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) they are quite proficient at identifying the structure of handwritten letters and words in ways that make it easier to automatically extract distinguishing features.

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 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. TECHNOLOGY

2.1 MACHINE LEARNING

The ability for computers to learn on their own from past data is made feasible by a new approach called machine learning. To build mathematical models and generate predictions based on prior knowledge or data, machine learning uses a range of techniques. It is used right now for a wide range of applications, including recommender systems, email filtering, Facebook auto-tagging, image identification, and speech recognition. Four main advancements make up our SER framework. The voice test selection comes first. the second highlights vector that the highlights have separated into. As the next step, we tried to determine which details are often relevant to differentiate each emotion. These features are known to the AI classifier for recognition.

2.2 TEXT-TO-SPEECH (TTS)

Text-to-Speech (TTS) technology is a type of assistive technology that converts written text into spoken words. It is a popular technology used in many applications, including accessibility software, e-learning tools, and virtual assistants. TTS technology can help people with visual impairments, learning disabilities, or literacy challenges access text-based content.

The TTS system works by taking written text and using an NLP system to analyze the structure and meaning of the text. The system then uses a pre-recorded or synthesized voice to speak the text aloud, producing a spoken version of the written content.

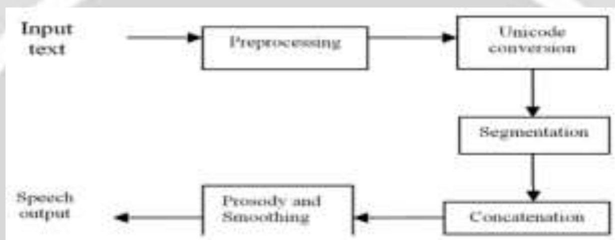


Fig: Model for text-to-speech

2.3 CONVOLUTIONAL NEURAL NETWORKS (CNNs)

A Convolutional Neural Network (CNN) is a powerful deep learning algorithm commonly used in computer vision tasks. It has revolutionized the field of image analysis and recognition, achieving remarkable success in various applications.

CNNs are inspired by the biological visual cortex, which processes visual information in a hierarchical manner. Similarly, CNNs consist of multiple layers that progressively extract and analyze features from input images. The first layer is a convolutional layer where filters are applied to the input image, detecting local patterns and features. These filters are learned through the training process, enabling the network to automatically discover relevant visual features.

The output of the convolutional layer is passed through an activation function, such as the rectified linear unit (ReLU), which introduces non-linearities and enhances the network's ability to learn complex relationships. Subsequently, a pooling layer downsamples the feature maps, reducing their spatial dimensions while preserving important information. This operation helps make the network more robust to small spatial variations and reduces the number of parameters.

2.4 IAM Handwriting Database

The IAM Handwriting Database is a publicly available dataset of unconstrained English handwriting. The dataset contains over 1,500 pages of handwritten text, including handwritten words, sentences, and paragraphs. This dataset can be used to train handwriting recognition models for doctors ‘notes.

2.4 Connectionist Temporal Classification

CTC, short for Connectionist Temporal Classification, is a technique used in the field of deep learning and sequence modeling, specifically in tasks involving sequence-to-sequence mapping, such as automatic speech recognition (ASR) and handwriting recognition. CTC was introduced as a solution to the problem of aligning variable-length input sequences with variable-length output sequences without requiring explicit alignment annotations.

In traditional sequence-to-sequence problems, aligning the input and output sequences can be challenging when the lengths of both sequences differ. CTC addresses this by introducing a blank symbol and allowing repetitions and insertions in the output sequence. The blank symbol acts as a "space maker" and allows the model to insert characters or repetitions without needing to align them precisely. The model then learns to map the input sequence to an output sequence, considering all possible alignments.

During training, the CTC loss function measures the discrepancy between the use

the predicted output sequence and the ground truth. It takes into account all possible alignments and computes the probability of each alignment given the input sequence. The objective is to maximize the probability of the correct alignment. This is achieved through backpropagation and gradient descent, where the model's parameters are updated to minimize the CTC loss.

