Explaining the Chronological Attribution of Greek Papyri Images

John Pavlopoulos[1,2]*, Maria Konstantinidou[3], Georgios Vardakas[4], Isabelle Marthot-Santaniello[5], Elpida Perdiki[3], Dimitris Koutsianos[1], Aristidis Likas[4], and Holger Essler[6]

[1] Athens University of Economics and Business, Greece
[2] Stockholm University, Sweden ioannis@dsv.su.se
[3] Democritus University of Thrace, Greece {mkonst,eperdiki}@helit.duth.gr
[4] University of Ioannina, Greece g.vardakas@uoi.gr, arly@cs.uoi.gr
[5] University of Basel, Switzerland i.marthot-santaniello@unibas.ch
[6] Ca'Foscari University of Venice, Italy holger.essler@unive.it

Abstract. Greek literary papyri, which are unique witnesses of antique literature, do not usually bear a date. They are thus currently dated based on palaeographical methods, with broad approximations which often span more than a century. We created a dataset of 242 images of papyri written in "bookhand" scripts whose date can be securely assigned, and we used it to train machine and deep learning algorithms for the task of dating, showing its challenging nature. To address the data scarcity problem, we extended our dataset by segmenting each image to the respective text lines. By using the line-based version of our dataset, we trained a Convolutional Neural Network, equipped with a fragmentation-based augmentation strategy, and we achieved a mean absolute error of 54 years. The results improve further when the task is cast as a multiclass classification problem, predicting the century. Using our network, we computed and provided precise date estimations for papyri whose date is disputed or vaguely defined and we undertake an explainability-based analysis to facilitate future attribution.

Keywords: Chronology Attribution · Computer Vision · Greek Papyri.

1 Introduction

No autographs of classical Greek authors survive today. Our knowledge of such works (along with post-classical literature and the first Christian works including the New Testament) relies on manuscripts postdating the original compositions.

* Corresponding author. This version of the contribution has been accepted for publication, after peer review but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: TBD. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms

Of these, the most chronologically proximal are a few thousand papyri excavated mainly in Egypt in the last two centuries. Due to physical damage, these papyri usually preserve only small portions of the texts in question unlike medieval manuscripts which tend to transmit them in full-length, but both papyri and manuscripts represent copies of copies of the original works.

1.1 Background

Despite their fragmentary nature, papyri are crucial witnesses for innumerable texts, not to mention that they occasionally preserve literary works that would be otherwise lost. They are also invaluable evidence for our understanding of book culture in Antiquity, as well as for philology, the evolution of writing scripts and book production. One of the most important aspects of such research is to determine the date of the papyri involved.

Unlike their documentary counterparts (i.e. papyri preserving official and everyday documents), literary papyri bear no date before the introduction of colophons in the Middle Ages (9th century CE). We customarily employ palaeographical methods to assign an approximate and broad (often spanning more than a century) date for their production. Apart from their content, the two categories, documentary and literary papyri, are also usually written in distinctly different scripts: unformal cursive writings for the former opposed to elegant bookhand for the latter. There are some exceptions on both sides, i.e. literary texts written in cursive and documentary texts with surprisingly elegant scripts. To this day, we lack an exhaustive list of the first category (literary texts in cursive script) which does not allow us to use the numerous dated documents to date these literary papyri by script comparison. However, a few specimen of the second group (documentary texts written in bookhand) have been collected in the CDDGB (see below). Palaeographers rely on the evidence-backed assumption that handwriting styles are typical of certain periods and change over time, much like fashions and trends in anything else. The subjectivity and authoritativeness of these methods are increasingly acknowledged among scholars [12,3,16,15] and further assistance for more reliable and/or accurate ones is highly desirable.

In traditional dating, papyrologists employ comparative dating. They use the—admittedly very few—objectively dateable papyri specimens to draw comparisons with non-dated ones and estimate the latter's place on a notional timeline. The comparison is performed on the basis of the form and features of single letters, or the script overall, also used for other palaeographical tasks such as identifying scribes, or classifying styles and types of scripts. The characteristics used for such studies may focus on size (small/large, short/long), shape (round/angular), specific parts of letters (arches/loops/serifs/decorations), speed of writing, *ductus* (the number, directions and sequence of strokes required to draw a letter), formality etc. Although the same features are regularly invoked by many palaeographers, each researcher is free to focus (and they often do so) on every conceivable aspect of the writing. Hence, there is no formally established methodology, set of features to be taken into account, or even terminology that managed to reach consensus.[22,23] Even for the commonly used and agreed upon features, it is rarely possible for scholars to measure them or objectively calculate their significance towards a conclusion. Research in digital palaeography quantifying script features such as angle and direction of writing (for instance [1,2]) usually provides one such feature as the base for performing computationally palaeographical tasks. In our study, we aim at performing such a task (in this case dating) without any input in the form of human-perceived features. Instead, we attempt to identify any clues or features that lead our models to a specific date for a papyrus image.

