Hard and Soft Labeling for Hebrew Paleography: A Case Study

Ahmad Droby[1(✉)], Daria Vasyutinsky Shapira[1], Irina Rabaev[2], Berat Kurar Barakat[1], and Jihad El-Sana[1]

[1] Ben-Gurion University of the Negev, Beer-Sheva, Israel
{drobya,dariavas,berat,el-sana}@post.bgu.ac.il
[2] Shamoon College of Engineering, Beer-Sheva, Israel
irinar@ac.sce.ac.il

Abstract. Paleography studies the writing styles of manuscripts and recognizes different styles and modes of scripts. We explore the applicability of hard and soft-labeling for training deep-learning models to classify Hebrew scripts. In contrast to the hard-labeling scheme, where each document image has one label representing its class, the soft-labeling approach labels an image by a label vector. Each element of the vector is the similarity of the document image to a certain regional writing style or graphical mode. In addition, we introduce a dataset of medieval Hebrew manuscripts that provides complete coverage of major Hebrew writing styles and modes. A Hebrew paleography expert manually annotated the ground truth for soft-labeling. We compare the applicability of soft and hard-labeling approaches on the presented dataset, analyze, and discuss the findings.

Keywords: Digital paleography · Medieval Hebrew manuscripts · Script type classification · Soft-labeling · Convolutional neural network

1 Introduction

Paleographic analysis of a historical document can determine the place and date when the manuscript was written. In some cases, it is even possible to identify the scribe, verify the authenticity of a manuscripts, or obtain other essential information. The continuing digitization of manuscript collections held by various libraries resulted in the availability of a large number of digital manuscripts. A professional paleographer can only process a limited number of manuscripts, and there are still manuscript collections lacking even a basic catalogue. Hence, the processing must be automated, and for the development, evaluation, and comparison of algorithms, benchmark datasets are required.

The survival rates of Hebrew manuscripts are much lower in comparison with that of Latin, Greek or Arabic ones. There are about one thousand fragments of manuscripts in Hebrew scripts that survived from the Middle ages (for this matter, beginning of the 10th century - 1540). At the current state of research it

S. Uchida et al. (Eds.): DAS 2022, LNCS 13237, pp. 492–506, 2022.
https://doi.org/10.1007/978-3-031-06555-2_33

is impossible to estimate their number more precisely. The ongoing digitization of Hebrew manuscripts is very advanced, because of the continuous efforts of most of the world's leading libraries. Alongside it, the Institute for Microfilmed Hebrew Manuscripts at the National Library of Israel has already assembled most of the known Hebrew manuscripts on microfilms and digital images. Thus, the automatic recognition and analysis of Hebrew manuscripts and historical documents is an urgent desideratum.

Among the surviving Hebrew manuscripts, about three thousand are dated, and these are included into the SfarData database[1] of Hebrew paleography and codicology, completed by Malachi Beit-Arié and his team.

In this paper, we present our research on automatic classification of Hebrew manuscripts into fourteen categories according to the script types and graphical modes. To train a deep neural network, we compiled a dataset of manuscripts where all of these categories are present. The margins between categories of writing styles are sometimes fuzzy and overlap on visual appearances level. To categorize the document, paleographers examine the visual appearance of the handwriting as well as the codicological data, e.g., the media on which the document was written. Since we are working with digital images only, we are unable to utilize the codicological data. We hypothesize that hard-labeling may not be the ideal way for training the deep-learning model to recognize the writing category. Therefore, for each page image we decided to add an additional level of labeling - a soft label. The soft label is a label vector, where each element indicates the similarity of the document's script to a specific script type or mode. An expert in Hebrew paleography manually annotated the soft label for each document.

The main contributions of this paper include: (1) We experiment with two different ground truth labeling schemes for training a deep-learning model and analyze the obtained results. We also discuss the issues of paleographic analysis of Hebrew writings, as well as their specificities in the context of automated processing. (2) We present a benchmark dataset of Hebrew manuscripts compiled especially for developing and evaluating machine learning algorithms. To the best of our knowledge, this is the first dataset in Hebrew that includes samples of major Hebrew writing types and modes to address digital paleography community. The dataset contains page images from 171 different manuscripts covering fourteen categories of writing, and accompanied by hard and soft labels. The dataset can be downloaded from Zenodo repository https://zenodo.org/record/6387471. We believe that this dataset will help to leverage automatic Hebrew historical documents processing, and the historical document processing in general.