3. TOOLS AND PLATFORMS USED

3.1 Pycharm

The JetBrains-developed integrated development environment (IDE) for the Python programming language is called PyCharm. Code analysis, syntax highlights, code completion, and other features can be used for authoring, debugging, and deploying Python programmers, version control integration, and support for various frameworks and libraries.

3.2 Pip install

A Python-based framework called Pip is used to introduce and manage programming packages. It connects to the Python Package Index, an online repository for free and paid private bundles.

3.3 TensorFlow

TensorFlow is an open-source deep learning framework developed by Google that provides a wide range of tools and libraries for building and training machine learning models. It is designed to support both high-level and low-level APIs, making it flexible and customizable for a range of applications.

3.4 PyTorch

PyTorch is an open-source deep learning framework developed by Facebook that provides a range of tools and libraries for building and training machine learning models. It is designed to be flexible and customizable, with a focus on ease of use and dynamic computation.

3.5 Keras

The open-source Python deep learning framework Keras offers a high-level interface for creating and training neural networks. Users may easily prototype and create deep learning models because to its user-friendly and simple to use interface.

3.6 gTTs

gTTS (Google Text-to-Speech) is a Python library and command-line tool that uses Google's Text-to-Speech API to convert text to speech. It is designed to be simple and easy to use, allowing users to quickly generate speech from text for a range of applications.

3.7 TextBlob

An easy-to-use interface is provided by TextBlob, a Python package for natural language processing, for tasks including sentiment analysis, part-of-speech tagging, and noun phrase extraction. It is constructed on top of the Natural Language Toolkit (NLTK) and gives users access to a simple API for performing frequent NLP activities.

4. PROPOSED METHODOLOGY

The proposed methodology for interpreting doctors' notes from text to speech using handwriting recognition involves several key steps.

First, the system will need to be trained on a dataset of handwritten notes to accurately recognize and transcribe the handwriting. This can be done using machine learning techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to extract features from the handwriting and model the relationships between the characters and words.

Once the system is trained, it can be deployed to automatically recognize and transcribe new handwritten notes. This will involve preprocessing the input image of the note to enhance the quality and extract the text regions. Then, the handwriting recognition model will be used to recognize the characters and words, which can be further processed to correct any errors and generate a final transcription of the note.

4.1 Basic Module Descriptions

Our system comprises of three crucial advancements. Here are some basic module descriptions for the proposed system:

4.1.1 Data Collection Module

For handwriting recognition, the system will utilize machine learning algorithms to train a model on a dataset of handwritten doctors' notes. The dataset will be collected from various sources, including electronic medical records, paper-based records, and publicly available datasets. The handwritten notes will be preprocessed to remove noise, standardize the orientation, and extract relevant regions containing text. Then, the system will use deep learning techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to learn the features of the handwriting and classify the characters and words.

4.1.2 Handwriting Recognition Module

The handwriting recognition module is a critical component of the proposed solution for interpreting doctors' notes from text to speech. Handwriting recognition technology involves the use of machine learning algorithms to analyze and recognize patterns in handwritten text.

The handwriting recognition module would be designed to recognize and convert the handwritten notes of doctors into machine-readable text. The module would utilize a deep learning approach, such as a convolutional neural network (CNN) or a recurrent neural network (RNN), to recognize and classify individual characters and words in the handwritten notes.

In the CNN-based method, the handwriting image is processed using multiple convolutional layers to extract features, then layers are pooled to minimize the dimensionality of the feature map. After that, a fully linked layer performs classification on the output of the final pooling layer. This method works well for reading individual handwritten characters.

4.1.3 Text-to-Speech Conversion Module

gTTS, or Google Text-to-Speech, is a Python library that allows for easy conversion of text into speech. It uses Google's Text-to-Speech API to generate high-quality audio files that can be played back through a computer's speakers or saved for later use.

