INTERPRETING DOCTORS NOTES FROM TEXT TO SPEECH USING HANDWRITING RECOGNITION

NANDINI PRASAD. K S[1], ABHINAV KUMAR[2], ABHISHEK ANAND[3], AYUSH KUMAR SINHA[4], BHAGYALAXMI N GAVAROJI[5]

[12345] *hodise@dsatm.edu.in, abhinavdsatm@gmail.com, abhishekanand76317@gmail.com,ngavaroji@gmail.com, inc.ayush29@gmail.com*

[12345] *7thsem, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bangalore-82, Karnataka*

[6] *Hod Department of Information Science and Engineering Dayananda Sagar Academy of Technology and Management, Bangalore-82, Karnataka*

ABSTRACT -The ability to express language, thoughts, and ideas through handwriting. It is a widely established fact that the majority of doctors have unreadable cursive handwriting. We show the Convolutional Neural Network (CNN)-based Handwriting Recognition System in action, which was developed to recognize the text in pictures of prescriptions written by doctors and to show how cursive handwriting may be transformed into legible text. The most successful method for resolving handwriting identification issues is to use convolutional neural networks (CNNs), and they are quite proficient at identifying the structure of handwritten letters and words in ways that make it easier to automatically extract distinguishing features. Handwriting recognition (HWR), is the ability of a computer to take and interpret comprehensible handwritten input from sources such as paper documents, images, touch displays, and other devices (HTR).

Keywords: Deep Learning, Machine Learning, Handwritten Digit Recognition (HTR), Convolution Neural Network (CNN), Text-to-speech, Voice Processing.

1. INTRODUCTION

Man's ability to convey words, thoughts, and ideas through handwriting. It is a well established fact that doctors frequently have unintelligible cursive handwriting. As a res
of clinical notes that are poorly written and have unclear substance, which has led to several instances of medical blunders, it has been demonstrated that even pharmacists who are responsible for dispensing the medications ordered for patients struggle to understand doctor's handwriting. Although researchers have worked in this area, they have not yet managed to reach 100% accuracy. The handwriting of many persons can be recognized by our eyes, but a machine cannot do this as readily. Optical Character Recognition (OCR) is one of the solutions to this issue. One method for transforming a scanned or printed picture document into an editable text document is optical character recognition (OCR). Our interest in mobile applications, which are expanding in the software sector, is rising nowadays. Today's world desperately needs
the technology known as HTR.

A significant amount of data has been lost throughout history as a result of the conventional technique of storing data. People may now save data on machines where it is simpler to organize, store, and retrieve it thanks to modern technology. Data that was previously kept can be accessed and stored more easily by using handwritten text recognition software. Additionally, it increases the data's security.

2. RELATED WORK

Deep learning, machine learning, and artificial intelligence have all gained from the enormous research and development effort that has been performed in response to the humanization of machines. Machines are getting more and more complex with time. From doing simple math operations to identifying the optic nerve, they have increased the safety and use of our real activities. The procedure of converting handwritten text into machine text is called handwriting recognition. This can be useful in a variety of contexts, including interpreting doctors' notes. There are several approaches to using deep learning for handwriting recognition. Handwritten characters could well be identified using a neural network. In this approach, the RNN is trained on a large dataset of handwritten text and learns to recognize patterns in the text that correspond to different words as a result, accuracy rate. Another study by Kim et al. used a combination of Information extraction using handwriting recognition and processing of natural language techniques .In addition, a study by Cho et al. implemented a variety of handwriting recognition and deep learning techniques to classify and extract information from handwritten medical notes

3.METHODOLOGY

Comparing Algorithms which is being used like
Optical Character Recognition Algorithms,
Natural Language Processing Algorithms,
Handwriting Recognition Algorithms, Speech Synthesis Algorithms is grounded on the map for individual algo dataset, number of epochs, algorithm complexity, exactness of individual algo, gadget specs for executing programs and runtime Algorithm under ideal conditions.

4.COMMON ALGORITHM WILL BE USED

4.1 OPTICAL CHARACTER RECOGNITION (OCR) ALGORITHMS

Tesseract OCR: Tesseract is a widely used opensource OCR engine developed by Google. It is known for its high accuracy in recognizing printed text . The engine is designed to recognize text in over 100 languages, including various scripts such as Latin, Cyrillic, Arabic, and Chinese. Tesseract uses advanced algorithms and machine learning techniques to analyze and recognize text in images. It can detect text orientation and language automatically, making it highly adaptable to different document types and languages. Tesseract is available as a command-line tool and can be integrated with various programming languages such as Python, Java, and C++. It also has a graphical user interface (GUI) called TesseractOCR which simplifies the process of OCR for nonprogrammers.

ABBYY FineReader: ABBYY FineReader is an OCR software developed by ABBYY, a software company that specializes in document conversion and language-based technologies. FineReader is designed to extract text from digital documents,

including scanned paper documents, PDF files, and images.FineReader is known for its high accuracy in recognizing text and retaining the formatting and layout of the original document. It uses a OCR algorithms to recognize various fonts, including handwritten text, and can extract text in over 200 languages.In addition to text recognition, FineReader includes features for document conversion, editing, and collaboration. The software can convert scanned documents into editable formats such as Microsoft Word, Excel, and searchable PDFs. It also includes tools for correcting and editing recognized text, as well as collaborative features for sharing and commenting on documents.

