezgin, Mehmet, and Bülent Sankur. 2004. "Survey over Image Thresholding Techniques and Quantitative Performance Evaluation." *Journal of Electronic Imaging* 13(1): 146–166. DOI: <https://doi.org/10.1117/1.1631315>
- Stokes, Peter A. 2007. "Palaeography and Image-Processing: Some Solutions and Problems." *Digital Medievalist* 3. DOI: <https://doi.org/10.16995/dm.15>
- Su, Bolan, Shijian Lu, and Chew Lim Tan. 2010. "Binarization of Historical Document Images Using the Local Maximum and Minimum." In *DAS '10: Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, Boston, Massachusetts, edited by David Doermann, Venu Govindaraju, Daniel Lopresti, and Prem Natarajan, 159–166. New York: ACM. DOI: <https://doi.org/10.1145/1815330.1815351>
- Surinta, Olarik, Mahir F. Karaaba, Lambert R. B. Schomaker, and Marco A. Wiering. 2015. "Recognition of Handwritten Characters Using Local Gradient Feature Descriptors." *Engineering Applications of Artificial Intelligence* 45: 405–414. DOI: <https://doi.org/10.1016/j.engappai.2015.07.017>
- Taylor, John R. 1997. *Introduction to Error Analysis, the Study of Uncertainties in Physical Measurements*. New York: University Science Books.
- Thorpe, Deborah E. 2015. "Tracing Neurological Disorders in the Handwriting of Medieval Scribes: Using the Past to Inform the Future." *Journal of the Early Book Society for the Study of Manuscripts and Printing History* 18: 241–248.
- Thorpe, Deborah E., and Jane E. Alty. 2015. "What Type of Tremor Did the Medieval 'Tremulous Hand of Worcester' Have?" *Brain* 138(10): 3123–3127. DOI: <https://doi.org/10.1093/brain/awv232>

- Thorpe, Deborah E., Jane E. Alty, and Peter A. Kempster. 2019. "Health at the Writing Desk of John Ruskin: A Study of Handwriting and Illness." *Medical Humanities* 46(1): 31–45. DOI: <https://doi.org/10.1136/medhum-2018-011600>
- Toledo, Juan Ignacio, Sounak Dey, Alicia Fornés, and Josep Lladós. 2017. "Handwriting Recognition by Attribute Embedding and Recurrent Neural Networks." In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto, Japan, 1038–1043. Los Alamitos, CA: IEEE. DOI: <https://doi.org/10.1109/ICDAR.2017.172>
- Vapnik, Vladimir. 1998. *Statistical Learning Theory*. New York: Wiley.
- Wang, Wenchao, Jianshu Zhang, Jun Du, Zi-Rui Wang, and Yixing Zhu. 2018. "DenseRAN for Offline Handwritten Chinese Character Recognition." In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Niagara Falls, New York, 104–109. Los Alamitos, CA: IEEE. DOI: <https://doi.org/10.1109/ICFHR-2018.2018.00027>
- Weka. 2021a. "Weka 3: Machine Learning Software in Java." Accessed October 11. <https://www.cs.waikato.ac.nz/ml/weka/>.
- . 2021b. *Class NaiveBayes*. University of Waikato. Accessed May 19. <http://weka.sourceforge.net/doc.dev/weka/classifiers/bayes/NaiveBayes.html>.
- . 2021c. *Class SMO*. University of Waikato. Accessed May 18. <https://weka.sourceforge.io/doc.dev/weka/classifiers/functions/SMO.html>.
- . 2021d. *Class MultilayerPerceptron*. University of Waikato. Accessed March 22, 2022. <https://weka.sourceforge.io/doc.dev/weka/classifiers/functions/MultilayerPerceptron.html>.
- Witten, Ian H., Eibe Frank, Mark A. Hall, and Christopher J. Pal. 2016. *Data Mining: Practical Machine Learning Tools and Techniques*. 4th ed. San Francisco: Morgan Kaufmann Publishers Inc.
- Xiong, Wei, Jingjing Xu, Zijie Xiong, Juan Wang, and Min Liu. 2018. "Degraded Historical Document Image Binarization Using Local Features and Support Vector Machine (SVM)." *Optik* 164: 218–223. DOI: <https://doi.org/10.1016/j.ijleo.2018.02.072>
- Zamora-Martínez, Francisco, Volkmar Frinken, Salvador España-Boquera, María José Castro-Bleda, Andreas Fischer, and Host Bunke. 2014. "Neural Network Language Models for Off-line Handwriting Recognition." *Pattern Recognition* 47(4): 1642–1652. DOI: <https://doi.org/10.1016/j.patcog.2013.10.020>
- Zhi, Naiqian, Beverly Kris Jaeger, Andrew Gouldstone, Rifat Sipahi, and Samuel Frank. 2017. "Toward Monitoring Parkinson's Through Analysis of Static Handwriting Samples: A Quantitative Analytical Framework." *IEEE Journal of Biomedical and Health Informatics* 21(2): 488–495. DOI: <https://doi.org/10.1109/JBHI.2016.2518858>