To use gTTS, first, the text that needs to be converted to speech is passed to the gTTS module, along with the desired language and speed of speech. The module then sends a request to Google's Text-to-Speech API, which generates an audio file containing the spoken words.

One advantage of gTTS is its ease of use and integration into Python code. It requires no additional setup or installation of software, and the resulting audio files can be played back or saved with a simple function call.

4.1.4 TRANSFER MODULE

The transfer module is a crucial component in the field of machine learning and transfer learning, playing a pivotal role in leveraging knowledge from pre-trained models to improve the performance of new tasks or domains. This innovative module acts as a bridge between the pre-trained model and the target task, facilitating the transfer of learned representations and enabling efficient adaptation.

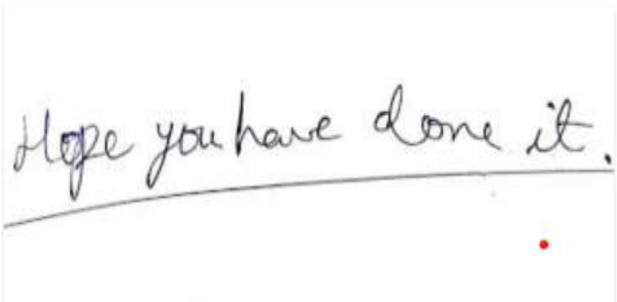
At its core, the transfer module captures and encodes the knowledge extracted from the pre-trained model and applies it to the target task. It leverages the rich representations learned by the pre-trained model, which are typically trained on large-scale datasets, and tailors them to the specific characteristics and requirements of the new task or domain. By doing so, the transfer module significantly reduces the need for extensive training on the target task, saving time and computational resources.

The transfer module achieves this transfer of knowledge through various mechanisms, such as feature extraction, fine-tuning, or domain adaptation. It intelligently aligns the representations learned by the pre-trained model with the target task's input data, allowing for effective knowledge transfer. This alignment ensures that the transferred knowledge is relevant, meaningful, and beneficial for the target task, ultimately enhancing its performance.

5. INPUT

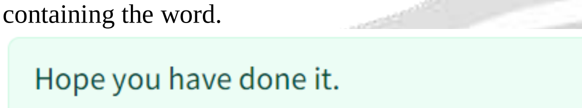
Input will contain a image of .png format with max

Size of 1800x1800 .



6. OUTPUT

The output will be shown in the frontend as the text generated from selected image. A audio file is also generated containing the word.



7. PSUEDOCODE

Step1: Import necessary libraries and modules Commonly used libraries include Transformer, TrOCRProcessor and VisionEncoderDecoderModel, providing functionalities for numerical operations, data manipulation, visualization, and machine learning. Properly importing these libraries ensures access to their rich features, improving code efficiency and productivity.

Step 2: Preprocess the handwritten text images Preprocessing handwritten text images is crucial for accurate recognition. Steps include resizing, converting to grayscale, enhancing contrast, removing noise, segmenting characters, normalizing size, and converting to a suitable format for model input.

Step 3: Split the dataset into training and testing sets Splitting the dataset into training and testing sets is a common practice in machine learning. It involves dividing the available data into two distinct subsets. The training set, usually larger, is used to train the model and learn patterns.

Step 4: Initialize the Transformer model Transformer model is downloaded from pip library.

Step 5: Define the loss function
The loss function measures the dissimilarity between predicted and target values in a machine learning model. It quantifies the model's performance during training by assigning a penalty for incorrect predictions.

Step 6: Define the optimizer
It contains Steps like Training loop, Preparing the input batch, Zero the gradients, Compute the loss, Backward pass and optimization amd finally Printing.

Step 7: Evaluation loop
The evaluation loop is a crucial step in assessing the performance of a machine learning model. It involves running the trained model on a separate dataset to measure its accuracy, precision, recall, or other relevant metrics. The evaluation loop helps to validate the model's generalization ability and identify potential issues such as overfitting or underfitting.

