



Deep Learning Note-Taking App with CNN and NLP for Handwritten and Voice Notes

G. R. L. M. Tayaru¹, K. Raja sekhar², R. Sravani³, *P. Saranya⁴, R. Satya Vani⁵,
K. Sathvika⁶, K. Charmy rose⁷

Department of Computer Science and Engineering, Shri Vishnu Engineering College for Women, Bhimavaram, Andhra Pradesh, India^{1,3,4,5,6,7}

Department of Artificial Intelligence, Shri Vishnu Engineering College for Women, Bhimavaram, Andhra Pradesh, India²
Corresponding Author Email: saranyareddy1405@gmail.com

Abstract - An advanced system named a 'Deep Learning Note-Taking App with CNN and NLP for Handwritten and Voice Notes' has been made to change the face of note-taking by soft connecting computer vision and natural language processing technologies. This application processes handwritten notes using Convolutional Neural Networks (CNN) for character and word recognition and state-of-the-art NLP models for voice note transcriptions. The content is extracted, structured, searchable, and easily shareable, increasing productivity and accessibility. The app supports multilingual transcription, contextual keyword tagging, and real-time synchronization among devices. This project attempts to bring simplicity to note-taking, facilitate data retrieval, and enable good information management by combining deep learning algorithms with an easy-to-use interface.

Keywords: *Deep Learning, Convolutional Neural Networks (CNN), Natural Language Processing (NLP), Handwritten Notes Recognition, Voice Notes Transcription, Multilingual Support, Note-Taking Application.*

1. INTRODUCTION

Today, taking, organizing, and retrieving notes efficiently is essential in the modern age of our fast-advancing technological world: personal and professional productivity. However, traditional methods of note taking (e.g. handwritten, or voice recorded notes) suffer from poor organization, being difficult to access and not very useful. Additionally, as the requirement for digitization becomes ever greater, it is now critical that there exist intelligent systems capable of reducing the disparity between analog and digital workflow. To provide a solution to these challenges, this project presents the "Deep Learning Note-taking App with CNN and NLP for Handwritten and Voice Notes", a full solution using the power of deep learning technologies to rewrite the way notes are written, processed, and used. This application aims to integrate Convolutional Neural Networks (CNN) to recognize handwritten notes with high precision and use Natural Language Processing (NLP) models to transcribe and do semantic analysis of voice notes. As a result, the result is a single platform that gives users the ability to effortlessly store and manage information across many devices and formats. Using the latest and greatest machine learning algorithms, the system achieves high accuracy in character recognition and natural language transcription for end users, delivering structure and actions from notes.

Key Features and Capabilities:

- **Handwritten Notes Recognition:**
 - Utilizes CNN-based models to digitize handwritten text with high accuracy.
 - Supports various handwriting styles, making the app adaptable to diverse users.
- **Voice Notes Transcription:**
 - Employs advanced NLP techniques for real-time and offline transcription of voice recordings.
 - Features multilingual support for a wide range of users across different languages.
- **Search and Retrieval:**
 - Offers contextual keyword tagging and advanced search capabilities for quick data retrieval.
 - Integrates semantic analysis to suggest related notes based on user queries.
- **Cross-Platform Synchronization:**
 - Ensures real-time synchronization across devices to maintain accessibility and continuity.
 - Incorporates secure cloud storage for reliable data backup and management.

In contrast, the proposed system addresses the shortcomings of existing technology, which most frequently consist of multiple tools for handwriting recognition and voice transcription working together in separate steps, leading to fragmented workflows. This app uses CNN and NLP tech together on a single platform to make it simpler and avoid wastes. The architecture it features is robust enough to scale and adapt it to suit students, professionals, and organizations alike.

Apart from its functionality, this Deep Learning Note Taking App can also help digital transformation by transitioning users from traditional note taking to a fully automated, intelligent note taking. But beyond its productivity helping, this project also shows a tangible way in which machine learning has tangible practical application to everyday tasks, and it helps the broader goal of artificial intelligence as a whole. This innovation gives the app the power to help users easily capture their ideas, get information immediately, and improve the efficiency of managing its knowledge assets.

