JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

PARAGRAPH READER USING CNN

[1]Dr. Saroja T.V., [2]Prathamesh Mayekar, [3]Aditi Barhate, [4]Diksha Telgote
[1]Associate Professor, Shivajirao S. Jondhale College of Engineering, Thane, India
[2-4]Student, Department of Computer Engineering, Shivajirao S. Jondhale College of Engineering, Thane, India,

Abstract: Convolutional Neural Networks (CNNs) and deep learning have found a wide range of applications across various fields due to their ability to effectively learn from large amounts of data and extract meaningful features. In this Project we have used the power of CNN and the Google Vision Api to create a comprehensive model which can understand the text, numbers, and special characters written in form of handwritten paragraph and convert this handwritten paragraph into digital text. This is accomplished by first uploading the image of the handwritten paragraph and then giving it as an input to the trained CNN model and the finally getting the digital text as an output, which can be copied from the clipboard present on the user interface. This project also focuses on the data security part where user's sensitive data in stored in encrypted form in the database. Also, we have created an end-to-end pipeline using which all users can track their upload history and also give a feedback score to the output given by the model.

Keywords: CNN, handwritten text, data security**.**

INTRODUCTION

Handwritten Text Recognition (HTR) using Convolutional Neural Networks (CNNs) is a transformative technology that enables computers to interpret and transcribe handwritten text into machine-readable form. At its core, HTR with CNNs involves the extraction of meaningful features from handwritten text images, followed by the interpretation of these features to recognize characters and words. The process begins with preprocessing the input images to enhance their quality and make them suitable for recognition. This may include operations such as resizing, normalization, and noise reduction to improve the clarity of the text.

Once pre-processed, the handwritten text images are fed into a CNN architecture, which is specifically designed to capture the intricate patterns and shapes present in handwriting. The convolutional layers of the CNN apply filters to the input images, extracting features that represent various characteristics of handwritten text, such as strokes, loops, and curves. These features are then processed by subsequent layers to further refine their representations. In some cases, recurrent neural networks (RNNs) or long short-term memory (LSTM) networks are integrated with CNNs to model the sequential nature of handwritten text and improve recognition accuracy.

During the training phase, the CNN-RNN model learns to map the extracted features to their corresponding text labels through a process of optimization, adjusting its parameters to minimize the difference between predicted and ground truth labels. This training is typically performed on a large dataset of labelled handwritten text images. Once trained, the model can be deployed to recognize handwritten text in new images, producing a sequence of characters or words that represent the transcribed text. Post-processing techniques such as language modelling and spell checking may be applied to refine the output and improve its accuracy further. HTR with CNNs has a wide range of applications, including document digitization, handwritten notes transcription, and automatic form processing, offering significant benefits in terms of efficiency and accuracy in various domains.

1.1 Needs

Document Digitization: Many historical documents, manuscripts, and archival materials exist only in handwritten form, making it essential to digitize them for preservation, accessibility, and dissemination. CNN-based Handwritten Text Recognition (HTR) systems enable the conversion of handwritten text into machine-readable digital text, facilitating the preservation and sharing of valuable cultural and historical artifacts.

Data Entry Automation: In industries such as finance, healthcare, and logistics, handwritten forms, invoices, and records are still prevalent. Manually transcribing handwritten text into digital format is time-consuming, error-prone, and resource-intensive. By leveraging CNN-based HTR systems, organizations can automate the process of data entry,significantly reducing labour costs and improving operational efficiency.

• **Accessibility**: Handwritten documents pose challenges for individuals with visual impairments or disabilities, as traditional optical character recognition (OCR) methods may struggle to accurately transcribe handwritten text. CNN- based HTR systems offer improved accuracy and reliability in converting handwritten text to digital format, enhancing accessibility for individuals who rely on assistive technologies such as screen readers

• Natural Language Processing (NLP): Handwritten text often contains valuable information for analysis and processing, but traditional NLP techniques primarily focus on digital text data. By converting handwritten text into digital format using CNN-based HTR systems, researchers and practitioners can apply advanced NLP algorithms for tasks such as sentiment analysis, information extraction, and language modelling, unlocking insights from handwritten documents.

• Personalization and Customization: Handwritten notes, letters, and messages carry unique stylistic elements and nuances that convey personal expression and sentiment. CNN-based HTR systems enable the conversion of handwritten text into digital format while preserving the individuality and authenticity of the original handwriting. This allows for personalized digital communication and customization of content in various applications, such as handwritten letters, greeting cards, and personalized gifts.

1.2 Applications

• Automatic Check Processing: Banks and financial institutions employ HTR to process handwritten checks and financial documents efficiently. HTR systems extract relevant information such as account numbers, amounts, and payee details, accelerating check processing workflows.

• Prescription and Medical Records Processing: In healthcare settings, HTR assists in digitizing handwritten prescriptions, medical records, and patient forms. This improves the accuracy of medical data entry, reduces errors, and enhances patient care coordination.

• Architectural Drawings and Engineering Diagrams: HTR is used to transcribe handwritten annotations, labels, and notes on architectural drawings, engineering diagrams, and blueprints. This facilitates collaboration, documentation, and revision tracking in design and construction projects

• Automatic Form Processing: In industries such as finance, insurance, and government, HTR automates the extraction of information from handwritten forms, surveys, and applications. This streamlines data entry processes, reduces errors, and improves efficiency in processing paperwork.

• Postal Services and Address Recognition: Postal services use HTR to automatically recognize handwritten addresses on mail and parcels, enabling efficient sorting and delivery. This technology helps streamline mail processing operations and improves accuracy in address recognition.

