Yakovchuk, Oleg; Vasin, Maksym

Article

Increasing the accuracy of handwriting text recognition in medical prescriptions with generative artificial intelligence

Technology audit and production reserves

Provided in Cooperation with:
ZBW OAS

Oleg Yakovchuk,
Maksym Vasin

INCREASING THE ACCURACY OF HANDWRITING TEXT RECOGNITION IN MEDICAL PRESCRIPTIONS WITH GENERATIVE ARTIFICIAL INTELLIGENCE

The object of the research is a system for recognizing handwritten text in medical prescriptions. The peculiarities of handwriting, the variety of calligraphy styles, as well as the specificity of medical prescriptions, create many problems and challenges for recognition algorithms, causing errors and reducing recognition accuracy.

The work presents a new system with additional components of post-processing the recognition results to increase the accuracy of the final results. An algorithm for combining words into lines and blocks is proposed, which makes it possible to group words while preserving contextual connections between them. Also, a generative neural network with a large language model is used to analyze the recognition result and correct possible errors. The results of the testing show an improvement in recognition accuracy by 0.13 %. Successful cases of generative artificial intelligence usage are analyzed, as well as examples of the results deterioration, that are related to grammatical errors in the initial input data.

The obtained results show the use of generative artificial intelligence as an additional step for processing the recognition results really can improve the accuracy of text recognition systems. The results of the study can be used for further experiments to improve recognition results in other tasks related to text recognition and in related fields.

Keywords: *handwriting recognition, generative artificial intelligence, recognition algorithms, deep neural networks.*

1. Introduction

Digitization of medical prescriptions is an important step in the development of modern medicine, which allows for improving the quality of medical services and providing faster and more convenient access to the necessary information. Receiving, transferring, and storing prescriptions in a digital way, reduces transcription errors, speeds up processing, and keeps medical records more secure than paper. The development of handwriting recognition systems for the digitization of medical prescriptions is a relevant and promising task that can increase the efficiency of the medical system and provide convenience for both patients and medical specialists [1].

Despite the handwriting recognition is already a well-known task, it is still not fully solved and is challenging for researchers and developers. Handwritten text itself is much more complex and visually diverse than printed text. This creates difficulties for recognition systems when they meet other languages, handwriting styles, and new ways of writing the same letters. To ensure high recognition accuracy, the system has to be adapted to a variety of writing styles, which requires a large amount of training and testing data. A separate problem is the connected handwriting when several characters or all characters in a word can be written without lifting the pen from the surface. Such writing creates new, unique connections between letters that do not occur when each letter is written separately. Therefore, modern recognition systems also have to cope with such situations [2].

Also, each task of handwriting recognition has its own specifics. In our case, handwritten medical prescriptions have clearly expressed differences from other handwritten documents – a certain pattern of the prescription, the presence of specific concepts and medical terms, and other features that can be used to improve recognition accuracy. They will be described in more detail in the next chapter.

The aim of this research is to investigate the existing problems of handwriting text recognition in medical prescriptions and to develop an improved approach based on generative artificial intelligence (AI), which will make it possible to correct part of the errors and increase the final recognition accuracy. The proposed approach can be used to improve the quality of existing or new similar systems, both in the medical domain and in other areas where handwriting recognition is used.

2. Materials and Methods

The object of the research is a system for recognizing handwritten text in medical prescriptions. The developed system consists of several components.

The main module is responsible for detecting and recognizing handwritten text. As input, it takes an image containing handwritten text and returns all recognized words in text format, along with the coordinates of each word's boundaries. The next components are postprocessing modules: merging of words in lines and blocks taking into account the peculiarities of the structure and format of the prescription, and a module for the correction of possible recognition errors using generative AI.

2.1. Handwriting recognition module. For the experiments, as a handwriting recognition module, the Handprint recognition system was used, that comprises of such services as Amazon-Rekognition, Amazon-Textract, Google, Microsoft [3]. The datasets used to train these services far exceed the available data in the public domain in terms of quantity and variety. The general system architecture is not tied to a specific implementation of the recognition module, so the selected service can be replaced by others.

A common feature of most modern approaches to handwriting recognition is the use of recurrent layers in deep neural network architectures. Recurrent networks use sequential data or time-series data, which can significantly improve accuracy for a natural language recognition task because the order of the words in the sequence is very important. A feature of recurrent neural networks is their «memory», as they have to take into account previous input data in order to change the current input and output data [4]. While traditional deep neural networks assume the inputs and outputs are independent of each other, the output of recurrent neural networks depends on the previous elements in the sequence.