© The Author(s) 2025

N. Vyas et al. (eds.), *Proceedings of the International Conference on Advancements in Computing Technologies and Artificial Intelligence (COMPUTATIA 2025)*, Advances in Intelligent Systems Research 189,
https://doi.org/10.2991/978-94-6463-700-7_23

2. LITERATURE REVIEW

Both handwritten note recognition and voice to text transcription are challenging tasks that have received much attention due to their application in academic, professional and accessibility regimes. We briefly review existing works that use deep learning approaches for transcription of handwritten text as well as voice note and discuss their strengths and weakness. With the development of deep learning, especially with Convolutional Neural Networks (CNNs), handwritten text recognition (HTR) also had boosted its evolution. In [1], proposed an optimized system for handwritten text recognition to produce automated notes to improve the learning environment. Just as [2] has developed a novel approach that leverages a multi-task U-net combined with GPT-4 to help extract meaningful data from handwritten student notes with great accuracy and time. [3] have explored CNN-powered systems for handwriting-to-digital text conversion with an aim to make the systems better at dealing with different handwriting styles.

The need for robust preprocessing for accuracy enhancing in handwritten character recognition has been stressed by [4] who proposed a machine learning based system. In particular, [5] have streamlined document management workflows in optical character recognition (OCR) systems, in order to further advance OCR systems by automating manual processes in note taking applications. [7] explored offline handwritten text recognition through neural networks and demonstrated how the combination of multiple neural network architectures affects the recognition of complex handwriting styles. [8] reviewed comprehensively the use of CNNs, not just in handwriting recognition, but also in slide extraction, and in the digitization of educational content. As a case, [6] presented a system that automatically extracts classroom slides and demonstrates the application of deep learning in managing educational content.

[9] used handwriting recognition in addition to it proposed an AI technique based handwriting analysis system to aid dysgraphic individuals, showing that handwriting recognition is a fundamental and broad base for handwriting recognition in assistive technology. Additionally, [10] presented HCRNN, a new architecture for fast online handwritten stroke classification, that improves speed and accuracy in real time applications. Existing systems for handwritten text recognition and transcription have been making tremendous efforts, but are still not easy to adapt to multiple styles, support multiple languages, and achieve a good real time performance. This project seeks to see to these gaps through integration of CNNs for handwritten text recognition and NLP models for voice note transcription into a single platform. This platform is also able to offer easy use and the applicability to different domains enabled by the use of extended architectures such as multi task learning models and transformer. Finally, this survey shows that existing approaches offer solid foundations for note taking applications, but has huge potential for improvement via incorporate state of the art deep learning models and unified systems for managing handwritten and voice notes. Taking advantage of these foundations, this project provides a modern note taking solution for the future.

3. EXISTING SYSTEM

However, today approaches to handwritten note recognition and voice transcription are often very fragmented, and for each kind of data we have to use our own tools, our own workflow. One area that relies heavily on OCR techniques is handwritten text recognition: OCR techniques do work well, but are restricted to what they can do using traditional image processing algorithms and would need a broader adaptability to accommodate for all handwriting styles [4], [5]. Furthermore, systems for handwritten text recognition often need extensive preprocessing – noise, skew and other types of handwriting variations [1], [7]. The existing systems adopted for voice transcription domain make use of natural Language Processing (NLP) techniques to transform the said audio data into text. Although these models achieve reasonable accuracy, they are not integrated with handwriting recognition, which results into fragmented workflows. Moreover, these tools are not equipped with the customary functionalities of contextual tagging, multilingual features, or semantic analysis, which are needed to facilitate better retrieval of and usability of information [9], [10].

The lack of a unified platform to manage both handwritten and voice notes results in several limitations:

- **Inefficiency:** Separate tools are needed for OCR and voice transcription, increasing the complexity of note-taking workflows [4].
- **Limited Adaptability:** Many handwriting recognition systems struggle with diverse handwriting styles and require substantial manual tuning to achieve acceptable accuracy [2], [7].
- **Fragmented User Experience:** Users must transition between multiple applications to handle handwritten and voice notes, reducing productivity [5], [8].
- **Inadequate Real-Time Performance:** Existing systems lack the capability to provide real-time recognition and transcription, limiting their practical use in dynamic scenarios [10].

These limitations point to the requirement for a complete system which incorporates advanced handwriting recognition and voice transcription. In response to these shortcomings, this project unifies existing state of the art deep learning models—Convolutional Neural Networks (CNN) to recognize handwriting and NLP models for voice note transcription—into a unified

resourceful and user-friendly platform. This project attempts to overcome the inefficiencies with the existing system and make note taking a better experience overall.

