2025 International Conference on Electrical, Computer and Communication Engineering (ECCE), 13-15 February 2025, CUET, Chattogram-4349, Bangladesh

Egyptian Hieroglyphics Handwriting Character Classification Using CNN with Explainable AI: SHAP and LIME

Md. Ismiel Hossen Abir
Dept. of Computer Science & Engineering
International Standard University
Dhaka, Bangladesh
Ismielabir286@gmail.com

MD. Shaek Ahmed
Dept. of Computer Science & Engineering
International Standard University
Dhaka, Bangladesh
shaekahmed69@gmail.com

Nurul Bashar
Dept. of Computer Science & Engineering
International Standard University
Dhaka, Bangladesh
nurulbashar21@gmail.com

Hasibul Islam Peyal
Dept. of Computer Science & Engineering
Rajshahi University of Engineering & Technology
Rajshahi, Bangladesh
hasibpeyal9@gmail.com

***Abstract*— Egyptian hieroglyphics, one of the oldest writing and human communication systems. In our modern life, it's challenging to interpret this language due to its complex visual symbols. In this study, we propose a Convolutional Neural Network (CNN) model for classifying Egyptian hieroglyphic handwriting characters. The dataset provides 18 different class of handwritten hieroglyph character images. To enhance model interpretability, we apply explainable AI techniques, specifically SHAP and LIME, to identify regions that influence model predictions. The results that our custom CNN model achieves train accuracy 90.62%, validation accuracy 88.25%, and test accuracy of 84.5%, with specific characters showing high classification performance. Also, AUC score reflects 0.99 to 1.00 for each of the classes.**

Keywords*— *Lightweight Convolutional Neural Networks, CNN, SVM Classifier, Potato, Explainable AI (XAI) Techniques, SHAP, Lime,

I. Introduction

Hieroglyphic one of the oldest forms of written language was used in ancient Egypt. Hieroglyphic symbols show objects but usually represent certain sounds or groups of sounds [1]. Despite its rich historical significance, hieroglyphics are challenging to interpret in modern times due to the complex visual patterns of each character or glyph. That makes the script difficult to read and understand. To solve this problem, we require a systematic approach to recognizing each glyph individually. In this study, we propose a hieroglyphic handwriting character classification method using a Convolutional Neural Network (CNN) model to identify each of the glyph. We also use explainable AI technique such as SHAP and LIME to enhance our custom CNN model explain ability. For this work, we use hieroglyph handwriting character classification dataset where there are 18 different classes [2].

A few studies and techniques are used in the past research to classify Egyptian Hieroglyphics character recognition. Some works are classification, segmentation and text translation hieroglyphs to English language related. Barucci et al. proposed various CNN architectures such as ResNet-50, Inception-v3, Xception, and a custom model named Glyphnet, their proposed Glyphnet model achieved best accuracy which is 97.6 by using training from scratch [3]. Elnabawy et al. proposed an Optical Character Recognition (OCR) approach using Histogram of Oriented Gradients (HOG) for segmenting and recognizing hieroglyphic characters [4]. Guidi et al. utilized the Mask R-CNN model to segment ancient Egyptian hieroglyphs from varied image sources, achieving best performance [5]. Nederhof developed a proof-of-concept OCR tool specifically tailored to transcribe handwritten Ancient Egyptian hieroglyphic texts, successfully recognizing signs and encoding them [6]. Guo et al. proposed a hierarchical representation combining Gabor-related low-level and sparse-encoder mid-level features, achieving improved accuracy when integrated with CNNs [7].

The motivation for this study comes from the observation that there are only a few works related to Egyptian hieroglyphics. There are still lots of research gap and opportunity in this arena. Therefore, we use a CNN model for hieroglyphic handwriting character classification and explainable AI techniques to enhance the interpretability of our model.

Key contribution of this work:

- Proposed a lightweight custom CNN model to classify 18 classes of hieroglyph characters.
- SHAP and LIME were used to clarify models performance.

This research organized into several key sections. In section II we described methodologies part including dataset preprocessing and model architecture. In section III we describe experimental results and section IV we discussed comparative analysis and section V we conclude our research. In section VI list of the reference cited.

