Graphic Compensation in Ancient Greek Documentary Hands: A Computational Paleographic Analysis from Handwritten Character Recognition

Historical scripts are structured visual systems whose internal organization has traditionally been explored through qualitative paleographic analysis. However, many of the principles shaping their evolution remain difficult to test empirically across corpora. One such principle, formulated by Jean Irigoin, is the hypothesis of graphic compensation, according to which graphic systems tend to maximize visual distinctiveness between letters that share similar phonetic functions while allowing greater visual proximity in the other cases. In this work, we introduce a computational approach for exploring the visual structure of historical scripts and apply it to the cursive Greek writing of Hellenistic papyri. Using representation learning on a corpus of character images, we analyze the classification behavior of neural networks in order to extract the similarity relations that structure the script. Rather than treating classification errors as poor technical results, we interpret them as paleographically meaningful patterns that reveal the internal organization of the graphic system. The resulting similarity matrix provides quantitative support for Irigoin's hypothesis and suggests that the principle of graphic compensation continues to operate even in the less controlled domain of cursive handwriting. More broadly, this study proposes a methodological framework for computational paleography, in which model behavior can be used to extract interpretable descriptions of script structure from large collections of cultural heritage material. Finally, we suggest that the learned relationship between phonetic difference and visual confusion may itself be used as paleographic background knowledge and incorporated in future HTR pipelines, helping recognition systems resolve visual ambiguity when visual evidence and linguistic expectations compete during decoding.

ACM Reference Format:
. 2026. Graphic Compensation in Ancient Greek Documentary Hands: A Computational Paleographic Analysis from Handwritten Character Recognition. 1, 1 (March 2026), 15 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

In the last three decades, a massive effort has been undertaken by institutions holding Cultural Heritage artifacts (usually referred to as GLAM for Galleries, Libraries, Archives and Museums) to make their collection accessible online. This is resulting in a still-expanding number of digital images available to researchers, which allows them to investigate on a large scale aspects such as material features that used to be beyond their reach in such proportions. Among Cultural Heritage items, ancient written artifacts and their study can benefit from many advances in Computer Science initially developed to study modern materials, especially in the field of Document Analysis and Recognition. Still, reading historical handwriting, understood as deciphering the meaning of an ancient message encoded in a given writing system, presents a significant challenge even when the fields of paleography and computer science are combined. In alphabetic scripts like ancient Greek, a key step is recognizing individual characters that form words. For centuries, human experts have noticed that the reading of ancient scripts is not merely a matter of optical recognition, but rather a complex cognitive task that is successful due to the balance between visual disambiguation and linguistic context (Paul Schubert

Author's Contact Information:

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.
Manuscript submitted to ACM

[18] spoke of "To and fro between the eye and the mind"). In paleographic studies, this dynamic was described by Jean Irigoin, who proposed the theory of graphic compensation to avoid character confusion, especially ambiguity between two vowels or two consonants [10, 11]. This paleographic theory was mainly built however upon the observation of highly uniform and standardized calligraphic hands in manuscripts, where professional and experienced copyists practiced letter structure design. A critical question that remains is whether and to what extent graphic compensation also appears in cursive scripts, such as those found in documentary papyri, where the writing is less controlled and readability less immediate.

To answer this, as well as to explore the basis and application of Irigoin's hypothesis, we turn to computational paleography, using representation learning as a window into the internal mechanics of historical graphic systems. We chose a dataset of ancient Greek cursive writing on papyri carefully annotated at the character level, and a deep learning framework developed in the scope of research on character temporal classification. It is based on a ResNet18 architecture enhanced with a novel lacunae-based fragmentation augmentation and a supervised contrastive loss with dynamically learned weights. In this study, we shift our focus from technical benchmarking to paleographic discussion. Our purpose is to present an original methodology for interpreting classification errors as a paleographically meaningful similarity signal which could be integrated into optical handwritten character recognition approaches. Thanks to supervised contrastive loss, our model learns representations based on visual similarity, and the resulting dynamic similarity matrix mirrors how visually similar two letters are perceived.

Moving beyond theory validation and expansion, this study proposes a new approach to the design of Character Recognition Systems when applied to cursive handwritten texts. We suggest that the similarity matrix extracted from our network should not be viewed as a map of errors of the model, but rather as a "paleographic handbook". By mapping the learned relationship between phonetic role difference and visual confusion, this knowledge can be integrated into future recognition systems. Rather than forcing models to make a prediction when faced with uncertainty in historical scripts, algorithms can use this "handbook" to guide the decoding process.

We summarize our contributions as the following key points:

- **A quantitative validation of paleographic theory**. We provide the first computational validation of Irigoin's hypothesis of graphic compensation; moreover, we prove that maximization of visual differences between phonetically similar letters applies not only to manuscript hands but also to documentary papyri.
- **A methodology for representing paleographic knowledge**. Using a representation learning framework, we extract a letter confusion matrix that represents the specific visual confusions in the graphic system of a defined group of papyri (in our case, from the Hellenistic period), which we call a "paleographic handbook. More broadly, we propose a general methodology through which optical handwritten character recognition can be used to extract and interpret paleographic patterns from handwritten cultural heritage material.
- **A novel approach to historical handwritten OCR**. We propose a practical bridge between paleography and computer science, highlighting how the extracted "handbook" can be integrated as model background knowledge.

2 Related Work

2.1 Readability and Graphic Compensation

To address the visual ambiguities introduced by the morphological evolution of handwriting and explain how scripts maintain legibility, Jean Irigoin [10, 11] proposed a theory to interpret the visual structure of handwritten Greek based on graphic compensation of phonetic similarity. By examining the so-called Greek book hands, i.e. highly calligraphic

and formalized handwriting styles such as the biblical majuscule script, he observed that letter structure follows a limited number of geometric shapes (circles, triangles, squares). This reduced geometric variety inevitably leads to visual similarities. However, Irigoin hypothesized that historical graphic systems compensated for this by ensuring maximal morphological opposition between letters that share the same phonetic function (e.g. vowels vs vowels), while permitting visual similarity between phonetically distinct letters (e.g. consonants vs vowels). Therefore, visual confusion occurs more often between a vowel and a consonant, and less between letters within the same phonological group. According to Irigoin, for instance, when Byzantine scribes transitioned from uppercase to lowercase systems, they introduced specific uppercase forms to the lowercase scripts in order to eliminate similar forms that constrained legibility and keep the maximal opposition of letters representing phonemes of the same order; for instance, in order to avoid confusion between lowercase Lambda and Nu (visually similar and transcribing sounds that are close), they would use uppercase Nu (N) and lowercase Lambda (λ), or lowercase Nu (ν) and uppercase Lambda (Λ). His theories ignore the writing movement or ductus, as criticized by Pratesi [16]. However, Irigoin focused on the readability of scripts and not on their production, thus this criticism is irrelevant to our discussion, as already noted by Canart [4] who confirmed Irigoin's findings. While these theories have been the basis for paleographic discussion, they have traditionally largely relied on book hands. Our research bridges this gap by computationally validating Irigoin's theory on cursive scripts of papyri.