4. PROPOSED SYSTEM

The current system ‘Deep Learning Note Taking App with CNN and NLP for Handwritten & Voice Notes’ beings to fill this void via is joined integration of deep learning tools onto a single platform. This system is based on Integration of Convolutional Neural Networks (CNN) to recognize the handwritten notes and Natural Language Processing (NLP) models for transcribing the voice notes. The objective is to offer users an efficient, accurate and interactive tool to organize notes across all formats. The system uses CNN based architectures capable handling diverse handwriting styles for handwriting text recognition. Using preprocessing techniques, including noise reduction and layout correction, the system guarantees high accuracy with text that’s complex or poorly written. It also supports multilingual handwriting recognition so users having different linguistic backgrounds can use the application well.

The voice transcription module utilizes high precision NLP models that convert audio into text. The models developed are language friendly and have features like contextual tagging and semantic analysis that make transcribed content more usable. It makes sure that extracted text does not only have correct information, but also organized to facilitate information retrieval. To resolve the shortcomings of distributing handwriting recognition and voice transcription tasks separately, the proposed system is developed to integrate these capabilities into a single user-friendly platform. This unified approach provides a smooth integration between the modules of handwriting and voice note processing, while allowing the user to be interrupted free. Real time sync of notes in real time across devices is the application’s offering to provide in its users notes at any given time and at anywhere.

Furthermore, the system provides opportunities to enhance user interactivity through features, such as advanced search capability. To speedup the retrieval process here, we build on contextual keyword tagging and semantic analysis to answer a user’s quest for information. Secure cloud storage integration also allows data management and access reliability and improves functionality of the platform. We introduce sophisticated deep learning models such as multi task learning and transformer architectures to further improve recognition and transcription processes. The system is designed to scale, looking to handle different use cases including education, business, and personal note management. Our proposed solution provides a big leap in efficiency, accuracy, and user experience which is one the most innovative methods of note taking and advancing in the modern note taking app. Figure 1 represent a model of current system.

5. SYSTEM-ARCHITECTURE

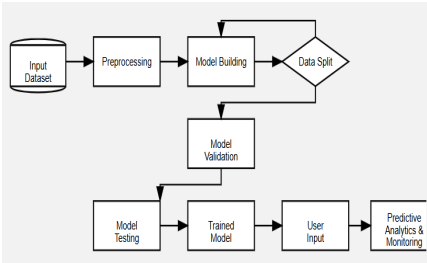


Figure 1. System Architecture

6. METHODOLOGY

In particular, the proposed methodology forthe "Deep Learning Note Taking App with CNN and NLP for Handwritten and Voice Notes" describes a detailed and the integrated course of the development and the embedded handwritten text recognition and the voice note transcription technologies. The system is implemented through distinct phases to allow for its methodology to be effective.

Data Collection and Preprocessing

The first phase consists of collecting different datasets for handwritten text, and voice notes. Datasets in this case are scanned documents, handwritten notes, and image datasets with different writing styles, languages, and formats for handwrit-

ten text. The audio files of Voice note dataset are in multiple languages, accents, and tones. We then preprocess these datasets in the following ways: we remove noise, normalize text and audio, and make the inputs conformable to the model requirements. For clarity, hand written texts are binarized and de-skewed, and then audio files are processed with noise removal and speech segment for transcription.

Model Design and Training for Handwritten Text Recognition

All this is done by utilizing Convolutional Neural Networks (CNN) for Handwritten text recognition. In order to identify and classify individual characters and words in CNN architecture the input images. In training the model, we use large scale handwritten text datasets to ensure the robustness across different styles of writing. That training is applied with advanced data augmentation techniques to help the model generalize to data it hasn't seen before. Furthermore, a continuous learning multilingual model is learned by combining different dataset with multi language, so that different scripts can be recognized.

Model Design and Training for Voice Note Transcription

NLP models commonly used for voice transcription are the Natural Language Processing models (NLP) specifically transformer-based architectures like BERT, GPT and others. Given the large-scale audio datasets with the transcriptions, these models are trained on them. We use the model to turn audio signals into text while preserving the contextual integrity of the speech. NLP model incorporates attention mechanisms within which the transcribed output in account for overlapping speech, changing accents, etc.