2 Related Work

Throughout the last decade, various computer vision algorithms have been used for paleography analysis. Earlier techniques relied on hand-crafted features,

[1] http://sfardata.nli.org.il/.

which were often based on textural and grapheme-based descriptors, and their combination [13–15]. During the recent years, deep learning approaches have set new benchmarks in a variety of academic fields, and have been also adapted for paleographic analysis [5,6,11,16,27]. Keglevic *et al.* [18] propose to use a triplet CNN to measure the similarity of two image patches. Abdalhaleem *et al.* [1] investigate in-writer differences in manuscripts. Their methodology is built on Siamese convolutional neural networks, which are trained to recognize little differences in a person's writing style. Studer *et al.* [25] explored the effect of ImageNet pre-training for various historical document analysis tasks, including style categorization of Latin manuscripts. They experimented with VGG19 [22], Inception V3 [26], ResNet152 [12], DenseNET12 [17], and additional well-known architectures. The models trained from scratch achieved 39%–46% accuracy rate, whereas the pre-trained models achieved a 49%–55% accuracy rate.

Two major competitions on the categorization of medieval handwritings in Latin script [7,8] were organized in 2016 and 2017. The goal of the competitions was to classify medieval Latin scripts into 12 categories based on their writing styles. The findings reveal that deep learning models can accurately recognize Latin script types with more than 80% accuracy on homogeneous document collections and about 60% accuracy on heterogeneous document collections.

There have been few works on Hebrew document paleography. Wolf *et al.* [29] explored handwriting matching and paleographic classification, focusing on the documents from the Cairo Genizah collection. Dhali *et al.* [9] use textural and grapheme-based features with support vector regression to determine the date of ancient texts from the Dead Sea Scrolls collection. Ben Ezra *et al.* [24] train a model for establishing the reading order of the main text by detecting insertion markers that indicate marginal additions. They used a corpus of 17 manuscripts of Tannaitic Rabbinic compositions dated from the 10*th* to 15*th* centuries. The international Israeli and French team [21,28] work on a project that combines handwritten text recognition of Medieval Hebrew documents with a crowdsourcing-based process for training and correcting the HRT model. Their project focuses on a subset of rabbinic works dated to 1-500 CE. The aforementioned projects work on different datasets and different kind of manuscripts, and each project is solving a different part of the puzzle. These projects complement each other for the final goal of recognition of the handwritten text in historical documents. In this work, we train a deep-learning model to classify medieval Hebrew scripts into fourteen classes.

3 Hebrew Paleography

Manuscripts are studied by means of paleography and codicology, that explores the writing and the material on which manuscripts are written, respectively. The theoretical basis of Hebrew paleography and codicology are formulated in the works of Malachi Beit-Arié, Norman Golb, Benjamin Richler, Colette Sirat [2–4,19,20,23,30].

Hebrew manuscripts refers to manuscripts written in Hebrew characters, as the language was often adopted from the host societies (Ladino, Judeo-Arabic, Yiddish etc.). Geographically, the spread of the Hebrew manuscripts was larger than Latin, Greek or Arabic manuscripts. Hebrew scripts themselves were influenced by the local traditions and often resemble the manuscripts of the host societies in scribal manner, material and ways of production.

There are six main types of the Hebrew script: Oriental, Sefardic, Ashkenazi, Italian, Byzantine and Yemenite. The writing styles of Hebrew manuscripts could be classified into two branches based on their geographic origin. Oriental, Sefardic and Yemenite styles developed in Islamic regions and were influenced by the Arabic calligraphy, while Ashkenazi and Italian styles evolved in Europe and were somewhat influenced by Latin scripts. The Byzantine type displays hybrid influences and probably the influences of Greek scripts.

Our project aims to recognize the main types of the Hebrew script, and their modes (square, semi-square, cursive). Paleographically, the backbone of our research is the SfarData. Malachi Beit-Arié and his team met with our team, discussed our project, gave us their full support, and allowed us to use their database in its entirety. Our team's paleographer, who is herself a student of Malachi Beit-Arié, handpicked digitized pages from the manuscripts described in the SfarData as the raw material for our project.

4 VML-HP-ext Dataset Description

The Hebrew paleography dataset is a valuable resource both for creating a large-scale paleographic examination of Hebrew manuscripts, and assessing and benchmarking scripts classification methods. In this paper we present an extended VML-HP-ext (Visual Media Lab - Hebrew Paleography Extended) dataset. The initial version of the VML_HP dataset was presented in [10]. It consists of pages excerpted from about 60 manuscripts, their corresponding hard labels, and the official split intro training and two testing sets. Compared to the first version, the extended dataset includes sample pages from three times more manuscripts. Every manuscript was carefully selected by our team's paleographer. The majority of the manuscripts used in this dataset are kept in the National Library of Israel, the British Library, and the Bibliothèque nationale de France. Almost all manuscripts in the Oriental square script belong to the National library of Russia (we used b/w microfilms from the collection of the Institute for Microfilmed Hebrew Manuscripts at the National Library of Israel). We only included pages with one script type and one script mode per page. For example, Sephardic square only, and not main text in Sephardic square and comments in Sephardic cursive. The main challenge when compiling the dataset was the limited amount of available digitized manuscripts. For some script types (Italian, Byzantine) the shortage was more pronounced; for others (Ashkenazi, Sephardic) we had manuscripts in abundance. Keeping the dataset balanced was a challenge by itself.

