

<https://doi.org/10.31891/csit-2026-3-1>

UDC 004.93

**Andrii IVASECHKO**

Phd. Student, Lecturer  
West Ukrainian National University  
<https://orcid.org/0009-0002-8623-7828>  
[a.ivasechko@wunu.edu.ua](mailto:a.ivasechko@wunu.edu.ua)

**Khrystyna LIPIANINA-  
HONCHARENKO**

Doctor of Technical Sciences, Associate  
professor  
West Ukrainian National University  
<https://orcid.org/0000-0002-2441-6292>  
[kh.lipianina@wunu.edu.ua](mailto:kh.lipianina@wunu.edu.ua)

Received: 21/07/2026

Accepted: 02/09/2026

Published: 30/09/2026

© Copyright  
2026 by the author(s)



This is an Open Access article distributed  
under the terms of the [Creative Commons  
CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

## AI-ASSISTED PIPELINE FOR SEMI- AUTOMATIC CONSTRUCTION OF HTR CORPORA FROM HISTORICAL DOCUMENTS

*This paper presents an improved software application for the semi-automatic annotation of handwritten words from historical documents using a CRNN deep learning model. The proposed approach integrates an artificial intelligence module directly into the annotation process, automatically generating a preliminary transcription for each image. It implements the Human-in-the-Loop concept, where the prediction serves as a recommendation that can be accepted, corrected, or replaced by the annotator, preserving expert control over the final annotation while reducing repetitive manual input.*

*The application provides automated ordering of images according to metadata, structured storage of transcriptions and document information, text normalization, detection of already processed samples, and resumption of interrupted annotation sessions. The AI prediction module is integrated directly into the graphical interface, allowing the annotator to review and edit the generated transcription without switching between separate tools.*

*The CRNN model was trained on a custom corpus of 2,763 segmented handwritten word images with diverse graphical and linguistic characteristics. Training employed image augmentation, AdamW optimization, CTC loss, Gradient Clipping, adaptive learning-rate reduction, and Early Stopping. The main focus of the study is the evaluation of annotation performance and the effect of AI-assisted predictions on user productivity.*

*Experimental results showed that the average processing time decreased from 8.4 s per word with manual transcription to 4.7 s when an AI prediction was available. This corresponds to a 1.79-fold increase in processing speed and a 44.0% reduction in processing time. Annotator productivity increased from 7.1 to 12.8 words per minute. These results demonstrate the practical effectiveness of integrating HTR prediction into an expert annotation workflow and its potential to accelerate the construction of training corpora for historical handwritten text recognition systems while maintaining expert verification of the final annotations.*

*Keywords: handwritten text recognition, HTR, CRNN, Human-in-the-Loop, data annotation, historical documents, deep learning, handwritten corpora, AI-assisted annotation, annotation performance.*

### Introduction

The digitization of historical handwritten documents is an important area of cultural heritage preservation and research. One of the key tasks in this process is Handwritten Text Recognition (HTR), which enables images of handwritten documents to be converted into structured text. Unlike the recognition of modern printed documents, HTR is considerably more challenging due to handwriting variability, non-standard orthography, damage to historical documents, overlapping characters and words, and the limited availability of high-quality annotated training data.

An important stage in the development of HTR systems is the creation of training corpora. This requires images of handwritten documents to be segmented, structured, and provided with corresponding textual transcriptions. Traditional manual annotation provides a high level of control over data quality but requires considerable time. This process is particularly challenging for historical documents, where accurate transcription often requires a specialist capable of taking into account handwriting characteristics, language, and historical orthography. As the corpus size increases, the time required to process individual samples becomes one of the key factors limiting the speed of training data generation.

To reduce the amount of manual work, modern research employs deep learning models, including Convolutional Recurrent Neural Networks (CRNN), transformer-based architectures, as well as augmentation and synthetic handwritten data generation methods. Such models can automatically generate predicted transcriptions that can be used as a basis for further processing. However, fully automated annotation may result in the accumulation of errors, negatively affecting the quality of the resulting training corpus. Therefore, for historical documents, it is appropriate to use automatic recognition not as a replacement for the expert, but as a means of reducing the amount of manual work required.