4.2 NATURAL LANGUAGE PROCESSING (NLP) ALGORITHMS

Stanford CoreNLP : Stanford CoreNLP is a natural language processing toolkit developed by the Stanford Natural Language Processing Group. It provides a set of tools for analyzing and processing natural language text in multiple languages. The toolkit includes various modules for tasks such as part-of-speech tagging,

named entity recognition, sentiment analysis, coreference resolution, dependency parsing, and more. It also includes a pipeline that allows users to apply these tools to text in a specific order and extract various linguistic features from the input. CoreNLP is built on top of the Java programming language and provides APIs for integrating with other programming languages such as Python and Ruby. It also includes a graphical user interface (GUI) for visualizing the output of the various processing tools.The toolkit is widely used in academic research and industry applications, including sentiment analysis, machine translation, and chatbot development. It is opensource and available under the GNU General Public License.

Stanford CoreNLP is one of the most widely used NLP toolkits due to its accuracy and versatility in various languages.it extracts text from images and documents without a text layer and outputs the document into a new searchable text file,PDF, or most other popular formats.

spaCy: spaCy is an open-source natural language processing (NLP) library for Python. It is designed to be efficient, fast, and extensible, and provides a set of tools for various NLP tasks such as part-ofspeech tagging, named entity recognition, dependency parsing, text classification, and more.One of the key features of spaCy is its efficiency in processing large volumes of text. It is optimized for speed and can process thousands of documents per second. It also includes pretrained models for various languages, making it easy to get started with NLP tasks without the need for extensive training data.Another key feature of spaCy is its ease of use and extensibility. It provides a simple and intuitive API that allows users to quickly build custom pipelines for their specific NLP tasks.

4.3 HANDWRITING RECOGNITION ALGORITHMS

Convolutional Neural Networks (CNNs):

Convolutional Neural Networks (CNNs) are a type of artificial neural network that are designed for processing and analyzing images and other multidimensional data. They are widely used in computer vision tasks such as image classification, object detection, and segmentation.CNNs are composed of multiple layers, each of which performs a specific operation on the input data. The first layer of a CNN is typically a convolutional layer, which applies a set of

filters to the input image to detect local features such as edges and corners.This is followed by a pooling layer, which reduces, the spatial size of the feature maps by down-sampling the image. The process of convolution and pooling is repeated in multiple layers, with each layer learning increasingly complex features that capture higherlevel information about the image. The final layers of a CNN are typically fully connected layers.

Long Short-Term Memory (LSTM) : it is a type of recurrent neural network (RNN) architecture that is designed to handle the vanishing gradient problem that occurs in traditional RNNs. The vanishing gradient problem occurs when the gradient of the loss function becomes very small during backpropagation, When it comes to deep learning networks, long-term dependencies can pose a challenge for effective learning. However, one type of neural network called LSTM addresses this issue by utilizing a unique memory cell that enables the retention of information over extended periods of time. Additionally, this memory cell allows for the selective modification and discarding of data as needed. The memory cell consists of three gates, namely, the input gate, the forget gate, and the output gate. Each gate is controlled by a sigmoid activation function, which determines how much information is passed through.The input gate determines how much new information should be added to the cell state, while the forget gate determines how much information should be discarded from the cell state. The output gate controls how much of the cell state is used to produce the output.

4.4 SPEECH SYNTHESIS ALGORITHMS

Google Text-to-Speech: it is a technology developed by Google that allows users to convert written text into spoken words. It uses advanced machine learning algorithms to create natural

sounding voice synthesizers that can read text aloud in a variety of languages and accents.The Text-to-Speech technology is integrated into various Google products and services, such as Google Assistant, Google Translate, and Google Play Books. Users can also access Text-to-Speech as a standalone feature on their Android device, allowing them to have text messages, emails, and other written content read aloud.

google text to speech it has a wide range of applications, from helping people with visual impairments to improving the accessibility of digital content. It can also be used to create voiceovers for videos or podcasts, or to add a more human touch to automated customer service interactions. Overall, Google Text-to-Speech is a powerful tool that has the potential to make digital content more accessible and engaging for a broader audience.

Amazon Polly: Amazon Polly is a cloud-based text-to-speech service developed by Amazon Web Services (AWS). It allows users to convert text into lifelike speech using advanced deep learning technologies. With Amazon Polly, users can choose from a variety of natural-sounding voices, languages, and accents to create speech-enabled applications, including voice-overs for videos and podcasts, chatbots, and audiobooks. Amazon Polly
also allows users to control the speed and intonation of the speech to create a more personalized experience. One of the notable features of Amazon Polly is its ability to generate speech in real-time, as in call centers and customer support services. It also supports a wide range of formats, including MP3, OGG, and PCM, which allows users to easily integrate the audio output into their applications.

