The Landscape of Handwriting Recognition Research: A Comprehensive Review of Datasets, Methods, and Future Direction

1st Reza Budiawan
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
33221040@std.stei.itb.ac.id

2nd Rinaldi Munir
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
rinaldi@staff.stei.itb.ac.id

2025 Tenth International Conference on Informatics and Computing (ICIC) | 979-8-3315-7583-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICIC68054.2025.11309377

***Abstract*—Optical character recognition has been extensively studied in the field of computer vision. A machine must read and understand writing digitally in the modern digital world. This research area has been explored since the creation of the first input pen. Despite advances and the increasing needs regarding technology, handwriting recognition is still adequate for method exploration or hardware development. There are many techniques to suit this zone. The traditional approach to recognizing handwriting involves multiple stages, including preprocessing, feature extraction, classification, and postprocessing. But the use of these classical approaches has started to be replaced by deep learning techniques. Deep learning methods by means of computer vision have enjoyed fine accuracy results in this area (especially, the Convolutional Neural Networks have been serving this purpose). CNNs can integrate feature extraction and classification processes due to the existence of convolutional layers in CNNs. The integration leads to a more efficient approach with better results. This paper describes several data sets and approaches used in handwriting recognition research.**

Keywords—handwriting recognition, machine learning, deep learning, character recognition system

I. INTRODUCTION

Handwritten recognition is one of the branches of computer vision domain. It's the way your computer understands written text (handwriting, printed text) as an input. This input might be an image of a document, or values corresponding to the coordinates of points taken by some measurement device. Handwriting recognition systems may be used for diverse everyday purposes, for example, processing of handwritten input on a tablet, postal code reading, bank form digitization, document information retrieval, language translation machines, as well as physical documents to electronic versions conversion.

Handwriting recognition is typically performed on a character-by-character basis. Handwritten text is broken down into character units to be further processed and recognized in a machine-readable format. The distinct shape of each character makes the recognition process possible. However, in some cases, characters can appear like to another character. For instance, the letter 'l' and the digit '1' present a unique challenge for the system to differentiate. Furthermore, ambiguity exists in handwriting. Figure 1 illustrates the ambiguity in reading the handwritten word "true" or "false". Humans can resolve this ambiguity by considering the context of the writing, whereas machines require additional processes to do so.

Fig. 1. Ambiguity in handwriting (*word can be in interpreted as true or false*)

The difficulties associated with developing a handwriting recognition system extend beyond mere character resemblance. Someone could have differences in writing styles to another. Furthermore, each person's handwriting can exhibit variations influenced by several factors, including the type of writing surface and their physical or emotional state during the act of writing [1]. Those are the obstacle to handle to recognize character automatically by optical character recognition machine.

Research in handwriting recognition aims to minimize character reading errors. Traditionally, the recognition process consists of four stages: preprocessing, feature extraction, classification, and postprocessing. The preprocessing stage ensures that the input has a uniform format before its important features are extracted in the feature extraction stage. Subsequently, the extracted features are classified using a classifier and may be further processed in the postprocessing stage to be recognized as machine-readable characters. Some stages, such as feature extraction and classification, are mandatory. In contrast, the preprocessing and postprocessing stages are optional and depend on the dataset used and the desired output. Various feature extraction and classification methods have been developed and adapted for implementation in handwriting recognition systems to achieve high accuracy. Moreover, deep learning methods like Convolutional Neural Networks (CNNs), which are widely used in computer vision research, are also applied to handwriting recognition systems.

This article discusses the methods applied in the research topic of handwriting recognition systems, with the following structure. Chapter II provides an explanation of the handwriting recognition system. Chapter III discusses the previous work of research topics in the field of handwriting recognition systems. Chapter IV contains a discussion about challenges in handwriting recognition research. Chapter V presents the conclusion of this article.

II. HANDWRITING RECOGNITION SYSTEM

Research on handwriting recognition systems has different focus areas. This paper outlines several of them related to input type, approach methods, recognition units, recognition methods used, and the development of new datasets.

A. Types of Handwriting Recognition Systems

Based on the input provided to the system, there are two types of handwriting recognition systems: online and offline. An online handwriting recognition system receives input from a specific device, while an offline system receives handwritten images as input. The input format for online handwriting recognition varies depending on the device and can be an image, an XML file, a CSV file, or a series of numbers in a specific file. Figure 2 shows a sequence of characters resulting from online character writing in the form of a binary string.