Table 1. Summary of the extended VML-HP-ext dataset. Some scripts do not have semi-cursive or cursive modes. Mss = manuscripts, pp = pages.

Type	Mode					
	Square		Semi-square		Cursive	
	#Mss	#pp	#Mss	#pp	#Mss	#pp
Ashkenazi	14	56	12	48	12	48
Byzantine	7	49	12	48	–	–
Italian	5	50	11	44	5	50
Oriental	15	45	11	44	–	–
Sephardic	15	45	16	48	12	48
Yemenite	24	92	–	–	–	–

The enlarged VML-HP-ext collection contains 715 page images excerpted from 171 different manuscripts. We also provide the official split of the VML-HP-ext into training, typical test, and blind test sets. Typical test set includes unseen pages of the manuscripts from the training set. While training and typical test sets are disjoint on page level, they do share the same set of manuscripts. Therefore, we also provide the blind test set, which consist of manuscripts that do not appear in the training set. The blind test set imitates a real-life scenario, where scholar would like to obtain a classification for a previously unseen document. Tables 1 and 2 summarize the extended VML-HP-ext dataset.

5 Case Study

In the following section, we present and discuss our experiments on the extended VMP-HP dataset. We report the results using hard label classification model and compare it with our previous results. In addition, we explore the use of the newly introduced soft labels to train a regression model, which can be used for classification.

5.1 Hard-Label Classification

This experiment aims at evaluating classification models on the extended dataset. We trained and evaluated several architectures on the extended dataset.

Table 2. The VML-HP-ext dataset - official split. Mss = manuscripts, pp = pages

Set	# Mss	# pp
Train	130	400
Typical test	130	143
Blind test	41	172
Total	171	715

The models were trained until convergence using 50K patches extracted from pages in the train set. The model was trained using binary cross entropy loss function. The patches were extracted using the patch generation method proposed in our previous work [10], which extracts patches with uniform text scale and on average 5 lines in each patch.

Table 3 shows the precision, recall, F1, and accuracy measures of the models on the blind test set. As seen, ResNet50 outperforms all the other model on every metric, achieving an accuracy of 60% which is significantly higher than the accuracy we obtained on the old dataset, which was 42.1%. Obviously, showing the benefits of the extended dataset, which include more varying handwriting in each script type. Table 4 presents the precision, recall, and F1 measures of ResNet50 for each class. As seen, there are some classes such as Italian square, Byzantine semi-square, and Ashkenazi semi-square that are frequently classified incorrectly as Italian semi-square, Byzantine square, and Ashkenazi cursive respectively (see the confusion matrix in Fig. 1). Considering those classes share the same regional style, these results suggest that there may be ambiguity in the definition of the script types; i.e., we hypothesize that there is no clear-cut between such graphical modes, rather, they lie on a spectrum between square and cursive (Fig. 2).

Table 3. Evaluation results of several classification models on blind test set of the extended dataset.

Model	Avg. Precision	Avg. Recall	Avg. F1-score	Accuracy
DenseNet	58	53	53	53
AlexNet	55	52	52	51
VGG19	60	57	56	56
ResNet50	**63**	**60**	**59**	**60**
SqueezeNet	58	55	54	55

Table 4. Evaluation results of classification model with ResNet50 architecture on the extended dataset.

Label	Square			Semi-square			Cursive		
	P	R	F1	P	R	F1	P	R	F1
Ashkenazi	0.78	0.90	0.83	0.63	0.47	0.54	0.50	0.57	0.53
Byzantine	0.14	0.07	0.09	0.29	0.65	0.40		-	
Italian	0.51	0.14	0.21	0.33	0.42	0.37	0.68	0.97	0.80
Oriental	0.95	0.69	0.80	0.50	0.75	0.60		-	
Sephardic	0.86	0.88	0.87	0.87	0.45	0.59	0.93	0.73	0.82
Yemenite	0.90	0.67	0.77		-			-	