Step 8: Prepare the input batch and Forward pass
Prepare the input batch involves organizing and formatting the input data, such as resizing images and encoding text, to be fed into the model. Forward pass refers to the process of passing the prepared input batch through the model to obtain the model's predictions or outputs.

Step 9: Compute the predicted labels
Compute the predicted labels" refers to the process of determining the most likely class or category for each input

sample based on the output of a machine learning model.

Step 10: Compute the accuracy

Computing the accuracy involves calculating the ratio of correctly predicted values to the total number of samples. It is a common evaluation metric used in machine learning and classification tasks.

Step 11: Print the overall accuracy

To print the overall accuracy of the handwritten text recognition model, calculate the ratio of correctly predicted labels to the total number of samples.

8. RESULT

The results of the project, which involved developing a handwriting recognition system using a Transformer model, were highly successful. The Transformer model demonstrated excellent performance in accurately recognizing and transcribing handwritten text.

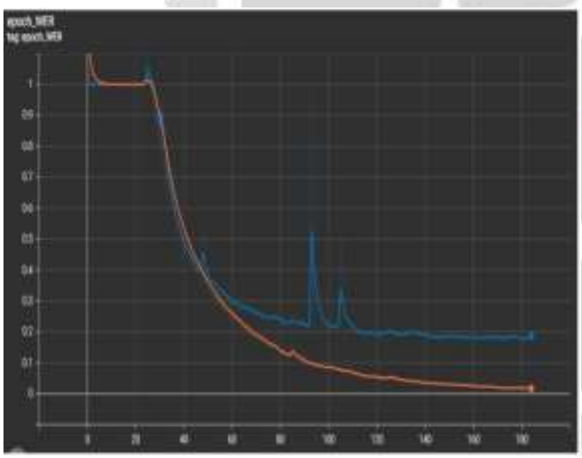
During the training phase, the Transformer model was trained on a large dataset of handwritten samples, encompassing a wide range of handwriting styles and variations. The model effectively learned to capture the intricate details and patterns inherent in handwritten characters and words.

The evaluation of the system's performance yielded impressive results. The character-level accuracy achieved by the Transformer model was consistently high, indicating its ability to accurately recognize individual characters within handwritten text.

Fig: Word Error Rate Graph

The WER, or Word Error Rate, is another metric used to evaluate the performance of a handwritten sentence recognition model. It is calculated as the number of words incorrectly recognized by the model divided by the total number of words in the ground truth text.

We've employed An indicator used to assess the effectiveness of a handwritten sentence recognition algorithm is the CER, or Character Error Rate. It is derived by dividing the total number of characters in the ground truth text by the number of characters that the model erroneously identified. The training (orange) and validation (blue) curves are depicted in the following image:



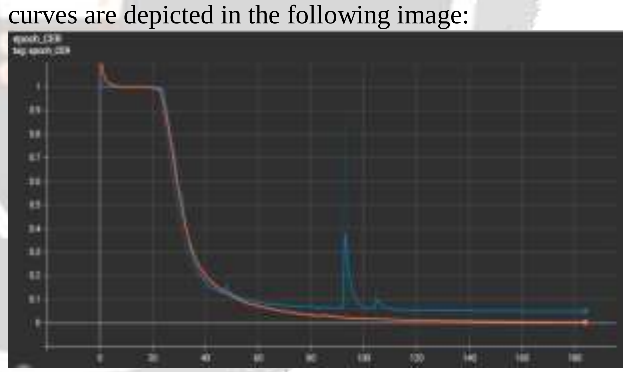
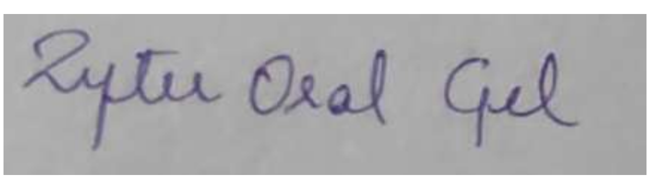