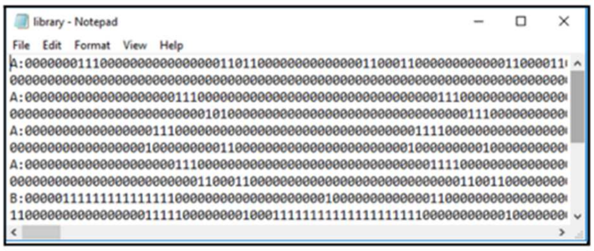

Fig. 2. Sequence characters of online handwriting recognition [2]

An offline handwriting recognition system takes an image as its input. The form and information contained within the image also vary depending on the dataset used. The image can be of a single character, a single line of text, or an entire handwritten document. This type of handwriting usually recognizes on historical document [3]. Figure 3 shows a historical document image that can serve as input for an offline handwriting recognition system.

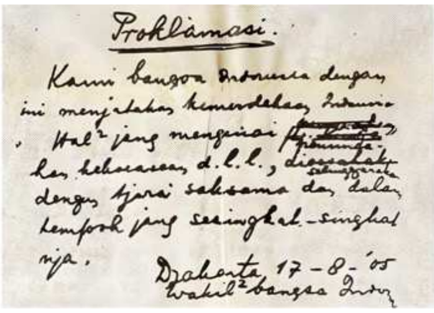

Fig. 3. Historical document with handwriting on proclamation text of Indonesia

B. Strategies to Recognize Handwriting

Handwriting recognition systems employ two main strategies to recognize handwriting: the analytic approach and the holistic approach [4]. The analytic approach focuses on deconstructing handwriting into smaller components through segmentation techniques. Conversely, the holistic approach perceives words as complete entities without disassembling them into smaller parts.

The analytic strategy utilizes segmentation to fragment the writing into individual character units. Each handwritten word is separated into its basic characters, which are then processed by a recognition engine. This engine identifies the characters, which are subsequently reassembled to create a fully machine-readable document. In contrast to the analytic method, which breaks down text into characters, the holistic approach recognizes handwriting at the word level. This method is often used for cursive writing and operates with a restricted vocabulary [4]. Figure 4 and Figure 5 shows illustration of analytic method and holistic method.

Fig. 4. Illustration of analytic method

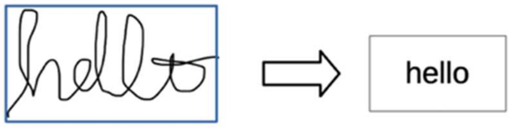

Fig. 5. Illustration of holistic method

C. Objects of the Handwriting Recognition System

Handwriting recognition research has a specific focus on the object being recognized. There are several level characters to be recognized. It focuses on these scenarios, namely: recognition of digit, strings of digits, single alphanumeric character, a string of characters, documents, and symbols.

Digit recognition limits the recognized object to only a single numerical character. This system is the simplest form of a handwriting recognition system. The main challenge in digit recognition is the variation in the way characters are written. A more complex form is the digit string. The object recognized in a digit string is a sequence of numerical characters. Challenges for this type of recognition include overlapping writing between numbers and differences in writing style. Figure 6 shows digit strings from the Computer Vision Lab (CVL) Handwritten Digit String (HDS) dataset.

Fig. 6. Sample digit string of dataset CVL-HDS [5]

Recognizing single alphanumeric handwriting character means translate the handwriting image content with written as single character of a-z/A-Z/0-9. Variation of the shape become the challenges on this scenario of recognition. This condition called allograph—a character which has one or more alternatives form, including capital, lowercase, or italic. For example, someone can make two different shapes for character lowercase “a”. Figure 7 shows the different shape of character “a”, digit “4”, and lowercase letter “l” written in a handwriting. This condition is one of a challenge of handwriting recognition system, because the intra-class variation could be bigger in several character.

