UDC 004.9
DOI: 10.15587/2706-5448.2025.343943

Mykyta Shupyliuk,
Vitalii Martovytskyi,
Yuri Romanenkov

ENHANCING WRITER IDENTIFICATION AND WRITER RETRIEVAL WITH CenSurE AND VISION TRANSFORMERS

The object of research is the process of writer identification based on handwritten text. Despite significant progress, existing methods for author identification from handwritten text have limitations that prevent them from achieving maximum accuracy and reliability.

This paper focuses on optimizing and improving the efficiency of writer identification from handwritten text by integrating image preprocessing methods, feature detection, and modern machine learning architectures. To this end, a functional model was developed that uses the CenSurE algorithm to detect key points and extract relevant image areas, and then the Vision Transformer model to identify the writer based on these extracted features. To reduce the variability of the results, experimental validation was conducted using a dual search and classification methodology. The use of the public CVL dataset increases reproducibility and helps in comparative analysis. The findings indicate that the implementation of the proposed approach leads to an improvement in the identification accuracy during retrieval, surpassing the results of other studies. This is evidenced by increased accuracy values of hard top k and soft top k by 1% and mean average precision by 2%. In addition, findings indicate significant performance improvement in the feature detection preprocessing stage. This improvement is quantitatively supported by reductions in both the average time per item and total processing duration by 39%, alongside the increase in total count of patches extracted by 70%.

The results obtained contribute to increasing the reliability of automated handwriting analysis systems, especially for the task of writer identification. This achievement is a valuable tool for graphologists and forensic document experts, supporting such critical tasks as the forensic authorship process.

***Keywords**: machine learning, writer identification, transformer, image, neural networks, handwriting, preprocessing.*

Received: 20.08.2025
Received in revised form: 24.10.2025
Accepted: 18.11.2025
Published: 29.12.2025

How to cite
Shupyliuk, M., Martovytskyi, V., Romanenkov, Y. (2025). Enhancing writer identification and writer retrieval with CenSurE and Vision Transformers. Technology Audit and Production Reserves, 6 (2 (86)), 6–14. https://doi.org/10.15587/2706-5448.2025.343943

1. Introduction

Technological advancements of humanity directly influence everyday human activities. Despite this, handwritten text continues to be used in various areas of human life. The uniqueness of handwriting is based on a person's health, age, abilities, personality traits, and character, which creates a theoretical basis for its analysis. Handwriting analysis has three main branches: graphology, handwriting expertise, and forensic linguistics [1]. Graphology studies the behavior of a person that can be recognized from handwriting [2], and its reliability is characterized by the fact that the measurable variables in handwriting are extremely consistent [1].

Writer identification during handwriting analysis is used in various areas of activity, including medicine, forensics, and legal proceedings [2]. One example of the use of handwriting analysis is its use in forensics when investigating various crimes [3]. Handwriting analysis is used to recreate a person's emotional state at the time of the crime, as handwriting can reveal a person's emotional state. This can be one of the key factors in determining the circumstances of a crime [4].

There are quite a few publications dedicated to the analysis of handwriting in various spheres of human activity. However, the widespread use of such methods is still limited by insufficient formalization of handwriting characteristics, subjective expert assessment, limited hardware, and imperfect algorithms. All of these factors prevent improvements in the accuracy and reliability of results.

Modern approaches to handwriting analysis often do not take into account handwriting variability and require modification to improve the accuracy of the analysis results. Most approaches do not take into account the physical and/or emotional state of the person, the type of instrument used to write the text, the speed of writing, and language characteristics. All these factors limit the use of handwriting analysis in critical areas such as forensics, document examination, and others.

Modern developments in optical technology, combined with information technology, are expanding the possibilities for handwriting analysis, thanks to the creation of high-resolution cameras. This makes it possible to identify small details in the handwriting characteristics of authors, which, combined with modern image processing methods based on artificial intelligence, opens up new possibilities for analysis [2].

The use of artificial intelligence, in particular machine learning methods, will enable the creation of more effective and reliable software for identifying handwriting authors. This software will not only identify the writer, but also provide experts with substantiated handwriting characteristics that confirm authorship. Such software tools will enable experts to make more objective conclusions when analyzing handwriting, which in turn will expand the practical application of handwriting analysis.

Therefore, taking into account all of the above, further research is needed to improve approaches that will overcome the limitations listed above and provide a better understanding of the mechanisms of handwriting formation and identification. One of the key tasks is to develop unified standards for handwriting analysis. Specifically, this involves integrating modern automated statistical methods, which will increase the objectivity and validity of the results [5].

Therefore, the relevance of developing and improving approaches for writer identification based on handwritten text is vital, since this will not only overcome existing limitations, but also open up new opportunities for practical application in various fields.

Analysis of studies in the field of writer identification showed that there are two approaches for identifying writer from handwriting: deep learning and conventional handcrafted algorithms [4].

Authors in [6] proposed to use densely computed RootSIFT descriptors to capture local properties of handwritten documents, use GMM supervectors as encoding method and use Exemplar-support vector machines (Exemplar-SVM) for similarity measure. This approach relies on handcrafted features and a complex pipeline involving GMM adaptation and Exemplar-SVMs, making it sensitive to parameter tuning and limiting its ability to generalize to unseen handwriting variations. Vision transformer (ViT) could potentially offer a more streamlined and adaptable solution removing reliance on handcrafted features.

In paper [7] authors proposed a method for unsupervised feature learning using convolutional neural network (CNN) activations, where surrogate classes are created via clustering to train a deep residual network, and the activations from the penultimate layer serve as features for subsequent classification tasks like writer identification. The proposed approach is highly dependent on the quality of clustering to create a set of handwriting feature classes. But this in turn greatly affects the completeness of handwriting feature discovery, unlike supervised machine learning methods. It is expected that this problem can be overcome by using the ViT model, which will allow detecting internal patterns and connections in handwriting.