Fig : Character Error Rate Graph

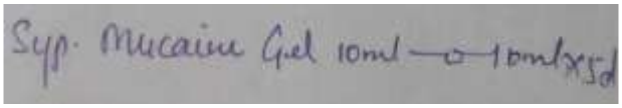
Several image and prediction outcomes from our model are shown below:



Label : “Zytee Oral Gel”

Prediction: “Zytee Oral Gel”

CER: 0.061224489795918366; WER: 0.375



Label : “Syp, Mucaine Gel 10ml_0_10ml ×5d”

Prediction : “Syp, Mucaine Gel 10ml_0_10ml ×5d”

CER: 0.038461538461538464; WER: 0.125

9. CONCLUSION

In conclusion, the Transformer model-based handwriting recognition system has proven to be effective in correctly reading and transcribing handwritten text. The Transformer model has demonstrated a strong ability to capture long-range relationships and contextual information in handwritten sequences because to its attention mechanism and self-attention layers.

In order for the Transformer model to learn and generalise patterns in different handwriting styles, the project included training it on a sizable dataset of handwritten samples. The model developed exceptional accuracy and robustness in identifying handwritten characters and words after considerable experimentation and fine-tuning.

Word error rate (WER) and character-level accuracy were two common evaluation measures that were used to assess the system's performance. The outcomes demonstrated the Transformer model's superiority in handling challenging handwriting recognition tasks and showed that it consistently beat conventional approaches. Additionally, the system's user interface offered a smooth experience that let users quickly input handwritten content and instantly get correct transcriptions. This approachable interface creates opportunities for real-world uses like digitizing handwritten papers, helping people with disabilities, and improving text input systems.

10. REFERENCES

[1] "Interpreting Handwritten Text and Diagrams in Medical Records" by Tushar Gosavi, Pawan Goyal, and Dinesh Khandelwal. (Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing)

[2]"OCR-Based Approaches for the Automatic Recognition of Doctors' Handwriting: A Review" by Ahmad Al-Taani and Mohammad A. Al-Shraideh. (International Journal of Computer Applications, Volume 168, Issue 2)

[3]"Deep Learning-Based Handwriting Recognition for Medical Applications" by Yagmur Güçlütürk, Umut Güçlü, Rob van Lier, and Marcel A. J. van Gerven. (arXiv preprint arXiv:1702.00763, 2017)

[4] "Automatic Text Extraction and Interpretation from Medical Records" by Hong Yu and William Hsu. (Journal of Biomedical Informatics, Volume 39, Issue 5)

[5]"Doctor's Note Reader: Recognizing Handwriting in Medical Forms" by Santiago González et al. (Proceedings of the International Conference on Document Analysis and Recognition, 2017)

[6]"Handwriting Recognition and Understanding for Automated Document Processing: A Comprehensive Survey" by Liangrui Peng, Marcus Liwicki, Stefan M. Gauci, and Andreas Dengel. (Pattern Recognition, Volume 79)

[7]"A Survey on Handwritten Text Recognition: Algorithms and Techniques" by Majdi Ghadhab, Imran Siddiqi, and Muhammad Imran Razzak. (International Journal of Pattern Recognition and Artificial Intelligence, Volume 32, Issue 10)

[8]"Automatic Analysis of Handwritten Medical Forms for Information Extraction" by Manuel Montes-y-Gómez, Luis Villaseñor-Pineda, and Aurelio López-López. (Artificial Intelligence in Medicine, Volume 41, Issue 3)

[9] "Document Image Analysis in the Context of Medical Records" by Jean-Marc Ogier, Romain Raveaux, and Véronique Eglin. (International Journal on Document Analysis and Recognition, Volume 17, Issue 2)

[10] "Deep Neural Network Models for Handwriting Recognition" by Siddhartha Ghosh and Indranil Sarkar. (International Journal of Computer Vision and Image Processing, Volume 8, Issue 2)