2.2 Early Quantitative Paleography Studies

The effort to quantify and computationally analyze historical scripts has deep roots in the tradition of paleography, reflecting a long-standing transition from human-centered expertise to machine-aided methodologies. Early efforts to categorize scribal hands in an objective way relied mostly on manual measurements of character dimensions. For instance, Gilissen [9] adopted the calculation of height-width and angle of writing to distinguish closely resembling Latin medieval hands. However, the variability that human handwriting presents posed significant challenges. Ornato [13] showed that the naturality of human handwriting can often render geometric shape measurement statistically insufficient for scribe identification. To overcome such limitations, scholars began integrating vision systems. Sirat [20], for instance, supported the use of visual processing to analyze Hebrew scripts, laying the ground for digital image analysis in paleography.

2.3 Bridging traditional Paleography and Handwritten Character Recognition

The intersection of traditional paleographic theories and modern HTR technology is a crucial area in the digital humanities. Historically, HTR systems transitioned from Hidden Markov Models (HMMs) with n-gram language models to architectures using Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). The latest surveys, e.g. [1], [2], [8] show that while current deep learning pipelines achieve remarkable accuracy on modern handwriting, they struggle when shifting to historical documents. Recent research in handwritten text recognition increasingly relies on transformer-based architectures, replacing earlier CNN-RNN pipelines, with attention models capable of capturing long-range visual dependencies in handwritten text ([8]). In particular, models based on Vision Transformers and encoder-decoder architectures such as TrOCR have become dominant in the field, as they combine a vision transformer encoder with a language model decoder to generate OCR transcript from images ([12]). Recent work applying such architectures to historical manuscripts demonstrates that domain-specific preprocessing and augmentation strategies intended to mimic material deterioration and writing variation can effectively reduce character error rates ([3]). At the same time, studies comparing modern HTR systems, such as PyLaia, TrOCR-based models and the Transformer-based

Titan system used in Transkribus, show that recognition accuracy varies significantly and depends on a wide number of factors including script complexity, abbreviation system and variation in orthography that is typical in historical written sources ([17]). A second emerging direction involves multimodal large language models (MLLMs), which use vision and language and are increasingly evaluated against domain-specific architectures. These models are often used for recognition or for post-correction of OCR output, with recent experiments showing that LLM-assisted pipelines can significantly reduce CER in challenging historical scripts ([6]).

Within the broader field of document analysis, several recent studies have addressed the problem of character detection and recognition in Greek papyri. A notable benchmark is the ICDAR 2023 Competition on Detection and Recognition of Greek Letters on Papyri [19], which introduced datasets and evaluation methods for this task. The competition stimulated the development of architectures especially designed to handle the fragmentary nature of papyrological material.

Among the leading approaches, [22] proposed a pipeline combining YOLOv8 for character detection with transformer-based classification models, integrating DeiT and SimCLR representations to improve recognition accuracy on papyrus fragments. Similarly, [23] introduced the PapyTwin network architecture, designed specifically for detecting Greek letters in Greek papyri. These studies show that deep learning methods can achieve promising results in identifying individual characters even under difficult conditions, such as material degradation and script variation.

Character-level annotation has also played a crucial role in enabling such research. The Ancient Lives project developed large-scale crowd-sourced annotation of Greek papyri [25] by allowing volunteers without knowledge of Greek to tag recognizable letters within papyrus images. Subsequently, these annotations supported machine learning experiments to explore character recognition [21]. Recent research has utilized small character images, cliplets, for various analytical tasks. For instance, NeuroPapyri [5] employs representations at the character-level to retrieve visually similar papyri, while other pipelines [24] incorporate character recognition as an intermediate step in paleographic dating models.

Additional studies have the potential of character-level data beyond recognition tasks. In the virtual reconstruction of Papyrus Bodmer I [15], manually annotated characters were combined to reconstruct missing Homeric text preserving natural variation in letterforms. Similarly, character-based approaches have been applied to writer identification [14]. These studies show that character-level representations can be used effectively for the study of historical handwriting.

Despite these advances, most approaches treat systems as tools that require further improvement, and recognition errors as statistical results of models rather than paleographically meaningful patterns. Recent studies support the view that quantitative evaluation metrics alone obscure meaningful patterns in recognition errors ([8]). Confusion patterns often reflect paleographic phenomena such as graphic similarity, scribal abbreviation systems and orthography variation rather than noise ([17]). Recent work in explainable artificial intelligence for document analysis also explores interpretability methods to understand which visual features affect recognition decisions in HTR models. These approaches demonstrate that model predictions often rely on local visual features corresponding to specific letter strokes, which suggests that machine-learning models may implicitly learn structural information within historical scripts. Building on this perspective, the present study investigates whether systematic character confusions produced by an HTR system correspond to paleographically meaningful relationships between letterforms, specifically those described in Irigoin's analysis of Byzantine scribal practices.

3 Datasets and Methods

3.1 The Hellenistic Papyri Dataset

To train and evaluate our model, we relied on "Hell-Date" (EGRAPSA Hellenistic Dated Papyri Dataset), an extensive and carefully curated dataset that was recently published [7]. This dataset brings together 194 images from Greek papyri that are securely dated to a specific year within the Hellenistic period (310 BCE - 3 BCE) based on internal evidence. It is particularly valuable for digital paleography because each document has been annotated with rich metadata, including the date of writing, geographical provenance and nature of the textual content. Moreover, it focuses on fast, cursive handwriting, thus allowing us to shift the focus from so-called book hands to ligatured and less controlled scripts.

Among the 194 available images, 171 contain character-level annotations that identify individual letters. The annotations include 29 character categories. In addition to the 24 letters of the Greek alphabet, the dataset includes three numeral characters (*stigma*, *koppa*, and *sampi*) as well as a general *Symbol* category for non-alphabet characters. An additional class with the label *Unknown* exists for uncertain characters, but appears empty within the dataset. Each character is also annotated with a Base-Type label (BT) ranging from BT1 (best preserved) to BT5 (badly preserved) that describes the degree of preservation of the character shape.