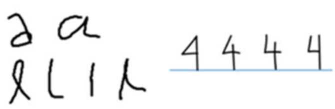

Fig. 7. Variation shape of letter and digit in handwriting

Another challenge that had to overcome is the resemblance of a character to another. For example, digit 0, lowercase character “o”, and uppercase character “O” are look-alike to each other. Another example in this condition is the similarity shape of letter “u” and “v” either on lowercase or uppercase. Figure 8 shows the similarity of these characters in a handwriting.

Fig. 8. The similarity shape of a letter to another in handwriting

Other objects that are recognized include characters, documents, and symbols. Recognition of these objects involves different approaches and methods depending on the research objective. Each object, whether it's a digit, digit string, or character, typically uses a single, uniform language. However, there is also research that explores the use of multiple languages (multi-language) in handwriting recognition.

D. Phase of the Handwriting Recognition System

A handwriting recognition system recognizes handwriting through two main stages: training and testing [6]. In the training stage, the system learns to recognize input patterns, resulting in a model. This model is then used by the system to recognize or determine patterns in new input. The input used during the training stage is different from the input used during the testing stage. An illustration of the handwriting recognition stages is shown in Figure 9.

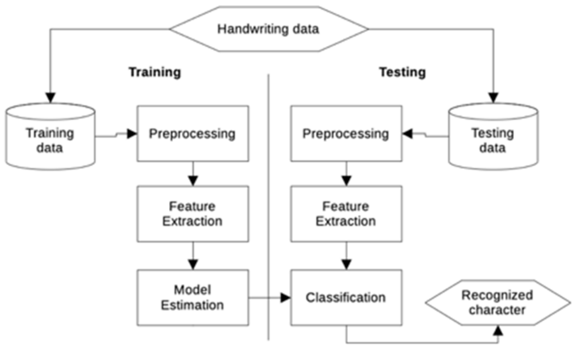

Fig. 9. General stage of handwriting recognition method [6]

To ensure the input has a uniform pattern for recognition, a preprocessing stage is performed. Several processes are carried out in this stage, such as segmentation, image normalization through resizing, cropping, thinning, and convert it into binary format. These processes are through to ensure the uniformness of the image. At the next stage, the important features will be extracted in the feature extraction stage. The methods used in this stage vary depending on the research objectives. Some of the methods that can be used include Wavelet [7], PCA [8], Zoning [9], Zernike Moment [10], or Neighborhood Rough Set [11].

III. Related Works – The Development of Research Topic in Handwriting Recognition

Research on handwriting recognition has been ongoing for a long time, and the methods used have become more advanced with the development of computer vision algorithms and supporting hardware. Several studies have created standard datasets for other researchers to use in their handwriting recognition research. Additionally, various methods have been developed in previous research. This section will review the datasets and methods commonly used in handwriting recognition research.

A. Datasets of Handwriting Recognition

Essential topic of handwriting recognition research area falls within the domain of pattern recognition. It means, the process involves learning and recognizing patterns from previous learning experiences. This pattern learning process requires standard, pre-processed datasets. In some datasets, the preprocessing stage has already been applied to obtain inputs with "uniform" characteristics. This uniformity can be seen in similarities in image size, the range of colors to be feature-extracted, thickness, and more. This allows researchers to focus on developing feature extraction and classification methods. There are several datasets commonly used in handwriting research, such as the MNIST & EMNIST, IAM, CASIA, and MADBase datasets. These datasets have different characteristics from one another.

1) NIST Special Database 19 Dataset [12]: This dataset was created by the National Institute of Standards and Technology (NIST). One version of this data is the NIST Special Database 19, which consists of a corpus of handwritten documents. A total of 3,600 writers were involved in writing (in isolated form—not cursive) a total of 810,000 characters. Each character was extracted from a Handwriting Sample Form (HSF).

2) MNIST Dataset [13]: The Modified National Institute of Standards and Technology (MNIST) dataset consists of numerical characters (digits) taken from the NIST dataset. This dataset is composed of 60,000 training images and 10,000 testing images, with each image sized at 28x28 pixels in grayscale. This dataset consist of 10 classes which represent the digit form 0 to 9. Each image of MNISTdataset has uniform characteristic, and it become the benchmark standard of image recognition.

3) EMNIST Dataset [14]: The MNIST dataset has an extended version called EMNIST. Similar to MNIST, the images in this dataset have a uniform size of 28x28 pixels. This dataset consists of handwritten characters, including letters and numbers. The EMNIST dataset contains 240,000 training images and 40,000 testing images.