The handwriting recognition module also uses Connectionist Temporal Classification (CTC) – a type of neural network output data and its associated evaluation function [5]. CTC was created as a new method for labeling non-segmented data sequences. The idea is to interpret the output of the neural network as a conditional probability distribution over all possible marks given the input sequence. Knowing the distribution, it is possible to get an objective function that maximizes the probabilities of correct labeling. Since the objective function is differentiated, the neural network can be trained by the standard method of backpropagation of the error in time [5].

2.2. Input data description. To evaluate the accuracy of the system, an in-house dataset was collected from publicly available data (for example, [2]) and labeled by ourselves. The test dataset contains 40 images of different quality with handwritten prescriptions in English. The average number of words in one image is about 50. An example of one of these images is shown in Fig. 1.

Since the medical prescription may be a pre-printed form to be filled out, the image may also contain printed symbols. However, even if the prescription is written on a blank paper and contains only a single drug, such text will most likely consist of several lines, since the doctor must also write the dosage, the period of taking the medication, put the date, and signature. Background noise is present in most images – the paper is usually not completely white, there may also be template marking with lines or cells, there may be shadows from various objects on the paper, and the image may contain not only a prescription but also external objects on the sides.
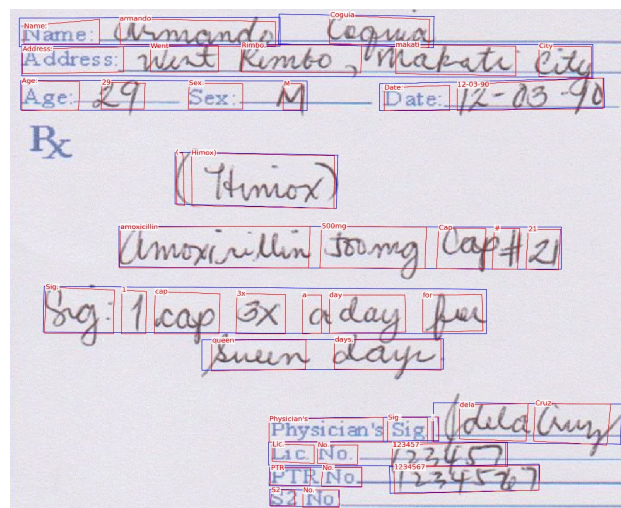

Fig. 1. An example of a medical prescription and the result of the main module of handwriting recognition. Red boxes indicate the coordinates of recognized words, blue boxes indicate the result of combining words into lines and blocks [2]

A prescription often contains additional information, such as the address and contact details of the medical facility, standard general warnings, and precautions for patients, advertising information, etc., that is not critical information for the common patient for prescription usage. Therefore, it is possible to select the major data that should be recognized with maximum accuracy – these are prescribed drugs, dosage, time of medication, further steps for the patient, and additional information related to the medication.

The specificity of medical terminology is a separate problem when using the base handwriting recognition approaches for solving the given task. A large number of terms that are used exclusively in medicine significantly reduces the effectiveness of using common language models and auto-correction.

2.3. Postprocessing modules. Most often, the prescription text consists of many lines and is not always arranged consecutively. The services in the recognition module are designed to work with different text files and do not take into account the specific peculiarities or structure of certain documents. In the example (Fig. 1), it can be seen that with the aligned image position, all the word boxes (highlighted in red) for each line are located approximately on the same y-coordinate. However, the recognizer extracted the phrases «PTR No» and «1234567» as two different strings (blue boxes). Such splitting of a single line into several lines creates problems during further processing steps, as the semantic connection between the words belonging to the same phrase is lost. Therefore, an algorithm was developed and utilized to combine words into strings and blocks to save their semantic connections.