II. LITERATURE SURVEY

Viswanatha V et al. [1] aim to recognize the handwritten digits using CNN by training the model on the MNSIT dataset of images of handwritten digits. The accuracy of the model is 97% with CTC loss of 3%. This was the first paper we studied which helped was understand about how CNN architecture works.

P. Rajeshwari et al. [2] discusses about how text can be extracted from images in general. This paper help us gain understanding the existing tools and libraries used in text extraction from the images like OpenCV, Pytesseract and TensorFlow.

G.R. Hemanth et al. [3] focuses on integrating the CNN architecture with RNN architecture to make the process of Handwritten text recognition more intricate and improve the overall accuracy of the model. This paper provides us with great overview of how image segmentation is performed in use cases where we are dealing with handwritten paragraphs.

J. Sueiras et al. [6] proposes the use of a new neural network architecture that combines a deep convolutional neuralnetwork with an encoder decoder, called sequence to sequence.

III. METHODOLOGY

In our Handwritten Text Recognition System, we have leveraged the power of Convolutional Neural Networks which are widely used when it comes to image feature extraction and categorization. Since using purely only CNN didn't give us that much accuracy, we integrated our CNN model with the Google Vision Api. This integration not only helped us improve the accuracy of our model but also enabled our system with the capabilities like detecting the nature of the text, like whether it is offensive or harmful in nature. Furthermore, this also enabled us to recognize and categorize different handwriting styles from the same language. We have also focused a lot on improving the user experience by providing the end user with a fully-fledged production grade web application created in Flask. The image data which we are storing in our database is also taken care of very securely, as in such use application, security and integrity of the data becomesvery necessary.

The Convolutional Neural Network (CNN) algorithm is a type of deep learning algorithm commonly used for image recognition, classification, and other tasks involving structured grid data. The Convolutional Neural Network is generally a feed forward neural network, meaning, each layer in the CNN architecture receives some processed input from the previous layers. The steps involved in CNN when we are using it for Handwritten Paragraph recognition is as follows:

Input Layer:
The input layer receives handwritten paragraph images, which are typically grayscale or colour images containing multiple lines of handwritten text.

Convolutional Layer:
Convolutional filters analyse local patterns and structures within the handwritten paragraph images, detecting features like strokes, curves, and shapes of individual characters. Multiple filters are applied to capture different aspects of the handwriting.

Activation Layer:
Activation functions like ReLU introduce non-linearity, allowing the network to model complex relationships between features in the handwritten text.

Pooling Layer:
Pooling layers down sample the feature maps, reducing the spatial dimensions while retaining important information about the handwriting. Pooling helps the network focus on the most salient features while reducing computational complexity.

Flattening Layer:
The flattened representation prepares the extracted features from the handwriting for input into fully connected layers.

Fully Connected Layer:
Fully connected layers learn higher-level representations of handwriting, capturing relationships between different characters and words. These layers enable the network to understand the context of the handwritten paragraph.

Output Layer:
The output layer produces the final predictions, where each neuron corresponds to a class label (e.g., individual characters, words, or entire paragraphs). Activation functions like soft-max are used for multi-class classification, enabling the network to output the most likely sequence of characters or words.

Loss Function and Optimization:
The loss function measures the discrepancy between the predicted text and the ground truth labels Optimization algorithms such as SGD or Adam adjust the network's weights to minimize this loss during training.

Training:

The CNN is trained on a dataset of handwritten paragraphs with corresponding labels. Through backpropagation, the network learns to recognize patterns and structures in the handwriting, improving its accuracy over time. Training continues until the model achieves satisfactory performance on validation data, ensuring robust recognition of handwritten paragraphs.

IV. DATASET: THE IAM HANDWRITINGDATASET

Our project utilized a subset of the IAM Handwriting Dataset, sourced from Kaggle due to accessibility issues with the original dataset. This dataset comprises images of handwritten text along with corresponding transcriptions, offering a diverse array of handwriting styles for model training and evaluation. The initial phase involved exploratory data analysis (EDA) to understand the dataset's characteristics, such as the distribution of word lengths, the variability of handwriting styles, and common preprocessing needs like binarization and normalization.

Why Choose the IAM Handwriting Dataset?

1. Richness and Diversity: The IAM Handwriting Dataset is renowned for its diversity and richness in handwriting samples. It encompasses writings from hundreds of individuals, offering a wide range of handwriting styles, including cursive and block letters, variations in letter sizes, slants, and spacings. This diversity is crucial for developing a model that can generalize well across different handwriting styles, rather than overfitting to a narrow subset of handwriting characteristics.

2. Extensive Annotations: Each sample in the dataset comes with detailed annotations, including the transcription of handwritten text and the segmentation of text into lines and words. This level of annotation is invaluable for training and evaluating machine learning models, as it provides a direct correspondence between the image data and the textual information the model needs to learn to predict.

3. Standard Benchmark: The IAM Dataset is widely recognized and used as a benchmark in the handwriting recognition research community. By employing this dataset, the project aligns with an established standard, allowing for meaningful comparisons with other state-of-the-art handwriting recognition systems and contributing to the collective knowledge base in this domain.

4. Support for Different Recognition Tasks: The dataset's structure supports various handwriting recognition tasks, including recognizing entire lines of text and individual words. This flexibility allows for experimenting with different levels of granularity in recognition, which can be pivotal in addressing specific application requirements or research questions.