4) IAM Dataset [15]: The IAM dataset consists of 13,353 images of handwritten text contributed by 657 writers. This dataset comprises 1,539 pages of handwritten documents containing 115,320 words. The IAM dataset has both online and offline handwriting versions.

5) CASIA Dataset [16]: The CASIA dataset is provided by the Institute of Automation, Chinese Academy of Sciences. The data provided consists of Chinese characters with two versions of the handwriting dataset: CASIA-OLHWDB (online handwriting) and CASIA-HWDB (offline handwriting). The data was collected using a special device, an Anoto Pen, from 1,020 writers.

6) MADBase Dataset [17]: MADBase is a modification of the Arabic Handwritten Digit Database. This dataset consists of Arabic numerical characters that are compatible with the MNIST data format.

In addition to the datasets mentioned above, there are various other datasets that contain specific languages , such as UCOM for the Urdu language, CENPARMI for the Farsi language, and ISI which contains Bangla numerals, among others [18].

B. *Research on Stages/Phases of Handwriting Recognition*

Handwriting recognition using machine learning methods involves several stages in its recognition process. Traditional approach uses these processes, namely: preprocessing, feature extraction, classification/recognition, and postprocessing whilst modern approach using deep learning that automate the feature extraction method. Figure 12 shows the sequence of the handwriting recognition process.

The preprocessing and postprocessing stages are optional, depending on the dataset and the research objective. In contrast, the feature extraction and classification stages are mandatory. Preprocessing is done to "normalize" the data before it is further processed. Some processes performed at this stage include resizing/cropping/thinning to ensure uniformity of image size, dimensions, and boundaries. There are also processes for converting the image to a binary format, thinning to adjust the thickness of the recognized handwritten image, and smoothing [19].

The feature extraction stage involves extracting important features from the handwriting to be recognized. One of the basic methods for feature extraction is zoning. This technique divides the handwritten character into several areas, either statically or dynamically [9]. Another technique that can be used is Vertical & Horizontal Projection. This technique was applied by Sahlol [11], [20] to recognize handwritten Arabic script, combined with feature selection. Geometric feature extraction was applied by Masrani et al. [21] to recognize handwritten letter images. Additionally, there is the sinusoidal modeling technique applied by Choudhury et al. [22] in recognizing online handwriting.

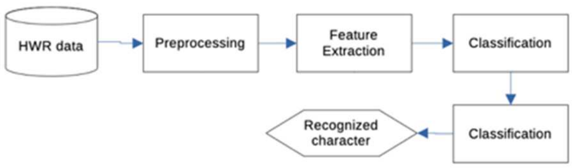

Fig. 10. Stages of handwriting recognition process

For modern approach, handwriting recognition use deep learning to recognize each character. The processes similar to traditional approach. The differences, feature extraction stage automated by the machine. The deep learning method such as Convolutional Neural Network (CNN) act as both a feature extractor and a classifier in the classification/recognition stage of a handwriting recognition system. This condition is possible due to the presence of convolutional layers in the CNN. It utilizes a convolution process with learnable filters/kernels to obtain representative information from the image to be recognized [23].

Several handwriting recognition studies also use the CNN method in their research. Hossain & Ali [23] created a numerical character (digit) recognition system using the CNN method. The character data used was the MNIST dataset, implemented on MatConvNet. Another study by Garg [24] applied CNN to the recognition of handwritten text. The method used in this research combines 5 CNN layers, 2 Recurrent Neural Network (RNN) layers, and CTC decoding. In addition to applying the CNN method, other studies related to this method focus on finding an efficient architecture and optimal parameter values to achieve maximum accuracy. Biswas & Islam [25] conducted research to determine the best architecture for recognizing numerical characters. This study used two datasets: MNIST and Kaggle. The experiment involved 5 architectures, with the best result achieved by the fifth architecture with an accuracy of 99.53%. This architecture consists of 4 convolutional layers, 2 pooling layers, and 2 fully connected layers. Prihatiningsih [26] also conducted an experiment to find the appropriate number of iterations for the CNN method on the MNIST dataset. The experiment results showed an increase in accuracy for iterations between 0-20.