A promising approach to this problem is the Human-in-the-Loop concept, in which the automatic model does not replace the annotator but acts as an intelligent assistant. In this approach, the system generates an initial transcription hypothesis, after which the user verifies, edits, or rejects the proposed result. This makes it possible to combine the speed of automatic processing with expert quality control. At the same time, the practical effectiveness of such an approach is determined not only by the quality of the model predictions but also by the extent to which they reduce the time required by the user to obtain the final transcription.

Thus, there is a need for software tools that directly integrate HTR models into annotation environments and make it possible to evaluate the impact of automatic predictions on user productivity. This paper proposes an intelligent application for handwritten word annotation based on a CRNN model, which provides automatic transcription prediction, expert verification, and storage of structured annotations.

### **Analysis of recent research and problem formulation**

The development of modern Handwritten Text Recognition (HTR) systems is largely associated with the transition from heuristic methods to deep learning models. Carlos Garrido-Munoz et al. conducted a comprehensive review of the evolution of HTR approaches and systematized existing methods according to the recognition level, ranging from individual characters and words to text lines and complete documents. The authors also identified the main challenges in the field, including significant handwriting variability, the complexity of historical documents, and the insufficient availability of high-quality annotated training corpora [1]. The latter problem is particularly important, as the creation of HTR corpora largely depends on manual work by experts.

An important stage in the preparation of HTR data is manuscript segmentation. Irina Rabaev and Marina Litvak conducted a systematic review of text-line segmentation and baseline detection methods for historical manuscripts based on 49 studies. The authors emphasize the importance of accurate text-region extraction for subsequent recognition and highlight the need for more universal methods for processing historical documents [2]. Similar problems are addressed in the work of Wissam AlKendi et al., devoted to HTR of historical French documents, where handwriting variability, overlapping characters, and the lack of specialized training data are identified among the key challenges [3].

Methods for data augmentation and synthetic generation of handwritten images are also being investigated to improve the quality and volume of training data. Y. H. Rassul et al. analyzed 848 scientific publications covering traditional augmentation methods, GANs, diffusion models, and transformer-based architectures. The authors demonstrate the significant potential of synthetic data for improving HTR model performance and reducing dependence on large amounts of manually annotated data [4]. Mahdi Bonyani et al. investigated the use of DenseNet and Xception architectures combined with data augmentation for Persian handwriting recognition, demonstrating the effectiveness of modern deep neural networks [5]. Mohammed Al-Hitawi and M. A. Alazaizi developed the HuHTR-Toolkit for generating synthetic handwritten data in Hungarian. Using the proposed toolkit, more than three million image-text pairs were generated, demonstrating the potential for significant scaling of the HTR data preparation process [6].

Another important direction of current research is the use of artificial intelligence directly in the process of annotation verification and creation. Sana Al-Azzawi, Elisa Barney, and Marcus Liwicki proposed the two-level CER-HV framework, which uses a CRNN model to automatically detect potential transcription errors followed by expert verification. This approach demonstrates the effectiveness of combining automated analysis with human supervision according to the Human-in-the-Loop concept [7]. Silvia Cascianelli et al. investigated a modified CRNN architecture incorporating Deformable Convolutions, aimed at improving recognition robustness to irregular handwriting and geometric distortions [8].

Another research direction involves the development of comprehensive software pipelines that combine segmentation, recognition, and user interaction. Yee Fan Tan et al. proposed an integrated pipeline covering text localization, segmentation, recognition, and subsequent semantic analysis of documents. Pranav Guruprasad et al. developed an interactive pipeline for HTR dataset creation in which automatic detection, segmentation, and recognition are combined with a web interface for user verification of the results [9]. Such systems demonstrate a general trend toward transferring individual stages of data preparation from fully manual to semi-automated processing.

