End term T12 - Project Phase II

On

IDENTIFICATION OF HUMAN SENTIMENTS USING AI WITH HELP OF PERSONALIZED HUMAN HANDWRITING

Project Members

Name	PNR No.
Abhishek Chebolu	1032180316
Ahbaz Memon	1032180046
Prajakta Chaudhari	1032180317
Snehalraj Chugh	1032181182

Under the Internal Guidance of

Prof. Varsha Naik

School of Computer Engineering and Technology
MIT World Peace University, Kothrud,
Pune 411 038, Maharashtra - India

2021-2022

Dr. Vishwanath Karad
MIT WORLD PEACE UNIVERSITY | PUNE
TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

SCHOOL OF COMPUTER ENGINEERING AND TECHNOLOGY

C E R T I F I C A T E

This is to certify that,
Abhishek Chebolu, Ahbaz Memon, Prajakta Chaudhari, Snehalraj Chugh

of BTech.(Computer Science & Engineering) have completed their project titled "*IDENTIFICATION OF HUMAN SENTIMENTS USING AI WITH HELP OF PERSONALIZED HUMAN HANDWRITING"* and have submitted this Capstone Project Report towards fulfilment of the requirement for the Degree-Bachelor of Computer Science & Engineering (BTech-CSE) for the academic year 2020-2021.

[Prof. Varsha Naik]	**[Dr. Vrushali Kulkarni]**
Project Guide	Head of School
School of CET	School of CET
MIT World Peace University, Pune	MIT World Peace University, Pune

Internal Examiner:

External Examiner:

Date:

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present our capstone project titled "Medical Imaging for Disease Detection and Analysis" undertaken at Dr. Vishwanath Karad MIT World Peace University, Pune, in the School of Electronics and Communication Engineering during the Final Year of our Bachelors of Technology degree. This project has broadened our knowledge horizons and will be an invaluable asset to our careers ahead as engineers.

Bearing in mind the above, we would like to extend our gratitude to Dr. Varsha Naik for her mentorship and guidance throughout our capstone project, under whom we were able to build new skillsets, solve problems by thinking critically and ensure the successful completion of our project. Her willingness to share her vast knowledge in the domains of Artificial Intelligence and Machine Learning made us understand the nuances of this project meticulously.

We would also like to thank Dr. Uma Pujeri and Dr. Vinayak Musale for being approachable and compassionate project coordinators, who assisted us to navigate through our capstone project in the most efficient way possible.

We are grateful to Dr. Vishwanath Karad MIT World Peace University, Pune and Head of School - Electronics and Communications Engineering, Dr. Vrushali Kulkarni for giving us this opportunity to conduct research on a topic that we are truly passionate about, at our esteemed institution.

We would like to extend our heartfelt gratitude to the entire institution, teaching faculty, ancillary staff and laboratory staff for their resources, time, constant support and encouragement. They have helped us to bring out the best from this capstone project opportunity and our personal abilities over the past 6 months during Trimester XI and Trimester XII.

Finally, last but by no means least; a big thanks to our peers and family have been our pillars of strength throughout.

Thank You!

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1. Project Statement

We want to develop a personal module to assist and better analyse an **individual** through their **handwriting** using **artificial intelligence**, than what has been done so far which is quite **generic** and limiting. We are combining four domains, .i.e., **Deep Learning, Computer Vision, NLP** for our research.

- Our **focus** is to invent **new** methods for **letter segmentations** in a sentence
- Applying own algorithms to develop an OCR system with the capability of extracting words and characters
- Creation of an OCR model which will help in extracting text from handwritten text documents
- Developing an AI based classification model to predict handwritten characters to generate text
- Applying own modules to develop an OCR system with the capability of extracting words and characters
- A module will be integrated to **help deaf** and **blind** people **understand handwritten text**.
- The machine can understand the **cancelled-out/scribbled words** by itself as to analyse the page more human-like.
- We will be particularly developing AI based model to:
    - Detect words in page
    - Localizing the characters
    - Predict the characters using deep learning techniques

1.2. Area

The number of ways individuals can take photos using modern technology has increased dramatically in the contemporary age. It is conceivable that these photos include the essential written type of data that have to be edited or preserved electronically.

It's indeed possible to browse for and identify information in photographs, as well as papers, with aid of said Handwriting Analysis Architecture. Using a computer application, a written message is produced by interpreting the letters of an image or digitized document.

Recurrent learned algorithms and faster processors have made it possible for a much more comprehensive writing identification system that really can translate photos of a few segmental letters or indeed a phrase.