The computer can pinpoint areas of the images which push predictions towards either extreme and/or alter these images (and predict the corresponding date) in a controlled manner. Nevertheless, it cannot provide explanations in real-life terms, nor identify features perceivable by humans. At the same time, human experts instinctively date scripts in terms of certain characteristics, however subjective, but are unable to measure each such feature's significance towards assigning a date. In this preliminary examination, our aim is to detect patterns (not necessarily semantically clear at this stage) in the application of saliency maps.

1.2 The contributions of this work

- C1. We developed two datasets of images of Greek papyri from Egypt, along with the dates assigned to them by experts: one with whole papyri fragments; the other with lines of writing extracted from the full-size images.
- C2. We proposed a Convolutional Neural Network (CNN), which we call fCNN, that is based on a fragmentation-based augmentation strategy and which predicts the date of text-line images with a mean absolute error of 54 years, using a regression head, and a macro-average F1 of 61.5%, using a classification head, setting the state of the art for Greek papyri image dating.
- C3. We used fCNN to precise the dating of the lines of eleven papyri, whose previous dates based on objective criteria are ranging across two centuries, and we share our predictions: https://github.com/ipavlopoulos/palit

2 Related work

Although researchers have suggested algorithms for the automated segmentation of papyri images to text-lines [19], and although the benefits of text-line segmentation are already known in the field of writer identification [4], no published work to date has investigated dating computationally Greek literary papyri by focusing on text line images. The baseline is set by a CNN that is fed with whole Greek literary papyri images, which achieved a mean absolute error of more than a century [18]. Our study shows that data segmentation to text lines leads to a much smaller error, with augmentation-enhanced CNNs providing the best-performing solution. In the absence of other related work for Greek papyri image dating, we summarise, next, the published work regarding dating in general [17].

Image-based regression

CNNs outperform approaches based on feature engineering on writer identification [14,6] and similar findings are reported in dating. In [7], the authors used pre-trained CNNs to date images of medieval Dutch charters from 14CE to 16CE, by focusing on image crops. The authors reported a mean absolute error of 10 years, a number beyond our reach with papyrus data where an approximation of 50 years is accepted. Regression using pre-trained CNNs on random crops was also suggested in [25], for the dating of medieval Swedish charters. Besides feature extraction with deep learning, earlier work approached the task with regression on top of extracted features, such as scale-invariant [8] or hinge and fraglets [9].

Dating from other modalities

Besides images, other modalities have also been used as input. In [11], for example, textual features were used to infer the date. Although reasonable in general, this is not a feasible approach for Greek literary papyri and manuscripts, the text of which may be of much older authors, such as Homer. A different approach was suggested in [20], where ordinal classification was combined with multispectral imaging, tracking spectral responses of iron-gall ink (of historical letters, 17-20CE) at different wavelengths. Although rich, this data representation is very expensive in time and resources to establish, which also explains why datasets in this form are very rare. Besides, papyri are mostly written with carbon-based and not iron-gall ink, which is to the present more difficult to date.

3 Data

3.1 The nature of the papyri

As already mentioned, papyri bearing literary texts do not carry a date and for the vast majority of them papyrologists assign a date based on the affinity of their script with objectively (not palaeographically) dated specimens. These specimens, referred to as 'objectively dated' ones, are dated using external indications (not contained in the literary text on the papyrus) [24]. Occasionally, it is archaeological evidence or even radiocarbon dating suggesting a more secure date, but most importantly, papyri were often re-used after they exceeded their lifespan and literary texts are often found on papyri that have dated documents on the opposite side.

3.2 Digitised papyri

The images included in our dataset come from a number of collections and online resources, whereas five or six of them were scanned from images in printed volumes. Their digitisation took place during a period of more than two decades, under substantially different imaging protocols. As a result, they vary greatly in their properties, most importantly in scaling to actual size, colour capturing, resolution and bit depth. For a few of them it was not possible to extract text lines, due to very low resolution, and they returned empty files during the segmentation stage.