Average	P	R	F1	**Accuracy**	0.60
	0.63	0.60	0.59		

P: precision, R: recall, F1: F1-score

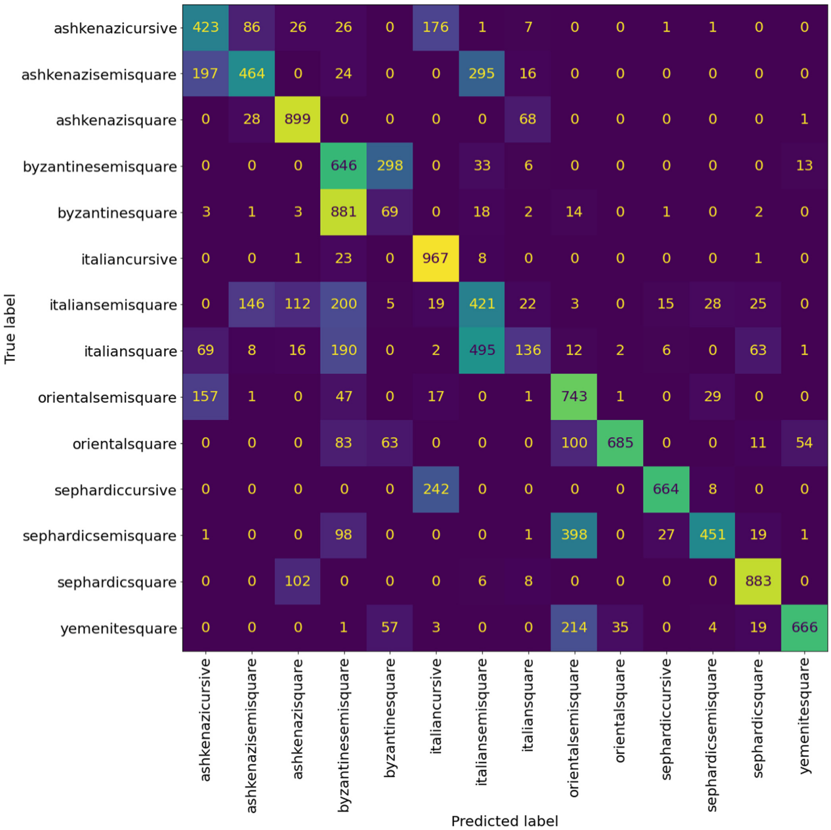

Fig. 1. Confusion matrix of classification model with ResNet50 architecture trained using the hard-labels.

5.2 Soft-Label Regression

As we have mentioned in Sect. 1, the margins between categories of writing styles are blurred, and there is an overlap between characteristics of different styles of writing. To categorize the document into writing category, paleographers rely both on visual appearance and codicological data (such as the media on which the document was written). However, we deal with only digital images and can not utilize codicological information. We hypothesize that hard labels may not be the best way to characterize the writing style of a document. Therefore, we added a second level of labeling - a soft label - for each page. The soft level is a label vector, where each element specifies the degree of similarity between the processed document and the certain script type or mode. The soft-labeling were done by an Expert Hebrew paleographer.

In a soft-labeling scheme, we label each manuscript using a vector of size eight. The first six elements of the vector express the degree of similarity of the manuscript to belong to certain regional type (Ashkenazi, Italian, Sephardic,

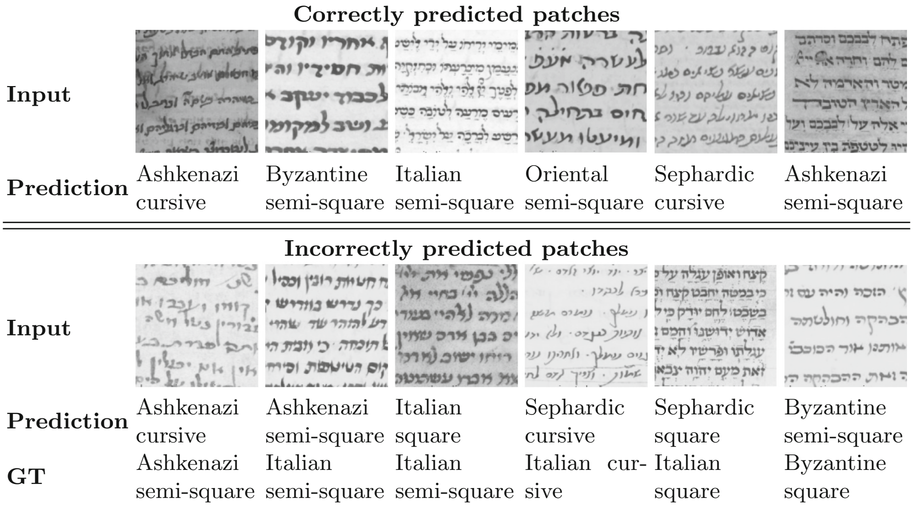

Fig. 2. Sample results from the ResNet50 classification model

Oriental, Byzantine and Yemenite) and the last two elements are the degrees of similarity to certain graphical mode, square and cursive (similar values for both square and cursive indicate the semi-square mode). Similar to the previous experiment, we extracted $50K$ patches and assign each patch a vector of probability values corresponding to a regional and graphical types. We trained a regression model with a ResNet50 backbone on the mentioned $50K$ patches with mean squared error loss function. The model was trained until convergence, which happened after 10 epochs.