The method has a wide range of potential uses. Capabilities include analysing financial cheques, identifying information from cards, assisting the sighted with identifying information on writings, as well as portraying emotions associated with the research process. For the moment, our program solely recognises text in English. Other languages could be implemented in the long term.

1.3. Project Introduction and Aim

Need of project

Programs that identify characters on a screen whenever data is scanned from printed material have become in high demand these days. Saving the data accessible inside the documents together in a computerised unit and utilising it subsequently via search has become more popular in today's world. For the most part, there is a need in the market for a computer that can replicate people's style of writing and help others who have problems reading one's style.

Aim

We aim to research on the various deep learning and AI based technologies available in the industry for handwritten text recognition. This research will help us gain insights for our main aim that is to develop our own OCR module without help of pretrained OCR models such as tesseract.

The project aims to cover different types of methodologies required to make a robust OCR system with minimized errors and also with the recognition of cursive handwriting. We aim on doing this by employing our own algorithms and creation of our training models.

In the study, the notion of saving the contents of handwritten documents in a computerised storage place and afterwards extracting text from them, analysing, and understanding the handwritten text is being explored in more depth and detail.

Picture enrichment processes, including noise removal, normalisation, binarization, and other similar techniques, are used to improve the overall quality of the raw image. Images with information that perhaps the user is unable to modify are used. However, in order to repurpose this data, it will need to get acquainted with the writer's normal handwriting style as well as the manner in which he composes the messages. Then, when it has been trained, it can also be used in a variety of diverse applications with the maximum level of accuracy.

CHAPTER 2

LITERATURE SURVEY

Paper Name	Consortium	Highlights of the paper	Uses	Research Gaps
Comparative Analysis of Text Extraction from Color Images using Tesseract and OpenCV	IEEE	Analyses the effect of unprocessed images and preprocessed images on the performance of tesseract	Text Extraction Image Preprocessing	Complicated or cursive handwriting
Image to Multilingual Text Conversion for Literacy Education	IEEE	It translates text just by capturing an image with the user's smartphone camera and translation instantly appears on the user's mobile screen in the language selected by the user.	Multilingual Text Conversion Character Recognition	Generalized approach
Real time license plate detection using OpenCV and tesseract	IEEE	The CV2 OpenCV library using Python language is used for image processing and Tesseract is used for text extraction from the processed image.	License Plate detection	Generalized approach
Text-Based Handwritten Recognition Through an Image Using Recurrent Neural Network	Springer	The main objective of this paper is to propose the design of an expert knowledge-based neural network system for handwritten recognition that can effectively recognize the text from the given input image using a recurrent neural network approach.	Word segmentation	Character level segmentation
Image Character Recognition using Convolutional Neural Networks	IEEE	This paper studies the use of CNN in detecting and recognizing handwritten text images with higher accuracy. The CNN model is tested on English handwritten characters and validated on its performance. The model performs feature extraction from images through multiple layers. These are later used for training the model and thereby recognizing characters.	Character Recognition	Cursive handwriting
Optical Character	IEEE	the paper presents the design and procedure of the OCR WebApp, which	Character Recognition	Generalized approach

Recognition using Tesseract and Classification		consists of three sections that are: Image-to-Text, Real-time OCR (using webcam), and Handwritten Text Recognition. In this project, OCR uses Tesseract as an engine to display the text to the user and HTR uses a Deep learning model to classify the letters and display them to the user.		
Optical Character Recognition for English Handwritten Text Using Recurrent Neural Network	IEEE	The framework introduces a Recurrent neural network for recognizing English handwritten text.	Letter classification	Cursive handwriting
A Novel machine learning approach for Scene Text Extraction	IEEE	Image based text extraction is a popular and challenging research field in computer vision in recent times. For text identification, contrast enhancement is done by applying LUV channel on an input image to get perfect stable regions. In text recognition, text regions are recognized and labelled with a novel CNN network. The CNN output is stored in a text file to make a text word.	Bounding boxes, Error Correction, Letter Classification	Cursive Writing, Cross-Relationship
Analysis on Preprocessing Techniques for Offline Handwritten Recognition	Springer	The paper emphasizes on various techniques for pre-processing an image that aids in the further process of Image recognition.	Image preprocessin g	Character level segmentation
An Efficient Digit Recognition System with an Improved Preprocessing Technique	Springer	In this paper the experimentation is done on the classification of different hand written English numbers with preprocessing of the image obtained from which digits are to be extracted.	Image preprocessin g	Implements word-level segmentation instead of character segmentation
Handwriting Recognition of Diverse Languages	IEEE	Online Handwriting Recognition for Diverse Languages is a system which is used to recognize digital as well as handwritten inputs. The two methods are K-Nearest Neighbour(KNN) and Tesseract OCR. After that they describe in detail the working of both methods, compare them according to their recognition accuracy.	Letter Segmentation , Bounding Boxes	Homogenous Database
Survey on handwriting-based	Springer	Graphology is the field of graphology to analyze personality based on handwriting. According to graphology, there is a vast	Graphology, Analysis of	Forgery, Loss/Exclude d features