3.3 Our new dataset

Our dataset comprises images of Greek papyri from Egypt and their respective dates, from the 1st to the 4th CE.[7] Images of papyri from other centuries were few, hence we did not consider them in this study. The papyri included were selected from CDDGB, the only available collection of (somewhat) securely dated literary papyri available, which includes also a few documentary texts in bookhand. The data it contains can be dated based on various objective dating criteria, such as the presence of a document that contains a date on the reverse side, internal evidence in the text (mostly for the few documentary ones and the 9th c. manuscripts having colophons), radiocarbon dating, or a dateable archaeological context associated with the manuscripts. In the CDDGB database, most records contain sampled images and we had to manually trace full-sized ones from the respective collections. We release our dataset in two forms, one where images contain whole fragments and one where they contain text lines.

The Papyri Literary Fragments (PLF) dataset consists of 242 images of publicly available papyri fragments, from 1BCE to 9CE. As shown in Figure 1, most fragments come from the 2nd or the 3rd CE, followed by the 9th and the 1st CE. When multiple fragments of the same manuscript were available, we included all of them. The date provided for most fragments is not specific. Typically, the minimum date range assigned to a literary papyrus spans 50 years, but it may reach up to two centuries. Most often, the latter cases concern a date between the two most frequent centuries (noted in Figure 1 as '2,3' CE). Our study focused on the four first centuries, from 1 to 4 CE, comprising 168 images of literary papyri. Nine images were empty, which led us to 159 images in total. The final distribution across the four centuries (1-4CE) was 20, 61, 60, and 18 respectively. We converted our images to grayscale to reduce the dimensionality and to facilitate machine learning experiments.

The Papyri Literary Lines (PLL) dataset extends PLF so that images of the text lines of the fragments are provided instead of images of the whole fragments. The 159 images were segmented automatically using the Transkribus HTR platform,[8] yielding 4,655 line images. For this segmentation step, we used the default settings in Transkribus and did not train a specific baseline model, due to the multiformity of our material. We interfered minimally, by manually correcting text regions where none or very few lines were captured in the automatically generated segmentation. We also manually corrected a small number

[7] There is a very small number of exceptions which reflect the complexity of our documentation: one text is in Coptic, a few don't come from Egypt but the Near-East and another few are written on parchment, not papyrus. In this study, we collectively call them 'papyri'.

[8] https://readcoop.eu/transkribus/

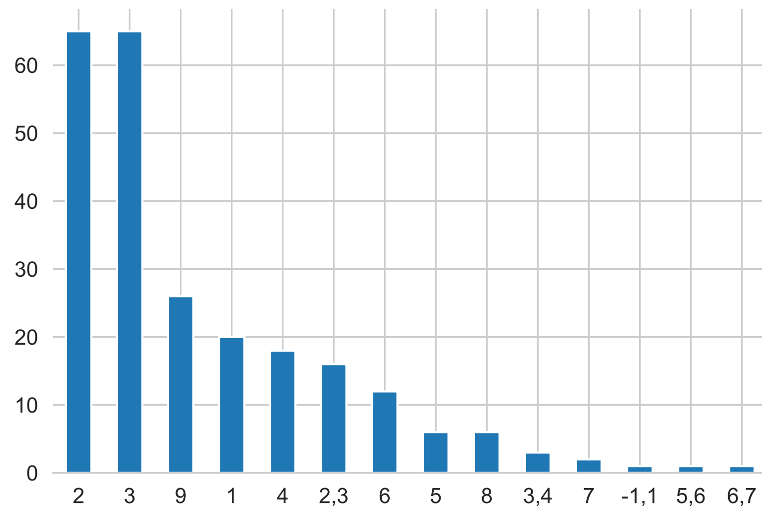


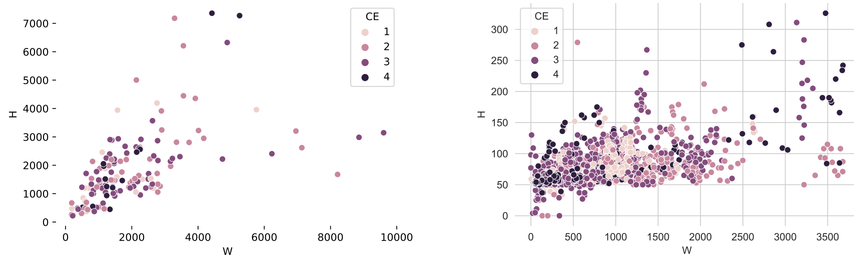
Fig. 1. The number of PLF images (vertically) per century (horizontally) or century range (when the date ranges between centuries), sorted by frequency.

of base lines and line regions (appr. 1-2%), when no or insignificant amount of writing was captured, or when substantial and useful writing areas were obviously excluded. Even so, a considerable number of possibly useful lines were not added and in several cases the automatic segmentation captured multiple lines in an instance, or substantial amount of background with minimal writing. As a result, the dataset would benefit from more interventional curation. We did not eliminate lines with noise, such as damaged papyrus surface, gaps in the writing material (holes), and lines bordering the edge of the papyrus. As a result, several line images still contain noise.