In publication [8], the authors used the weighted label smoothing regularization (WLSR) method to expand the training set with unlabeled data and increase the resolution of the CNN network (ResNet-50). This method showed good results due to the use of unlabeled data. But such use of the CNN network limits the ability of the entire approach to take into account global dependencies in the writing style, and the performance of the approach also depends on the efficiency of WLSR. ViT model can learn more complete and robust representations without the need for explicit data augmentation, as in the case of WLSR.

In [9], a deep adaptive learning method was proposed for writer identification based on images of individual words using multi-task learning. An auxiliary task is added to the training process to ensure the emergence of reusable features. The proposed method transfers the advantages of the learned features of a convolutional neural network from an auxiliary task, such as explicit content recognition, to the main task of author identification in a single procedure. However, such a method strongly depends on the relevance of the selected auxiliary tasks and the effectiveness of the adaptive convolutional layer of the network in detecting generalizing features. ViT can improve on this approach by learning more complete and robust representations directly from word images, without the need for predefined auxiliary tasks.

In paper [10], a convolutional neural network (CNN) was presented for solving the problem of writer identification. The performance of CNN-derived features when fed to a support vector machine (SVM) classifier is also evaluated, considering the traditional classification approach. This approach may be sensitive to specific parameters of the texture generation process and CNN architecture. In contrast, the ViT model is able to directly extract relevant details from the original manuscript images and form deeper representations of the writing style.

In [11], a combination of deep and handwriting descriptors is used to learn patterns from handwritten images. First, local regions are extracted from the handwritten images. Then, these regions are simultaneously fed to deep and handwriting descriptors to generate local descriptions. The extracted local features are then aggregated to create a whole description matrix. Next, by applying vector locally aggregated descriptor (VLAD) encoding to the description matrix, a one-dimensional feature vector is obtained to represent the author's pattern. The disadvantages of this approach include the fact that using only local fragments without taking into account their spatial relationships may limit the possibility of detecting deeper stylistic patterns. Also, the combination of manually created features and deep descriptive features may negatively affect the effectiveness of the entire approach, since it depends very much on the choice of specific descriptors and their impact on the final result.

In contrast, ViT model can learn more robust representations by capturing the global relationships between patches. This allows it to build a more complete and holistic understanding of an individual's writing style.

In publication [12] authors presented a SEGmentation-free model based on CNN with use of a weakly supervised region selection mechanism to characterize the writer of a given sample. While this approach is efficient, it has limitations. Its effectiveness depends on using a fixed percentage of top regions for analysis and is sensitive to the quality of the CNN's probability map. In contrast, the ViT model offers a more flexible method, as it can dynamically select the most informative regions by learning their importance from the context.

Authors in [13] proposed a deep neural network, FragNet, with two pathways: feature pyramid to extract feature maps and fragment pathway to predict writer. FragNet's focus on isolated word images limits its ability to capture broader contextual information and makes it sensitive to image quality variations, potentially leading to overfitting and difficulty handling diverse handwriting styles. While computationally demanding, a ViT could improve writer identification by using its self-attention mechanism to capture long-range dependencies within larger text samples and offer greater robustness to image variations.

In work [14] in contrast to CNN based approaches presented a ViT based method with usage of SIFT algorithm for keypoint detection, and demonstrates strong performance on multiple datasets, including forensic data, with high top-1 accuracy and analysis of script and writing style impact. However, usage of SIFT may be suboptimal because it can fail to detect a sufficient number of reliable keypoints for the ViT to analyze.

In publication [15] authors proposed to use Resnet-34 trained to capture deep features from small patches, patches are extracted using the FAST algorithm and Harris Corner (HC) detector, while machine learned features are encoded with VLAD and triangulation embedding. While effective, the reliance on FAST keypoint detector and HC detector might miss crucial stylistic information present in other parts of the handwriting, because they rely on simple localized features like corners. Using SIFT or CenSurE keypoint detectors might further improve patch extraction because they consider a wider context around each keypoint and are more robust to variations in scale, rotation, and illumination.

Authors in [16] proposed global-context residual recurrent neural network (GR-RNN) where spatial relationship between the sequence of fragments is modeled by the recurrent neural network (RNN) to strengthen the discriminative ability of the local fragment features and the complementary information between the global-context and local fragments is leveraged. However, this method needs the word or subword image segmentation, which requires extra prepossessing steps for applying it on other documents.

In paper [17] authors proposed to use the Siamese network to generate encoded representation and extract identification label, input data is segmented into two inputs and then is processed by Xception feature extractor. While this approach shows astonishing accuracy, the biggest downside is training computation time of five to seven days.

The analysis of scientific publications [6–17] showed that despite significant progress, existing methods of writer identification have significant limitations. On the one hand, CNN-based approaches are often unable to effectively capture the global context and long-term

 ISSN-L 2664-9969; E-ISSN 2706-5448

dependencies in handwriting. On the other hand, newer architectures, such as deep convolutional networks or recurrent neural networks, although they demonstrate high potential for analyzing complex patterns, their effectiveness is largely offset by high computational costs for training and critical sensitivity to input data preparation methods.

Thereby, the combination of these problems and shortcomings indicates that there is a need for additional research in field of writer identification to create an optimized and comprehensive approach to author identification.

Thus, *the object of research* is the process of writer identification based on handwritten text. *The aim of research* is to optimize and improve efficiency of writer identification based on handwritten text by integrating methods for image preprocessing, feature detection methods and modern machine learning architectures. This will result in an increase in the reliability of automated systems for identifying the writer by handwriting, which will strengthen the analytical capabilities of specialists in the field of graphology and handwriting examination.

To achieve the aim, the following objectives were set:
– to build a functional model of an improved approach for writer identification;
– to conduct an experimental study of the proposed approach.

2. Materials and Methods

The main hypothesis of research is that the use of modern machine learning architectures, in particular visual transformers and new feature extraction methods, will ensure high accuracy and reliability of writer identification from handwritten text. This, in turn, will provide graphologists and handwriting experts with an effective automated tool to increase the objectivity, accuracy, and completeness of their expert conclusions.