personality trait identification		range of features of handwriting strokes which carry psychological characteristics of the writer. Psychologically supported handwriting features help to understand personality traits. The paper relates these features and encourages the use of computer-based graphology for personality prediction.	individual personality	
Improved Lane Line Detection Algorithm Based on Hough Transform	Springer	They propose an algorithm directly identifying lane line in Hough space. The image is conducted with Hough transform, and the points conforming to the parallel characteristics, length and angle characteristics, and intercept characteristics of lane line are selected in Hough space.	Line detection	Character recognition of natural scene had limited success
A Method of Workpiece Coherent Line Detection Based on Progressive Probabilistic Hough Transform	ACM	This paper discusses an Improved PPHT method which performs edge detection combine with original PPHT algorithm to find lines of workpiece object. After discarding noise lines, this method divide the detected lines into several groups by finding collinear candidates.	Line detection	Cursive handwriting
Natural scene text detection based on YOLO V2 network model	IOP	The main works in the paper include the following: prepare the datasets; we train the YOLO v2 with the optimum parameters, carry out the regression analysis of the coordinate parameters and categories of bounding boxes, obtain the detection result; according to different detection models.	Object detection	Character level segmentation

CHAPTER 3

PROBLEM STATEMENT

Previously, individuals utilised paper and pen to record every piece of information that was accessible at the moment. Papers were used to write each and every piece of literature and literary text. It is now necessary to digitalise all of the papers that were previously only available on paper.

The majority of the materials are written on paper. When an author needs to begin writing things with their own handwriting, they would have to utilise a few apps that require people to start writing in a blueprint and then use that layout as a typeface. However, the copy of the original will still be created using their beautifully written font, regardless of the implementation they use. In this case, they have difficulty since it still seems to be an exact digital replica of a typeface that appears like human writing, but it does not convey almost any of the emotions that an author's written paper would convey.

The learning from previously accessible written manuscripts or pages that provide the model with a notion of the varied fonts a student writes, which can be recognised by actual forms of writing, is what our suggested model is centred on. This revolutionary approach is capable of identifying certain unique types of images, such as a small blurred movement-image and also those that contain letters that have been deleted, as well as recognising words inside in a cleaner and more effective way than a human being would be capable of.

3.1. Project Scope

In this study, the primary goal was to design a system that would aid in the categorization and identification of handwritten letters and numbers that could subsequently be utilised in a variety of applications.

In today's digital world, the ability to recognise letters and figures is critical, particularly in organisations and individuals with written papers that must be analysed using systems.

Systems that are utilised for writing categorization and identification assist institutions and individuals in completing complicated tasks by gaining a grasp of the author's point of view and also their feelings while composing the text. The present system processes and reads written letters and numbers using neural network models, which are currently under development. The method relied on convolutional neural networks (CNNs), which were trained on training examples to recognise characters and digits with relative ease. The use of CNN enabled the computer to become more responsive to various aspects of things, analogous to the human sensory systems.

Because of the learning data contained in the state's databases, it was straightforward to categorise and recognise various handwritten characters, text, and numbers. Photographic capture, digitisation, pre-processing (including classification), feature extraction (including identification), and detection are all aspects of handwriting recognition.

The finished system met or exceeded all of the defined standards for precision as well as for classification results. The new study's findings may be applied to identifying possible in other dialects, which is a promising development. According to the present study, ML models may be used to transform novels, papers, handwritten notes, or newspaper articles into electronic textual form utilising neural network models.

3.2. Project Assumptions

With this tool, past studies' flaws, such as their inflexible mathematical formalism and contradictory theoretical premises when coping with inaccurate inputs, are no longer present. Despite the sophisticated mathematical analysis, the most common objections may be divided into two categories:

1) Constraints imposed by the computational mathematics assumption that were chosen. Inside the theory, a few of these hypotheses is the mathematical distributions of the condition lifetime, which is one of the hypotheses. As a consequence, several fundamental characteristics of the architecture of communication are not well represented in the systems that are now accessible.