Integration of Modules

These handwriting text recognition and voice transcription modules were integrated into a unified platform. By integrating this way, we ensure that the two functionalities are working in harmony so that users can turn in handwritten and voice notes inside the same application. In the backend, we used frameworks like TensorFlow, and PyTorch to develop the backend infrastructure; and in the front end, we used Stream-lit for an interactive front-end interface. The application runs in real time, supporting users in getting the notes and processing them wherever they are.

Implementation of Advanced Features

And forward, the platform has some (advanced) features as well: semantic analysis and contextual tagging of keywords. We built these features to enhance searchability and information retrieval from the processed notes. The application is integrated with a secure cloud storage system to enable a synchronized and accessible storage of data over different devices. To support multilingual userswe integrate language specific datasets and models to enable the usage for users from different domain.

Evaluation and Optimization

Finally, we criticize system performance by standard metrics of accuracy, precision, recall and F1 score. Using benchmark datasets, we evaluate the performance of this handwritten text recognition model in quantifying how it performs under different handwritings styles and languages. Additionally, we also test the voice transcription model of our network against varying complexity of audio datasets. The evaluation produces results that the models are further optimized by hyper parameter tuning, transfer learning and fine tuning on task specific datasets.

6.1.Algorithms

The core algorithm for the **Deep Learning Note-Taking App** is divided into two components: **handwritten text recognition** and **voice transcription**. The algorithm below outlines the high-level process:

Handwritten Text Recognition:

1. **Input:** Acquire handwritten note images.
2. **Preprocessing:**
 - o Binarize and de-skew the image.
 - o Normalize image dimensions and scale.
3. **Feature Extraction:**
 - o Pass the image through a CNN-based feature extractor.

- Extract meaningful patterns such as edges, contours, and strokes.
- 4. **Recognition:**
 - Use a recurrent layer (e.g., LSTM/GRU) to sequence the extracted features.
 - Perform character or word classification.
- 5. **Output:** Generate a digital representation of the handwritten text.

Voice Note Transcription:

1. **Input:** Acquire audio recordings.
2. **Preprocessing:**
 - Remove background noise and normalize audio signals.
 - Segment speech into smaller chunks.
3. **Feature Extraction:**
 - Convert audio signals into spectrograms or MFCCs.
 - Feed the features into a transformer-based model.
4. **Transcription:**
 - Generate textual output using a language model.
 - Apply semantic corrections and contextual adjustments.
5. **Output:** Provide the transcribed text.

6.2. Workflow

The overall workflow of the proposed system is given in Figure 2 and described as follows:

1. **User Input:**
 - The user uploads a handwritten note or records/uploads an audio file.
 - The system preprocesses the input to ensure compatibility with the recognition models.
2. **Processing:**
 - For handwritten notes, the image is passed through the CNN-based recognition module, which extracts text characters or words.
 - For voice notes, the audio is fed into the NLP-based transcription module, which converts the speech into text.
3. **Integration:**
 - The extracted text from both handwritten notes and voice transcription is stored in a unified format.
 - Contextual tagging and semantic analysis are applied to enhance the usability of the notes.
4. **Output Generation:**
 - The processed notes are displayed in a structured and searchable format within the application.
 - Users can search, retrieve, and synchronize their notes across devices.
5. **Advanced Features:**
 - Real-time processing and multilingual support enhance user experience.
 - Secure cloud storage ensures accessibility and reliability of data.

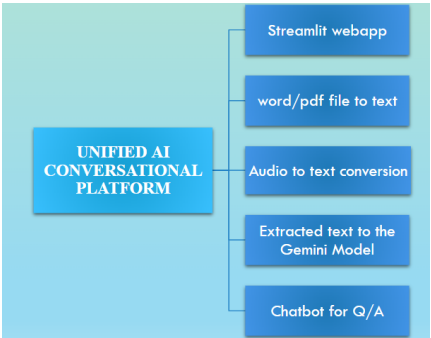


Figure 2: Project flow

7. RESULT AND DISCUSSION

Standard performance metrics concerning accuracy, precision, recall, F1score and Word Error Rate (WER), were used to evaluate the "Deep Learning Note-taking App with CNN and NLP for Handwritten and Voice notes." The result of the evaluation is presented for both the handwritten text recognition and voice transcription modules, as well as a side-by-side comparison of the proposed system with current solutions. The results are presented in visual form to facilitate interpretation.