This study utilized the publicly available Computer Vision Lab (CVL) dataset as the source for input data [18]. This dataset contains images of cursive handwritten texts in both German and English, sourced from various literary works. It contains contributions from 310 distinct authors, with each author providing seven unique text samples. The entire dataset, encompassing all 310 subjects, was employed for the subsequent analysis.

Machine learning models require the dataset to be split into the following parts: training, validation, and testing. To verify model's ability to predict writer of given text, a double validation approach was used. This approach involves performing validation using retrieval and classification methods.

Image retrieval approach is based on calculating distances or similarities between images and on the use of informative invariant rapidly computed features [19]. Retrieval methods get the requested image and calculate the distance (similarity) to other images in a dataset. Images with the closest distances are considered to be alike and returned as results.

Image classification approach is based on assigning predefined categories/labels based on their visual content [2]. Classification methods classify unlabeled data and recognize patterns because of training on categorized/labeled.

The Vision transformer (Google, USA) [20] served as the foundational machine learning model for this work, leveraging attention mechanisms as its primary learning paradigm. Fig. 1 provides a comprehensive overview of its architecture.

At its core, the ViT is built upon the transformer encoder, an architecture comprising two critical layers: multi-head self-attention and a feed-forward neural network. This encoder is engineered to process input tokens, generating context-sensitive representations through the interplay of these layers. Specifically, the multi-head self-attention mechanism performs several key functions: it effectively captures intricate relationships among diverse patches within the input sequence and computes a weighted sum of patch embeddings. This weighting strategically emphasizes salient patches by integrating both global and local contextual information. Complementing this, the feed-forward neural network introduces non-linearity, enabling the model to learn complex, non-linear relationships across patches. Following each of these computational stages, the outputs are channeled through a normalization and residual connections sub-layer. This sub-layer is crucial for stabilizing and accelerating the learning process by normalizing inputs to each layer and by mitigating the well-known issue of vanishing gradients.

To identify a writer, handwriting analysis uses a method of individual personal identification. The essence of method is full exploration of the individual characteristics of the document being examined and the comparison of these with the known manifestations of these characteristics in other documents (comparative material) [21].

Taking into account handwriting features, machine learning architectures and the object of study, the following simplifications were adopted in the study:
– the analysis was restricted to a standardized dataset written in a single language and script, using a closed set of known writers;
– only graphological features are important thus word and sentence meanings are not taken into account.

The modeling was carried out using the following technologies: Python, OpenCV and PyTorch libraries.

The training/evaluation stand used for the research had the following characteristics: Intel 12th gen CPU, NVIDIA GeForce RTX 3070Ti Mobile, DDR4 32 GB.

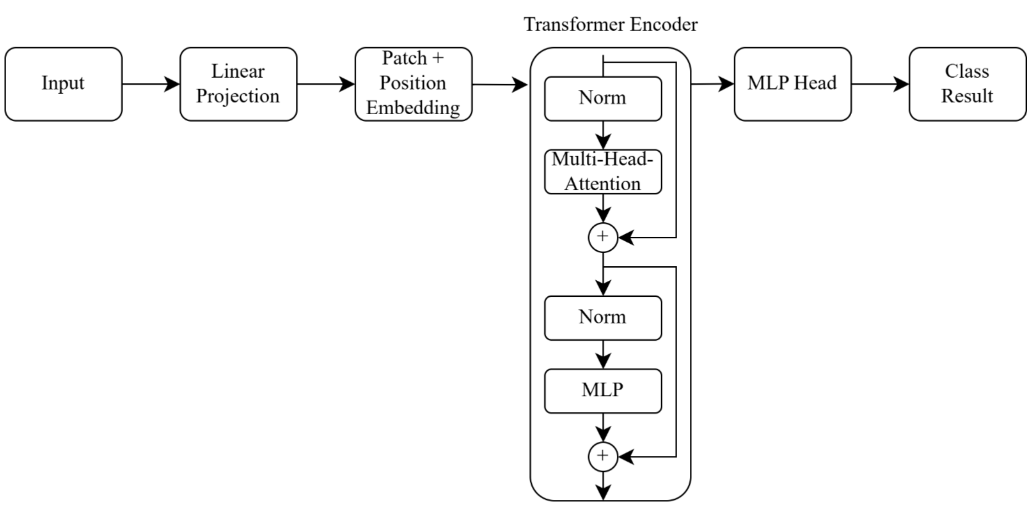

Fig. 1. Visual transformer model

Evaluation metrics for retrieval and classification performance include: accuracy, soft top k, hard top k, completeness on k, accuracy on k, average accuracy.

3. Results and Discussion

3.1. Writer identification functional model

After analysis of approaches used in the literature for writer identification based on analysis of handwritten text, common parts used for writer identification with machine learning can identified. Common parts include:

1) data preprocessing;
2) graphological features extraction;
3) writer identification.

According to these steps a high-level model for writer identification based on handwritten text image using machine learning is built and shown in Fig. 2.

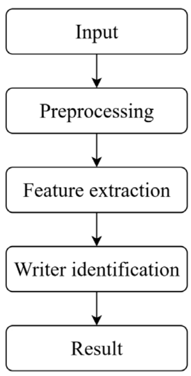

Fig. 2. High-level model of the writer identification approach

According to the high-level model, a functional model for writer identification using machine learning was built. This model includes a preprocessing step and combined steps of graphological feature extraction and writer identification.

Preprocessing step is a crucial part for the writer identification approach, as it fulfills several purposes at once.

Processing multi-colored images presents inherent complexities and demands significant computational resources. To mitigate these challenges, a fundamental pre-processing step involves converting the image to grayscale.