2) The rigid numerical framework, which really is incompatible with the imperfect quality of writing and spoken data.

The reality that our approach is built solely on information concerning pen motions results in the limitation of pen-hand paradigm that they have created. Furthermore, the theories are predicated on the notion of a fixed location on the printed sheet, upon which the hand-pen mechanism is supposed to rotate as a finite string of letters is created by the pens. Clearly, this hypothesis might be validated by research that makes use of the 3D motion monitoring system.

We feel that if characteristics of the real personal notes are accounted for, a much more realistic version than that which has been described in several studies might be constructed. In this regard, the strategy that will be discussed in this section may quickly prove to be beneficial.

For the most part, 3D motion monitoring will be valuable in graphemic studies in either evaluation studies where hypotheses about the link between major axes in calligraphy and aspects of handwritten imprints need to be tested or handwriting imprints themselves.

3.3. Project Limitations

Text recognition from photos faces a number of difficulties, including system integration concerns, text distortion, picture quality, cursive text interpretation, and low-resolution sensors, among others. Lots of studies are being conducted in this field at the moment.

The transformation of text from lesser-known or voting bloc languages is indeed a significant area of study. Processing the visual and audio material necessitates the use of a significant amount of storage. In addition, decreasing the dimensionality of the problem while maintaining the output quality is a significant issue when putting the system into operation.

3.4. Project Objectives

The primary goal of this project is to develop an intelligent system for handwritten character recognition that is based on a neural network. Other aims are as follows:

Handwritten character recognition systems are extremely unreliable, and researchers are working to improve their accuracy by building a system that would employ efficient and unique technologies to recognise handwritten letters and phrases from visual sources, being our idea to contribute through this research

Research and demonstration of the use of neural network tech for the creation of efficient handwritten character recognition systems as well.

CHAPTER 4

PROJECT REQUIREMENTS

4.1. Resources

Human Resources

- *Handwritten Notes of a Writer*
- *Capstone team members*
- *Team Supervisor*

Reusable Software Components

- *NLP*
- *Neural Network*
- *OpenCV*

Hardware Requirements

- *Windows/MAC laptop*
- *RAM 8GB*
- *NVIDIA GPU*

Software Requirements:

- *Python 3*
- *Tensorflow*
- *Google Colab*

4.2. Requirements Rationale

- This research will be directed at people who has handwritten notes and would want to transform them into an electronic copy.
- Although digital technology is increasingly being used in organizations, handwriting continues to be a component of people's day-to-day activities.
- Writers write down their thoughts, intentions, and suggestions on their notepad in their very own writing. Thus, in the research, this will assist our technology in doing the learning part.
- Instructors and students may have personal notes that they would like to share with one another, and this model might be used in that situation as well, since handwritten notes with paper and a pen cannot be completely substituted by digital technology.

4.3. Risk Management

Risk	Risk Type	Probability of Occurrence	Impact	Priority	Risk Management	Plan
IMAGE CAPTURING ERROR	Technical	20%	2-critical	Medium	Redoing the capturing process	Modifying using software
CORRUPTED IMAGE	Module specific	60%	4-marginal	High	Adding layers to distinguish and analyse the corrupted files	Reduction or discarding if required
VERSION ISSUE	Technical	10%	3-medium	Medium	By changing the versions and integrating new model versions	Keeping track of the working of the versions and updating it.

4.4. Functional Specifications

Interfaces

External interfaces required – Pages with handwritten notes

Internal interfaces required - Python

Communication interfaces - Django

Graphical User Interfaces – React.js

Interactions

User will have to upload few pages of their handwritten notes and then the function starts training and understanding to get its required output. Our selected module will display the particular output to the user with the requirements specified in that module.

CHAPTER 5

SYSTEM ANALYSIS PROPOSED ARCHITECTURE

5.1. Design Consideration

Future work

5.2. Assumption and Dependencies

Assumption

The user is already having the handwritten notes filled more than 5 pages.

Dependencies

Our module and web interface are interdependent

5.3. General Constraints

Text identification from images has several challenges including system integration issues, text warping, image stabilization, cursive text understanding and low resolution sensors.