In this study, we rely on this annotated portion but reorganize it to better fit the requirements of character-level recognition experiments. We refer to the resulting subset as "Hell-Char". Specifically, for each image, we sampled up to five random images from each character category to limit the dominance of highly frequent letters and reduce sampling bias across the papyri. All non-alphabetical categories (numerals, symbols, unknown) were ignored. Additionally, only characters labelled BT1 and BT2 (perfectly preserved letters and damaged but unambiguous letters) were retained, excluding severely deformed or damaged characters that cannot be reliably read out of context. After this preprocessing, the resulting dataset comprises 13,014 character images originating from 157 papyri. Table 1 shows the letter frequency distribution in Hell-Char, where one can easily notice that, despite the balancing procedure, and due to the natural frequency of letters in the Greek language, rare instances, such as Psi and Zeta, appear in less than half a percent of the instances.

The choice of dataset was suited to our research question. Hellenistic era cursive writing is a worthy stress-test for any legibility evaluation task. During the Hellenistic period, everyday writing evolved rapidly, moving away from the clear-shape letters of book hands, resulting in the creation of strongly connected shapes (ligatures). The visual boundaries between different letters become blurred, making identification challenging without the aid of linguistic context. Therefore, our dataset offers an ideal testing ground to computationally evaluate the degree of graphical variation that letterforms exhibit within the visual system of the script.

3.2 Network Architecture

As backbone architecture for feature extraction we deployed three models; ResNet18, ConvNeXt-V2 and ViT16S. All three produced highly similar results and led to consistent conclusions. For clarity, here we present the results of one representative model, ResNet18, a model that is widely accepted as a baseline in image analysis. ResNet architecture, utilizing residual connections, allows for the effective training of deep neural networks by preventing the problem of vanishing gradients. In our experiment, the model was initialized with pretrained weights (ImageNet) and adapted to classify the 24 classes of the Greek alphabet, functioning as a basic mechanism for character embedding production.

Although modern architectures based on attention-based models, such as Vision Transformers, are capable of identifying global dependencies and often achieve better classification results, a network such as ResNet18 proves to

Example Glyph	Letter	Samples
	Iota	704
	Alpha	696
	Tau	692
	Sigma	690
	Epsilon	688
	Omicron	680
	Nu	669
	Rho	666
	Upsilon	665
	Pi	636
	Kappa	636
	Mu	630
	Omega	628
	Eta	620
	Lambda	583
	Delta	565
	Gamma	527
	Theta	431
	Chi	427
	Phi	417
	Beta	333
	Xi	235
	Zeta	110
	Psi	86

Table 1. Letter frequency distribution in our Hell-Char subset. The glyphs are given as mere examples. They were binarized (the scale not being preserved) from P.Mich. III 173 (169 BCE), University of Michigan Library, inv. P. 4225 (TM8337).

be more suitable for this specific paleographic purpose, because its filters are by definition designed to focus on local

morphological features such as curves, strokes and line intersections. In paleography, the essence of letter shapes lies precisely in these local details. In contrast, attention models tend to overfit to broader patterns or to the background noise of the papyrus.

3.3 Lacuna-based Fragmentation (LF)

To improve model adaptation to Greek cursive handwriting from papyri, we do not resort to classic data augmentation methods. Instead, we adopt the method called Lacuna-based Fragmentation. This augmentation method works by computationally simulating realistic damage to the papyrus. During training, it randomly fragments parts of the letter image. This process enables the model not to memorize misleading features, but to learn more structural representations.

3.4 Dynamically-learned Supervised Contrastive Loss (DSCL)

The heart of the methodology applied lies in the dynamically-learned custom loss function. Traditional loss functions, such as Cross-Entropy Loss, punish each misclassification equally, as if all errors share the same weight. In historical cursive handwriting, however, confusion between an Alpha (Α) and a Lambda (Λ) is visually justified due to their structure, whereas this is not true in the case of confusion between an Alpha (Α) and a Chi (Χ). The basic supervised contrastive loss method attempts to bring together samples of the same class and push samples of different classes apart. However, this is not sufficient for capturing the subtle visual details of the Greek letters. This is why the adopted loss function is also enriched with dynamically learned weights. As it is trained, the network dynamically learns which character classes tend to overlap visually. Hence, the model is not severely penalized when it places morphologically similar letters close to each other in its representation space. The final result is the creation of a dynamic weight matrix that accurately reflects the visual ambiguity of the script, allowing us to measure how the graphic system organizes visual distinctions between letters and verify the theory of graphic compensation (Figure 3).

4 Results

4.1 Classification Performance

Table 2 shows per-letter classification performance of the methodology used, evaluated on our dataset. The network achieves an overall Accuracy of 0.83 across 2,603 instances. Despite the class imbalance, with strongly represented characters such as Iota (141 instances) and heavily underrepresented ones like Psi (17 instances), the macro-average and weighted-average F1-scores are both steady at 0.82 and 0.83 respectively. This indicates that the methodology successfully prevented the model from collapsing towards the majority class. The model also demonstrates strong results for several letter classes. For instance, classes such as Beta, Eta and Kappa achieve F1-scores of 0.92, while Omicron class reaches 0.90. On the other hand, the model struggles with specific letters of the alphabet. Alpha and Lambda are the most under-performing categories, both with an F1-score of 0.63. For Alpha, a very low Recall is observed (0.55), meaning the model misses frequently true Alpha instances; the high variation of Alpha-shapes within Hell-Char may explain this result (Figure 1). Zeta also underperforms (0.70 F1), though this can be partially attributed to its low support (22 instances).

Some letters present high divergence between Precision and Recall. Mid-performing Sigma (0.81 F1) is a good example of this. Sigma maintains a high Recall (0.90), but suffers from low Precision (0.74), which suggests that the model tends to predict the letter in case of visual ambiguity. The same is true for high-performing Nu (F1: 0.89; Recall: 0.97; Precision:

0.83). On the contrary, mid-performing Iota (F1: 0.79; Recall: 0.72; Precision: 0.89) presents a low Recall but a high Precision.

Table 2. The classification report on Hell-Char, per letter, of ResNet18 enhanced with our LF augmentation and our SCL with dynamically-learned weights.

Class	Precision	Recall	F1-Score	Support
Alpha	0.73	0.55	0.63	139
Beta	0.92	0.91	0.92	67
Chi	0.97	0.82	0.89	85
Delta	0.87	0.84	0.85	113
Epsilon	0.84	0.89	0.86	134
Eta	0.92	0.91	0.92	128
Gamma	0.83	0.75	0.79	105
Iota	0.89	0.72	0.79	141
Kappa	0.89	0.94	0.92	127
Lambda	0.63	0.64	0.63	136
Mu	0.83	0.87	0.85	126
Nu	0.83	0.97	0.89	134
Omega	0.84	0.82	0.83	123
Omicron	0.88	0.92	0.90	126
Phi	0.88	0.84	0.86	83
Pi	0.80	0.91	0.85	127
Psi	0.83	0.89	0.86	17
Rho	0.86	0.91	0.89	133
Sigma	0.74	0.90	0.81	138
Tau	0.73	0.85	0.79	139
Theta	0.73	0.79	0.76	86
Upsilon	0.82	0.73	0.77	133
Xi	0.76	0.83	0.80	47
Zeta	0.78	0.64	0.70	22
Accuracy			0.83	2603
Macro (avg)	0.84	0.82	0.82	2603
Weighted (avg)	0.83	0.83	0.83	2603

4.2 The Confusion Matrix

To understand the factors contributing to the model's performance decline on specific letters, a deeper analysis of the error distribution is required. As illustrated in the confusion matrix (Figure 2), the misclassifications made by the model are not distributed in a random way across the alphabet, but rather illuminate specific confusion pairs.