Input images in datasets like CVL often include substantial dimensions, with some examples averaging 2500 × 1500 pixels. Such high resolutions pose a significant challenge for machine learning models due to the resulting large number of tokens generated. This is particularly problematic for architectures like vision transformers, where the computational complexity of self-attention scales quadratically with the token count. To mitigate this computational complexity, a preprocessing step of patch extraction is introduced. This involves dividing the image into fixed-size patches, typically 32 × 32 pixels. This patch size is commonly used for very high-resolution imagery as it effectively reduces the total number of tokens the model needs to process, thereby managing computational complexity.

Given that fields with empty space are present in CVL dataset images, a targeted patch extraction strategy is necessary. To get most relevant information from the image, an additional patch extraction preprocessing step using Center Surround Extremas (CENSURE, also known as OpenCV STAR) [22] was added. This step transforms image data into features at the extrema of the center-surround filters over multiple scales as a result local object coordinates are obtained. These local object coordinates are used as center points for extracting 32 × 32 pixel patches.

To effectively separate objects from their background during input data preprocessing, an automatic thresholding algorithm [23] was used. This algorithm binarizes grayscale images by selecting an optimal threshold that maximizes the inter-class variance in gray levels. Pixels below this threshold are set to 0 (black), and those above are set to 1 (white), clearly segmenting foreground from background.

The model of preprocessing of handwritten text input data is shown in Fig. 3.

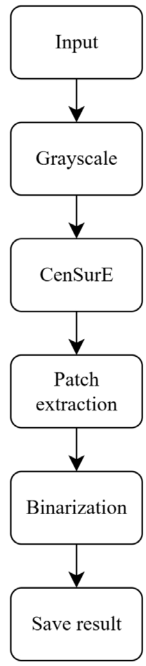

Fig. 3. Model of preprocessing of handwritten text input data

The proposed functional model for writer identification using machine learning is shown in Fig. 4.

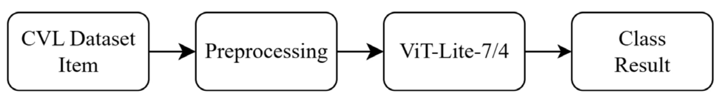

Fig. 4. Proposed functional model for writer identification

The preprocessed input data was fed into a vision transformer, specifically the small-scale learning model detailed in [24]. This model was selected because it is highly suitable for training on small data sets like CVL, which are considerably smaller than conventional computer vision datasets like ImageNet. The architecture of the employed model, ViT-Lite-7/4, features 7 transformer encoder layers and 4 attention heads.

3.2. Results of experimental study on the proposed approach

An experimental study conducted to verify the proposed approach included several parts. The first part is related to Feature detection methods. The second part is related to writer identification.

Feature detection methods were applied to images from the CVL database to extract important features. To demonstrate the difference between baseline method and proposed method detected areas of images were highlighted.

Fig. 5 demonstrates the keypoints identified by the SIFT feature detection algorithm on a sample image from the dataset. The highlighted areas correspond to the features deemed most significant by the detector.

Fig. 6 demonstrates the keypoints identified by the CenSurE feature detection algorithm on a sample image from the dataset. The highlighted areas correspond to the features deemed most significant by the detector.

ISSN-L 2664-9969; E-ISSN 2706-5448

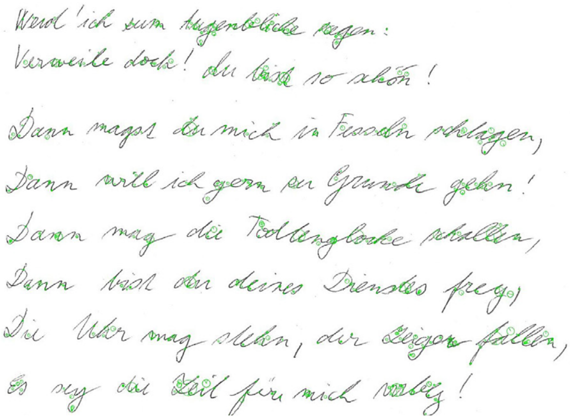

Fig. 5. Demonstration of feature detection using the baseline method

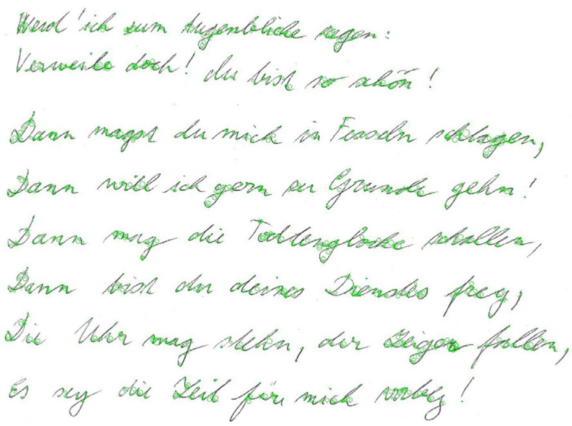

Fig. 6. Demonstration of feature detection using the proposed method

To quantitatively assess the performance of various feature detection methods, comparative analysis was conducted focusing on computational efficiency and feature yield. Each method was tested on an entire dataset to process every item under uniform conditions. During the experiment, key performance metrics were measured: the average time to indicate keypoints per item, the total time required for the complete dataset, and the total number of patches extracted from the detected keypoints. The results of this assessment are presented in Table 1. This data provides a direct comparison, highlighting the trade-offs between the processing speed and feature detection density of each method.

Table 1

Comparison of feature detection methods

Method	Avg per item, s	Total time, s	Total patches
Baseline (SIFT)	0.8665	1390.93	2555409
SURF	0.4633	744.50	3209245
ORB	0.1662	267.28	801000
Proposed (CenSurE)	0.52	848.44	4368819

The second part related to writer identification included training of several models. Two models differ in dataset composition and verification approach. Dataset composition were following:

1. *The dataset was partitioned into two equal halves.* The first half was allocated for model training and validation, while the second half was reserved exclusively as a hold-out test set.
2. *A per-subject partitioning strategy was implemented.* For each of the 310 subjects, their seven handwritten text samples were allocated as follows: two samples for training, one for validation, and the remaining four for testing.