5.4. Block Diagrams

Navigation

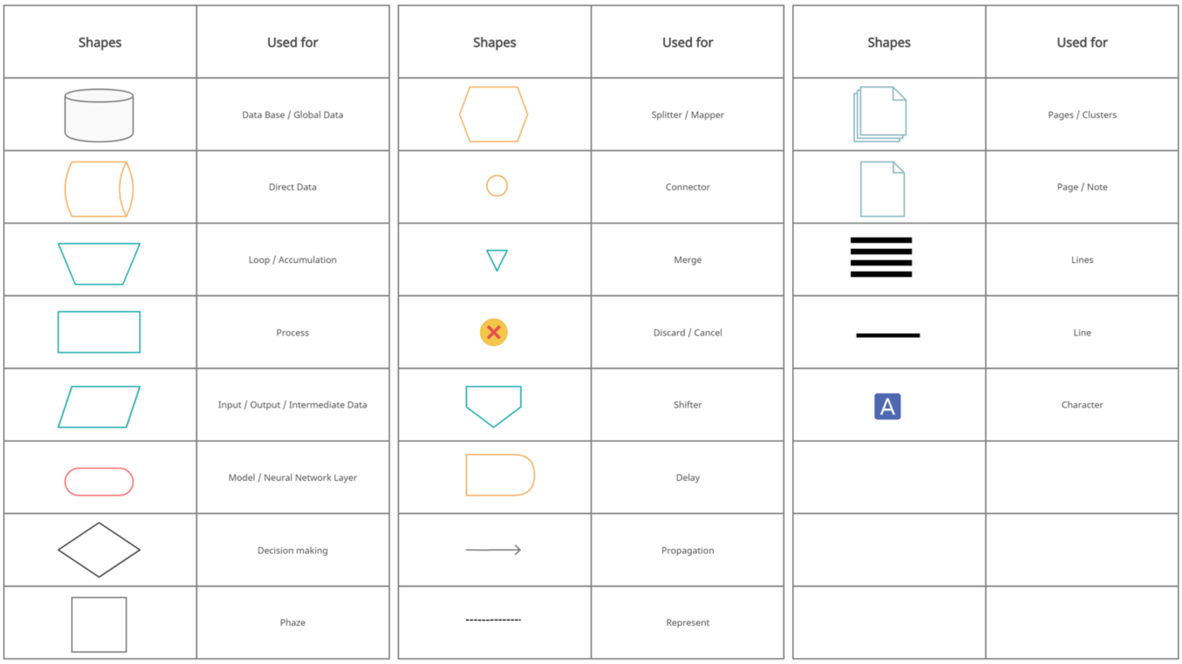


Figure 1: Navigation through each diagram giving an idea onto denoting the shapes as well as their usage