Handwriting research also applies transformer-based methods. Momeni & BabaAli [27] introduce 2 architectures of transformer using attention mechanism to overcome the problem of linguistic rules in arabic handwritten text. Li et.al [28] propose a simple and data-efficient approach for handwritten text recognition, with minimal modifications on the ViT. Zhang et.al. [29] overcome the issue of lack of data on ViT using novel framework named Improved Multi-task Transformer based on Localization Mechanism Network (IMTLM-Net).

C. *Developing a New Dataset*

In addition to the development of methods applied to existing datasets, research on creating new datasets is also quite common in the field of handwriting recognition. These new datasets are created to complement handwriting recognition research for languages not covered by existing datasets.

Felix et al. [30] proposed a new dataset created using a STABILO digipen. This dataset is used for online handwriting recognition and consists of 31,275 handwritten characters grouped into 52 classes, covering 52 uppercase and lowercase letters and 10 numerical characters. There were 119 writing respondents in the data collection process. The digipen used captures signals through several sensors: 2 accelerometers, 1 gyroscope, 1 magnetometer, and 1 force sensor. The collected data was tested using machine learning (Support Vector Machine—SVM) and deep learning (Convolutional Neural Network—CNN) algorithms. The accuracy of the CNN model reached 85% in recognizing this dataset.

Research to create datasets for offline handwriting recognition is usually done for non-Roman alphabets. One such effort is the development of a dataset named Hijja, which contains Hijaiyah letters, conducted by Altwaijry & Al-Turaiki [31]. The difference from other Arabic alphabet datasets is that every character collected was written exclusively by children aged 7-12. The respondents involved were 591 children, with 47,434 characters stored in the dataset. The Hijja dataset was classified using a CNN and its results were compared with the Arabic Handwritten Character Dataset (AHCD). The accuracy for recognizing the Hijja dataset was 88%. This value is relatively low compared to the accuracy of the AHCD, which reached 97%. The model created predicted adult handwriting better than children's handwriting. This indicates the uniqueness of children's handwriting compared to adults' and expands the potential for research specifically focused on children's handwriting.

Research to build a dataset for the Cyrillic alphabet was conducted by Potanin et al. [32]. This dataset contains manuscripts of the Russian emperor Peter the Great, who lived in the 17th century. This manuscript is handwritten in the form of image-text pairs. The dataset has three parts: 6,237 training pairs, 1,930 validation pairs, and 1,527 test pairs. The model used for this dataset consists of a 7-layer CNN, 2 bidirectional GRUs, and a CTC-loss function with evaluation metrics such as character error rate (CER), word error rate (WER), string accuracy (ACC), and time inference. Another dataset was developed for the Kazakh alphabet, named KOHTD—Kazakh Offline Handwritten Text Dataset [33]. This dataset is used for an offline handwriting recognition system, a project carried out by Toiganbayeva et al.. It contains 3,000 handwritten documents in the form of exam papers, 140,335 segmented images, and approximately 922,010 symbols.

IV. Challenges & Opportunities

The application of handwriting recognition is not limited to its methods. The development of this research can also be seen in its application cases. There have been several developments, such as recognition for multiple languages, applications in the field of education, and for document information retrieval.

Several research studies on handwriting recognition have addressed the domain of multi-language recognition. The use of multiple languages increases the number of classes for the classification machine, which can result in longer response times. Furthermore, there are similarities between characters across different languages. Zouari et al. [34] proposed a framework to identify scripts and recognize online handwriting. This system uses a Time Delay Neural Network (TDNN) clustering algorithm based on parameters from a beta-elliptical model. Another study related to multi-language handwriting recognition was conducted by Chen et al. [35], who introduced MuLTReNets, a framework used to identify scripts and recognize handwriting. Research on recognition methods focused on multi-language is also supported by research to build corresponding datasets. This study was conducted by Jiang [36], who built a new dataset of numerical characters for handwriting consisting of various languages. This dataset is named MNIST-MIX and is a collection of 13 different datasets in 10 languages (Arabic, Bangla, Devanagari, English, Farsi, Kannada, Swedish, Telugu, Tibetan, and Urdu). The data was created by adapting the MNIST format, in the form of grayscale images with a size of 28x28 pixels representing the numbers 0-9 from various languages. In the experiment conducted, an accuracy of 90.22% was achieved using the LeNet model.