The balance of the dates followed that of PLF, with 439, 2,116, 1,797, and 303 images from the 1st, 2nd, 3rd, and 4th CE respectively. As can be seen in Figure 2(b), most images are higher than 50 pixels but width is characterised by a greater variety. Figure 2(a) presents the scatterplot of PLL, where lines comprise texts of various lengths, from a single word to more than ten. We filtered out images with a height lower than 50 pixels and ones with a width less than 300 pixels, which resulted in 2,774 images in total (40% reduction).

4 Method

Our method, called fCNN, is a 43m-parameter CNN that exploits augmentation so that it is robust to fragmented input, often met in papyri.



(a) PLF (b) PLL

Fig. 2. Scatter plot of the width (shown horizontally) and height of the images.

4.1 fCNN

The network consists of two Conv2D layers to represent the image of each text line, of 32 and 64 channels respectively, followed by a 3-layer feed-forward neural network (FFNN) with a single output neuron to yield the date. We used a convolutional kernel of size 5, single stride, zero padding, and max-pooling (2x2). The FFNN receives a flat representation from the Conv2D which is reduced to 1024 and then to 512 neurons before the date is estimated. A ReLU activation function is used per layer.

Synthetic fragmentation is a possible augmentation channel during training. Papyri are very often fragmented, leading to partial information in the image to be dated. We exploited this pattern as part of our augmentation strategy, by erasing randomly (0.5 probability) image fragments, setting their pixel values to 0.5. Images were transformed with Gaussian blur (kernel size of 3) and random affine (up to 3 degrees). The actual letter size as well as the image ratio to actual size in our dataset greatly varies, hence, to assist the network's robustness, we also randomly cropped and resized each image by keeping the 1:6 aspect ratio.

4.2 The baseline

We used the state-of-the-art in regression, which is achieved by ensembles [5], including Extremely Randomized Trees (XTR) and XGBoost (XGB). We experimented with both these regressors, using patches of 50*300 windows cropped from the center of each image, which was also represented with PCA-extracted 500-dimensional features. In our preliminary experiments, PCA led to better results compared to image binarisation using Canny edge detection and Otsu, which have been reported beneficial in writer identification [13]. We used the implementation provided by SKLEARN setting all hyper-parameters to default values, besides the objective of XGB, which was set to the squared error.

5 Experiments

We approached dating, which is a regression task, with fCNN, a CNN that uses a fragmentation-based augmentation strategy. We experimented with algorithms on the PLL dataset, using as input the images of the lines of the papyri and as output the date of the respective papyri. We also show results when we cast the problem as a classification task, predicting the century as one out of four labels.

5.1 Experimental details

We used Adam optimisation [10] with a learning rate of 1e-3, batch size of 16, 200 epochs, early stopping with patience of 20 epochs. The regression variant was trained with a mean squared error loss and the classification variant with a cross entropy loss. We used PyTorch and we release our code in our repository.[9]

The benchmark

A majority baseline (BLM), which always predicts the 2nd CE, achieved an MAE of approx. 0.632 and an MSE of 0.772. XTR and XGB perform better than this weak baseline, with a considerable difference when looking at MSE. The latter penalizes greater distances more, which means that papyri of the 1st and 4th CE were better handled by XGB and XTR. Our fCNNr performs considerably better than all the baselines, achieving an average absolute error of 54 years.

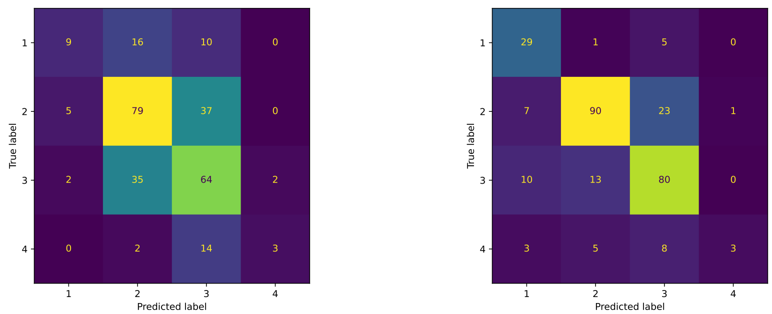
Table 1. Mean absolute and squared error of dating along with their standard error of the mean in parenthesis.