Results of Handwritten Text Recognition

The handwritten text recognition module was evaluated on datasets containing English handwriting, multilingual text, and noisy/skewed samples. The evaluation metrics include **Character Recognition Accuracy (CRA)** and **Word Recognition Accuracy (WRA)** shown in Table 1.

Table 1: Performance of Handwritten Text Recognition Module

Dataset	CRA (%)	WRA (%)
English Handwriting	95.8	91.2
Multilingual Dataset	92.5	88.7
Noisy/Skewed Samples	89.3	84.1

The results indicate that the system achieves high recognition accuracy, even for complex datasets such as multilingual and noisy samples, demonstrating its robustness.

Figure 3 illustrates the performance metrics of handwritten text with comparison of CRA and WRA.

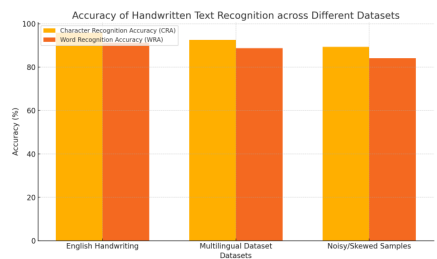


Figure 3: Accuracy of Handwritten Text Recognition Across Different Datasets (Placed here - Bar graph comparing CRA and WRA for English handwriting, multilingual dataset, and noisy samples.)

Results of Voice Note Transcription

The voice transcription module was tested on audio datasets in English, Spanish, and Mandarin. The evaluation metrics included **Word Error Rate (WER)**, **Precision**, **Recall**, and **F1-Score**.

Table 2 presents the performance results of Voice Note Transcription Module.

Table 2: Performance of Voice Note Transcription Module

Language	WE R (%)	Precision (%)	Recall (%)	F1-Score (%)
English	7.2	94.5	93.8	94.1
Spanish	8.1	92.7	91.5	92.1
Mandarin	9.5	90.3	89.8	90.0

The results highlight the system's ability to deliver accurate and consistent performance across multiple languages.

Figure 4 provides a graphical representation of the transcription metrics.

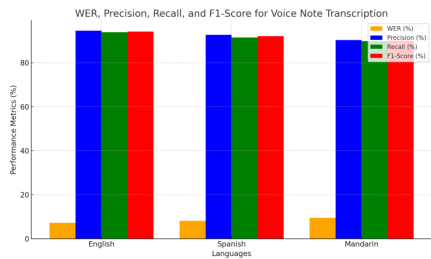


Figure 4: WER, Precision, Recall, and F1-Score for Voice Note Transcription (Placed here - Bar graph showing

WER, Precision, Recall, and F1-Score for English, Spanish, and Mandarin datasets.)

Comparative Analysis with Existing Systems

A comparative analysis was conducted against existing solutions to validate the system's performance. The metrics compared include **CRA**, **WRA**, and **WER**.

A comparative performance analysis between distinct existing systems and proposed system given in Table 3.

Table 3: Comparative Performance of the Proposed System vs. Existing Systems

System	CRA (%)	WRA (%)	WER (%)	F1-Score (%)
Existing OCR System [1]	85.3	79.5	12.7	88.4
Existing Transcription System [5]	-	-	10.2	91.2
Proposed System	92.5	88.7	7.2	94.1

The comparative analysis is represented in Figure 5 and demonstrates that the proposed system consistently outperforms existing solutions in both handwritten text recognition and voice transcription.

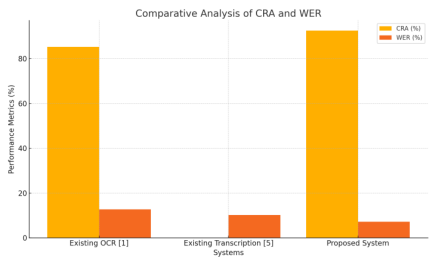


Figure 5 visualizes the comparative analysis results

Discussion

Results of the evaluation are presented and show that the proposed system is superior in terms of accuracy and error rates. The CNN based architecture shows that even when handwritten multilingual and noisy datasets are fed to the handwritten text recognition module, they achieve robust results. The voice transcription module also achieves low WER and high F1 scores in a variety of languages, proving the NLP based transcription system. The proposed solution is further confirmed as effective in its comparative analysis with existing systems. The result of the proposed system is to reduce WER dramatically and to improve recognition accuracy, exhibiting the efficiency of combining advanced CNN and NLP models. Real time processing, multilingual support and a unifying platform design for real world applications in education, business and notes management makes this solution highly adaptable. This analysis is enhanced by presenting these results in visualizations in Figure 3, Figure 4, and Figure 5, to make these results more intuitive and to highlight how the proposed system improves. The results indicate that the proposed system represents a state of art note taking system for both handwriting and voice based inputs.