How We Used the Dataset

1. Data Preprocessing: Given the inherent variability in the handwritten text, preprocessing steps were crucial to standardize the images before feeding them into our CNN model. This included resizing images to a uniform dimension to ensure consistency, converting colour images to grayscale to reduce computational complexity, and normalizing pixel values to a range that is more suitable for neural network training.

2. Exploratory Data Analysis (EDA): Before diving into model training, we conducted an EDA to understand the characteristics of the dataset more deeply. This involved visualizing sample images, analysing the distribution of word lengths, and identifying common patterns and anomalies in the data. The insights gained from EDA helped in tailoring the preprocessing steps and model architecture to better suit the data's nuances.

3. Data Augmentation: To make our model more robust to variations in handwriting not covered by the dataset and to prevent overfitting, we applied data augmentation techniques. These included rotating, scaling, and shifting the images, artificially increasing the diversity of handwriting styles the model was exposed to during training.

4. Model Training and Evaluation: We split the dataset into training, validation, and test sets to train our CNN model and evaluate its performance. The training set was used to teach the model to recognize handwritten text, the validation set to tune the hyperparameters and prevent overfitting, and the test set to assess the model's generalization ability on unseen data.

5. Benchmarking and Iterative Improvement: Utilizing the IAM Dataset allowed us to benchmark our model's performance against other handwriting recognition systems and iterate on our approach. Through continuous testing and comparison, we identified areas for improvement, leading to refinements in our model architecture, training process, andpreprocessing pipeline.

V. EXPLORATORY DATA ANALYSIS (EDA)AND DISTRIBUTION ANALYSIS

Exploratory Data Analysis

The EDA phase was crucial for informing our preprocessing and modelling strategies. We analysed the dataset to identify patterns, anomalies, and typical characteristics of handwritten text. Techniques such as image visualization, histogram analysis for pixel intensity distributions, and examining the variance in text sizes across the dataset were employed. This analysis helped in designing preprocessing pipelines that included resizing images to a uniform scale,converting them to grayscale to reduce complexity, and applying image augmentation techniques to enhance the model's robustness to variations in handwriting.

Understanding the Dataset Structure

Before delving into specific analyses, the first step was to understand the structure and content of the IAM HandwritingDataset. This involved reviewing the dataset's documentation to learn about the format of the images, the annotations available (such as the segmentation of lines and words, and the transcriptions), and how the data is organized (e.g., by writer, form, or sentence).

Visual Inspection of Handwriting Samples

A crucial early step in EDA was the visual inspection of a subset of handwriting samples. This allowed us to observe the variability in handwriting styles across different writers, including variations in cursive versus block letters, the presence of slants in writing, variations in letter and word spacing, and differences in line alignment. By visually inspecting these samples, we could start to appreciate the challenges involved in recognizing such a diverse array of handwriting styles.

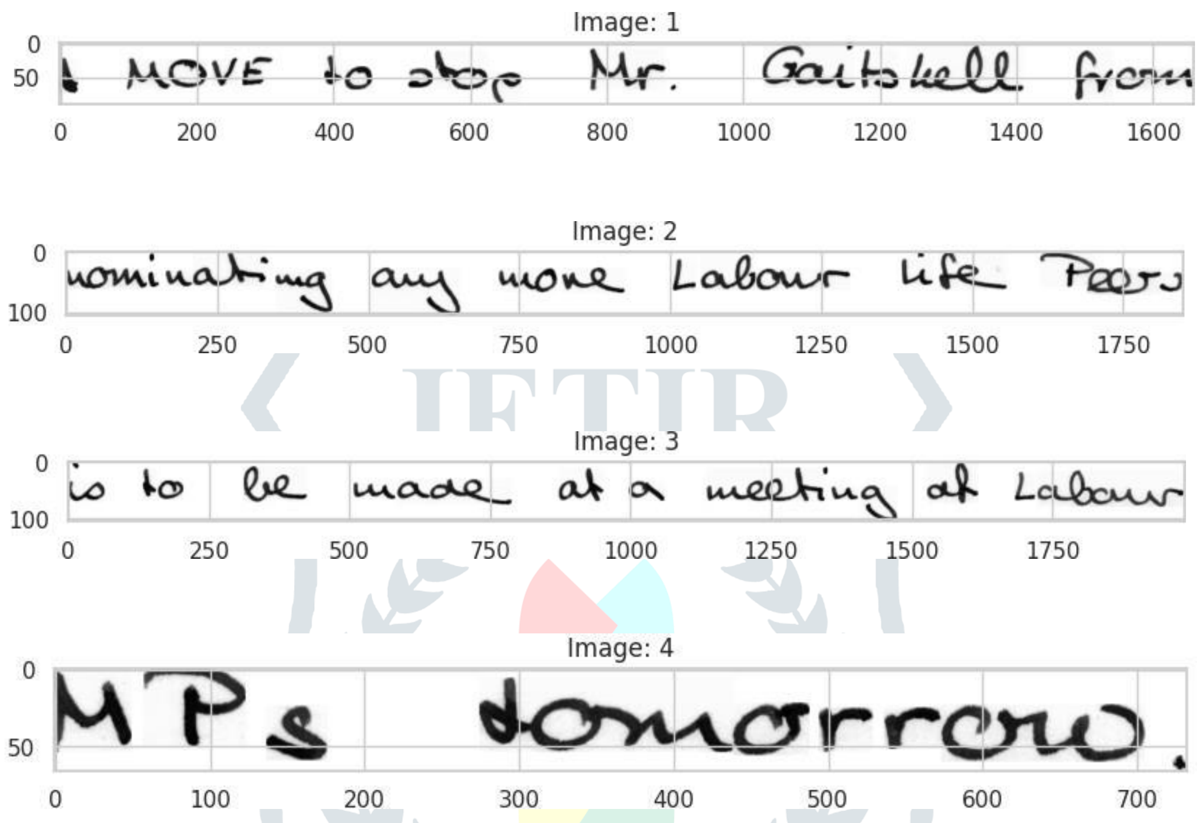


Image: 1
A MOVE to stop Mr. Gaitskell from
Image: 2
nominating any more Labour life Peers
Image: 3
is to be made at a meeting of Labour
Image: 4
MPs tomorrow.