979-8-3503-5750-9/25/$31.00 ©2025 IEEE

2025 International Conference on Electrical, Computer and Communication Engineering (ECCE) | 979-8-3503-5750-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ECCE64574.2025.11013519

II. METHODOLOGIES

In this research, we use CNN model with explainable AI such as SHAP and LIME to identify handwriting image classification of ancient Egypt hieroglyphics characters. We use 18 classes of different hieroglyphics character.

A. Dataset Collection and Splitting

In this dataset, we use 18 different classes of hieroglyphic characters. The dataset was created using handwritten images of hieroglyphic characters. Each of the class there are approximately 50 to 100 images. There are total 1288 images. We split this dataset as 80:20 rules wise where 1030 images used for training and 129 and 129 images used for test and validation purpose. According to the figure 1, it shows the 8 different classes of hieroglyph characters.

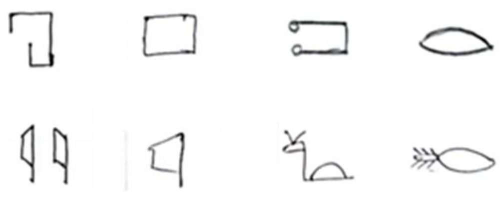

Fig. 1. Dataset Sample Images

B. Data Augmentation

Hieroglyph is a low resources language and this dataset also has less number of images. To overcome this problem, we use data augmentation techniques. In the training part of this dataset, we used random rotation, width shift, height shift, shearing, zooming, and horizontal flip techniques. Another reason for using data augmentation techniques is to handle real-world scenarios effectively. For the test and validation dataset, we don't use any augmentation method. We also resize the entire image into one size which is 100 X 100 pixels.

C. Proposed CNN Model Architecture

Our proposed CNN model represent in the figure 2.

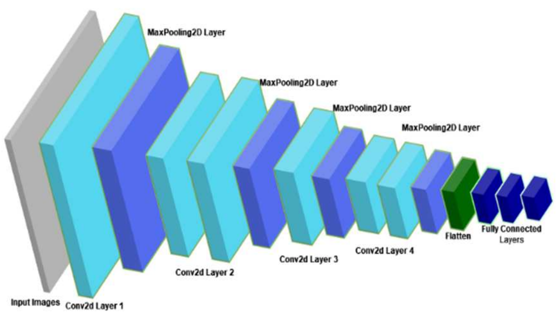

Fig. 2. Proposed CNN Model Architecture

This model consists of four convolutional layers with max-pooling layers, followed by a flatten layer, two fully connected layers, and a final fully connected layer for output. In this CNN model, we use 32 filters and 3 X 3 sizes with activation function Relu. In the second layer we use 64 filters and third and fourth layer we use 128 filters. We also used Maxpooling layers and Batch Normalization with each four parts of the model. We also used L2 regularization technique to improve model generalization and prevent overfitting.

D. Explainable AI

In this study, we use SHAP and LIME techniques. SHAP (SHapley Additive exPlanations) provides a unified framework for interpreting complex model predictions by assigning importance values to each feature that improved efficiency and interpretability [8]. LIME is an explanation technique that interprets the predictions of any classifier by locally learning an interpretable model [9].

E. Method Evaluation Metrics

To accurately classify the model we use several evaluation metrics such as accuracy score, precision, recall, f1-score etc.

$$Accuracy\ \frac{TP+TN}{TP+TN+FP+FN} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

$$Precision = \frac{TP}{TP+FP} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

$$Recall = \frac{TP}{TP+FN} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

$$F1\_Score = \frac{Precision\ x\ Recall}{Precision+Recall} \ldots\ldots\ldots\ldots\ldots\ldots(4)$$

Here, *TP, TN, FP, and FN* represent true positives, true negatives, false positives, and false negatives. We also used classification report, confusion matrix, ROC curve, AUC score as evaluation of our result parts.

III. EXPERIMENTAL RESULTS

In this experimental results part we divided it into three parts, in the below we discuss each of the parts results:

A. Quantitative Analysis

According to the table 1, this classification model for hieroglyphic characters achieved an overall accuracy of 84%, effectively mapping hieroglyphs characters. Hieroglyphic characters equivalent English letter 'z', 't', and 'sh' reached high F1-scores, reflecting strong precision and recall, while other characters such as 'r' and 's' showed lower recall values. According to the figure 3, it reflects the confusion matrix of this model. This confusion matrix shows the performance of the model in classifying hieroglyphic characters higher numbers of correct predictions. The model performs well for certain characters such as 'h', 't', and 'n', which have higher counts along the diagonal, while other characters, such as 'r' and 'qu,' show some misclassifications.