	MAE ↓	MSE ↓
BLM	0.632 (0.032)	0.772 (0.050)
XGB	0.612 (0.005)	0.558 (0.012)
XTR	0.610 (0.006)	0.544 (0.012)
fCNNr	**0.540 (0.001)**	**0.511 (0.009)**

From regression to classification

By rounding the predictions of our fCNNr, we created a confusion matrix, which is shown in Figure 3(a). Confusion regards mainly neighboring centuries. The model correctly detects images from the 2nd and 3rd while images from the 3rd may be predicted close to the 2nd, and vise versa. Difficulties in dating regard the two edges, because the 1st and 4th CE are more often predicted as of the 2nd and 3rd CE respectively.

Although our task in hand is a regression one in nature, we also trained and assessed a classification variant (fCNNc), which learns to disregard the order of



(a) fCNNr

(b) fCNNc

Fig. 3. Confusion matrices of fCNNr (rounded predictions) and fCNNc.

centuries and simply treat them as labels. In Figure 3(b), we observe that results improve across all centuries except from the 4th CE, where the difficulty remains approximately the same. Table 2) shows the F1 per century per fCNN variant, along with the benefit in absolute number when using the classification head instead of the regression one. We also trained an XGB and an XTR classifier, with the former performing better yet much worse than fCNNc.

	1CE	2CE	3CE	4CE
XGB	0.38	0.69	0.58	0.09
XTR	0.00	0.56	0.44	0.00
fCNNr	0.35	0.62	0.56	0.25
fCNNc	**0.69** (+0.34)	**0.78** (+0.16)	**0.73** (+0.17)	**0.26** (+0.01)

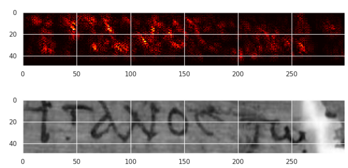
Table 2. F1 per century of fCNNr (predictions are rounded) and fCNNc, the absolute difference between the two is shown in parenthesis.

Despite the fact that both fCNN variants are trained on the same data, we note that we do not consider them as competitors. The regression-based fCNNr suggests a date, which can provide a very rough estimation of when the papyrus was written. If the predicted date was 280CE, then this is an indication that the papyrus is dated between the 3rd and the 4th CE, and that a year close to the latter is likelier. On the other hand, the classification-based fCNNc suggests a century and yields a score to indicate its confidence. If the predicted century was the 4th CE and the confidence was 80%, then this means that the network is confident that the date is 4th and no other. Although our task in hand is one of regression, both can generate useful explanations. Therefore, since our end goal is to assist and not supplement the expert, we used them both in our explainability study, discussed next.

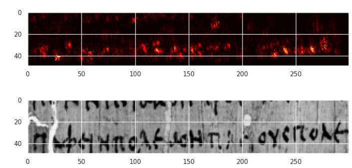
[9] https://github.com/ipavlopoulos/palit

Explainability

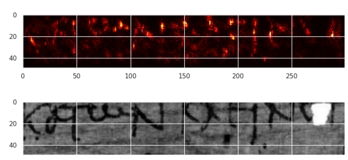
Saliency maps [21] reveal the parts of the image which are responsible for the network's prediction. We experimented with both variants, fCNNr and fCNNc, and we used both, gradient- and perturbation-based attribution. In this study, we opted for fCNNc using gradient-based attribution, but we observe that explanations by the two variants can be combined to yield richer explanations.



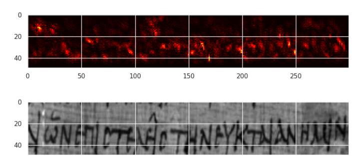
(a) 1CE



(b) 2CE



(c) 3CE



(d) 4CE

Fig. 4. Saliency maps for lines of papyri per century

We computed one heatmap per predicted line and we present a random sample of lines in Figure 4. The heated colours show that the network consistently focuses on the letters in order to yield its predictions for the date. This means that the model is basing its prediction on the shape of specific letters, the distance between them, the size, or the intensity of the ink. By contrast, it seems invariant from background noise and other attributes which may be often present in Greek literary papyri. For example, gaps (holes in the papyrus) such as those in Figures 4(a) and 4(c), do not get any attention by the model.

6 Assessing data sources limitations

CDDGB is not a product of targeted research on securely dated papyri, but rather a compilation of such examples mentioned in other papyrological works.[10]

[10] More reliable compilations are promised by current projects, but are still work-in-progress for the time being.

Hence, the collection is not comprehensive and the data included is not meticulously assessed by the compilers. Shortcomings concern the accuracy of some dates. Still and all, it is the same data of objectively dated papyri that papyrologists use as reference for palaeographical dating. In this study, we introduce the computational factor in assessing scripts in connection with their assigned dates. Also, by focusing on the explainability of dating images of handwritten text, we do not consider these shortcomings detrimental. The possible inaccuracies in dating and the wide-range of the assigned dates does not affect the explanations, which aim to provide pointers on features of the script.