Figure 5.1: Sample Image Visualization

Distribution Analysis

Pixel Intensity Distribution: Analysing the distribution of pixel intensities across the dataset helped us understand thecontrast levels within the handwritten notes. This analysis could inform preprocessing steps such as binarization (converting images to black and white) or contrast adjustment to make the handwriting more pronounced against the background.

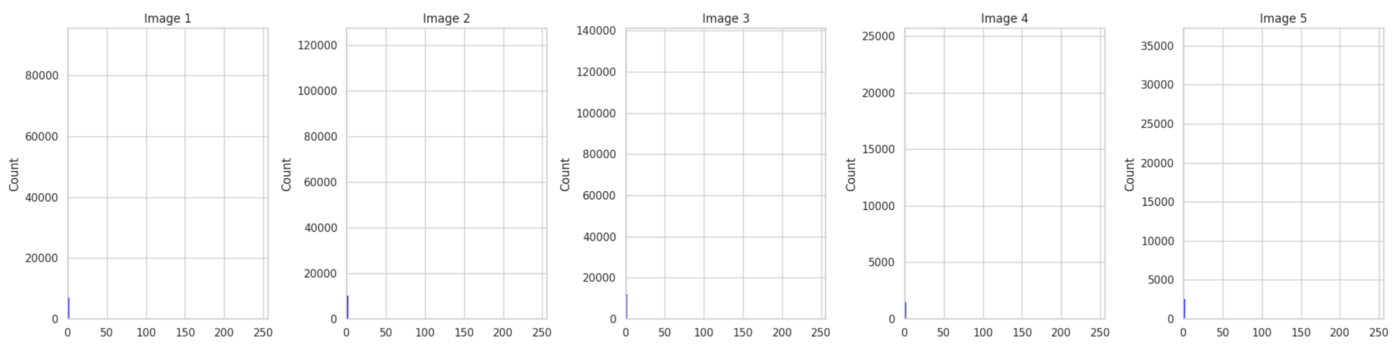


Figure 5.2: Pixel Intensity Distribution

Word Length Distribution: We analysed the distribution of word lengths (in terms of character count) within the dataset. This gave us insights into the complexity of the text samples we were dealing with and helped in designing the network architecture to accommodate the variability in input sizes.

Line Length Distribution: Similar to word length, analysing the length of the lines (in terms of both characters and pixels) provided valuable information for preprocessing steps, such as determining optimal image sizes for model inputand understanding the distribution of text densities across samples.

Preprocessing Requirements Identification

Through EDA, we identified several preprocessing requirements essential for standardizing the dataset for modeltraining:

- Resizing Images: To ensure that all input images to the model have a uniform size, we determined the most commondimensions for resizing based on our analysis of line and word lengths.

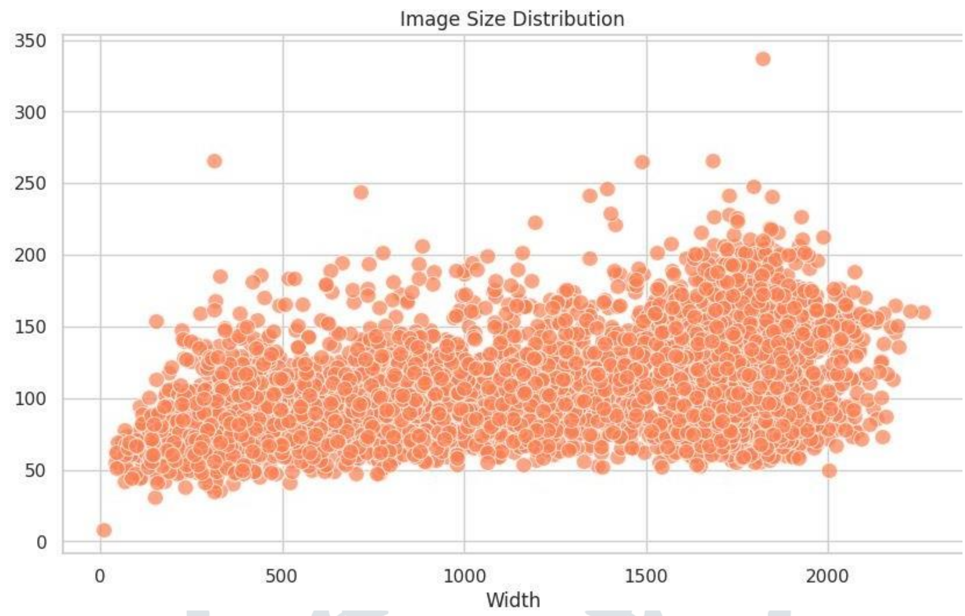


Figure 5.3: Image Size Distribution

- Grayscale Conversion: Given that colour information is not critical for handwriting recognition, converting images to grayscale was identified as a necessary step to reduce computational complexity.
- Normalization: We recognized the need to normalize the pixel values to a standard range (e.g., 0 to 1) to facilitatemore stable and faster model training.