Figure 3 reports sample results. To evaluate the model numerically, we calculated the Root Mean Square Error (RMSE) on the blind test set. RMSE is calculated according to the following formula:

$$RMSE = \sqrt{\frac{\sum_{i=0}^{N} ||y(i) - \hat{y}(i)||^2}{N}},$$

where $y(i)$ is the predicted label for patch i, and $\hat{y}(i)$ is its actual label.

The trained model achieved RMSE of about 0.24. Although, this might give us an indication that the model give good results (as can be seen in Fig. 3), it is not very meaningful and does not show how this model compare against other classification methods. Therefore, arose a need to convert the predicted soft-label to hard-labels. Next, we will explore two different conversion methods and report corresponding results.

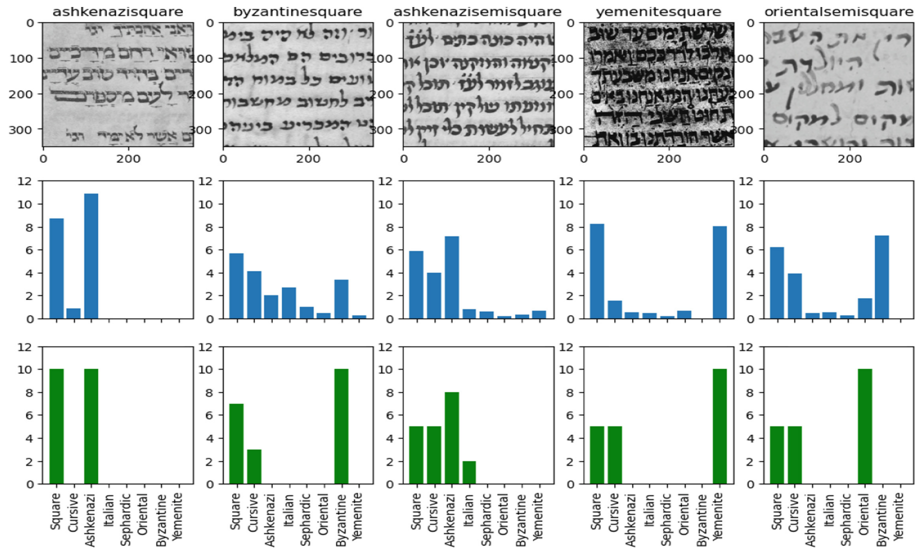

Fig. 3. Sample results of the regression model. Top row: the input patch with its ground-truth label. Second row: the predicted soft-label. Bottom row: the ground-truth soft-label.

5.3 Maximum Score Class Assignment

In this approach, a predicted soft-label s is converted to a hard label according to the following formula:

$$Regional(s) = argmax\{s(r)\};_{r \in \{\text{Ashkenazi, Byzantine, Italian, Oriental, Sephardic, Yemenite}\}}$$

$$Graphical(s) = \begin{cases} argmax_{g \in \{\text{square, cursive}\}}\{s(g)\}, & s(\text{square}), s(\text{cursive}) < T \\ \text{semi-square}, & \text{else} \end{cases}$$

In other words, the label is determined by taking the regional style and graphical mode with the maximum score unless both, the square and cursive, scores are under a predefined threshold T (we set $T = 0.3$), in which case the graphical mode is determined to be as semi-square.

Table 6 presents the evaluation results of the regression model after converting the soft-labels. It is important to note that this method introduces new labels that are not present in the dataset, such as Byzantine cursive, Oriental cursive, and Yemenite semi-square. The conversion method achieved an accuracy of 47%.

Table 5. The results of the regression model for the regional style classes only.

Label	Precision	Recall	F1-score
Ashkenazi	0.60	0.88	0.72
Byzantine	0.47	0.84	0.60
Italian	0.74	0.53	0.62
Oriental	0.92	0.37	0.53
Sephardic	0.85	0.68	0.76
Yemenite	0.83	0.64	0.72
Accuracy			0.66
Macro avg	0.74	0.66	0.66
Weighted avg	0.73	0.66	0.66

The regression model obtains an accuracy of 67% using the regional style labels only, as seen in Table 5. This indicates that the graphical style scores hinders the classification more than the regional ones. The model encounters the highest confusion between the Italian and Ashkenzi patches, as illustrated in the confusion matrix of the regional classification (Fig. 4). Most of the incorrect predictions are minor mistakes that a human paleographer could also have made. Neverthless, the confusion between Sephardic and Italian is a more serious error.