The main idea of the algorithm is that by adding new words to an existing line, a new average center and a new average height of the line are calculated. A word is added

to a line only if a significant part of the word belongs to that line. A word that does not fit into an existing line creates a new line. The formulas for calculating the new average center (1) and the average height of the line (2) are recursive, that, knows the previous value without the added word, it is possible to get the next one with the added word without extra recalculations.

$$\begin{cases} C_{n+1}^{x} = \left(C_n^x \cdot n + x_n \right) / \left(n+1 \right), \\ C_{n+1}^{y} = \left(C_n^y \cdot n + y_n \right) / \left(n+1 \right), \end{cases} \quad (1)$$

$$H_{n+1} = \left(H_n \cdot n + h_n \right) / \left(n+1 \right). \quad (2)$$

In a similar way, words can be combined into blocks, based on the geometric distances between the found word boxes, and taking into account the general structure of medical prescriptions, where the head block, the central part, and the signature block of the prescription are usually clearly distinguished [6]. This makes it possible to combine the recognized words into whole sequences of words without breaking semantic connections, which is important for the next module.

The next step of postprocessing the result is analyzing the text with generative AI. For the experiment, the Chat-GPT service was chosen as a ready-to-use large language model that can analyze the recognized text using the context provided to it and correct possible errors made previously by the recognition module.

GPT-3 uses the same model and architecture as GPT-2 [7], including modified initialization, pre-normalization, and reverse tokenization. However, GPT-3 utilizes dense and locally connected sparse attention patterns, that alternate in the transformer layers, similar to a sparse transformer [8, 9].

The model analyzes the input sequence, breaking it into tokens. Tokens can be words, parts of words, or even single characters. Most transduction models of neural sequences have an encoding-decoding structure. Specifically, the encoder maps an input sequence of symbol representations $\left(x_1, \ldots, x_n \right)$ to a sequence of continuous representations $z = \left(z_1, \ldots, z_n \right)$. Having z, the decoder generates an output sequence $\left(y_1, \ldots, y_m \right)$ of symbols one at a time. At each step, the model is autoregressive, using previously generated characters as additional input data for text generation [10].

In the developed system, Chat-GPT utilizes as a web service – by generating a request to correct possible errors in the prescription and receiving a response after the processing. One of the base models that have been used is the «text-davinci-003» model. A separate field in the request is the temperature – randomization of the generated answer, during the experiments the temperature was set to 0.

3. Results and Discussion

The result of the study is the measurement of word and letter recognition accuracy based on the conducted experiments. Char recognition rate (CRR) and Word recognition rate (WRR) were used as metrics for comparisons. The results of the study are given in the Table 1.

In the experiment, 3 different configurations of the system for recognizing handwriting in prescriptions were used. The first approach uses recognition via the main module without additional processing, other approaches use pro-

cessing of the obtained result using a generative neural network with different requests. Each approach uses an algorithm for combining words into lines and blocks. The following requests to ChatGPT were used:
– Full edit request:
The task is to analyze the provided text and to correct possible errors in it.
– Partial edit request with context providing:
Doctor wrote medical prescription, then it was read by neural network with possible errors. The task is to analyze the provided text and to correct possible errors in it. Do not change the structure, do not remove or add line breaks. Do not change names, surnames, addresses.

From the obtained data, it is possible to see that processing the recognition results by generative AI with a simple full edit request degrades the accuracy. Looking at the results, there are many cases where Chat-GPT corrects grammatical errors, and can even rearrange words in a sentence to change the sentence structure to a more correct one. However, this is not a valid correction according to the marked ground truth values.

Table 1

Results of testing the developed system for recognizing handwritten text in medical prescriptions

Configuration	CRR	WRR
No postprocessing	0.936	0.814
Postprocessing via generative AI with full edit request	0.844	0.727
Postprocessing via generative AI with partial edit request and context providing	0.949	0.823

The described problem was solved in the third approach by providing the context, describing the input data, and setting restrictions on editing certain parts of the prescription, as well as additional tiny corrections to the network based on individual images. Thus, it is possible to see that in comparison to the basic approach without additional processing, in this approach CRR is improved by 0.13 % and WRR by 0.09 %. This result is explained by successful corrections of single and complex errors in words when neighboring words were recognized correctly, which allows Chat-GPT to correctly understand the context and correct an obvious error in a word (Fig. 2).

Analyzing the results of the third approach, there are a large number of cases when Chat-GPT correctly changed the word, correcting an obvious error and adjusting the changed word to the context, but this did not increase the accuracy of recognition in this image, but rather decreased it. This is because all data in the test dataset is marked according to the text written by doctors, and therefore grammatical errors made during writing are considered correct writing at the verification stage. Such examples are shown in Fig. 2. These errors are difficult to distinguish from real recognition errors, since the generative network receives only text as input, without any information from the input images.