To ensure reproducibility of results and uniform data shuffling across all model modules, a fixed random seed was used. To avoid overfitting, an early stop mechanism was implemented that stops training if the loss function does not improve within 10 epochs.

The parameters selected when tuning the model are presented in Table 2.

To ensure that the models were trained correctly, the loss (loss_val, loss_train) and accuracy (accuracy_val, accuracy_train) metrics were monitored on the training and validation samples.

Fig. 7–10 illustrate the results of developing models for the classification (Fig. 7, 8) and retrieval (Fig. 9, 10) approaches.

Table 2

Model configuration parameters

Parameter	Value	Description
lr	0.0005	learning rate
num-epochs-warmup	5	warmup epochs
batch-size	128	batch size
num-workers	4	PyTorch workers
num-epochs-patience	10	the number of epochs after which training will stop if the validation loss no longer improves
num-epochs	60	number of epochs

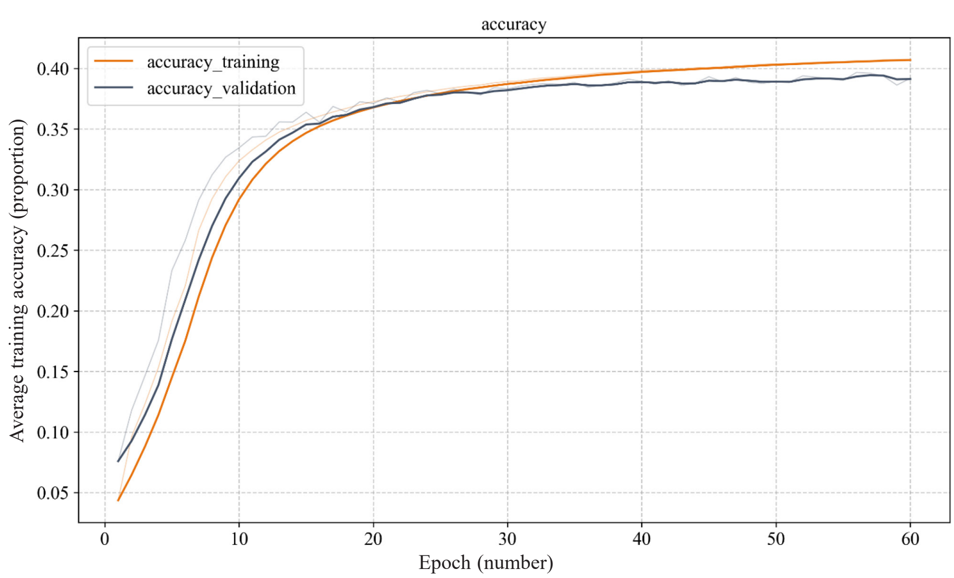

Fig. 7. Accuracy of validation and training of the proposed approach on the test dataset for the classification approach:
X-axis – epoch (number); Y-axis – average training accuracy (proportion)

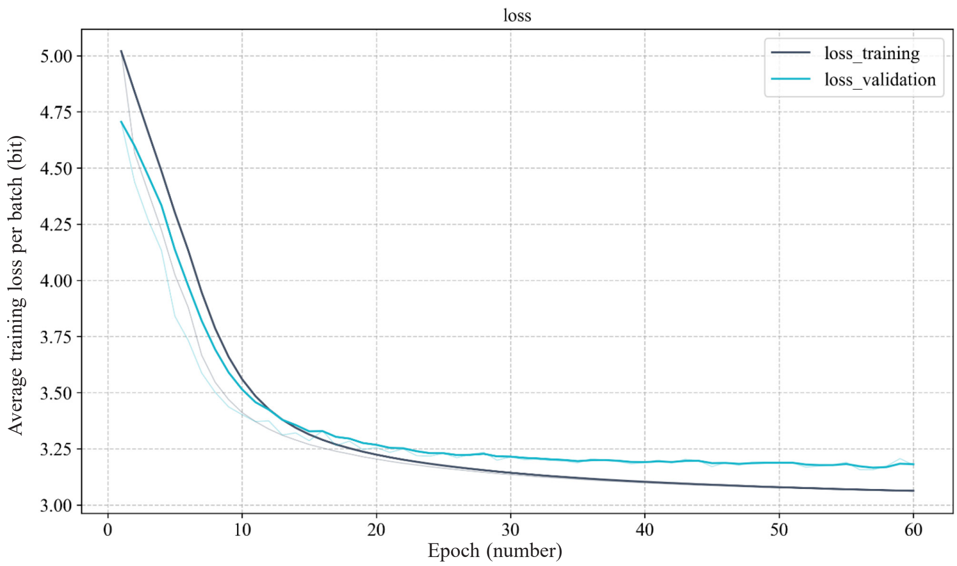

Fig. 8. Validation and training loss of the proposed approach on the test dataset for the classification approach: X-axis – epoch (number); Y-axis – average training loss per batch (bit)