Flowchart

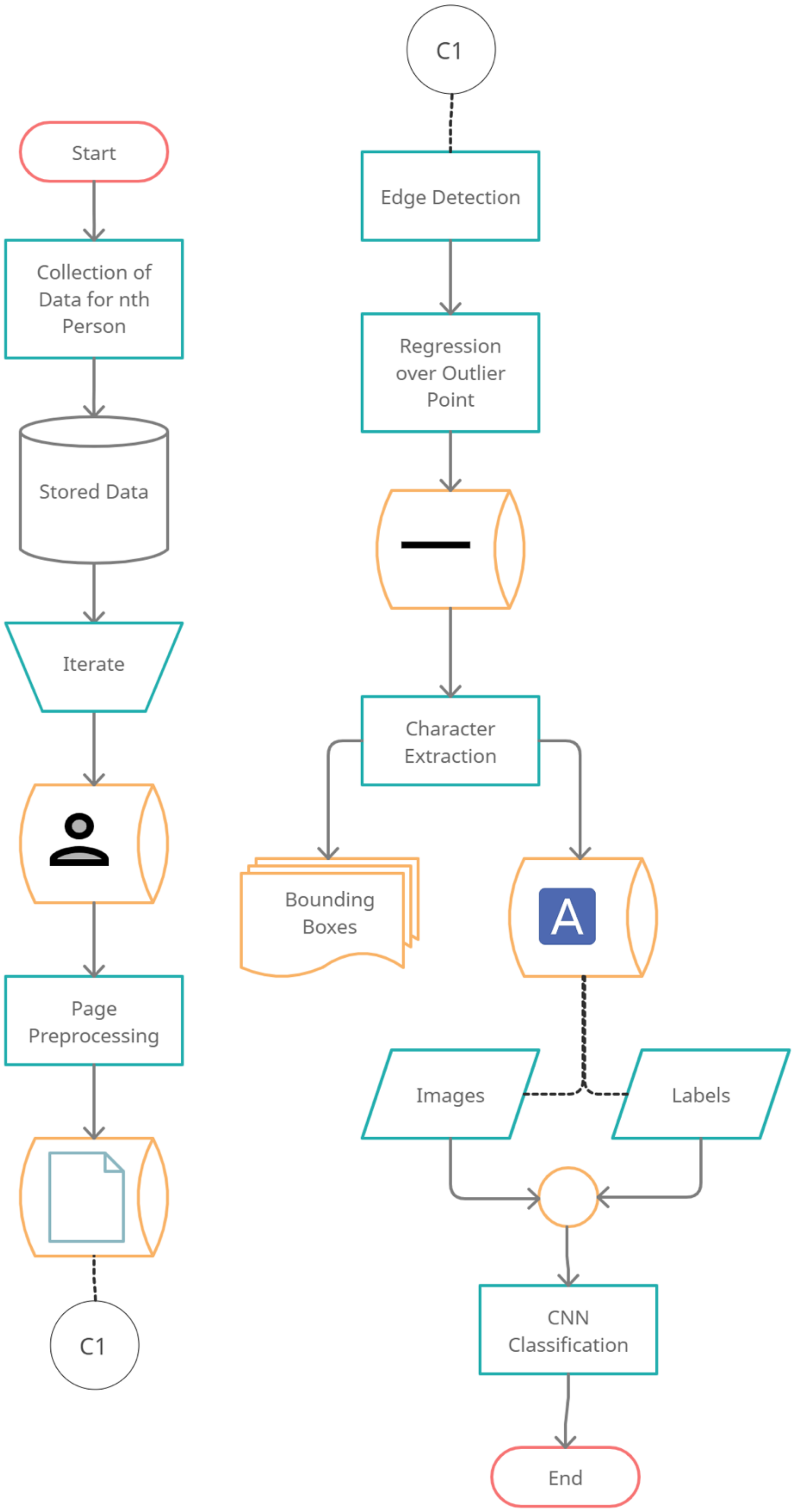


Figure 2: Flow chart describes our process of how each person's page is iterated and converted to characters. These extracted characters are then used for Classification.