The imbalance in the size of the fragments and quantity of lines is an inherent issue owing to the nature of the available material. A papyrus may contain three or four usable lines, whereas others may have more than fifty. This does not affect dating significantly because, although test lines may come from a manuscript not hidden during training, each line constitutes a completely different image pattern. The same issue could be an advantage regarding explainability, because possible features are brought out in a more controlled manner when multiple lines of the same manuscripts are involved. While some features, especially palaeographically insignificant characteristics, remain consistent (such as colour/intensity of the ink, texture and colour of the background, general size of script, scale, etc.), explanations can focus on pivotal ones.

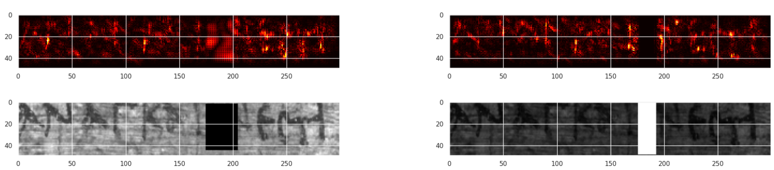
Our train and validation subsets are mutually exclusive at the line level but not necessarily at the fragment level. Although the former is straight forward, the latter is not due to the diversity of lines in the fragments. To experiment with the latter, we kept lines from papyri whose index modulo a value (13) was zero for validation and testing (in half), keeping the rest for training. Although introducing a distribution drift, presenting relatively fewer lines from 1CE during testing, this split met our restrictions. The error of fCNNr is slightly higher (0.612 ± 0.002), but remains the best. The F1 score remains approx. the same in classification, except from the 1st CE that drops to 0.4 but whose support in the test set is only 3 (out of 80) images. Future work will carefully compile more train, development, and test subsets, to investigate this issue further.

7 Error analysis

To go further in our understanding of the relevance of our experiment, we provide in this section an error analysis, followed by an experiment on the way the model handles the damages on the papyri by ablating input images before dating.

Analysis By studying fCNN's deviations from the ground truth, we observe that these concerned predictions toward the neighbouring century. Images from the two edge centuries, 1st and 4th CE, are scored up to the 2nd and 3rd CE respectively, the two most frequent centuries (Figure 1). Images from the 2nd and 3rd CE, on the other hand, were scored not far from each other, most often to the 3rd and the 2nd respectively. By looking at the saliency maps of the misclassifications, we observed that letter-shaped noise, present in the source images, received the model's focus.

Ablation Our error analysis revealed that fragments may deceive our model. In order to investigate the model's sensitivity, we fed fCNNc with test images, augmented with randomly-shaped black and white patches. We observe that the model's focus changes according to the colour of the patch. White boxes appear to be disregarded by our model, by contrast to black boxes, which are receiving attention. An example is shown in Figure 5, where the same line from a papyrus of the 3rdCE is altered in two ways. In Figure 5(b), the focus is everywhere except from the white patch. This is in line with our findings about the breaks, which are also depicted in white in the images (Figure 4). By contrast, the black patch of Figure 5(a) affects the prediction as if the model is guessing what character was missing and as if the black colour of the patch was ink.



(a) black (b) white

Fig. 5. Saliency maps of the same test line, from a papyrus of the 3rd CE, whose source image was transformed either with a black or a white patch before dating.

8 Dates in doubt: A computational estimate

fCNN can accurately predict the date of a text line image (Table 1) and, when the task is simply to predict the century and not an exact date, a classification variant that ignores the temporal relation of the labels yields even better results (Table 2). As was shown from our study of saliency-based explanations, fCNN focuses on the letters, that is the foreground and not the background (e.g., the blank parts of the papyrus sheet, the fibres, the holes and damages). In order to provide the experts with suggestions that could possibly improve the current dating,[11] we apply this network to loosely dated texts (across two centuries).

In our primary source, 11 papyri are dated either to the 2nd or the 3rd CE. Using fCNNc, we found that 87% of the lines are classified to the 2nd or

[11] Datings usually come from one expert, the editor of the text. Sometimes another expert makes a case that the dating should be modified and the correction may be accepted or provided as alternative dating.