VI. System Design

High Level Design

With reference to Fig 6.1The Handwritten Text Recognition system developed in this project can be broadly divided into three major components - The web UI, the backend and the database. The Web UI is the frontend part of our Handwritten Text Recognition System, where the user can interact and upload the images of the handwritten text and also it contains a canvas where the output can be displayed.The Backend component of our system contains the trained CNN model, which is integrated with the Google Vison Api. The Database component of our system is responsible for storing the user data securely using various encryption techniques. The below diagram depicts the High-Level Design of our system.

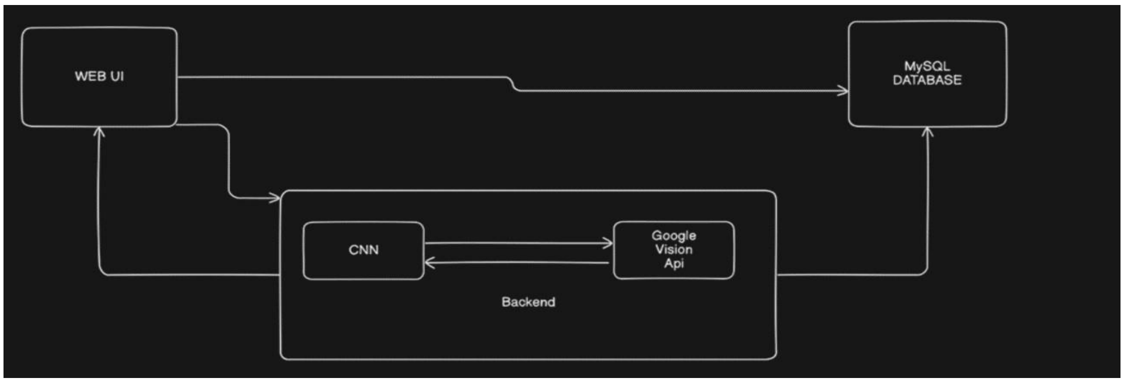


Figure 6.1: System Design

Low Level Design

To understand the low-level design of our system, we now delve deep into all the broad level components we described in the High-Level Design. The first component is the web UI which is made using the Flask Framework, as our UI is not very complex. The main components of the web UI include the following:

- Home Page
- Sign Up Page
- Login Page
- History Page
- Upload page

The low-level design architecture of the web UI components focuses mainly on enhancing the user experience and also giving the user the choice to give feedback for the accuracy of the model. This feedback given by the user helps us improve the model further. The UI of our Handwritten Text Recognition system is designed carefully so as to provide the user with a fast and productive environment while converting the handwritten text into digital text. Furthermore, more features like a dedicate canvas, on which the recognized handwritten text is displayed has the option copy to clipboard, giving the user the best user experience. The below figure depicts the low-level design of the Web UI component.

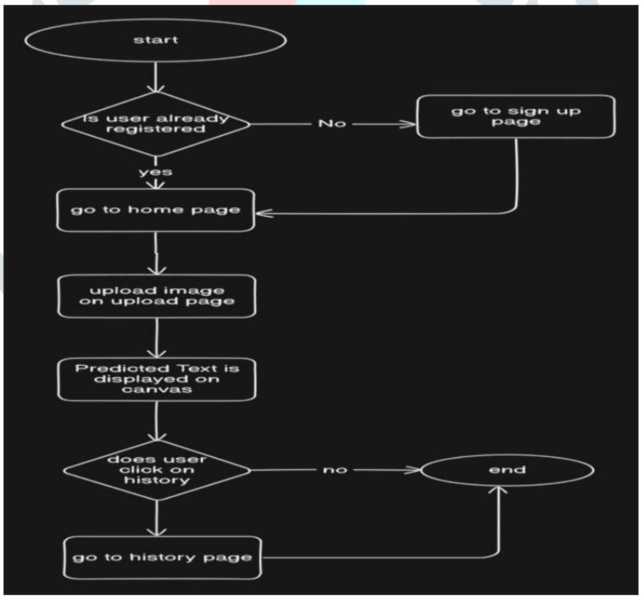

start
is user already registered
No
go to sign up page
yes
go to home page
upload image on upload page
Predicted Text is displayed on canvas
does user click on history
no
end
go to history page