At the same time, most of the reviewed studies focus primarily on recognition accuracy, segmentation quality, or increasing the volume of training data, while the practical impact of AI predictions on annotator productivity has received comparatively limited attention. In particular, an important question remains as to the extent to which

integrating an HTR model directly into an annotation interface can reduce the time required to process an individual sample compared with fully manual transcription.

In previous work by A. Ivasechko and K. Lipianina-Honcharenko, a semi-automatic pipeline for constructing HTR corpora from historical documents was proposed, focusing on the segmentation and annotation of individual characters [10]. In contrast, the present study develops an intelligent application for annotating handwritten words, in which a pretrained CRNN model is directly integrated into the annotator's workflow. The model automatically generates an initial transcription that can either be accepted or edited by an expert, thereby implementing the Human-in-the-Loop concept.

Thus, the aim of this study is to develop and experimentally evaluate an intelligent tool for accelerating the annotation of historical handwritten documents, in which automatic HTR prediction is used as an auxiliary tool for the expert. The primary focus is placed not only on the quality of the generated transcriptions but also on evaluating the performance of the annotation process, the processing time per sample, and the increase in user productivity resulting from the use of AI-generated suggestions.

### Presentation of the main material

The conducted analysis showed that, despite significant progress in the development of HTR models and data preparation tools, insufficient attention has been paid to the integration of modern recognition models into software tools for annotating historical manuscripts. This highlights the need for intelligent annotator support systems that implement the Human-in-the-Loop concept and combine automatic transcription prediction with expert verification. The present study addresses this problem by proposing an application for annotating handwritten words with an integrated CRNN model.

To improve the productivity of the handwritten text dataset creation process, the interactive annotation pipeline was enhanced to implement the Human-in-the-Loop concept. Unlike traditional manual annotation systems, the proposed software combines annotation process management tools with an automatic transcription prediction module based on a CRNN model. This approach makes it possible to use automatic recognition results as recommendations for the annotator while maintaining full expert control over the creation of reference annotations. The overall architecture of the developed application is presented in Figure 1.

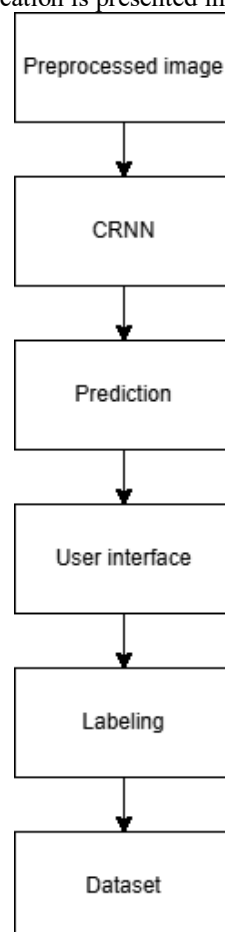


Fig. 1. General architecture of the application

The application operates on pre-segmented, binarized images of handwritten words with a fixed height of 64 pixels and an adaptive width, which makes it possible to preserve the proportions of the original image while providing

a unified representation of the input data. Each image has a unique filename whose structure contains information about its position within the original archival document. The filename format is `fond-257_1_1_ukr_page1_line1_word1`, where `fond-257` identifies the archival fond number; the first numerical field after the fond identifier specifies the file or archival storage unit within the corresponding fond; the second numerical field represents the operator index who performed the initial processing or created the corresponding dataset; `ukr` specifies the language or graphic category; and the `page1`, `line1`, and `word1` fields correspond to the page number, line number, and sequential word number within the line, respectively. This naming structure provides unambiguous image identification while preserving information about its hierarchical position within the document structure.

After startup, the application generates a list of images from the specified directory and automatically sorts them according to the structural attributes encoded in the filenames. For this purpose, a filename parsing mechanism based on regular expressions is used to extract the fond identifier, file number, operator index, page number, line number, and word number. The extracted values are used as sorting keys, allowing the images to be displayed in their natural sequence corresponding to the order of the text in the original document. This approach is particularly important for historical materials, as it preserves the relationship between an individual word and its original context within the document.