B. Qualitative Analysis

According to the figure 4, it shows the ROC curve. This ROC curve shows the model's performance among all 18 different hieroglyphic character classes. According to the table 2, most classes have an AUC close to 1.0, indicating excellent performance of the model. Some classes achieving a perfect score of 1.0, which reflects high accuracy in predictions.

TABLE I. CLASSIFICATION REPORT

Hieroglyphics Characters Equivalent English letters	Precision	Recall	F1-score
I	1.00	0.67	0.80
a	0.83	1.00	0.91
b	0.86	0.86	0.86
ch	1.00	0.71	0.83
f	0.86	0.86	0.86
h	0.78	0.93	0.85
i	0.78	1.00	0.88
k	0.64	1.00	0.78
kh	0.83	0.71	0.77
n	1.00	0.86	0.92
p	0.78	0.88	0.82
qu	0.83	0.71	0.77
r	1.00	0.57	0.73
s	1.00	0.50	0.67
sh	0.86	1.00	0.92
t	1.00	0.86	0.92
y	0.80	1.00	0.89
z	1.00	1.00	1.00
accuracy			0.84

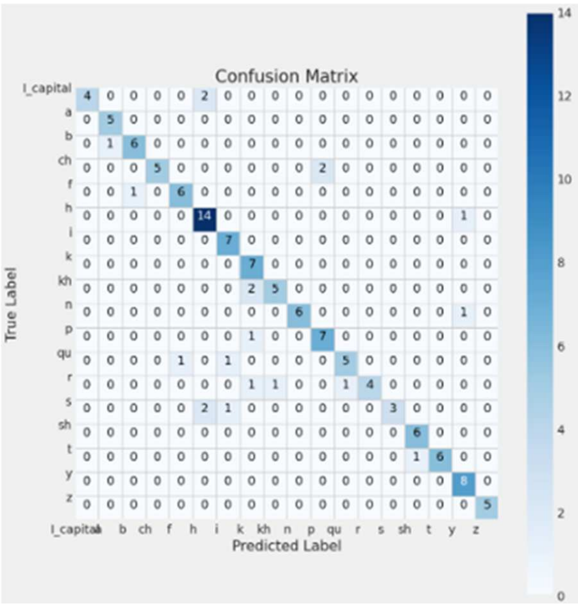

Fig. 3. Confusion Matrix

C. Explainability Analysis

According to the figure 5, it reflects the LIME. The left side of this image shows the original image and the right side of the image reflects the LIME explanation of hieroglyphic character. In the right side image, highlighting in yellow contributed most to the model's prediction of Class 5. The model almost correctly identifies the character. The LIME visualization helps identify which regions of the character was most influential, providing insight into the model’s interpretability by indicating the areas that led to the classification decision.

According to the figure 6 it reflects the several hieroglyph characters with SHAP explanation. There are original image and it SHAP explanation combined. Each of the SHAP explanation mainly reflected two colors one is red and another one is blue. Red color regions reflect increased the probability of a certain class. On the other hand, blue showing regions that decreased it. These visualizations allow us to see which parts of each character were most significant in the model’s classification. This identification provides insights into the model’s interpretability and impact decision-making process.

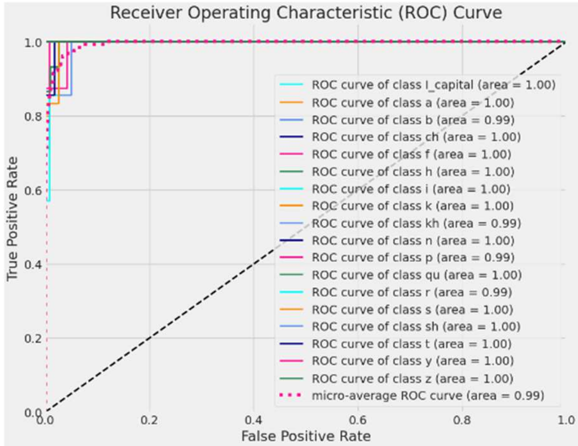

Fig. 4. ROC Curve

TABLE II. AUC SCORE

AUC For Class	Score
I	**1.00**
a	**0.99**
b	**1.00**
ch	**1.00**
f	**1.00**
h	**1.00**
i	**1.00**
k	**1.00**
kh	**0.99**
n	**1.00**
p	**0.99**
qu	**1.00**
r	**0.99**
s	**1.00**
sh	**1.00**
t	**1.00**
y	**1.00**
z	**1.00**
Micro-average AUC	**0.99**
Macro-average AUC	**1.00**