Table 6. Evaluation results of the regression model with maximum score class conversion method.

Label	Square			Semi-square			Cursive		
	P	R	F1	P	R	F1	P	R	F1
Ashkenazi	1.00	0.79	0.88	0.32	0.40	0.36	0.29	0.09	0.14
Byzantine	0.27	0.06	0.10	0.24	0.81	0.37		-	
Italian	0.00	0.00	0.00	0.22	0.66	0.33	0.18	0.18	0.18
Oriental	0.88	0.61	0.72	0.17	0.07	0.09		-	
Sephardic	0.98	0.36	0.52	0.32	0.64	0.43	0.99	0.15	0.25
Yemenite	0.83	0.31	0.45		-			-	

Average	P	R	F1	**Accuracy**	0.37
	0.50	0.37	0.34		

P: precision, R: recall, F1: F1-score

5.4 Nearest Neighbor Label Conversion

This approach utilizes the soft and hard labels in the training set. It calculates the distances between the predicted labels and the soft-labels in the training

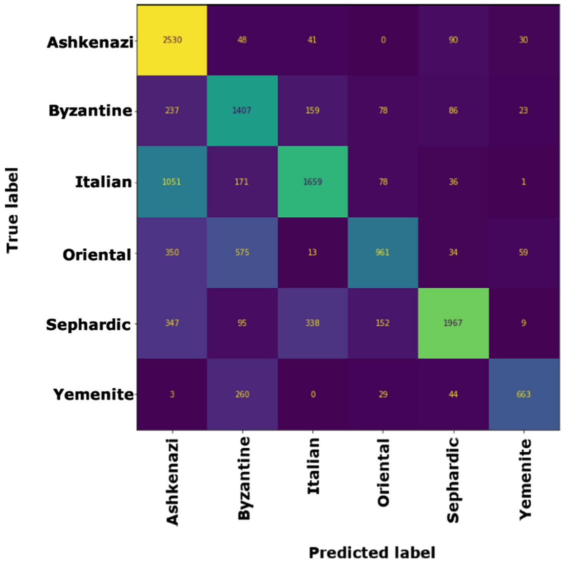

Fig. 4. The confusion matrix of the classification for the regional style classes.

set and converts each predicted soft-label to the nearest hard-label in the train set. Figure 5 presets sample results of this conversion, and Table 7 reports the classification accuracy using this method. The method obtains 46% accuracy, reaching results on par with the previous conversion method.

5.5 Comparison Between Soft and Hard-Label Classification

As we have reported earlier, the hard-label classification obtains higher accuracy in comparison with soft-labeling configuration. However, this does not tell the whole story as the soft-labeling regression model offers more insight on the script style. For examples, for a square graphical style text that has some cursive characteristics, using hard-label classification most probably will classify this text as square or semi-square, but a regression model will indicate “how