3rd CE. Exceptions were from 16 which were classified to the 1st and 1 which was classified to the 4th. Figure 6 presents the analytical results. Using fCNNr,

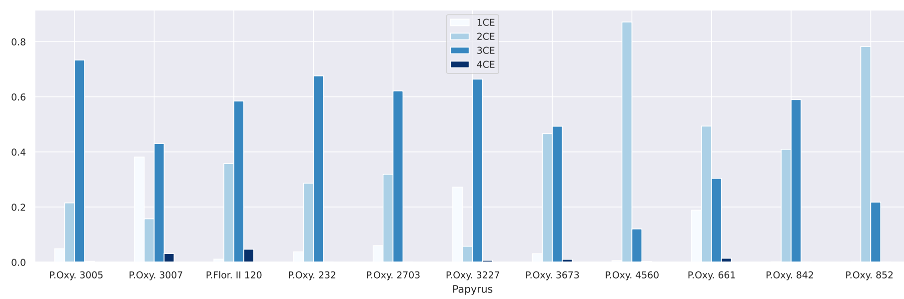


Fig. 6. Chronological attribution of fCNNc of lines in fragments dated between 2-3CE

we attempted then to estimate a more precise chronology for the lines in these papyri. Despite the fact that our regressor was trained on ground truth at the century level, our expectation is that it will have learned to yield a chronology that is closer to the objective date. Figure 7 presents our predictions, organised per papyrus. The predicted dates for the lines of P.Oxy 3005, which was classified by fCNNc on the 3rd, are diverse, with the majority falling on the late 2nd and early 3rd. Overall, our network's estimations agree with the range provided by the experts. The earliest prediction was 98CE, for a line in P.Oxy. 661. This papyrus comprises parts of a poem by Callimachus and is dated from 150 to 250 CE,[12] with the first editor arguing that it is the late 2nd CE.[13] On average, our predictions suggest 200CE, but some lines are predicted as early as 100CE while others as 250CE. The latest prediction is 270CE for a line in P.Flor. II 120,[14] dated from 250 to 261CE. In this papyrus, in very few lines our predictions agree with the experts, because on average our network dates it before the 200CE. In P. Oxy. 4560, only one line is used, and date is 100CE. In P. Oxy. 232, although lines are few, all our predictions date the papyrus between 100 and 150CE.

9 Conclusions

This work introduced two datasets of images of Greek literary papyri, one with whole papyri fragments (PLF) and one with lines of writing (PLL). Our experiments showed that an augmentation-enhanced CNN predicts the date of text-line images with a mean absolute error of 54 years, using a regression head, and a macro-average F1 of 61.5%, using a classification head, setting the state of the

[12] https://www.trismegistos.org/text/59375 (accessed: May 25, 2023).
[13] The Photographic Archive of Papyri in the Cairo Museum (accessed: May 25, 2023).
[14] https://papyri.info/ddbdp/p.flor;2;120 (accessed: May 25, 2023)

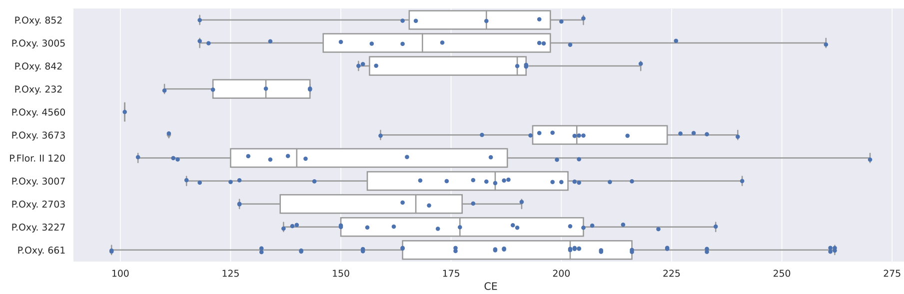


Fig. 7. Chronological attribution of fCNNr of lines in fragments dated between 2-3CE

art for Greek papyri image dating. An explainability study revealed that fCNN clearly focuses on letters to predict the date, following the palaeographer's path. Using fCNN, we predicted the date of the text lines in eleven papyri, whose objective date is ranging across two centuries, and we discussed our findings.