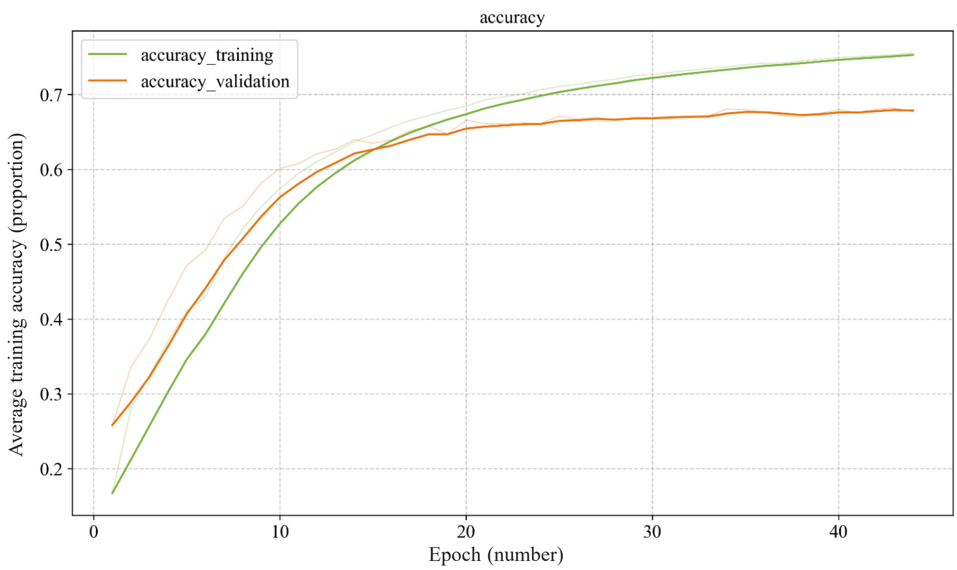

Fig. 9. Accuracy of validation and training of the proposed approach on the test dataset for the retrieval approach: X-axis – epoch (number); Y-axis – average training accuracy (proportion)

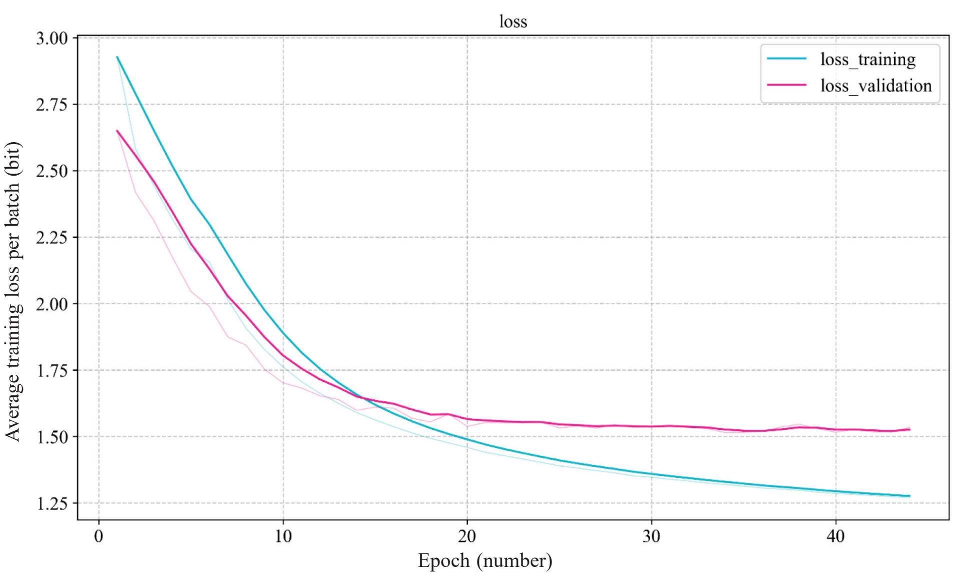

Fig. 10. Validation and training loss of the proposed approach on the test dataset for the retrieval approach: X-axis – epoch (number); Y-axis – average training loss per batch (bit)

 ISSN-L 2664-9969; E-ISSN 2706-5448

The main experiment evaluated the effectiveness of two approaches to writer identification. For the classification approach, the key indicator was the accuracy of the model, while for the retrieval approach, the accuracy of identification based on the squared Euclidean distance between feature vectors. For the purpose of comprehensive validation, a number of additional metrics relevant for evaluating machine learning models were also calculated.

The results of testing the classification and retrieval approaches are summarized in Tables 3 and 4, respectively.

Table 3

Classification approach performance metrics

Metric	*k*	Baseline	Proposed
Accuracy	1	0.98	0.98
	2	0.99	0.99
	3	0.99	0.99
	4	0.99	0.99
	5	0.99	0.99
	6	0.99	0.99
	7	0.99	0.99
	8	0.99	0.99
	9	0.99	0.99
	10	0.99	0.99

Table 4

Retrieval approach performance metrics using squared Euclidean distance

Metric	*k*	Baseline	Proposed
Hard Top K	1	0.95	0.96
	2	0.89	0.91
Soft Top K	1	0.95	0.96
	2	0.96	0.97
	3	0.96	0.97
	4	0.97	0.97
	5	0.97	0.98
mAP	–	0.87	0.89

As can be seen in Table 1, approach based on CenSurE demonstrates better results in identification of maximum number of informative points (Total patches) – over 4.3 million. Exceeding numbers, in comparison to other approaches, mean that CenSurE captures greater amount of handwriting features. In terms of processing speed CenSurE demonstrates balanced performance while being significantly faster than the baseline SIFT method.

Fig. 7 demonstrates that during training over 60 epochs classification model performed good and showed steady improvement. During initial stage (up to ~25 epoch) model shows gradual growth in accuracy on both the training (accuracy_training) and validation (accuracy_validation) samples. This indicates that model quickly and effectively captures patterns from samples. During later stage (from ~30 epoch and higher), validation accuracy can be seen reaching steady 39%. This indicates that model reached its optimal performance thus confirming good generalization ability and its ability to successfully work with new, unknown information.

Fig. 8 shows that during training over 60 epochs classification model loss function metrics demonstrated a constant reduction in prediction errors. During initial stage (up to ~15 epoch) model shows rapid decrease in loss values on both training (loss_training) and validation (loss_validation) samples. This indicates that model quickly and effectively captures patterns from samples (this correlates to what can be seen on accuracy graphs in Fig. 7). Simultaneous decrease and constant proximity of both loss functions values indicate model's good generalization ability and absence of overfitting. During later stage (from ~25 epoch and higher), training loss can be seen reaching steady values. This indicates that model reached optimal state, where error on new data is minimal.