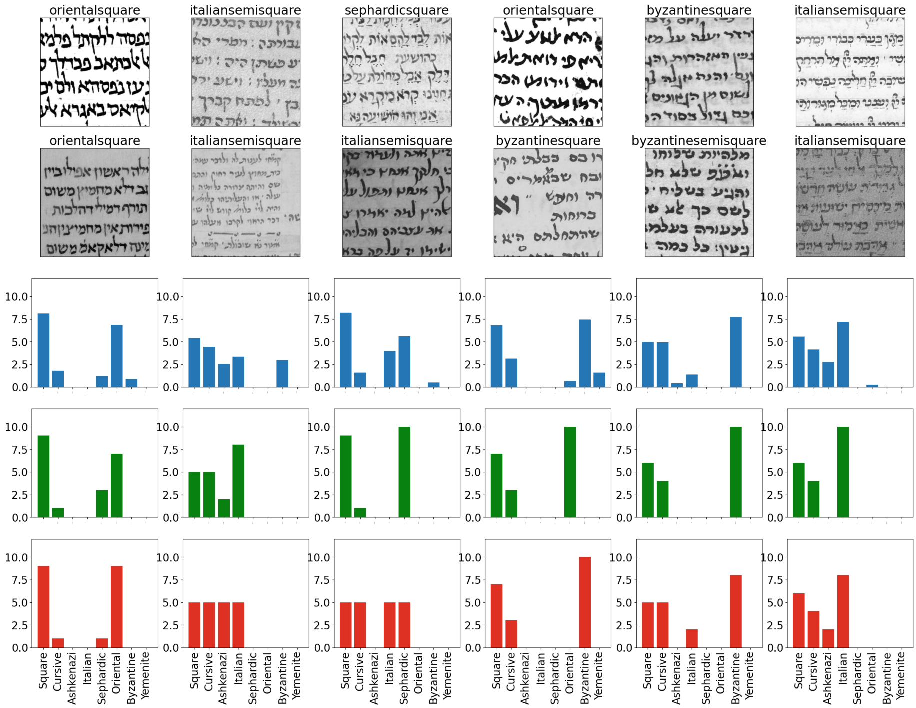

Fig. 5. Sample results of regression model with the nearest neighbor label conversion method. Top row: input patch with its ground-truth label. Second row: The nearest neighbor of the input patch. Third row: The predicted label of the input patch. Fourth row: The ground-truth soft-label of the input patch. Bottom row: The ground-truth soft-label of the nearest neighbor patch.

much cursive" this text is. Furthermore, such a regression model can offer paleography experts a tool to analyze the fluidity of the Hebrew script style, as a text can have multiple regional style characteristics while having varying level of squareness/cohesiveness.

Table 7. Evaluation results of the regression model with the nearest neighbor label conversion method.

Label	Square			Semi-square			Cursive		
	P	R	F1	P	R	F1	P	R	F1
Ashkenazi	0.98	0.57	0.72	0.43	0.67	0.52	0.50	0.01	0.03
Byzantine	0.03	0.01	0.01	0.25	0.87	0.38		-	
Italian	0.00	0.00	0.00	0.23	0.64	0.33	0.39	0.21	0.27
Oriental	0.99	0.65	0.79	0.29	0.06	0.10		-	
Sephardic	0.99	0.37	0.54	0.49	0.73	0.58	1.00	0.01	0.02
Yemenite	0.88	0.63	0.73		-			-	

Average	P	R	F1	Accuracy	0.40
	0.53	0.40	0.37		

P: precision, R: recall, F1: F1-score

6 Conclusion and Further Research

In this paper, we investigated the use of two types of labeling for Hebrew script types classification, hard and soft-labeling. Hard-labeling refer to the traditional labeling where each page is labeled with one script type. Soft-labeling assigns a vector of size eight to each page. The vector indicts how similar this page's writing style is to each geographical type and graphical mode. To perform the experiments, we compiled the VML-HP-ext dataset that covers major Hebrew script types. The dataset includes soft-labels for each page in addition to hard-labels.

We trained and evaluated several classification models on the hard-labeling configuration. ResNet50 topped the list with an accuracy of 60%. In addition, we experimented with soft-labeling, training a regression model to predict the similarity values of each image to each geographical and graphical type. Since such a model cannot be directly compared with regular hard-label classification, we proposed and evaluated two methods that convert soft labels to hard labels. We conclude that while the soft-labeling provides more information about the script style, e.g., how square or cursive it is, using the regression model with the conversion methods does not reach the accuracy of the models trained using hard-labeling.

In future work, we plan to experiment with additional ways to interpret the soft-labels and convert them to hard-labels. In addition, we want to experiment with unsupervised or semi-supervised classification, which may give us a more precise definition of the script type.

Acknowledgment. This research was partially supported by The Frankel Center for Computer Science at Ben-Gurion University.