5.5. System Architecture

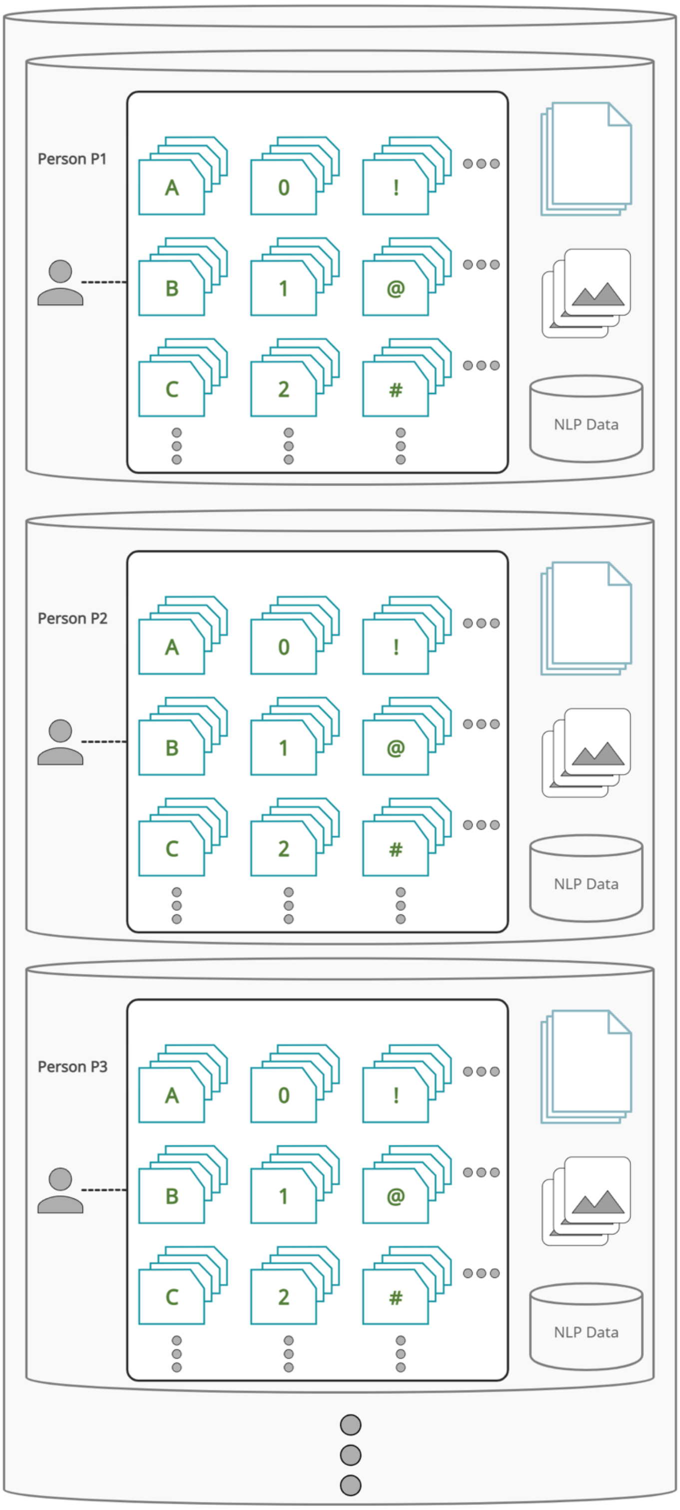


Figure 3: Person Specific Database created using the inserted pages

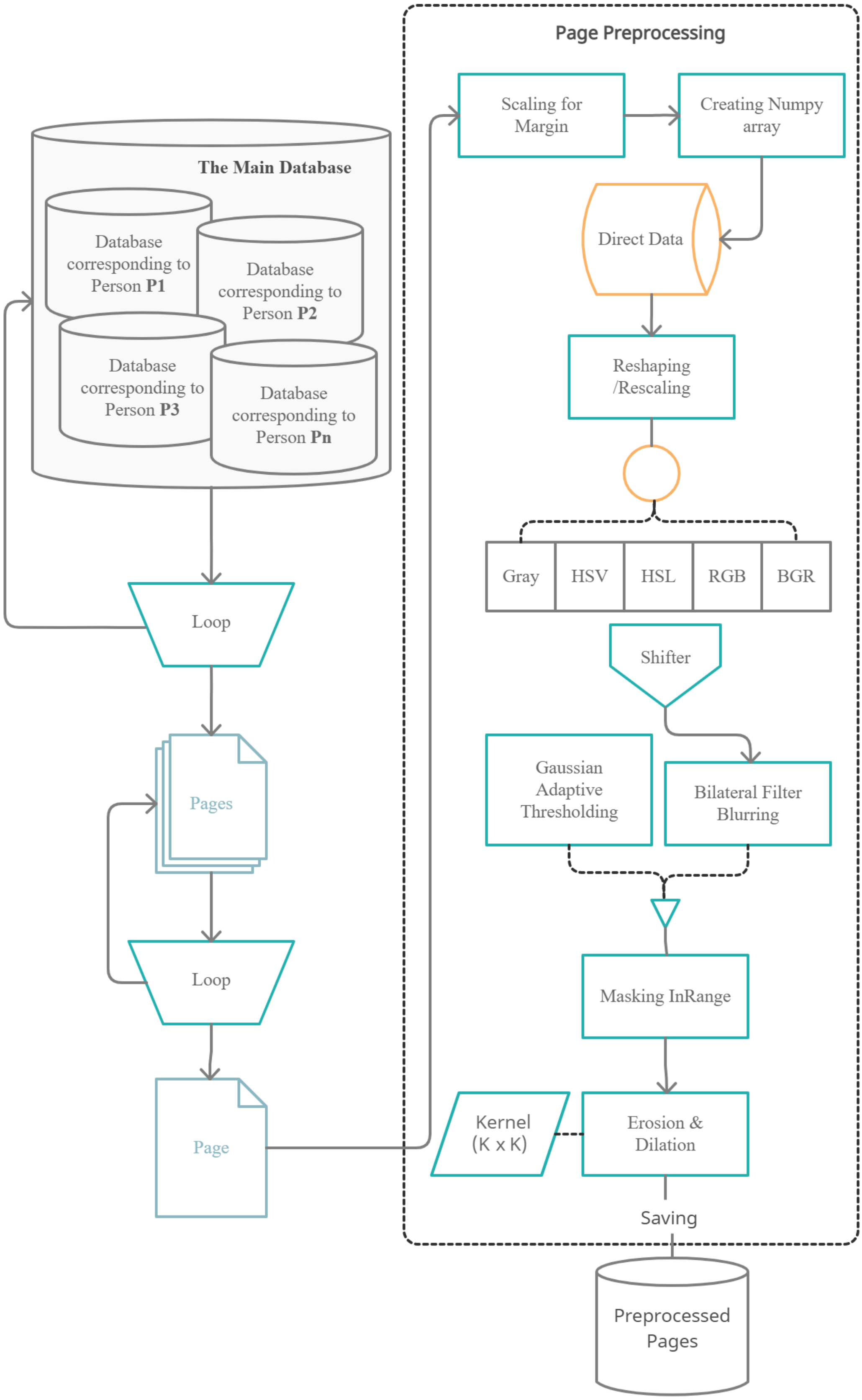