For each image, a corresponding record is created in the `words_metadata.csv` file. The metadata table contains the fields `word_id`, `image_path`, `transcription`, `normalized_text`, `language`, `writer_id`, `document_id`, `page_id`, `line_id`, `word_index_in_line`, `source_type`, `split`, and `is_valid`. The `word_id` field is used as a unique identifier for the textual sample, while `image_path` contains the path to the corresponding image. The `transcription` and `normalized_text` fields store the original transcription and its normalized representation, respectively. Additional fields describe the language category, author or operator, document, page, and line, as well as the position of the word within the line. The `source_type`, `split`, and `is_valid` fields are used to specify the origin of the sample, its assignment to the corresponding dataset partition, and the validity of the record.

To ensure that the annotation process can be resumed after an interruption, the system automatically loads previously created annotations from the `words_metadata.csv` file. The records are organized as a dictionary in which the relative path to the corresponding image is used as the key. This enables the system to quickly identify already annotated words, avoid duplicate records, and automatically open the first unprocessed dataset element. Thus, the annotation process can be performed incrementally without requiring previously annotated samples to be processed again.

A separate mechanism is implemented for storing the annotator configuration. The user identifier is automatically saved to a JSON file and restored during the next application launch. The presence of the identifier makes it possible to associate completed annotations with a specific operator, enabling subsequent annotation quality control, analysis of individual annotator productivity, and assessment of inter-annotator agreement.

Before being displayed in the graphical interface, textual values undergo a normalization procedure that includes removing unnecessary whitespace, converting characters to the canonical Unicode form, and standardizing input values. Images are automatically scaled to a height of 128 pixels while preserving their proportions, providing a unified representation of the material in the interface regardless of the original image size. At the same time, the original images are stored with a height of 64 pixels, corresponding to the data format used during CRNN model training.

One of the key components of the developed system is the automatic transcription prediction module, which generates an initial recognition hypothesis for a handwritten word. Its primary function is not to provide fully automated text recognition, but rather to provide the annotator with a recommended transcription that can accelerate the manual annotation process. The architecture of the module is presented in Figure 2.

The automatic handwritten text recognition module is based on the Convolutional Recurrent Neural Network (CRNN) architecture, which combines a convolutional neural network for visual feature extraction with a bidirectional recurrent network (BiLSTM) for sequential character modeling. After preprocessing, the input image is passed through a sequence of convolutional layers incorporating Batch Normalization and Max Pooling. As a result, a compact multi-level representation of visual features is generated, while the spatial resolution of the feature maps is progressively reduced.

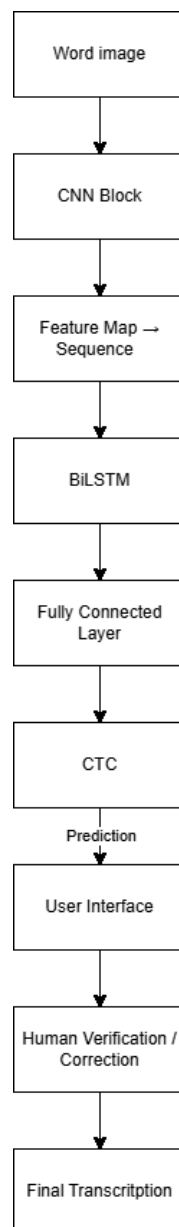
The resulting feature map is transformed into a temporal sequence of feature vectors, where individual time steps correspond to vertical regions of the word image. This representation makes it possible to formulate handwritten word recognition as a sequence analysis task. The resulting sequence is then fed into a bidirectional BiLSTM network, which captures contextual information in both forward and backward directions. This approach enables the model to account for relationships between neighboring characters and improves its robustness to handwriting variability, differences in character width, and local distortions characteristic of historical handwritten documents.