Figure 6.2: Low-Level Design

VII. Experimental Scenario

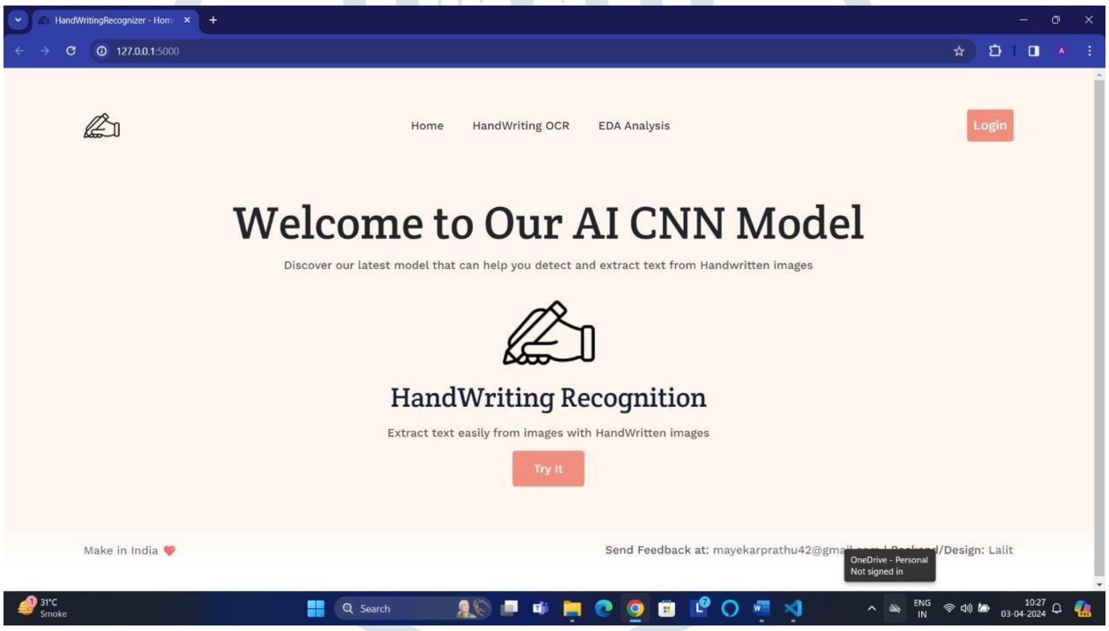
Initially we trained the model using purely the CNN architecture on Google Collab using 12 epochs on a T4 GPU which helped us train the model much faster. The accuracy in the case was found to 70%approximately. Hence, to further improve the accuracy of the model we increased the number of epochs
to 20, then 25. But there was no significant change observed in the accuracy of the model.

Table 7.1. Comparison of Accuracies w.r.t epochs

Accuracy	epochs
70%	12
72.14%	15
73.21%	20
73.65%	25

To improve the accuracy of the model, we integrated the already trained CNN model with Google Vision Api, and thisincreased the accuracy of our model significantly increased to almost 97%, with a very minimal loss. Further experimentation showed that the model does not predict handwritten text accurately when the text is not clearly visibleor it is in some other format. But in real world applications, this scenario rarely happens.

VIII. RESULTS



Welcome to Our AI CNN Model

Discover our latest model that can help you detect and extract text from Handwritten images