The most significant case of model failure is the Alpha and Lambda pair. The confusion here is asymmetrical; a true Alpha (A) is misclassified as Lambda (Λ) 40 times (28% of instances), whereas a true Lambda is misclassified as Alpha only 10 times (8% of instances). This shows that the Hellenistic era Alpha morphology significantly overlaps with the visual structure of Lambda, but the opposite is not true. As shown in Figure **??**, many Alpha forms in our dataset are reduced to two main strokes, often losing the horizontal bar or integrating it into a single motion. This transition creates a visual shape nearly identical to the Lambda shape. The inverse, a Lambda being perceived as an Alpha, is

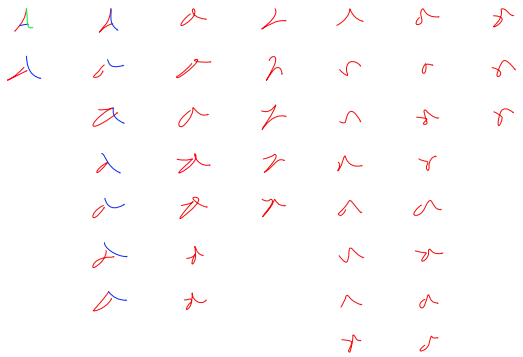

Fig. 1. Examples of Alpha forms in Hell-Char dataset. The colored segments represent the hypothetical reconstruction of the stroke sequence (red for first stroke, blue for second, green for third). The samples demonstrate a clear tendency toward structural simplification, with several forms being reduced to two main diagonal strokes.

not as often observed, as the former requires an additional structural element, a horizontal bar, that is absent from the Lambda structure.

Similar confusion patterns highlight other visually related pairs. The circular letters form a distinct cluster of visual ambiguity. In specific, true Theta instances are frequently misclassified as Omicron (9 times, or 10% of instances) and true Sigma instances are misclassified as Epsilon 11 times, or 8% of instances. It is worth mentioning that these confusions are not mutually dependent. While Sigma is confused with Epsilon, Epsilon is almost never confused with Sigma (only 1 instance). Other interesting confusion pairs include Upsilon being misclassified as Tau (21 times, or 15% of instances) and Iota being misread as Rho (8 times, or 5,5% of instances). Low support for letters also increases the chance of misclassification; for instance, Psi is confused with Phi 5 times, or 29% of all instances and Zeta is confused with Xi 5 times or 22% of all instances.

What is important to note here is that the system does not actually confuse these letters at random. For example, as already observed by Irigoin [10], Alpha (Α), Lambda (Λ) and Delta (Δ) all belong to the same geometric shape group: triangles. Our network's confusion within this group perfectly respects the phonetic boundaries of graphic compensation. While the vowel Alpha is frequently confused with the consonant Lambda (50 total misclassified instances in both directions) and less frequently with the consonant Delta (11 total misclassified instances), the two consonants of the group, Lambda and Delta, are only misclassified once (Lambda as a Delta). The same holds true for circular shapes: the two vowels Epsilon and Omicron are never confused and the two consonants Sigma and Theta are only confused 4 times in total. The system confuses letters that are both visually and phonetically similar only in case of low support, e.g. for Phi and Psi and Xi and Zeta.

This suggests a highly structured pattern in confusion and proves that the model captures script characteristics and morphological overlaps, inviting a paleographic interpretation of these results.