At the final stage, the outputs of the recurrent network are processed by a fully connected layer that produces probability estimates for the correspondence of each time step to characters from the target alphabet. The final transcription is obtained using Connectionist Temporal Classification (CTC), which allows the model to be trained and used without prior segmentation of the word image into individual characters.

In the developed application, the predictor module is invoked when a new image is opened. It uses the pre-trained CRNN model to automatically generate a predicted transcription. The resulting output is passed to the

graphical interface and displayed in the AI Prediction field. The prediction is used as a recommendation for the annotator: the proposed text can be fully accepted, partially edited, or completely replaced by the expert. Thus, the system implements the Human-in-the-Loop concept, in which the artificial intelligence model serves as an intelligent assistant, while the final decision regarding the correctness of the transcription remains with the user.

The model was trained on a custom dataset containing 2,763 images of handwritten words. The dataset is characterized by substantial linguistic and graphical diversity: 1,006 samples contain Cyrillic texts, 1,000 contain Latin texts, 471 represent texts written using historical Ukrainian orthography, 154 contain Ukrainian Latin transliteration, and 132 are Polish-language samples. The mean word length is 5.64 characters, while the median length is 5 characters. At the same time, the corpus contains both short words and single-character function words, as well as long words exceeding 10 characters. Such variability results in input sequences of different lengths and makes it possible to evaluate the model's ability to process textual units of varying complexity.



**Fig. 2. Architecture of the AI Prediction Module**

A distinctive feature of the dataset is the presence of historical texts in which characters belonging to different modern linguistic systems may occur within the same word, such as the Ukrainian letter “ї” and the Russian “ы”. Therefore, such samples are assigned the label “cyr”, which indicates that the text belongs to the Cyrillic writing system rather than to a specific modern language. This approach helps avoid ambiguity during the preparation of training data and focuses the model on recognizing the graphical characteristics of the handwriting.

To improve the model's generalization capability, several image augmentation techniques were applied during training, including random rotation, Gaussian noise addition, brightness and contrast adjustments, and motion

blur simulation. These transformations partially reproduce distortions that may occur during the digitization of historical handwritten documents and help reduce the risk of overfitting.

The network parameters were optimized using the AdamW algorithm with an initial learning rate of  $10^{-4}$  and Weight Decay regularization. The Connectionist Temporal Classification Loss (CTC Loss) was used as the loss function, eliminating the need for prior segmentation of words into individual characters. To stabilize the training process, Gradient Clipping was applied with a maximum gradient norm of 5.0.

$$\|g\|_2 \leq 5.0 \quad (1)$$

The learning rate was automatically reduced using the ReduceLROnPlateau scheduler when no improvement was observed on the validation set. Additionally, an Early Stopping mechanism was employed to terminate training after six consecutive epochs without improvement in the validation performance.

The recognition quality was evaluated using the Character Error Rate (CER) and Word Error Rate (WER) metrics. The CER is defined as the ratio of the minimum number of edit operations—insertions, deletions, and substitutions—required to transform the predicted transcription into the reference transcription to the total number of characters in the reference text:

$$CER = \frac{S+D+I}{N} \quad (2)$$

where S is the number of substituted characters, D is the number of deleted characters, I is the number of inserted characters, and N is the total number of characters in the reference transcription.

The Word Error Rate (WER) is calculated analogously at the word level:

$$WER = \frac{S_w+D_w+I_w}{N_w} \quad (3)$$

where  $S_w$ ,  $D_w$ ,  $I_w$  denote the numbers of substituted, deleted, and inserted words, respectively, while  $N_w$  represents the total number of words in the reference transcription. These metrics make it possible to evaluate both character-level recognition quality and the correctness of complete word transcriptions.

The CER measures the ratio of the minimum number of edit operations—insertions, deletions, and substitutions—required to transform the predicted transcription into the reference transcription to the total number of characters in the reference text. The WER is defined analogously at the level of complete words and represents the proportion of incorrectly recognized textual units. Additionally, Word Accuracy was calculated during validation to measure the proportion of words recognized completely without errors.