In the field of education, handwriting recognition research can be applied to measure the quality of produced handwriting. A study conducted by Gao et al. [37] assessed the quality of Chinese character handwriting. The recognized writing was divided into 5 grades from A-E and compared with manual judgments from 10 respondents. Based on the experimental results, a confidence rate of 91.22% was obtained.

There is a variety of focuses in this research, such as online handwriting with specific devices, the application of new methods, and the construction of specific datasets. It also includes research on other topics like handwriting in historic documents for Information Retrieval. Another related topic is Natural Language Processing as a post-processing method in handwriting recognition. The table in the appendix shows a summary of experiments and results achieved in previous research from several technical papers. The initial trend in handwriting recognition research involved using machine learning methods that focused on the feature extraction and classification stages. Subsequently, research development has shifted more towards the use of deep learning methods, either as a classifier or for both feature extraction and classification. However, the development of other methods continues. These new methods are compared with machine learning classifiers like SVM. The creation of new datasets also involves data validation using CNN and its variations, also using Transformers method.

V. Conclusion

Based on the description of cases for the handwriting research topic, there are several challenges, namely recognizing characters with similar shapes, differences in character styles, and also recognizing overlapping writing. For some languages, there is also the challenge of how to recognize connected letters (ligatures). The segmentation process is very influential in producing good input for the recognition engine. Experiments with optimal methods are needed to recognize handwriting, especially using transfer learning methods. In addition, experiments are needed to explore deep learning method, such as CNN or Vision Transformer, for handwriting recognition. It can be applied to single language or multiple languages. This represents a research opportunity in the topic of handwriting recognition.

The creation of new datasets exclusively written by specific groups is also a research opportunity. Furthermore, the discussion about assessing the quality of handwriting is also a case that can be investigated. The assessment of writing quality can be developed into research that creates learning aids for beginners to learn writing. The research topic of handwriting recognition has a wide variety of aspects that can be further explored, both in terms of methods and the focus of the cases discussed. Each case has its uniqueness with specific methods for its solution. Further development possibilities can also be pursued in line with hardware advancements that enable the application of deep learning to the topic of handwriting recognition.

References

[1] S. Rajwar and V. Singh, "Variation in Handwriting Under Different Writing Conditions," *Int. J. Creat. Res. Thoughts*, vol. 9, no. 4, 2021.

[2] M. H. Alkawaz, C. C. Seong, and H. Razalli, "Handwriting Detection and Recognition Improvements Based on Hidden Markov Model and Deep Learning," in *2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA)*, Langkawi, Malaysia: IEEE, Feb. 2020, pp. 106–110. doi: 10.1109/CSPA48992.2020.9068682.

[3] F. Lombardi and S. Marinai, "Deep Learning for Historical Document Analysis and Recognition—A Survey," *J. Imaging*, vol. 6, no. 10, p. 110, Oct. 2020, doi: 10.3390/jimaging6100110.

[4] J. Das Gupta, S. Samanta, and B. Chanda, "Ensemble classifier-based off-line handwritten word recognition system in holistic approach," *IET Image Process.*, vol. 12, no. 8, pp. 1467–1474, Aug. 2018, doi: 10.1049/iet-ipr.2017.0745.

[5] A. F. De Sousa Neto, B. L. D. Bezerra, E. B. Lima, and A. H. Toselli, "HDSR-Flor: A Robust End-to-End System to Solve the Handwritten Digit String Recognition Problem in Real Complex Scenarios," *IEEE Access*, vol. 8, pp. 208543–208553, 2020, doi: 10.1109/ACCESS.2020.3039003.

[6] M. Blom, N. Nobile, and C. Y. Suen, Eds., *Frontiers in pattern recognition and artificial intelligence*. in Series on language

processing, pattern recognition, and intelligent systems, no. volume 5. Singapore: World Scientific, 2019. doi: 10.1142/11362.

[7] A. M. Al-Shatnawi, F. Al-Saqqar, and A. Souri, "Arabic Handwritten Word Recognition Based on Stationary Wavelet Transform Technique using Machine Learning," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 21, no. 3, pp. 1–21, May 2022, doi: 10.1145/3474391.