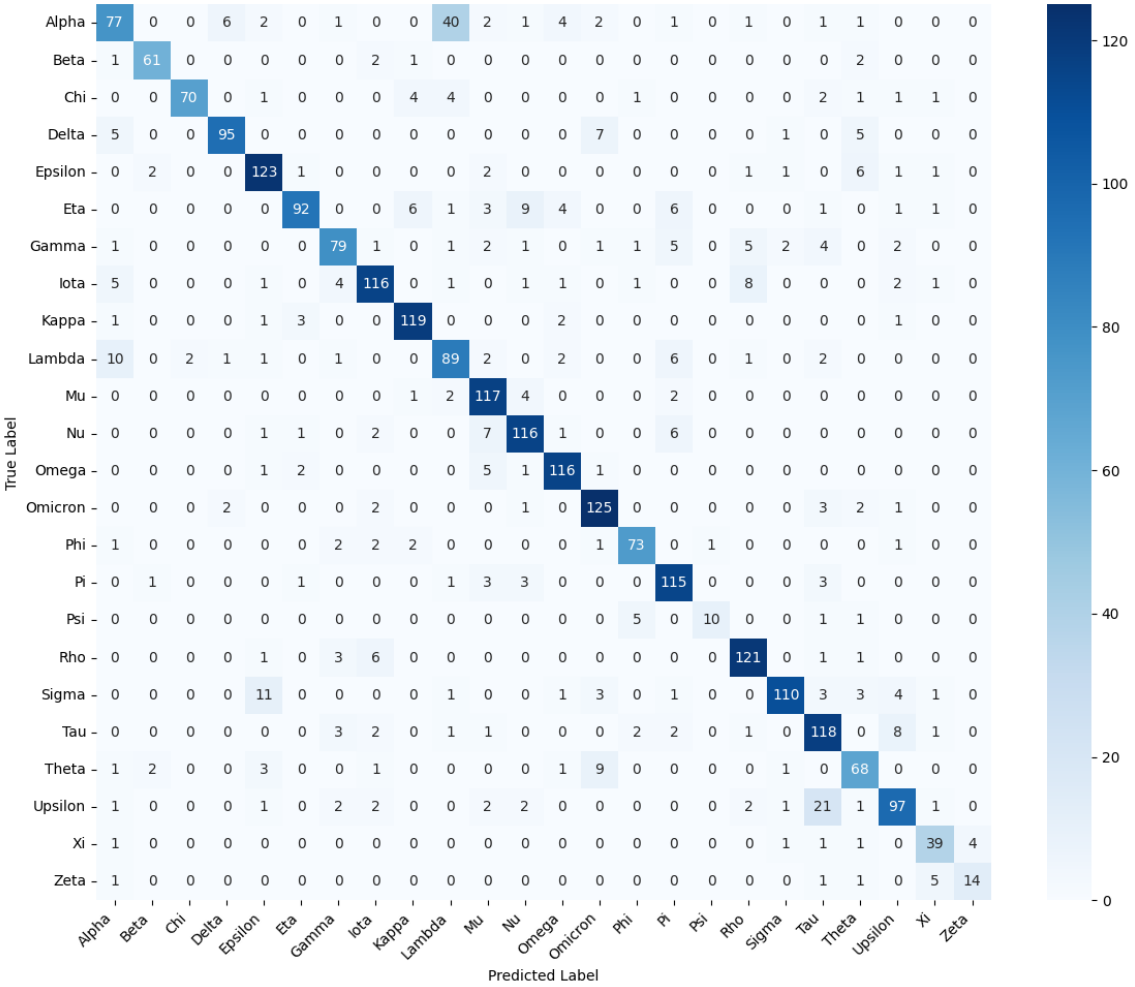

Fig. 2. Confusion matrix of ResNet18 with LF and SCL.

5 Discussion

5.1 Validating and Extending Irigoin's Hypothesis of Graphic Compensation: a New Methodology for Paleographic Analysis

The analysis of Hellenistic-era cursive handwriting style using deep neural networks provides us with a unique opportunity to observe the mechanisms of visually recognizing a character not through the eyes of a paleographer but through the "eyes" of a computational system that attempts to solve the problem of visual ambiguity from scratch. As Jean Irigoin has analyzed, in calligraphic scripts (e.g. biblical majuscule), the graphic system is in tension between the simplification of forms and the need for legibility. When letters are constructed from the same geometric structures, their visual similarity threatens the comprehension of the text. The solution provided by the graphic system itself,

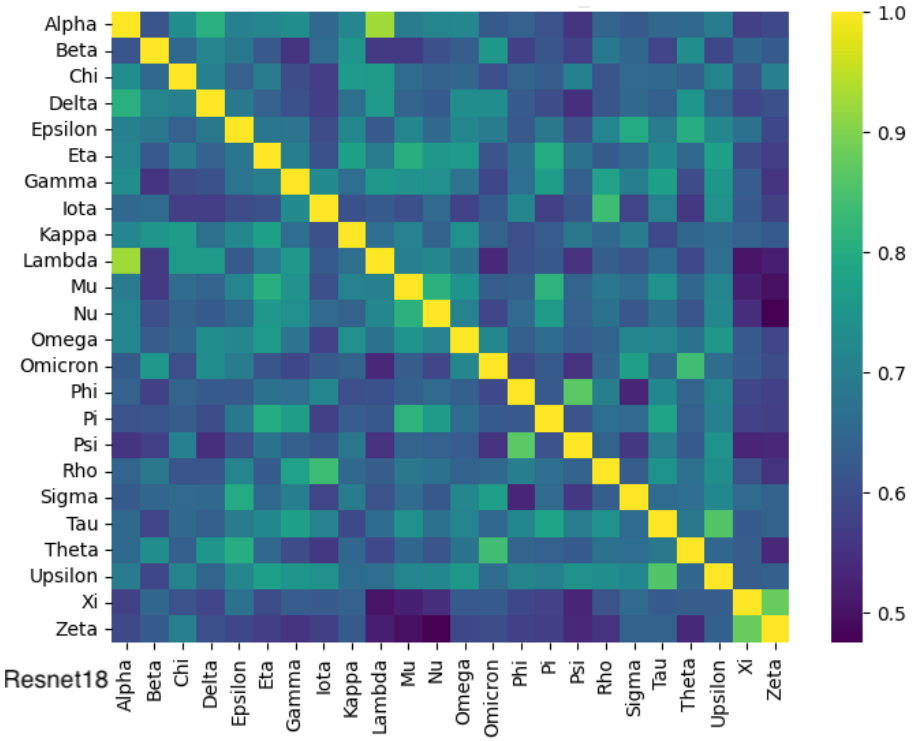

Fig. 3. Dynamic similarity matrix between letters learned during training. Light colours indicate high similarity, such as Alpha-Lambda, Theta-Omicron, Xi-Zeta, Phi-Psi.

according to Irigoin, is graphic compensation. In other words, *visual contrast must be maximized between letters that have the same phonetic function* (e.g. between two vowels), because that is where the greatest risk of confusion lies when reading a word. In contrast, visual confusion between a vowel and a consonant is acceptable because the reader can resolve the ambiguity through the phonological rules of the language.

The question posed by this study is whether this rule, which was proposed for the standardized context of book hands, also applies to the fluid cursive script of Hellenistic papyri, where the need for rapid writing simplifies letter shapes as much as possible. The results of our model provide computational confirmation for a number of letter pairs. Examining the Confusion Matrix (Figure 2), we observe that the model systematically fails in pairs of letters, where the tailored loss function allowed visual similarity (as indicated in the final extracted similarity matrix, Figure 3). Because the similarity matrix is learned during training and the confusion matrix is computed independently at inference time, their agreement simply confirms that letters identified as visually similar during training tend to be confused in the model's predictions. However, these pairs of letters are all phonetically distant (consonants VS vowels), and hence we can argue that their visual similarity was tolerated by ancient scribes and readers because their phonetic distance was obvious.

The most characteristic example is the intense confusion between Alpha (vowel) and Lambda (consonant). In the similarity matrix, this pair has one of the highest values (yellow indication with a value close to 0.9), which translates

into 40 false positives of Alpha as Lambda in the confusion matrix. We find similar patterns in other consonant-vowel combinations as well. Sigma (consonant) is confused with Epsilon (vowel) 11 times, and, in the similarity matrix, they have a strong light green indication. Theta (consonant) is misclassified as Omicron (vowel) 9 times, and Upsilon (vowel) is confused with Tau (consonant) 21 times. All these confusions in the confusion matrix arise among pairs of characters that were allowed to be confused during training, because of how the loss function and the related similarity matrix was constructed. These are also shapes that paleographers know can be similar.

Even more impressive is the behavior of the network when we analyze the relationships between letters of the same phonetic category. In the similarity matrix (Figure 3), the interactions between consonants (e.g., Delta with Lambda, Gamma with Delta, Kappa with Pi) appear in blue and purple shades with similarity values around 0.5; Pi with Mu is slightly higher, with a greenish shade around 0.75 that is still lower than the above-observed values. In practice, this means the network will rarely, if ever, confuse them, as reflected in the confusion matrix. The only exceptions to this tendency are the poorly supported consonant pairs, such as Xi and Zeta or Phi and Psi, which are formed with letters that are among the least frequent ones of the Ancient Greek language (less than 1 percent). This shows that the graphic system kept the consonant structure visually clear and distinct, just as paleographic theory predicted.