The best training result was obtained at epoch 36. On the validation set, the model achieved CER = 0.1298, WER = 0.4053, and Word Accuracy = 0.6024. These results indicate sufficient automatic transcription quality for using the model as an intelligent assistant for the annotator. In particular, even when the prediction is not completely correct, a partially generated transcription can be corrected by the user considerably faster than entering the entire transcription manually from scratch. Thus, integrating the CRNN model into the graphical interface provides a practical implementation of the Human-in-the-Loop approach and creates a basis for reducing the time required for manual annotation of historical handwritten materials.

Fig. 3. User Interface

The user interface shown in Figure 3 is designed to minimize the number of operations required to annotate a single word. At the top of the window, a progress indicator displays the current annotation status. Below it, a binarized image of the word with a height of 128 pixels is displayed, with options for full-screen viewing and zooming to enable detailed inspection. The metadata of the current sample is presented below the image, including the document identifier, page number, line number, word index, and information about the document language.

The automatic transcription prediction module is integrated directly into the interface, with its output displayed in the AI Prediction field. This is followed by a field for entering the annotator identifier and two main text fields: Transcription, used to enter the original transcription of the word, and Normalized text, containing its normalized form. If no normalized variant is provided, the system automatically uses the value from the Transcription field, reducing the number of repetitive operations required during annotation.

After the entered data are confirmed, the system generates a structured record containing the unique word identifier, path to the corresponding image, transcription, normalized text, linguistic and document metadata, annotator identifier, and auxiliary attributes, including the source type, dataset split, and validity flag. Before saving, the records are automatically ordered according to the hierarchy “document – page – line – word” and stored in a CSV file suitable for subsequent training of handwritten text recognition models.

To improve annotation efficiency, a system of keyboard shortcuts was implemented, allowing the main operations to be performed without using a mouse. Pressing Enter saves the current annotation and automatically moves to the next word. The Right and Left Arrow keys provide navigation between the next and previous words, respectively, Tab is used to skip the current sample, and Delete removes an invalid image. After any operation, the cursor automatically returns to the Transcription field, minimizing auxiliary user actions and increasing annotation speed.

The proposed application combines automatic transcription prediction with interactive expert verification, implementing an efficient Human-in-the-Loop approach to the construction of high-quality handwritten text datasets. Combining CRNN-based recommendations with the ability to promptly edit the predicted transcription makes it possible to substantially reduce annotation time while maintaining full control over the quality of the resulting data.

The experimental evaluation demonstrated that integrating the CRNN model significantly increased the speed of handwritten word annotation. The average processing time for a single word with manual transcription was 8.4 s, whereas the use of an AI-assisted prediction reduced this time to 4.7 s. Thus, the model accelerated the annotation process by a factor of 1.79 and reduced the average processing time per sample by 44.0%. Accordingly, annotator productivity increased from 7.1 to 12.8 words per minute. These results confirm the practical effectiveness of integrating automatic transcription prediction into the Human-in-the-Loop workflow and demonstrate that AI assistance not only increases the level of automation but also directly reduces the labor required to construct an annotated corpus.

### Conclusions

This paper presents the development of an intelligent handwritten text recognition module designed to support the annotation of historical documents. The proposed approach is based on the CRNN architecture, which combines convolutional visual feature extraction with sequential modeling using BiLSTM and CTC decoding. This architecture enables handwritten word recognition without prior segmentation into individual characters.

The experimental evaluation was conducted on a custom corpus containing 2,763 images of handwritten words, characterized by substantial graphical and linguistic diversity. To improve model robustness, image augmentation, the AdamW optimization algorithm, Gradient Clipping, adaptive learning-rate reduction, and Early Stopping were employed. The best validation result was obtained at epoch 36, with CER = 0.1298, WER = 0.4053, and Word Accuracy = 0.6024. These results demonstrate the ability of the model to generate automatic transcriptions suitable for practical use in the annotation of historical handwritten materials.