IV. COMPARATIVE ANALYSIS

In this research, we achieve 90.62% train, 88.25 validation, and 84.5% test accuracy. Only a few words are held based on the hieroglyph character classification. Some past research are based on segmentation and Hieroglyph text to translation English test. However, significant research gap is past studies are not including explainable AI techniques. In this research we used explainable AI techniques such as SHAP and LIME to visualization and increase model explain ability.

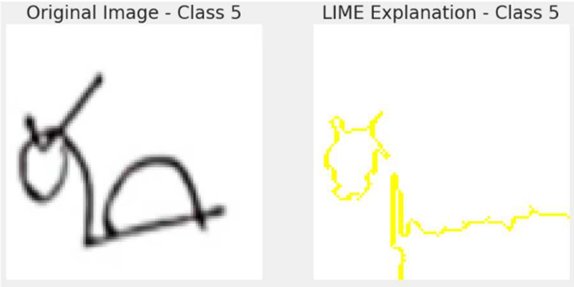

Fig. 5. LIME

Fig. 6. SHAP

V. Conclusion

In this research, we presented a CNN-based model for Egyptian hieroglyphic handwriting character classification. We tried to identify and interpret one of the world's oldest writing systems. Our approach shows that a lightweight CNN model, in combination with explainable AI techniques such as SHAP and LIME. We achieve effective classification accuracy. The model achieved train, validation, and test accuracies of 90.62%, 88.25%, and 84.5%. We also achieve high AUC scores close to 1.0 across all classes. Our work contributes to the field by addressing the challenges of interpreting complex visual symbols inherent in hieroglyphs. The use of SHAP and LIME allowed us to pinpoint specific image regions that were crucial in character prediction. However, there some areas for improvement, such as increasing the dataset size. In future work, we can explore hieroglyph to English translation. This study opens new possibilities for future research in ancient language interpretation using modern AI techniques and suggests that similar models could be adapted to other low-resource languages.

References

[1] The Editors of Encyclopedia Britannica, "Hieroglyph | writing character," Encyclopædia Britannica. Jan. 08, 2017. Available: https://www.britannica.com/topic/hieroglyph

[2] Md. I. H. Abir, Md. Jahid, and N. S. Liza, "Handwriting Hieroglyphs Alphabet Recognition," IEEE Dataport, Sept. 14, 2024. [Online]. Available: https://dx.doi.org/10.21227/qhw2-d161.

[3] A. Barucci, C. Cucci, M. Franci, M. Loschiavo, and F. Argenti, "A Deep Learning Approach to Ancient Egyptian Hieroglyphs Classification," IEEE Access, vol. 9, pp. 123438–123447, 2021, doi: https://doi.org/10.1109/ACCESS.2021.3110082.

[4] Reham Elnabawy, R. Elias, and M. Salem, "Image Based Hieroglyphic Character Recognition," Nov. 2018, doi: https://doi.org/10.1109/sitis.2018.00016.

[5] T. Guidi, L. Python, M. Forasassi, C. Cucci, M. Franci, F. Argenti, and A. Barucci, "Egyptian Hieroglyphs Segmentation with Convolutional Neural Networks," Algorithms, vol. 16, no. 2, p. 79, 2023. [Online]. Available: https://doi.org/10.3390/a16020079.

[6] M.-J. Nederhof, "OCR of handwritten transcriptions of Ancient Egyptian hieroglyphic text *." Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=1811b69e530e255def426381764b4aa0fbcaed03

[7J. Guo, C. Wang, E. Roman-Rangel, H. Chao, and Y. Rui, "Building Hierarchical Representations for Oracle Character and Sketch Recognition," vol. 25, no. 1, pp. 104–118, Jan. 2016, doi: https://doi.org/10.1109/tip.2015.2500019.

[8] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," arXiv:1705.07874 [cs, stat], Nov. 2017, Available: https://arxiv.org/abs/1705.07874.

[9] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," arXiv.org, Feb. 16, 2016. https://arxiv.org/abs/1602.04938.