Computational analysis reveals an interesting directionality in the visual structure of letter forms, which often escapes naked observation. In the Alpha-Lambda pair, we observe a clear asymmetry in the confusion matrix (Figure 2). While Alpha is misclassified as Lambda 40 times, Lambda is confused with Alpha only 10 times. This imbalance is not merely a statistical error, but reflects the paleographic reality. Due to the speed of cursive writing, the horizontal element of Alpha is merged with the oblique left part, which drives it to visually degenerate into the structural space of Lambda (Figure 1). Lambda, on the contrary, retained its structural consistency. The network functioning as a "computational paleographer" managed to capture not only the morphological similarity but also the direction of this morphological change.

In cursive handwriting, the speed of the writing tool naturally forces distinct letter shapes to partially merge, increasing visual similarity between different characters. However, our extracted similarity matrix shows that although cursive handwriting tends to increase visual similarity among letter forms, this does not result in a complete visual collapse. Letters belonging to similar phonetic categories remain visually distinct in the model's representation space and maintain a safe visual distance from each other. What our model suggests is that graphic compensation operates not as a static geometric rule, but as a dynamic, adaptive mechanism that keeps the script legible despite the variability introduced by rapid handwriting. From this perspective, if readability must be preserved, rapid handwriting cannot vary arbitrarily but must either avoid potentially ambiguous transformations or introduce new strategies of visual differentiation. Irigoin's hypothesis of graphic compensation is thus not only validated but also extended from meticulous, geometric calligraphic writing to rapid, cursive scripts. Through this extension, we provide a methodology for interpreting classification errors as a paleographically meaningful similarity signal giving us access to the visual structure of a script. This methodology could be extended beyond Greek handwriting.

5.2 Paleography-informed Character Recognition for Historical Scripts

The quantitative validation of Irigoin's graphic compensation theory has important implications for the future design of Character recognition systems applied to historical documents. Current language model decoders rely on general statistical knowledge and have no principled way to decide how much to trust the visual signal versus the linguistic signal when uncertainty is high. Our similarity matrix could provide exactly that calibration, not by adding linguistic context, but by making the weighting between visual and linguistic evidence script-specific and paleography-informed.

We propose that the final extracted similarity matrix from our representation learning framework should not be treated merely as a paleographic analysis, but as a paleographic "handbook," which can be integrated into future HTR model architectures as an independent weighting parameter.

In practice, this can be analyzed as a paleography-conditioned decoding mechanism. During the (sequence) prediction stage, when the model generates a highly uncertain probability distribution between two characters, the algorithm can query the paleographic "handbook" matrix. If the matrix indicates that the specific pair is a historically expected one and originated from graphic compensation visual ambiguity, the decoding algorithm will decrease the weight of the vision model and bias the decision in favor of the language model. The language model, which follows the phonological rules of the ancient Greek language, will naturally select the character that fits the sequence of surrounding characters. On the other hand, if the vision model is uncertain between two characters that the matrix dictates cannot be confused (e.g., Lambda and Delta), the algorithm treats this as a system error, preventing the language model from hallucinating a fix that violates the phonetic rules of the graphic system. By calculating matrices tailored to specific languages and scripts (e.g., cursive Hellenistic Greek vs. minuscule Medieval Greek), such an architecture can be implemented across all types of graphic environments.

While implementing such a novel decoding architecture falls beyond the scope of this study, the extraction and validation of the confusion matrix satisfy the essential requirements for its realization. By translating Irigoin's paleographic theory into a computational representation, we provide a way for Character Recognition systems to incorporate script-specific paleographic knowledge. Integration of this paleographic perspective allows recognition algorithms to address the visual ambiguity in historical handwriting in a way that better reflects how ancient readers perceived this script.

6 Conclusion

This study successfully bridges the gap between traditional paleography research and machine learning by demonstrating that the morphological evolution of historical handwriting is regulated by phonological rules. By applying representation learning to the Hellenistic era cursive script of papyri, we presented a quantitative validation of Jean Irigoin's hypothesis of graphic compensation and extend it to informal scripts. Our results show that the historical graphic system tolerated visual confusion between characters with different phonetic functions, while maintaining maximal distance between characters belonging to the same phonetic category. This finding suggests that visual organization of historical scripts reflects deeper linguistic constraints that shape the evolution of writing systems.

Moving beyond this theoretical validation, our work contributes a methodological framework for computational paleography. By extracting the learned similarity matrix into a paleographic "handbook", we provide a methodology to observe the visual structure of a script. We could interpret classification errors as a paleographically meaningful similarity signal giving us access to the complex relationship between legibility, shape simplification and phonetic structure of a language. This methodology could be fruitfully applied to other languages and scripts, to offer interpretable descriptions of the graphic system and can support the analysis of visual variation in large corpora of handwritten cultural heritage material.

Finally, this methodology could also be integrated in future HTR systems to strengthen their character recognition pipeline. The similarity matrices calculated could work as "handbooks" to condition the visual and linguistic signals during the decoding stage in cases of uncertainty in recognition.

The implications of this approach extend beyond the specific case of Hellenistic papyri. The proposed framework enables researchers to computationally investigate questions that have traditionally been addressed qualitatively in

paleography, such as the evolution of letterforms across long chronological spans, the identification of scribal habits or the comparison of graphic systems across different corpora. The methodology proposed here though may find applications beyond handwriting analysis. Any domain where visual forms encode historically structured conventions, such as art and textual heritage, may benefit from similar representation learning approaches that aim to model the internal organization of visual systems. By combining machine learning with the interpretative frameworks of the humanities, this work contributes to the emerging field of computational paleography and demonstrates how artificial intelligence can serve not only as a tool for optical recognition, but also as a method of generating new knowledge.