The obtained results prove the ability of generative AI to successfully analyze the results of handwriting text recognition and correct errors made at previous stages. The Chat-GPT service used in the experiments showed it is possible to improve the basic recognition results when it is given the necessary context of the given task, and by setting limitations on the changing parts of the input sequence.
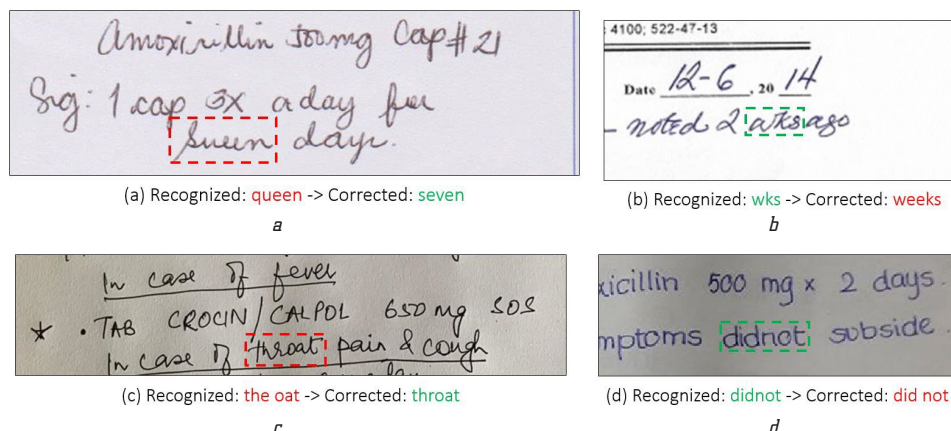
(a) Recognized: queen -> Corrected: seven

a

(b) Recognized: wks -> Corrected: weeks

b

(c) Recognized: the oat -> Corrected: throat

c

(d) Recognized: didnot -> Corrected: did not

d

Fig. 2. Examples of generative AI processing results with a partial edit request and context provided:
a, c — successful correction of the recognition error; *b, d* — failed correction of valid recognition

There were also cases of deterioration of the results, which is explained both by grammatical errors in the initial text, and therefore in the marked ground truth data, and by the peculiarities of the language, slang constructions, and possible abbreviations, when writing by hand. Other errors that were made in the initial recognition step and not corrected in the post-processing steps are usually related to complex handwriting, unusual writing styles, and overlapping text during handwriting. Such errors are caused by the limitations of the image handwriting recognition task itself, and cannot be corrected without additional information.

The results of the research can be used both to improve existing text recognition systems and to develop new systems, taking into account the possibilities of post-processing the results by generative AI. As can be seen from the results, the context that the generative network receives is really important, as well as the language models on which the network is trained that directly affect the obtained results. Despite the fact that the research was conducted under martial law in Ukraine, this did not affect the obtained results.

Further research of this approach may be directed to the development of specialized generative networks for more specific and distinct problems. For example, the analysis of the recognized text followed by selecting only the necessary information personalized for the user, in our task it can be the selection of information about drugs and their dosage for the patient. Such problem formulation removes some of the constraints imposed on the generative network and reduces the amount of redundant information that it must get from the input data.

4. Conclusions

In this work, the possibility of improving the recognition of handwritten text in medical prescriptions using generative AI was investigated. The main components of the developed system and the proposed approach of combining recognized words into lines and blocks were described. In addition, it was collected a dataset with handwritten prescriptions for conducting experiments. The system was tested with several variations of requests to the generative network. The result was an improvement in the accuracy of the character recognition rate by 0.13 % and the word recognition rate by 0.09 %. The accuracy increase is produced by the generative network that successfully corrected many of the recognition errors made by the core recognition module. The corrections are done by semantically analyzing the input data and using large language models to understand the context. Important to note, many cases where the generative network fails to correct grammatical errors are actually valid writing according to the displayed text in the prescription.

Based on the conducted research, it is possible to conclude about the great potential of generative AI for the task of text recognition, which, with the successful configuration of the network and the selection of language models that match the context of the task, can be successfully used to correct errors made at the previous stages of recognition.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this study, including financial, personal, authorship, or any other, that could affect the study and its results presented in this article.

Financing

The study was conducted without financial support.

Data availability

The manuscript has no associated data.