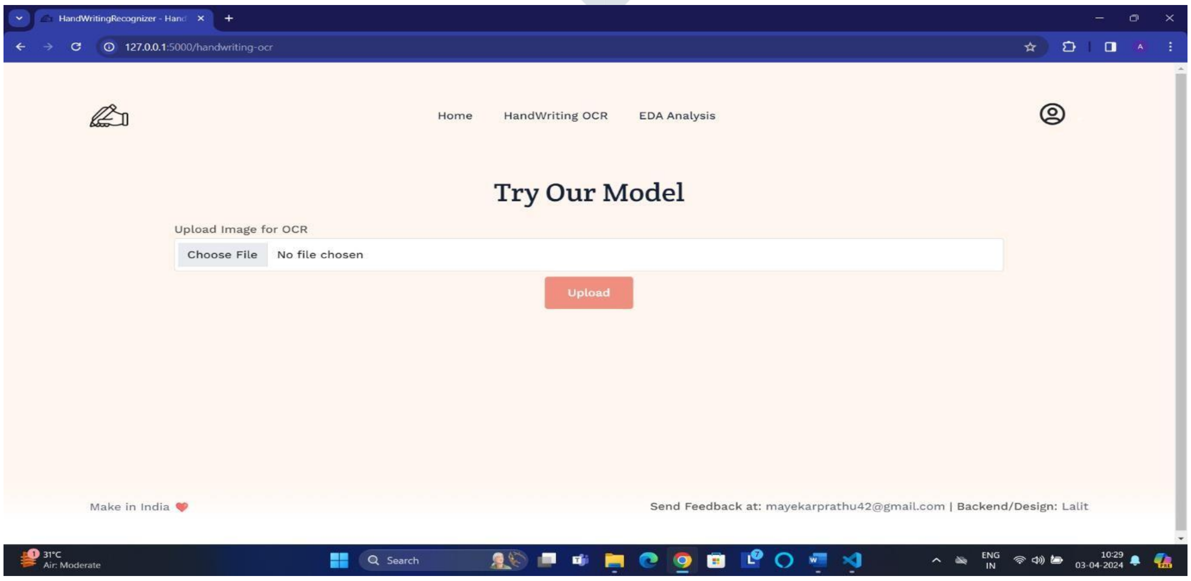
HandWriting Recognition

Extract text easily from images with HandWritten images

Make in India

Send Feedback at: mayekarprathu42@gmail.com | Backend/Design: Lalit

Figure 8.1: Home Page



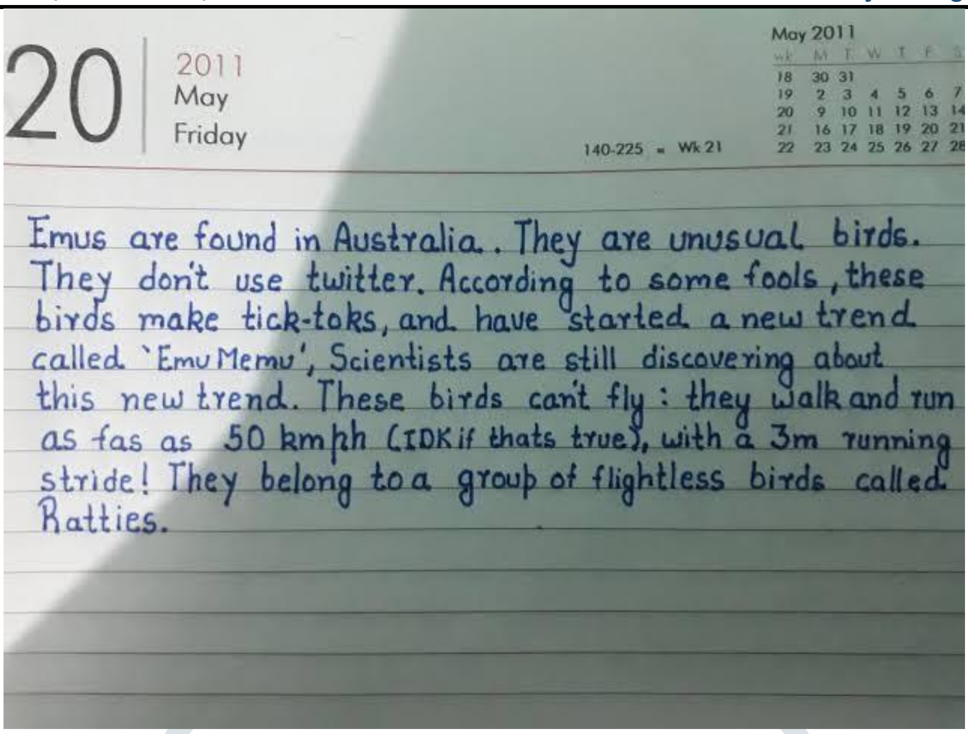
Try Our Model

Upload Image for OCR

Make in India

Send Feedback at: mayekarprathu42@gmail.com | Backend/Design: Lalit

Figure 8.2: Upload Image Page



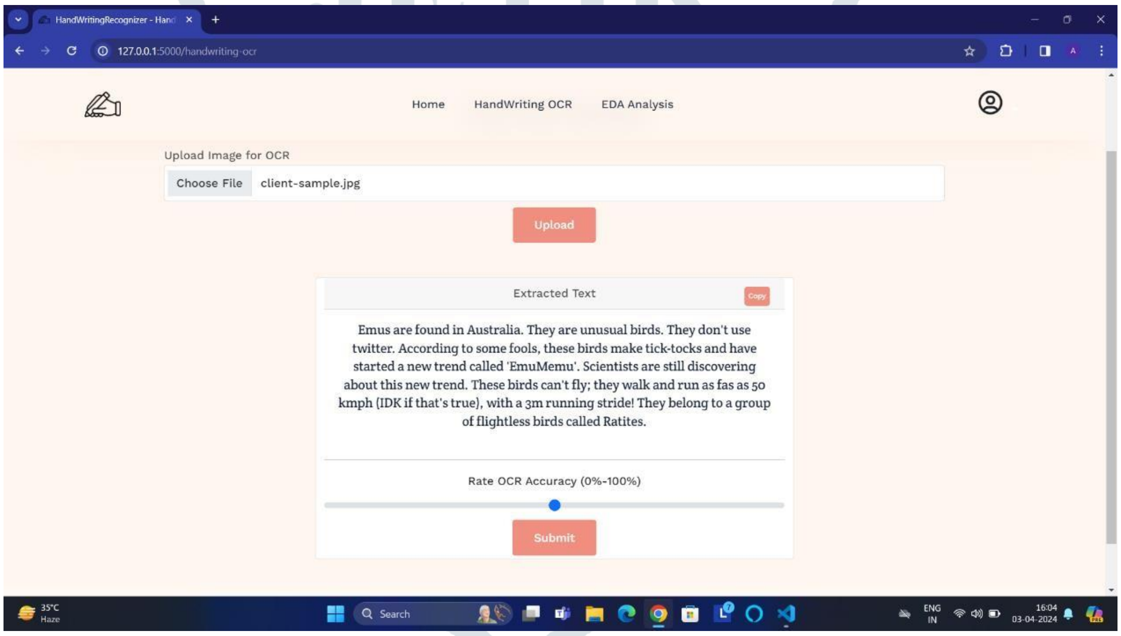
20 2011 May Friday

May 2011

140-225 Wk 21

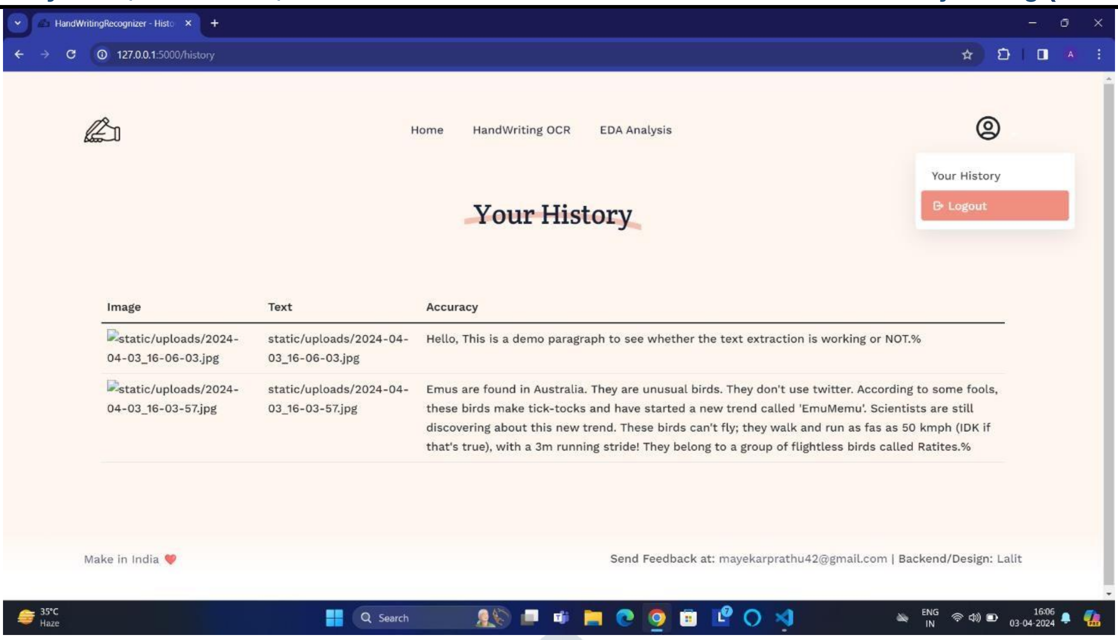
Emus are found in Australia. They are unusual birds. They don't use twitter. According to some fools, these birds make tick-toks, and have started a new trend called 'Emu Memu', Scientists are still discovering about this new trend. These birds can't fly: they walk and run as fas as 50 kmph (IDK if thats true), with a 3m running stride! They belong to a group of flightless birds called Ratties.

Figure 8.3: Sample Image of Handwritten Paragraph



Emus are found in Australia. They are unusual birds. They don't use twitter. According to some fools, these birds make tick-tocks and have started a new trend called 'EmuMemu'. Scientists are still discovering about this new trend. These birds can't fly; they walk and run as fas as 50 kmph (IDK if that's true), with a 3m running stride! They belong to a group of flightless birds called Ratites.

Figure 8.4: Model Predicting the Handwritten Image Text



Your History

Image	Text	Accuracy
static/uploads/2024-04-03_16-06-03.jpg	static/uploads/2024-04-03_16-06-03.jpg	Hello, This is a demo paragraph to see whether the text extraction is working or NOT.%
static/uploads/2024-04-03_16-03-57.jpg	static/uploads/2024-04-03_16-03-57.jpg	Emus are found in Australia. They are unusual birds. They don't use twitter. According to some fools, these birds make tick-tocks and have started a new trend called 'EmuMemu'. Scientists are still discovering about this new trend. These birds can't fly; they walk and run as fas as 50 kmph (IDK if that's true), with a 3m running stride! They belong to a group of flightless birds called Ratites.%

Make in India ❤

Send Feedback at: mayekarprathu42@gmail.com | Backend/Design: Lalit

Figure 8.5: User History Page

IX. CONCLUSION