The practical value of the proposed solution lies in integrating the recognition model directly into the annotation workflow. Unlike a fully automated approach, the system implements the Human-in-the-Loop concept, in which the CRNN prediction serves as a recommendation, while the final transcription is determined by the expert. This approach combines the speed of automatic processing with the ability to perform expert verification of complex historical texts.

The experimental evaluation of annotation performance demonstrated a substantial reduction in processing time when AI predictions were used. The average annotation time decreased from 8.4 s to 4.7 s per word, corresponding to a 1.79-fold increase in processing speed and a 44.0% reduction in annotation time. As a result, annotator productivity increased from 7.1 to 12.8 words per minute. These findings confirm the practical effectiveness of integrating CRNN-based predictions into the annotation workflow and demonstrate that AI assistance can reduce the manual effort required to construct handwritten text corpora while preserving expert control over the final annotations.

Further research should focus on increasing the size and balance of the training corpus, optimizing the CRNN architecture, investigating the impact of different augmentation strategies, and evaluating the effectiveness of AI-assisted annotation with different annotators and document types. An additional promising direction is the

development of adaptive prediction and user-interface mechanisms aimed at further reducing annotation time while maintaining the reliability of expert-verified transcriptions.

#### ADDITIONAL INFORMATION

#### AUTHOR CONTRIBUTIONS

Conceptualization, K.L.-H.; methodology, A.I.; validation, K.L.-H.; formal analysis, A.I.; investigation, A.I.; data curation, K.L.-H.; writing-original draft preparation, A.I.; writing-review and editing, K.L.-H.; visualization, A.I.; supervision, K.L.-H.; project administration, K.L.-H. All authors have read and agreed to the published version of the manuscript.

#### DECLARATION ON THE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE TOOLS

In preparing this work, the author used DeepL Translate and Grammarly for: grammar and spelling checks, paraphrasing, and rephrasing. After using these tools/services, the author reviewed and edited the content and takes full responsibility for the content of this publication.

#### REFERENCES

- [1]C. Garrido-Munoz, A. Rios-Vila and J. Calvo-Zaragoza, "Handwritten Text Recognition: A Survey," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 48, no. 4, pp. 4367-4387, April 2026, <https://doi.org/10.1109/TPAMI.2025.3646002>
- [2]Rabaev, I., Litvak, M. Recent advances in text line segmentation and baseline detection in historical document images: a systematic review. *IJDAR* 29, 3–39 (2026). <https://doi.org/10.1007/s10032-025-00526-w>
- [3]AlKendi, W., Gechter, F., Heyberger, L., & Guyeux, C. (2024). Advancements and Challenges in Handwritten Text Recognition: A Comprehensive Survey. *Journal of Imaging*, 10(1), 18. <https://doi.org/10.3390/jimaging10010018>
- [4]Rassul, Y. H., Ahmed, A. M., Fattah, P., Hassan, B. A., Abdulkareem, A. W., Rashid, T. A., & Lu, J. (2025). Advancing offline handwritten text recognition: A systematic review of data augmentation and generation techniques. *arXiv preprint arXiv:2507.06275*.
- [5]Bonyani, M., Jahangard, S., & Daneshmand, M. (2021). Persian handwritten digit, character and word recognition using deep learning. *International Journal on document analysis and recognition (IJDAR)*, 24(1), 133-143.
- [6]Al-Hitawi, M., & Alazaizi, M. A. (2026). Toolkit for Generating and Augmenting Hungarian Handwritten Text Recognition Dataset. *Dijlah Journal of Engineering Sciences ISSN: 3078-9664, e-ISSN: 3078-9656*, 3(1).
- [7]Al-azzawi, S., Barney, E., & Liwicki, M. (2026). A human-in-the-loop Label error detection framework applied to Arabic-script HTR datasets. *arXiv preprint arXiv:2601.16713*.
- [8]Cascianelli, S., Cornia, M., Baraldi, L., & Cucchiara, R. (2022). Boosting modern and historical handwritten text recognition with deformable convolutions: S. Cascianelli et al. *International Journal on Document Analysis and Recognition (IJDAR)*, 25(3), 207-217.
- [9]Tan, Y. F., Connie, T., Goh, M. K. O., & Teoh, A. B. J. (2022). A pipeline approach to context-aware handwritten text recognition. *Applied Sciences*, 12(4), 1870.
- [10]Ivasechko, A., & Lipianina-Honcharenko, K. (2025, November). Semi-automatic pipeline for constructing HTR corpora from Ukrainian-language historical documents. In *BAITmp* (pp. 140-151).