5. COMPARATIVE STUDY OF VARIOUS ALGORITHM WHICH WILL BE USED

Sr. No	Algorithms	Pre-Processing	Result/Accuracy
1	Supervised layer wise training of a deep convolution neural network (SL-DCNN)	The utilized dataset was subjected to image resizing, wherein the images were adjusted to have a uniform dimension of 30 pixels.	The TCNN model demonstrates a lower error rate of 9.67% in contrast to the DCNN's error rate of 15.96%.
2	Artificial Neural Network (ANN)	The dimensions of each character are uniformly adjusted to 30 pixels in width and 20 pixels in height.	They were able to achieve a character recognition accuracy rate of 90.19%.
3	Convolutional Neural Network (CNN) architecture.	Resized all images to 28 X 28 pixels.	They have achieved an accuracy of 99.40% for character recognition.
4	Convolution Neural Network (CNN)	A convolution layer takes as input an image with dimensions M x M x C, where M represents the height and width of the image, and C represents the number of channels per pixel.	They achieved an accuracy of 94.9%.
5	Deep Neural Network (DNNs)	Shui character image normalized to 52 * 52 pixels. The class label is added to the clustering results.	Accuracy is around 93.3% achieved.
6	Deep Convolution neural network (DCNN)	To convert an image of size h x w x c to a size of m x m x c, a normalization procedure must be implemented. This procedure ensures that the image is resized to the desired dimensions, where 'm' represents both the height and width of the new image.	Using a DCNN layer-wise training model, they obtained 98% recognition accuracy.

Table- I: Comparisons of algorithm uses for character recognition.

6. CONCLUSION

Here,we will analyze and design three models in this research and test using the Pre – trained model dataset, identify handwritten characters. The models were based on algorithms for deep and learning techniques, including support vector machines. The models were compared based on their characteristics, including accuracy and execution time, to determine which was the most accurate. It was found that the CNN model had the highest accuracy for handwritten digit recognition, and that extending the total number of time periods without changing the algorithm's parameters did not improve performance due to the limitations of the model. The model also tended to overfit the dataset a certain number of periods later, leading to biased predictions. As a result, it was concluded that the CNN model was the most suitable for predicting image data and other types of prediction problems.

7.FUTURE ENHANCEMENT

There are virtually no limits to the potential progress of applications built on deep and machine learning algorithms. In future Some potential areas of focus to [1] Improved accuracy Currently, these systems are able to accurately interpret a large proportion, but not all, of handwritten notes. Future work could focus on improving the accuracy of these systems, potentially by using more sophisticated deep learning models or by incorporating additional sources of data (e.g., electronic medical records), [2] Incorporating natural language processing (NLP) techniques could allow these systems to better understand the context and meaning of the handwritten notes. This could enable them to extract more nuanced information and make more accurate interpretations and [3] Personalization Currently, these systems will be trained on a large dataset of handwritten notes.

REFERENCES

[1] Abhishek Bala and Rajib Saha, An Improved Method for Handwritten Document Analysis using Segmentation, Baseline Recognition and Writing Pressure Detection ,6th International Conference On Advances In Computing Communications, ICACC 2016, 6-8 September 2016, Cochin, India, Elsevier-2016**.**

[2] Kanchan Keisham and Sunanda Dixit, Recognition of Handwritten English Text Using Energy Minimisation, Information
Systems Design and Intelligent Applications, Advances in Intelligent Systems and
Computing, Bangalore, India, Springer-2016.

[3] Nibaran Das,Sandip Pramanik,Subhadip Basu,Punam Kumar Saha, Recognition of handwritten Bangla basic characters and digits using convex hull based feature set, 2009 International conference on Artificial intelligence and pattern recognition(AIPR09).

[4] Subhadip Basu, Nibaran Das, Ram Sarkar, Mahantapas Kundu, Mita Nasipuri, Dipak Kumar Basu, A hierarchical approach to recognition of handwritten Bangla characters, Elsevier -2009.

[5] Namrata Dave, Segmentation Methods for Hand Written Character Recognition, International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 8, No. 4 (2015), pp. 155-164.

[6] Nafiz Arica, Student Member, IEEE, and Fatos T. Yarman-Vural, Senior Member, IEEE, Optical Character Recognition for Cursive Handwriting, IEEE transactions on pattern analysis and machine intelligence, vol. 24, no. 6, june 2002 .

[7] G. Louloudisa, B.Gatosb, I.Pratikakisb, C.Halatsisa, Text line and word segmentation of handwritten documents a Department of Informatics and Telecommunications, University of Athens, Greece Computational Intelligence Laboratory, Institute of Informatics and Telecommunications, National Center for Scientific Research Demokritos, 15310Athens,Greece .

[8] AyushiTrivedi, NavyaPant, Pinal Shah,

SimranSonik and Supriya Agrawal,"Speech to text and text to speech recognition systems" , IOSR Journal of Computer Engineering (IOSRJCE) eISSN: 2278-0661, p-ISSN: 2278-8727, Vol-20, Issue-2, Ver-I, March.April-2018.

[9] Hussain Rangoonwala, Vishal Kaushik, P Mohith and DhanalakshmiSamiappan "Text to speech module International Journal of Pure and Applied Mathematics" Vol-115 No. 6 2017.