Fig. 9 demonstrates that during training over 40 epochs retrieval model accuracy metrics display a consistent gain. During initial stage (up to ~20 epoch) model shows fast increase in accuracy values on both training (accuracy_training) and validation (accuracy_validation) samples. This indicates that model is able to correctly identify patterns in samples. During later stage (from ~20 epoch and higher), validation accuracy can be seen reaching as high as 69%. This indicates that model training results into a complete and persistent perception of handwriting data from samples thus confirming the ability to accurately identify relevant details from the training set.

Fig. 10 shows that during training over 40 epochs retrieval model loss metric reflected a stable prediction error decrease. During initial stage (up to ~15 epoch) model shows gradual decrease in loss values on both training (loss_training) and validation (loss_validation) samples. This indicates that model steadily improved over time hence making better predictions (this correlates to what can be seen on accuracy graphs in Fig. 9). During later stage (from ~20 epoch and higher), validation loss can be seen converging at approximately 1.5. This indicates that model performance hit a ceiling, where it learned all that it could from training samples. Small difference between both loss functions values and absence of loss increase indicate that model has no overfitting problem and performs well on unseen data.

Performance metrics of the proposed approach shown in Tables 3 and 4 indicate an improvement in writer identification in both classification and retrieval tasks. The major improvement is seen in retrieval tasks. Metrics such as Hard Top K, Soft Top K, and mean Average Precision (mAP) show an increase compared to the baseline approach.

The advantages of the proposed approach can also be seen in comparison with other well-known approaches. Unlike methods that depend on complex, multi-stage pipelines and handcrafted features like RootSIFT descriptors [6] or a combination of deep and handcrafted features [11], proposed approach uses a more simplified and computationally effective feature extraction method. Usage of such feature extraction methods also helps to avoid sensitive parameter tuning and removes necessity in balancing and combining features.

Furthermore, proposed approach addresses known drawbacks of other architectures such as those based on CNNs. Known drawbacks of ResNet-50 [8] and FragNet [13] models include limited ability to capture global context and difficulty handling diverse handwriting styles. Since proposed approach is based on ViT it utilizes self-attention mechanism to capture long-range dependencies and relationships between patches. These characteristics not only address the known shortcomings of CNN architectures, but also serve as advantages over patch-based methods [11, 12].

Finally, because ViT works directly with images, the proposed approach shows great results without using auxiliary tasks [9] or complex segmentation preprocessing [16], thus avoiding potential complications associated with the use of additional preparatory steps.

It is also important to note that the proposed approach maintains high accuracy values for the classification approach. This, together with improved values for the retrieval approach, allows to conclude that the proposed approach demonstrates high effectiveness in writer identification tasks.

Therefore, proposed approach to writer identification based on handwritten text contributes to improving the efficiency of automated handwriting analysis systems. It also serves as a valuable tool for graphologists and forensic document examiners through optimization of their professional activities.

Although the use of a public dataset contributes to the reproducibility of the experiment, its limitation to English and German languages affects the generalizability of the proposed approach. At the same time, a key condition for reproducing the results of the search method is to use exclusively squared Euclidean distance, since other metrics may yield different results.

Further research is possible to develop and evaluate new image processing methods to further improve writer identification accuracy and lower computational complexity, as well as to improve existing preprocessing methods including algorithms for feature detection, grayscale and binarization.

4. Conclusions

1. The paper presents a functional model of an improved approach for writer identification. The developed functional model establishes a process for analyzing handwritten text. This process works by first fully investigating the individual characteristics of a document. It then compares these unique features to known examples from other documents, a method which ultimately allows for the accurate identification of the writer. The results were achieved using a two-step process. First, the CenSurE algorithm identifies key points in the handwriting and extracts small image sections, known as patches, from those points. Then, a vision transformer model classifies these patches to determine the writer's identity. In order to ensure the robustness and reliability of the results, a two-component experimental validation methodology was applied, combining retrieval and classification approaches. In addition, the use of a public dataset ensures transparency, reproducibility, and comparability of the obtained results in the context of modern scientific research.

2. An experimental study of the proposed writer identification approach was conducted, which showed the advantages of the proposed writer identification approach. In particular, higher accuracy was achieved in retrieval tasks compared to existing methods, which is confirmed by increased accuracy values of hard-top-k and soft-top-k by 1% and mean average precision by 2%. At the same time, a significant increase in computational efficiency was recorded at the feature extraction stage. This is quantitatively confirmed by the simultaneous reduction in the average processing time of one element and the total duration of the process by 39%, alongside with an increase in the number of highlighted informative areas by 70%.

Conflict of interest

The authors declare that they have no conflict of interest in connection with this research, including financial, personal, authorial or any other interest that could affect the research and its results presented in this article.

Financing

The research was conducted without financial support.

Data availability

The data will be provided upon reasonable request.

Use of artificial intelligence

Gemini 2.5 Pro was used in the "Introduction" section to search for sources from the last 5 years. The authors verified the results provided by reviewing the original sources and their content, and confirm that it did not influence the conclusions of the research.

Authors' contributions

Mykyta Shupyliuk: Software, Writing – original draft; ***Vitalii Martovytskyi***: Conceptualization, Validation, Writing – reviewing and editing; ***Yuri Romanenkov***: Investigation, Validation, Writing – reviewing and editing.