References

1. Baniulyte, G., Rogerson, N., Bowden, J. (2023). Evolution – removing paper and digitising the hospital. *Health and Technology, 13 (2),* 263–271. doi: https://doi.org/10.1007/s12553-023-00740-8
2. Dhar, D., Garain, A., Singh, P. K., Sarkar, R. (2020). HP_DocPres: a method for classifying printed and handwritten texts in doctor's prescription. *Multimedia Tools and Applications, 80 (7),* 9779–9812. doi: https://doi.org/10.1007/s11042-020-10151-w
3. Hucka, M. (2022). *Caltechlibrary/handprint: Release 1.5.6 (v1.5.6).* CaltechDATA. doi: https://doi.org/10.22002/D1.20059
4. Schmidt, R. (2019). *Recurrent Neural Networks (RNNs): A gentle Introduction and Overview.* doi: https://doi.org/10.48550/arXiv.1912.05911
5. Graves, A., Fernández, S., Gomez, F., Schmidhuber, J. (2006). Connectionist temporal classification. *Proceedings of the 23rd International Conference on Machine Learning – ICML'06.* doi: https://doi.org/10.1145/1143844.1143891

6. Yakovchuk, O., Cherneha, A., Zhelezniakov, D., Zaytsev, V. (2020). Methods for Lines and Matrices Segmentation in RNN-based Online Handwriting Mathematical Expression Recognition Systems. *2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP)*. doi: https://doi.org/10.1109/dsmp47368.2020.9204273

7. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. (2019) *Language Models are Unsupervised Multitask Learners.* Available at: https://www.semanticscholar.org/paper/Language-Models-are-Unsupervised-Multitask-Learners-Radford-Wu/9405cc0d6169988371b2755e573cc28650d14dfe

8. Child, R., Gray, S., Radford, A., Sutskever, I. (2019). *Generating Long Sequences with Sparse Transformers.* doi: https://doi.org/10.48550/arXiv.1904.10509

9. Vaswani, A., Shazeer, N., Parmar, N. (2017). *Attention Is All You Need.* doi: https://doi.org/10.48550/arXiv.1706.03762

10. Brown, B., Mann, B., Ryder, N., Subbiah, M. (2020). *Language Models are Few-Shot Learners.* Available at: https://arxiv.org/pdf/2005.14165.pdf

✉ *Oleg Yakovchuk, Assistant, Postgraduate Student, Department of System Design, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine, ORCID: https://orcid.org/0000-0002-9842-9790, e-mail: olegyakovchuk@gmail.com*

Maksym Vasin, Department of System Design, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine, ORCID: https://orcid.org/0009-0005-1903-9874

✉ *Corresponding author*

**Bohdan Hirianskyi,
Bogdan Bulakh**

A REVIEW OF PRACTICE OF USING EVOLUTIONARY ALGORITHMS FOR NEURAL NETWORK SYNTHESIS AND TRAINING

The object of this research is the application of evolutionary algorithms for the synthesis and training of neural networks. The paper aims to select and review the existing experience on using evolutionary algorithms as competitive methods to conventional approaches in neural network training and creation, and to evaluate such existing solutions for further development of this field.

The essence of the obtained results lies in the successful application of genetic algorithms in conjunction with neural networks to optimize parameters, architecture, and weight coefficients of the networks. The genetic algorithms allowed improving the performance and accuracy of neural networks, especially in cases where backpropagation algorithms faced difficulties in finding optimal solutions.

These results can be attributed to the fact that genetic algorithms are efficient methods for global optimization in parameter space. They help avoid local minima and discover more reliable and stable solutions. The obtained findings can be practically utilized to enhance the performance and quality of neural networks in various classification and prediction tasks. The use of genetic algorithms enables the selection of optimal weight coefficients, network connections, and identification of significant features from the dataset. However, they come with the limitation of additional time costs for evaluating the entire population according to the selection criteria.

It is worth noting that the application of genetic algorithms is not a universal method for all tasks, and the algorithm parameters should be individually tuned for each specific problem. Further research could focus on refining the combination methods of genetic algorithms and neural networks, as well as exploring their application in new domains and tasks.

Keywords: *neural networks, evolutionary algorithms, genetic algorithms, hybrid approach, optimization neural network architecture.*

1. Introduction

The increasing demand for high-precision systems, intelligent applications, and smart technologies in the modern world poses a challenge for researchers and engineers to explore innovative approaches for neural network (NNs) synthesis and training. NNs, as powerful tools for modeling and understanding complex tasks, find diverse applications in various fields, such as medicine, finance, industry, data analysis, and robotics. They have also gained significant