[8] Zhihong Hu, Qian Tian, Hongbo Wang, and Zhenyu He, "Handwritten number recognition based on PCA and neural network," presented at the 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China: IEEE, pp. 1570–1574. doi: 10.1109/IMCEC51613.2021.9482278.

[9] S. S. Rosyda and T. W. Purboyo, "A Review of Various Handwriting Recognition Methods," vol. 13, no. 2, 2018.

[10] C. Singh and A. Aggarwal, "An Effective Approach for Noise Robust and Rotation Invariant Handwritten Character Recognition Using Zernike Moments Features and Optimal Similarity Measure," *Appl. Artif. Intell.*, vol. 34, no. 13, pp. 1011–1037, Nov. 2020, doi: 10.1080/08839514.2020.1796370.

[11] A. T. Sahlol, M. Abd Elaziz, M. A. A. Al-Qaness, and S. Kim, "Handwritten Arabic Optical Character Recognition Approach Based on Hybrid Whale Optimization Algorithm With Neighborhood Rough Set," *IEEE Access*, vol. 8, pp. 23011–23021, 2020, doi: 10.1109/ACCESS.2020.2970438.

[12] Patrick J Grother and Kayee K Hanaoka, "NIST Special Database 19 Handprinted Forms and Characters Database." National Institute of Standards and Technology, Sept. 13, 2016.

[13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[14] G. Cohen, S. Afshar, J. Tapson, and A. van Schaik, "EMNIST: an extension of MNIST to handwritten letters," Mar. 01, 2017, *arXiv*: arXiv:1702.05373. doi: 10.48550/arXiv.1702.05373.

[15] U.-V. Marti and H. Bunke, "The IAM-database: an English sentence database for offline handwriting recognition," *Int. J. Doc. Anal. Recognit.*, vol. 5, no. 1, pp. 39–46, Nov. 2002, doi: 10.1007/s100320200071.

[16] C.-L. Liu, F. Yin, D.-H. Wang, and Q.-F. Wang, "CASIA Online and Offline Chinese Handwriting Databases," in *2011 International Conference on Document Analysis and Recognition*, Beijing, China: IEEE, Sept. 2011, pp. 37–41. doi: 10.1109/ICDAR.2011.17.

[17] A. M. Hafiz and G. M. Bhat, "Boosting OCR for Some Important Mutations," in *2015 Second International Conference on Advances in Computing and Communication Engineering*, Dehradun, India: IEEE, May 2015, pp. 128–132. doi: 10.1109/ICACCE.2015.120.

[18] J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," *IEEE Access*, vol. 8, pp. 142642–142668, 2020, doi: 10.1109/ACCESS.2020.3012542.

[19] M. AbdElNafea and S. Heshmat, "Efficient Preprocessing Algorithm for Online Handwritten Arabic Strokes," in *2019 International Conference on Innovative Trends in Computer Engineering (ITCE)*, Aswan, Egypt: IEEE, Feb. 2019, pp. 64–69. doi: 10.1109/ITCE.2019.8646643.

[20] Ahmed. T. Sahlol, C. Y. Suen, M. R. Elbasyoni, and A. A. Sallam, "Investigating of Preprocessing Techniques and Novel Features in Recognition of Handwritten Arabic Characters," in *Advanced Information Systems Engineering*, vol. 7908, C. Salinesi, M. C. Norrie, and Ó. Pastor, Eds., in Lecture Notes in Computer Science, vol. 7908. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 264–276. doi: 10.1007/978-3-319-11656-3_24.

[21] Herviana Masrani, Ilhamsyah, and Ikhwan Ruslianto, "APLIKASI PENGENALAN POLA PADA HURUF TULISAN TANGAN MENGGUNAKAN JARINGAN SARAF TIRUAN DENGAN METODE EKSTRAKSI FITUR GEOMETRI," *Coding J. Komput. Dan Apl.*, vol. 6, no. 2, July 2018, doi: 10.26418/coding.v6i2.26674.

[22] H. Choudhury and S. R. M. Prasanna, "Handwriting recognition using sinusoidal model parameters," *Pattern Recognit. Lett.*, vol. 121, pp. 87–96, Apr. 2019, doi: 10.1016/j.patrec.2018.05.012.