References

1. Hengl, M. (2014). Comparison of the Branches of Handwriting Analysis. *Chasopys Natsionalnoho universytetu "Ostrozka akademiia". Seriia: Pravo, 2 (10).*
2. Shupyliuk, M., Martovytskyi, V., Bolohova, N., Romanenkov, Y., Osiievskyi, S., Liashenko, S. et al. (2025). Devising an approach to personality identification based on handwritten text using a vision transformer. *Eastern-European Journal of Enterprise Technologies, 1 (2 (133)),* 53–65. https://doi.org/10.15587/1729-4061.2025.322726
3. Aliyev, E. (2024). Forensic Handwriting Analysis to Determine the Psychophysiological Traits. *International Journal of Religion, 5 (6),* 511–530. https://doi.org/10.61707/2r6bmr11
4. Pandey, N., Singh, B., Singh, S. (2024). Review on handwriting examination on unusual surface. *IP International Journal of Forensic Medicine and Toxicological Sciences, 8 (4),* 125–131. https://doi.org/10.18231/j.ijfmts.2023.028
5. Romanenkov, Y., Pronchakov, Y., Zieiniiev, T. (2020). Algorithmic Support for Auto-modes of adaptive short-term Forecasting in predictive Analytics Systems. *2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT),* 230–233. https://doi.org/10.1109/csit49958.2020.9321875
6. Christlein, V., Bernecker, D., Hönig, F., Maier, A., Angelopoulou, E. (2017). Writer Identification Using GMM Supervectors and Exemplar-SVMs. *Pattern Recognition, 63,* 258–267. https://doi.org/10.1016/j.patcog.2016.10.005
7. Christlein, V., Gropp, M., Fiel, S., Maier, A. (2017). Unsupervised Feature Learning for Writer Identification and Writer Retrieval. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR),* 991–997. https://doi.org/10.1109/icdar.2017.165
8. Chen, S., Wang, Y., Lin, C.-T., Ding, W., Cao, Z. (2019). Semi-supervised feature learning for improving writer identification. *Information Sciences, 482,* 156–170. https://doi.org/10.1016/j.ins.2019.01.024
9. He, S., Schomaker, L. (2019). Deep adaptive learning for writer identification based on single handwritten word images. *Pattern Recognition, 88,* 64–74. https://doi.org/10.1016/j.patcog.2018.11.003
10. Helal, L. G., Bertolini, D., Costa, Y. M. G., Cavalcanti, G. D. C., Britto, A. S., Oliveira, L. E. S. (2019). Representation Learning and Dissimilarity for Writer Identification. *2019 International Conference on Systems, Signals and Image Processing (IWSSIP),* 63–68. https://doi.org/10.1109/iwssip.2019.8787293
11. Sulaiman, A., Omar, K., Nasrudin, M. F., Arram, A. (2019). Length Independent Writer Identification Based on the Fusion of Deep and Hand-Crafted Descriptors. *IEEE Access, 7,* 91772–91784. https://doi.org/10.1109/access.2019.2927286
12. Kumar, P., Sharma, A. (2020). Segmentation-free writer identification based on convolutional neural network. *Computers & Electrical Engineering, 85.* https://doi.org/10.1016/j.compeleceng.2020.106707
13. He, S., Schomaker, L. (2020). FragNet: Writer Identification Using Deep Fragment Networks. *IEEE Transactions on Information Forensics and Security, 15,* 3013–3022. https://doi.org/10.1109/tifs.2020.2981236
14. Koepf, M., Kleber, F., Sablatnig, R. (2022). Writer identification and writer retrieval using Vision Transformer for forensic documents. *Document Analysis Systems.* Cham: Springer, 352–366. https://doi.org/10.1007/978-3-031-06555-2_24
15. Semma, A., Hannad, Y., Siddiqi, I., Djeddi, C., El Youssfi El Kettani, M. (2021). Writer Identification using Deep Learning with FAST Keypoints and Harris corner detector. *Expert Systems with Applications, 184,* 115473. https://doi.org/10.1016/j.eswa.2021.115473
16. He, S., Schomaker, L. (2021). GR-RNN: Global-context residual recurrent neural networks for writer identification. *Pattern Recognition, 117.* https://doi.org/10.48550/arXiv.2104.05036
17. Suteddy, W., Agustini, D. A. R., Atmanto, D. A. (2024). Offline Handwriting Writer Identification using Depth-wise Separable Convolution with Siamese Network. *JOIV: International Journal on Informatics Visualization, 8 (1),* 535–541. https://doi.org/10.62527/joiv.8.1.2148
18. Kleber, F., Fiel, S., Diem, M., Sablatnig, R. (2013). CVL-DataBase: An Off-Line Database for Writer Retrieval, Writer Identification and Word Spotting. *2013 12th International Conference on Document Analysis and Recognition,* 560–564. https://doi.org/10.1109/icdar.2013.117
19. Smelyakov, K., Sandrkin, D., Ruban, I., Vitalii, M., Romanenkov, Y. (2018). Search by Image. New Search Engine Service Model. *2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T),* 181–186. https://doi.org/10.1109/infocommst.2018.8632117

ISSN-L 2664-9969; E-ISSN 2706-5448

20. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T. et al. (2021). *An Image is Worth 16 x 16 Words: Transformers for Image Recognition at Scale.* arXiv 2021. https://doi.org/10.48550/arXiv.2010.11929

21. Shupyliuk, M., Martovytskyi, V. (2025). Analysis of personality detection and writer identification methods. *Control, Navigation and Communication Systems, 79 (1),* 138–142. https://doi.org/10.26906/SUNZ.2025.1.138-142

22. Agrawal, M., Konolige, L., Blas, M. (2008). CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching. *10th European Conference on Computer Vision.* Berlin, Heidelberg: Springer, 102–115. https://doi.org/10.1007/978-3-540-88693-8_8

23. Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics, 9 (1),* 62–66. https://doi.org/10.1109/tsmc.1979.4310076

24. Hassani, A., Walton, S., Shah, N., Abuduweili, A., Li, J., Shi, H. (2021). *Escaping the Big Data Paradigm with Compact Transformers.* arXiv:2104.05704. https://doi.org/10.48550/arXiv.2104.05704

Mykyta Shupyliuk, *PhD Student, Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine, ORCID: https://orcid.org/0009-0005-3457-9895*

✉***Vitalii Martovytskyi,*** *PhD, Associate Professor, Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine, e-mail: vitalii.martovytskyi@nure.ua, ORCID: https://orcid.org/0000-0003-2349-0578*

Yuri Romanenkov, *Doctor of Technical Sciences, Vice-Rector for Scientific Work, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine, ORCID: https://orcid.org/0000-0002-6544-5348*

✉*Corresponding author*