Андрій ІВАСЕЧКО, Христина ЛІП'ЯНІНА-ГОНЧАРЕНКО  
Західноукраїнський національний університет

### ПІДТРИМУВАНИЙ ШТУЧНИМ ІНТЕЛЕКТОМ ПРОЦЕС НАПІВАВТОМАТИЧНОГО ФОРМУВАННЯ HTR-КОРПУСІВ З ІСТОРИЧНИХ ДОКУМЕНТІВ

У статті представлено вдосконалений програмний застосунок для напівавтоматичного анотування рукописних слів з історичних документів із використанням моделі глибокого навчання CRNN. Запропонований підхід передбачає безпосередню інтеграцію модуля штучного інтелекту в процес анотування, завдяки чому для кожного зображення автоматично генерується попередня транскрипція. Реалізовано концепцію Human-in-the-Loop, відповідно до якої прогноз моделі використовується як рекомендація, що може бути прийнята, виправлена або повністю замінена анотатором. Такий підхід забезпечує збереження експертного контролю над кінцевою анотацією та водночас зменшує обсяг повторюваного ручного введення даних.

Програмний застосунок забезпечує автоматичне впорядкування зображень відповідно до метаданих, структуроване збереження транскрипцій та інформації про документи, нормалізацію тексту, виявлення вже опрацьованих зразків, а також можливість продовження перерваних сеансів анотування. Модуль прогнозування на основі штучного інтелекту безпосередньо

інтегрований у графічний інтерфейс, що дає змогу анотатору переглядати та редагувати згенеровану транскрипцію без необхідності переходу між окремими програмними засобами.

Модель CRNN було навчено на спеціально сформованому корпусі, що містить 2763 сегментовані зображення рукописних слів із різноманітними графічними та лінгвістичними характеристиками. Під час навчання застосовано методи аугментації зображень, оптимізатор AdamW, функцію втрат CTC, обмеження градієнта (Gradient Clipping), адаптивне зменшення швидкості навчання та механізм ранньої зупинки (Early Stopping). Основну увагу в дослідженні зосереджено на оцінюванні ефективності процесу анотування та визначенні впливу прогнозів, сформованих штучним інтелектом, на продуктивність користувача.

Експериментальні результати засвідчили, що середній час опрацювання одного слова зменшився з 8,4 с за ручної транскрипції до 4,7 с за наявності прогнозу, сформованого штучним інтелектом. Це відповідає збільшенню швидкості опрацювання у 1,79 раза та скороченню часу обробки на 44,0 %. Продуктивність анотатора зросла з 7,1 до 12,8 слова за хвилину. Отримані результати підтверджують практичну ефективність інтеграції HTR-прогнозування в експертний процес анотування та свідчать про потенціал запропонованого підходу для прискорення формування навчальних корпусів для систем розпізнавання історичних рукописних текстів зі збереженням експертної перевірки кінцевих анотацій.

Ключові слова: розпізнавання рукописного тексту, HTR, CRNN, Human-in-the-Loop, анотування даних, історичні документи, глибоке навчання, корпуси рукописних текстів, анотування за підтримки штучного інтелекту, ефективність анотування.