[23] Md. A. Hossain and Md. M. Ali, "Recognition of Handwritten Digit using Convolutional Neural Network (CNN)," *Glob. J. Comput. Sci. Technol.*, pp. 27–33, May 2019, doi: 10.34257/GJCSTDVOL19IS2PG27.

[24] V. Garg, "Handwritten Text Classification using Deep Learning," vol. 07, no. 05, 2020.

[25] A. Biswas and Md. S. Islam, "An Efficient CNN Model for Automated Digital Handwritten Digit Classification," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 7, no. 1, p. 42, Apr. 2021, doi: 10.20473/jisebi.7.1.42-55.

[26] S. Prihatiningsih, N. S. M, F. Andriani, and N. Nugraha, "ANALISA PERFORMA PENGENALAN TULISAN TANGAN ANGKA BERDASARKAN JUMLAH ITERASI MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK," *J. Ilm. Teknol. Dan Rekayasa*, vol. 24, no. 1, pp. 58–66, 2019, doi: 10.35760/tr.2019.v24i1.1934.

[27] S. Momeni and B. BabaAli, "A transformer-based approach for Arabic offline handwritten text recognition," *Signal Image Video Process.*, vol. 18, no. 4, pp. 3053–3062, June 2024, doi: 10.1007/s11760-023-02970-9.

[28] Y. Li, D. Chen, T. Tang, and X. Shen, "HTR-VT: Handwritten text recognition with vision transformer," *Pattern Recognit.*, vol. 158, p. 110967, Feb. 2025, doi: 10.1016/j.patcog.2024.110967.

[29] Q. Zhang, F. Liu, and W. Song, "IMTLM-Net: improved multi-task transformer based on localization mechanism network for handwritten English text recognition," *Complex Intell. Syst.*, vol. 11, no. 1, p. 125, Jan. 2025, doi: 10.1007/s40747-024-01713-8.

[30] F. Ott, M. Wehbi, T. Hamann, J. Barth, B. Eskofier, and C. Mutschler, "The OnHW Dataset: Online Handwriting Recognition from IMU-Enhanced Ballpoint Pens with Machine Learning," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 3, pp. 1–20, Sept. 2020, doi: 10.1145/3411842.

[31] N. Altwaijry and I. Al-Turaiki, "Arabic handwriting recognition system using convolutional neural network," *Neural Comput. Appl.*, vol. 33, no. 7, pp. 2249–2261, Apr. 2021, doi: 10.1007/s00521-020-05070-8.

[32] M. Potanin *et al.*, "Digital Peter: New Dataset, Competition and Handwriting Recognition Methods," in *The 6th International Workshop on Historical Document Imaging and Processing*, Lausanne Switzerland: ACM, Sept. 2021, pp. 43–48. doi: 10.1145/3476887.3476892.

[33] N. Toiganbayeva *et al.*, "KOHTD: Kazakh offline handwritten text dataset," *Signal Process. Image Commun.*, vol. 108, p. 116827, Oct. 2022, doi: 10.1016/j.image.2022.116827.

[34] R. Zouari, H. Boubaker, and M. Kherallah, "Multi-language online handwriting recognition based on beta-elliptic model and hybrid TDNN-SVM classifier," *Multimed. Tools Appl.*, vol. 78, no. 9, pp. 12103–12123, May 2019, doi: 10.1007/s11042-018-6764-0.

[35] Z. Chen, F. Yin, X.-Y. Zhang, Q. Yang, and C.-L. Liu, "MuLTReNets: Multilingual text recognition networks for simultaneous script identification and handwriting recognition," *Pattern Recognit.*, vol. 108, p. 107555, Dec. 2020, doi: 10.1016/j.patcog.2020.107555.

[36] W. Jiang, "MNIST-MIX: a multi-language handwritten digit recognition dataset," *IOP SciNotes*, vol. 1, no. 2, p. 025002, Sept. 2020, doi: 10.1088/2633-1357/abad0e.

[37] Yan Gao, L. Jin, and Nanxi Li, "Chinese handwriting quality evaluation based on analysis of recognition confidence," in *2011 IEEE International Conference on Information and Automation*, Shenzhen, China: IEEE, June 2011, pp. 221–225. doi: 10.1109/ICINFA.2011.5948991.