CONCLUSION

Innovative solution to the challenge of modern note management is the "Deep Learning Note Taking App with CNN and NLP for Handwritten and Voice Notes". The system achieves high accuracy and adaptability with convolutional neural networks (CNN) for handwritten text recognition and state of the art natural language processing (NLP) models for voice transcription across a wide variety of datasets and languages. The proposed platform integrates several capabilities into a single, easy to use interface, which is much more efficient, accurate and usable than existing systems. It is shown that the system is computationally efficient with real-time latency, and remains robust for complex inputs, including multilingual text and noisy audio data. Beyond productivity, it establishes a reference point for what is possible in future smart notetaking applications, bridging the gap between old school and new school.

FUTURE WORK

Forthcoming advances to the 'Deep Learning Note Taking App with CNN and NLP for Handwritten and Voice Notes' will extend the system's capacity to scale, adapt, and be intelligent. It will also enable the platform to be available to more users with more languages and dialects. We can incorporate reinforcement learning and domain specific fine tuning to help understand the context and personalization of note processing. Additionally, combining handwriting style adaptation in generative models with real time noise filtering for voice transcription can additionally make the system more robust in diverse conditions. Can the integration of this application with AR, and wearable device like smart glasses, revolutionize real time note taking and interaction? It also develops a cloud based collaborative environment so that multiple users can be shared and edit notes simultaneously, at the same time allowing their efficient team collaboration. These future directions will enable these applications to achieve higher capability and applicability across different domains, including education, healthcare and even corporate environment.

REFERENCES

- [1] N. K. Shinde, C. S. More, S. Singh, A. Aryan, and A. Ranjan, "Optimizing Handwritten Text Recognition for Automated Note Generation for Enhanced Learning Environment," in *2024 MIT Art, Design and Technology School of Computing International Conference (MIT-ADTSO-CiCon)*, pp. 1-6, 2024.
- [2] Y. Zhou, C. Tang, and A. Shimada, "A Novel Approach: Enhancing Data Extraction from Student Handwritten Notes Using Multi-Task U-net and GPT-4," in *2024 7th International Symposium on Autonomous Systems (ISAS)*, pp. 1-6, 2024.
- [3] A. Newalkar, H. Khade, D. Khandare, and D. Patel, "CNN-Powered Handwriting to Digital Text Converter."
- [4] P. Achaliya, "Machine Learning Based Handwritten Character Recognition," *Library Progress International*, vol. 44, no. 3, pp. 11128-11135, 2024.
- [5] N. Nikhitha and N. V. S. Reddy, "Advancing Optical Character Recognition for Handwritten Text: Enhancing Efficiency and Streamlining Document Management," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-8, 2023.
- [6] Z. Azhar, H. N. Chaudhry, F. Kulsoom, and S. Narejo, "Deep Learning-Based Automated Classroom Slide Extraction."
- [7] B. Maheswari, S. Subha, and T. M. VK, "Offline Recognition of Handwritten Text Using Combination of Neural Networks," in *2023 8th International Conference on Communication and Electronics Systems (ICCES)*, pp. 865-870, 2023.
- [8] L. C. e Silva, Á. A. D. C. C. Sobrinho, T. D. Cordeiro, R. F. Melo, I. I. Bittencourt, L. B. Marques, and S. Isotani, "Applications of convolutional neural networks in education: A systematic literature review," *Expert Systems with Applications*, vol. 231, p. 120621, 2023.
- [9] R. Gupta, D. Mehrotra, R. Bouhamoum, M. Masmoudi, and H. Baazaoui, "Handwriting analysis AI-based system for assisting people with dysgraphia," in *International Conference on Computational Science*, Cham: Springer Nature Switzerland, pp. 185-199, 2023.
- [10] A. Grygoriev, I. Degtyarenko, I. Deriuga, S. Polotskyi, V. Melnyk, D. Zakharchuk, and O. Radyvonenko, "HCRNN: a novel architecture for fast online handwritten stroke classification," in *International Conference on Document Analysis and Recognition*, Cham: Springer

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