References

1. Abdalhaleem, A., Barakat, B.K., El-Sana, J.: Case study: fine writing style classification using siamese neural network. In: 2nd International Workshop on Arabic and Derived Script Analysis and Recognition, pp. 62–66 (2018)
2. Beit-Arié, M.: Hebrew codicology. Tentative Typology of Technical Practices Employed in Hebrew Dated Medieval Manuscripts, Jerusalem (1981)
3. Beit-Arié, M.: Hebrew Codicology. ZFDM Repository (2021). https://doi.org/10.25592/uhhfdm.8849
4. Beit-Arié, M., Engel, E.: Specimens of mediaeval Hebrew scripts, vol. 3. Israel Academy of Sciences and Humanities (1987, 2002, 2017)
5. Christlein, V., Bernecker, D., Maier, A., Angelopoulou, E.: Offline writer identification using convolutional neural network activation Features. In: Gall, J., Gehler, P., Leibe, B. (eds.) Pattern Recognition, GCPR 2015. LNCS, vol. 9358, pp. 540–552. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24947-6_45
6. Christlein, V., Gropp, M., Fiel, S., Maier, A.: Unsupervised feature learning for writer identification and writer retrieval. In: 14th International Conference on Document Analysis and Recognition, vol. 1, pp. 991–997 (2017)
7. Cloppet, F., Eglin, V., Helias-Baron, M., Kieu, C., Vincent, N., Stutzmann, D.: ICDAR2017 competition on the classification of medieval handwritings in Latin script. In: 14th International Conference on Document Analysis and Recognition, vol. 1, pp. 1371–1376 (2017)
8. Cloppet, F., Eglin, V., Stutzmann, D., Vincent, N., et al.: ICFHR2016 competition on the classification of medieval handwritings in Latin script. In: 15th International Conference on Frontiers in Handwriting Recognition, pp. 590–595 (2016)
9. Dhali, M.A., Jansen, C.N., de Wit, J.W., Schomaker, L.: Feature-extraction methods for historical manuscript dating based on writing style development. Pattern Recogn. Lett. **131**, 413–420 (2020)
10. Droby, A., Kurar Barakat, B., Vasyutinsky Shapira, D., Rabaev, I., El-Sana, J.: VML-HP: Hebrew paleography dataset. In: Lladós, J., Lopresti, D., Uchida, S. (eds.) Document Analysis and Recognition – ICDAR 2021. LNCS, vol. 12824, pp. 205–220. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86337-1_14
11. Fiel, S., Sablatnig, R.: Writer identification and writer retrieval using the fisher vector on visual vocabularies. In: 12th International Conference on Document Analysis and Recognition, pp. 545–549 (2013)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
13. He, S., Samara, P., Burgers, J., Schomaker, L.: Discovering visual element evolutions for historical document dating. In: 15th International Conference on Frontiers in Handwriting Recognition, pp. 7–12 (2016)
14. He, S., Samara, P., Burgers, J., Schomaker, L.: Historical manuscript dating based on temporal pattern codebook. Comput. Vis. Image Underst. **152**, 167–175 (2016)
15. He, S., Sammara, P., Burgers, J., Schomaker, L.: Towards style-based dating of historical documents. In: 14th International Conference on Frontiers in Handwriting Recognition, pp. 265–270 (2014)
16. Hosoe, M., Yamada, T., Kato, K., Yamamoto, K.: Offline text-independent writer identification based on writer-independent model using conditional autoencoder. In: 16th International Conference on Frontiers in Handwriting Recognition, pp. 441–446 (2018)

17. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
18. Keglevic, M., Fiel, S., Sablatnig, R.: Learning features for writer retrieval and identification using triplet CNNs. In: 16th International Conference on Frontiers in Handwriting Recognition, pp. 211–216 (2018)
19. Richler, B.: Hebrew manuscripts in the Vatican library: catalogue, pp. 1–790 (2008)
20. Richler, B., Beit-Arié, M.: Hebrew manuscripts in the biblioteca palatina in parma: catalogue; palaeographical and codicological descriptions (2011)
21. Schor, U., Raziel-Kretzmer, V., Lavee, M., Kuflik, T.: Digital research library for multi-hierarchical interrelated texts: from 'Tikkoun Sofrim' text production to text modeling. In: Classics@18 (2021)
22. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
23. Sirat, C.: Hebrew Manuscripts of the Middle Ages. Cambridge University Press, Cambridge (2002)
24. Stökl Ben Ezra, D., Brown-DeVost, B., Jablonski, P.: Exploiting insertion symbols for marginal additions in the recognition process to establish reading order. In: Barney Smith, E.H., Pal, U. (eds.) Document Analysis and Recognition – ICDAR 2021 Workshops. LNCS, vol. 12917, pp. 317–324. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86159-9_22
25. Studer, L., et al.: A comprehensive study of imagenet pre-training for historical document image analysis. In: International Conference on Document Analysis and Recognition, pp. 720–725 (2019)
26. Szegedy, C., et al.: Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
27. Vidal-Gorène, C., Decours-Perez, A.: A computational approach of Armenian paleography. In: Barney Smith, E.H., Pal, U. (eds.) Document Analysis and Recognition – ICDAR 2021 Workshops. LNCS, vol. 12917, pp. 295–305. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86159-9_20
28. Wecker, A.J., et al.: Tikkoun sofrim: a webapp for personalization and adaptation of crowdsourcing transcriptions. In: Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, pp. 109–110 (2019)
29. Wolf, L., Potikha, L., Dershowitz, N., Shweka, R., Choueka, Y.: Computerized paleography: tools for historical manuscripts. In: 18th IEEE International Conference on Image Processing, pp. 3545–3548 (2011)
30. Yardeni, A., et al.: The Book of Hebrew Script: History, Palaeography, Script Styles, Calligraphy and Design. Carta Jerusalem, Jerusalem (1997)