This project represents a significant step forward in the field of handwriting recognition, showcasing the power of combining CNNs with advanced tools like Google Vision. Through meticulous dataset analysis, careful model design and training, and strategic technology integrations, we have developed a system that pushes the boundaries of what's possible in handwriting detection. As we continue to refine and enhance this system, its potential applications in document analysis, automated form processing, and historical text digitization are vast and promising.

X. FUTURE SCOPE

The project achieved an impressive accuracy of approximately 86% over 12 epochs, a testament to the effectiveness of the chosen methodologies and integrations. The system demonstrated a strong capability to recognize a wide range of handwriting styles, with Google Vision providing a significant accuracy boost in complex cases. Future work could explore further enhancements, such as employing more sophisticated CNN architectures (e.g., ResNet or Inception), exploring transfer learning, and expanding the dataset to include more diverse handwriting samples. Additionally, improving the system's real-time processing capabilities and exploring its deployment in practical applications would be valuable next steps.

REFERENCES

[1] Viswanatha V, Ramachandra A C, Satya Dev Nalluri, Sai Manoj Thota, and Aishwarya Thota, "Handwritten Digit Recognition Using CNN", International Journal of InnovativeResearch in Computer and Communication Engineering, Volume 11, Issue 1, January 2023

[2] P.Rajeshwari, D.Vinay Sekhar Reddy, G.Pranay, and Mohd Arbaz Mazharuddin, "TextExtraction from an Image using CNN", Journal Of Emerging Technologies And Innovative Research (JETIR), Volume 9, Issue 4, April 2022.

[3] G.R. Hemanth, M. Jayasree, S. Keerthi Venii, P. Akshaya, and R. Saranya, "CNN-RNN Based Handwritten Text Recognition", ICTACT Journal On Soft Computing, Volume 12,Issue 1, October 2021.

[4] A. De Sousa Neto, B. Bezerra, A. Toselli and E. Lima, "HTR-Flor: A Deep LearningSystem for Offline Handwritten Text Recognition", Proceedings of International
Conference on Graphics, Patterns and Images, pp. 1-8, 2020.

[5] A. Sampath and N. Gomathi, "Handwritten Optical Character Recognition by HybridNeural Network Training Algorithm", The Imaging Science Journal, Vol. 67, No. 7, pp. 359-373, 2019.

[6] J. Sueiras, V. Ruiz, A. Sanchez and J.F. Velez, “Offline Continuous Handwriting Recognition using Sequence to Sequence Neural Networks”, Neurocomputing, Vol. 289,pp. 119-128, 2018.

[7] Q. Vo, S. Kim, H. Yang and G. Lee, “Text Line Segmentation using a Fully
Convolutional Network in Handwritten Document Images”, IET Image Processing,Vol. 12, No. 3, pp. 438-446, 2018.