References

[1] Aguilar, S.T., Jolivet, V.: Handwritten text recognition for documentary medieval manuscripts. Journal of Data Mining and Digital Humanities (2023). https://doi.org/10.46298/jdmdh.10484

[2] AlKendi, W., Gechter, F., Heyberger, L., Guyeux, C.: Advancements and challenges in handwritten text recognition: A comprehensive survey. Journal of Imaging **10**(1), 18 (2024). https://doi.org/10.3390/jimaging10010018

[3] Barrere, K., Soullard, Y., Lemaitre, A., Coüasnon, B.: Training transformer architectures on few annotated data: An application to historical handwritten text recognition. International Journal on Document Analysis and Recognition (IJDAR) **27**(4), 553–566 (2024). https://doi.org/10.1007/s10032-023-00459-2

[4] Canart, P.: La paléographie est-elle un art ou une science? Scriptorium **60**, 159–185 (2006). https://doi.org/10.3406/scrip.2006.3939

[5] De Gregorio, G., Perrin, S., Pena, R.C.G., Marthot-Santaniello, I., Mouchère, H.: Neuropapyri: A deep attention embedding network for handwritten papyri retrieval. In: Mouchère, H., Zhu, A. (eds.) Document Analysis and Recognition – ICDAR 2024 Workshops. Lecture Notes in Computer Science, vol. 14936, pp. 63–78. Springer (2024). https://doi.org/10.1007/978-3-031-70642-4$_5$

[6] Evaggelatos, A., Palaiologos, K., Gatos, B., Kaddas, P., Christopoulou, A., Katsouros, V.: Old greek ocr result correction using llms. In: Proceedings of the 2025 ACM Symposium on Document Engineering (DocEng '25). pp. 1–4. Association for Computing Machinery (2025). https://doi.org/10.1145/3704268.3748676

[7] Ferretti, L., Serbaeva Saraogi, O., De Gregorio, G., Marthot-Santaniello, I.: Hell-Date. EGRAPSA Hellenistic Dated Papyri Dataset (2025). https://doi.org/10.5281/zenodo.15083590, https://zenodo.org/records/15083590

[8] Garrido-Muñoz, C., Rios-Vila, A., Calvo-Zaragoza, J.: Handwritten text recognition: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (2026). https://doi.org/10.1109/TPAMI.2025.3646002

[9] Gilissen, L.: L'expertise des écritures médiévales : recherche d'une méthode avec application à un manuscrit du XIe siècle : le lectionnaire de Lobbes, Coddex Bruxellensis 18018. Story-Scientia, Gand (1973)

[10] Irigoin, J.: Structure et évolution des écritures livresques de l'époque byzantine. In: Wirth, P. (ed.) Polychronion. Festschrift Franz Dölger zum 75. Geburtstag, pp. 253–265. C. Winter, Heidelberg (1966)

[11] Irigoin, J.: L'alphabet grec et son geste des origines au IXe siècle après J.-C. In: Sirat, C., Irigoin, J., Poulle, E. (eds.) L'écriture: le cerveau, l'œil et la main, pp. 299–305. Brepols, Turnhout (1990)

[12] Li, Y., Chen, D., Tang, T., Shen, X.: Htr-vt: Handwritten text recognition with vision transformer. Pattern Recognition **158**, 110967 (2025). https://doi.org/https://doi.org/10.1016/j.patcog.2024.110967

[13] Ornato, E.: Statistique et paléographie: peut-on utiliser le rapport modulaire dans l'expertise des écritures médiévales? Scriptorium **29**, 198–234 (1975). https://doi.org/10.3406/scrip.1975.1111

[14] Peer, M., Sablatnig, R., Serbaeva, O., Marthot-Santaniello, I.: Kairacters: Character-level-based writer retrieval for greek papyri. In: Antonacopoulos, A., Chaudhuri, S., Chellappa, R., Liu, C.L., Bhattacharya, S., Pal, U. (eds.) Pattern Recognition – ICPR 2024. Lecture Notes in Computer Science, vol. 15319, pp. 65–79. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-78495-8$_5$

[15] Perrin, S., Cudilla, L., Xie, Y., Mouchère, H., Marthot-Santaniello, I.: Homer restored: Virtual reconstruction of papyrus bodmer 1. In: Proceedings of the 7th International Workshop on Historical Document Imaging and Processing. p. 37–42. HIP '23, Association for Computing Machinery (2023). https://doi.org/10.1145/3604951.3605518

[16] Pratesi, A.: Paleografia greca e paleografia latina o paleografia greco-latina. In: Musca, G. (ed.) Studi storici in onore di Gabriele Pepe, pp. 161–172. Dedalo, Bari (1969)

[17] Romein, C.A., Rabus, A., Leifert, G., Ströbel, P.B.: Assessing advanced handwritten text recognition engines for digitizing historical documents. International Journal of Digital Humanities **7**(1), 115–134 (2025). https://doi.org/10.1007/s42803-025-00100-0

[18] Schubert, P.: Editing a Papyrus. In: Bagnall, R.S. (ed.) The Oxford Handbook of Papyrology, pp. 197–215. Oxford University Press, Oxford, New York, NY (2009)

[19] Seuret, M., Marthot-Santaniello, I., White, S.A., Serbaeva Saraogi, O., Agolli, S., Carrière, G., Rodriguez-Salas, D., Christlein, V.: ICDAR 2023 Competition on Detection and Recognition of Greek Letters on Papyri. In: Fink, G.A., Jain, R., Kise, K., Zanibbi, R. (eds.) Document Analysis and Recognition - ICDAR 2023. pp. 498–507. Springer Nature Switzerland (2023). https://doi.org/10.1007/978-3-031-41679-8$_2$9

[20] Sirat, C.: L'examen des écritures, l'oeil et la machine. Essai de méthodologie. Éditions du CNRS, Paris (1981)

[21] Swindall, M.I., Croisdale, G., Hunter, C.C., Keener, B., Williams, A.C., Brusuelas, J.H., Krevans, N., Sellew, M., Fortson, L., Wallin, J.F.: Exploring learning approaches for ancient greek character recognition with citizen science data. In: 2021 IEEE 17th International Conference on eScience (eScience). pp. 128–137 (2021). https://doi.org/10.1109/eScience51609.2021.00023

[22] Turnbull, R., Mannix, E.: Detecting and recognizing characters in greek papyri with yolov8, deit and simclr. International Journal on Document Analysis and Recognition (IJDAR) **28**, 277–285 (2025). https://doi.org/10.1007/s10032-024-00504-8

[23] Vu, M.T., Beurton-Aimar, M.: Papytwin net: a twin network for greek letters detection on ancient papyri. In: Proceedings of the 7th International Workshop on Historical Document Imaging and Processing. p. 43–48. HIP '23, Association for Computing Machinery (2023). https://doi.org/10.1145/3604951.3605522

[24] West, G., Swindall, M.I., Brusuelas, J.H., Maltomini, F., Gerhardt, M., D'Angelo, M., Wallin, J.F.: A deep learning pipeline for the palaeographical dating of ancient greek papyrus fragments. In: Proceedings of the 1st Workshop on Machine Learning for Ancient Languages (ML4AL 2024). pp. 177–185. Association for Computational Linguistics (2024). https://doi.org/10.18653/v1/2024.ml4al-1.18

[25] Williams, A.C., Wallin, J.F., Yu, H., Perale, M., Carroll, H.D., Lamblin, A.F., Fortson, L., Obbink, D., Lintott, C.J., Brusuelas, J.H.: A computational pipeline for crowdsourced transcriptions of ancient greek papyrus fragments. In: 2014 IEEE International Conference on Big Data (Big Data). pp. 100–105 (2014). https://doi.org/10.1109/BigData.2014.7004460