References

1. Brink, A.A., Smit, J., Bulacu, M.L., Schomaker, L.R.B.: Writer identification using directional ink-trace width measurements. Pattern Recognition **45**(1), 162–171 (2012). https://doi.org/https://doi.org/10.1016/j.patcog.2011.07.005
2. Bulacu, M., Schomaker, L., Vuurpijl, L.: Writer identification using edge-based directional features. In: Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. pp. 937–941 (2003). https://doi.org/10.1109/ICDAR.2003.1227797
3. Choat, M.: Dating Papyri: Familiarity, Instinct and Guesswork. Journal for the Study of the New Testament **42**(1), 58–83 (2019). https://doi.org/10.1177/0142064X19855580
4. Cilia, N.D., De Stefano, C., Fontanella, F., Marthot-Santaniello, I., Scotto di Freca, A.: Papyrow: a dataset of row images from ancient greek papyri for writers identification. In: Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10-15, 2021, Proceedings, Part VII. pp. 223–234. Springer (2021)
5. Fernández-Delgado, M., Sirsat, M.S., Cernadas, E., Alawadi, S., Barro, S., Febrero-Bande, M.: An extensive experimental survey of regression methods. Neural Networks **111**, 11–34 (2019)
6. Fiel, S., Sablatnig, R.: Writer identification and retrieval using a convolutional neural network. In: Computer Analysis of Images and Patterns: 16th International Conference, CAIP 2015, Valletta, Malta, September 2-4, 2015, Proceedings, Part II 16. pp. 26–37. Springer (2015)
7. Hamid, A., Bibi, M., Moetesum, M., Siddiqi, I.: Deep learning based approach for historical manuscript dating. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). pp. 967–972. IEEE (2019)
8. He, S., Samara, P., Burgers, J., Schomaker, L.: Discovering visual element evolutions for historical document dating. In: 2016 15th International conference on frontiers in handwriting recognition (ICFHR). pp. 7–12. IEEE (2016)

9. He, S., Sammara, P., Burgers, J., Schomaker, L.: Towards style-based dating of historical documents. In: 2014 14th International Conference on Frontiers in Handwriting Recognition. pp. 265–270. IEEE (2014)
10. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
11. Kumar, A., Baldridge, J., Lease, M., Ghosh, J.: Dating texts without explicit temporal cues. arXiv preprint arXiv:1211.2290 (2012)
12. Mazza, R.: Dating Early Christian Papyri: Old and New Methods –Introduction. Journal for the Study of the New Testament **42**(1), 46–57 (2019). https://doi.org/10.1177/0142064X19855579
13. Nasir, S., Siddiqi, I.: Learning features for writer identification from handwriting on papyri. In: Pattern Recognition and Artificial Intelligence: 4th Mediterranean Conference, MedPRAI 2020, Hammamet, Tunisia, December 20–22, 2020, Proceedings 4. pp. 229–241. Springer (2021)
14. Nguyen, H.T., Nguyen, C.T., Ino, T., Indurkhya, B., Nakagawa, M.: Text-independent writer identification using convolutional neural network. Pattern Recognition Letters **121**, 104–112 (2019)
15. Nongbri, B.: The limits of palaeographic dating of literary papyri: Some observations on the date and provenance of p.bodmer ii (p66). Museum Helveticum **71**(1), 1–35 (2014)
16. Nongbri, B.: Palaeographic analysis of codices from the early christian period: A point of method. Journal for the Study of the New Testament **42**(1), 84–97 (2019). https://doi.org/10.1177/0142064X19855582, https://doi.org/10.1177/0142064X19855582
17. Omayio, E.O., Indu, S., Panda, J.: Historical manuscript dating: traditional and current trends. Multimedia Tools and Applications **81**(22), 31573–31602 (2022)
18. Paparrigopoulou, A., Kougia, V., Konstantinidou, M., Pavlopoulos, J.: Greek literary papyri dating benchmark (2023). https://doi.org/10.21203/rs.3.rs-2272076/v2
19. Papavassiliou, V., Stafylakis, T., Katsouros, V., Carayannis, G.: Handwritten document image segmentation into text lines and words. Pattern recognition **43**(1), 369–377 (2010)
20. Rahiche, A., Hedjam, R., Al-Maadeed, S., Cheriet, M.: Historical documents dating using multispectral imaging and ordinal classification. Journal of Cultural Heritage **45**, 71–80 (2020)
21. Simonyan, K., Vedaldi, A., Zisserman, A.: Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034 (2013)
22. Stokes, P.A.: Computer-aided palaeography, present and future. In: Rehbein, M., Schaßan, T., Sahle, P. (eds.) Kodikologie und Paläographie im digitalen Zeitalter - Codicology and Palaeography in the Digital Age, vol. 2, pp. 309–338. BoD, Norderstedt (2009), https://kups.ub.uni-koeln.de/2978/
23. Stokes, P.A.: Scribal attribution across multiple scripts: A digitally aided approach. Speculum **92**(S1), S65–S85 (2017). https://doi.org/10.1086/693968
24. Turner, E.G.: Greek Manuscripts of the Ancient World.(P. J. Parsons, Ed.; Revised and Enlarged ed.). Institute of Classical Studies, London (1987)
25. Wahlberg, F., Wilkinson, T., Brun, A.: Historical manuscript production date estimation using deep convolutional neural networks. In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 205–210. IEEE (2016)