From: **ACM Journal on Computing and Cultural Heritage**
<onbehalfof@manuscriptcentral.com>
Date: Tue, 12 May 2026 at 11:10 PM
Subject: Journal on Computing and Cultural Heritage - Decision on Manuscript ID JOCCH-26-0237
To: <platanou@aueb.gr>, <lavinia.ferretti@unito.it>, <i.marthot-santaniello@unibas.ch>, <gdegregorio@cvc.uab.cat>, <spirosbarbakos7@gmail.com>, <konstantinidou.maria5@gmail.com>, <asiminapaparrigopoulou@gmail.com>, <annis.pavlo@gmail.com>

12-May-2026

Dear Author(s):

Manuscript ID JOCCH-26-0237 entitled "Graphic Compensation in Ancient Greek Documentary Hands: A Computational Paleographic Analysis from Handwritten Character Recognition" which you submitted to the Journal on Computing and Cultural Heritage, has been reviewed. The comments of the reviewer(s) are included at the bottom of this letter.

The reviewer(s) have recommended publication, but also suggest some major revisions to your manuscript. Therefore, I invite you to respond to the reviewer(s)' comments and revise your manuscript.

Because we are trying to facilitate timely publication of manuscripts submitted to the Journal on Computing and Cultural Heritage, your revised manuscript should be uploaded within 45 days, by the 26-Jun-2026. If it is not possible for you to submit your revision in a reasonable amount of time, we may have to consider your paper as a new submission.

To revise your manuscript, log into https://mc.manuscriptcentral.com/jocch and enter your Author Center, where you will find the option to submit a revision. Please revise your manuscript by tracking change, by using bold or coloured text either in Word or LaTex.

PLEASE NOTE: it is essential that you address the comments and improvements requested by the reviewers. These are meant to ensure the quality of your paper and its contribution.

When submitting your revised manuscript, you will be able to respond to the comments made by the reviewer(s) in the space provided. It is important that you use this space to highlight how the changes to the manuscript address all of the reviewer's comments. Please be as specific as possible in your response to the reviewer(s). We will not be able to consider your paper if the reviewers' comments are not properly addressed and proceed to reject it.

IMPORTANT: Your original files are available to you when you upload your revised manuscript so that reviewers can compare both version. Please delete any redundant files before completing the submission.

Once again, thank you for submitting your manuscript to the Journal on Computing and Cultural Heritage and I look forward to receiving your revision.

Sincerely,
Dr. Karina Rodriguez Echavarria
Editor in Chief, Journal on Computing and Cultural Heritage
K.Rodriguez@brighton.ac.uk

Editor(s)' comments to Author:

Comments to the Author:
Dear Authors,

Thank you for submitting your manuscript, JOCCH-26-0237: Graphic Compensation in Ancient Greek Documentary Hands: A Computational Paleographic Analysis from Handwritten Character Recognition , to the ACM Journal on Computing and Cultural Heritage (JOCCH) Special Issue on Visual Heritage 2026.

While the review committee sees potential in your submission, the reviewers have raised substantial concerns. A major revision is required to increase the level of the paper to meet our journal quality standards.

To help you make an informed decision and guide your potential revisions, the detailed reviewer feedback is attached below to this message. In addition, to provide further context for this outcome, please find here a summary of the key reviewers' feedback and the rationale behind our final decision.

Summary of Reviews and Final Decision Rationale:
—----------
The reviewers unanimously agree that this paper presents a valuable and promising interdisciplinary contribution to the field of computational paleography. The core idea, framing HTR model confusions as structured paleographic evidence and extending Irigoin's theory of graphic compensation to Hellenistic documentary papyri, is recognized as highly innovative. Reviewers commended the sound methodology, the clear presentation of the classification results, and the introduction of a quantitative framework for analyzing visual similarity among historical letterforms. This work opens an interesting new direction for historical HTR systems.

While the conceptual foundation is strong, the reviewers identified several critical issues regarding technical validation and presentation. A primary concern is that the current manuscript reads more like a methodological proposal than a fully validated technical contribution (R3). Both R1 and R3 noted that the paper's claims are significantly overstated and must be calibrated to match the empirical evidence.

Additionally, the experimental section is currently incomplete. Crucial elements such as an ablation study and the results for the mentioned ConvNeXt-V2 and ViT16S models are missing (R1). Reviewers also pointed out a lack of detail regarding the training process and data splitting, which hinders reproducibility. Finally, the manuscript requires a more thorough discussion of its technical limitations and careful proofreading to fix missing references and formatting inconsistencies.

Mandatory Modifications
To be considered for acceptance after revision, the authors must thoroughly address the following mandatory points:
- Calibrate Claims and Improve Presentation: Tone down overstated claims to strictly align with the provided evidence. Clarify ambiguous sections, remove redundancies, and shorten the manuscript where appropriate, as suggested by R3.
- Expand Experimental Validation: Include a comprehensive ablation study (R1, R3). You must also report the missing results from the ConvNeXt-V2 and ViT16S models (R1) and include a global statistical test over all letter pairs (R3).
- Ensure Reproducibility: Explicitly detail the training process (R1) and clearly report a leakage-safe split protocol (R3). Discuss your specific technical choices in greater depth.
- Discuss Limitations: Include a dedicated discussion on the limitations of the proposed methodology and framework (R1).
- Fix References and Formatting: Incorporate all missing references pointed out by the reviewers (R1, R3) and correct formatting inconsistencies, particularly regarding DOI suffixes (R3). Ensure the manuscript undergoes moderate copy editing to improve overall readability.
- Address Minor Comments: Review and address all minor comments provided in the individual reviewer reports (R1, R3).
—----------

As you consider whether to pursue this revision, please be aware of the three possible publication paths moving forward:
- Full Acceptance: You address the reviewers' changes properly within the timeframe, and the paper is accepted to both the JOCCH special issue and the GCH conference.
- Invitation to GCH: If, at the end of the second review cycle (July 21st, 2026), the revisions do not meet the reviewers' requests and the strict journal-level requirements, but the editors believe

the revised paper is suitable for the conference, we will invite you to prepare a version specifically for GCH.
- Direct Submission to GCH: If you evaluate the requested changes and decide you cannot realistically address them by the journal's revision deadline, you may opt to submit a version of your work directly to GCH by the June 1st, 2026 deadline. Please note that if you choose this path, the paper will undergo a new, separate review process for the workshop.

If you choose to proceed, please submit your revised manuscript and a comprehensive response letter detailing your changes by June 21st, 2026.

Thank you for sharing your work with us, and we hope you find the reviewers' feedback constructive.

Sincerely,

Ruggero Pintus

Guest Editors, ACM JOCCH Visual Heritage 2026