Figure 4: Main database looping and preprocessing of the pages

5.6. Low level Design

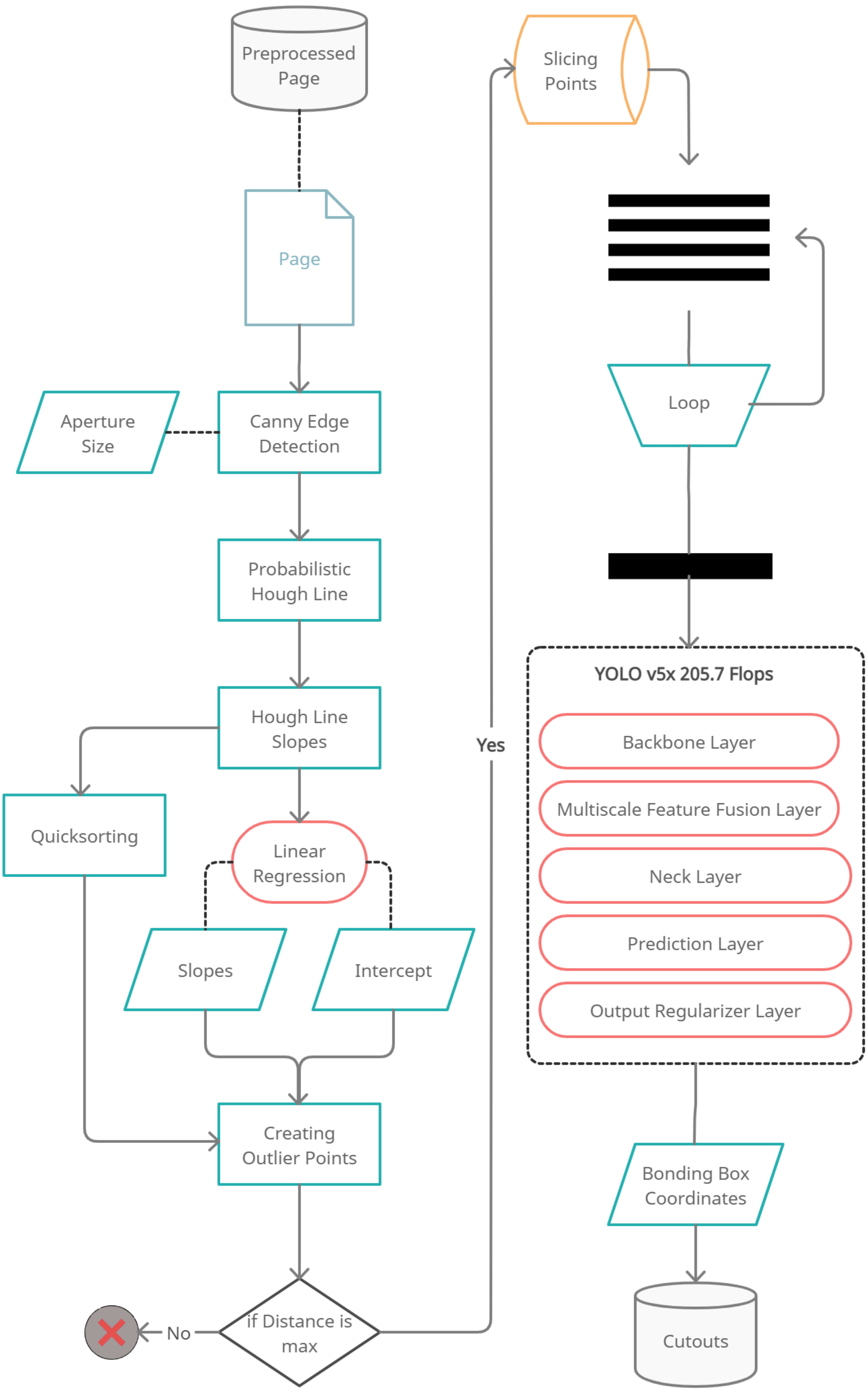


Figure 5: Preprocessed Pages for edge detection and character extraction using YOLO v5x FLOPS

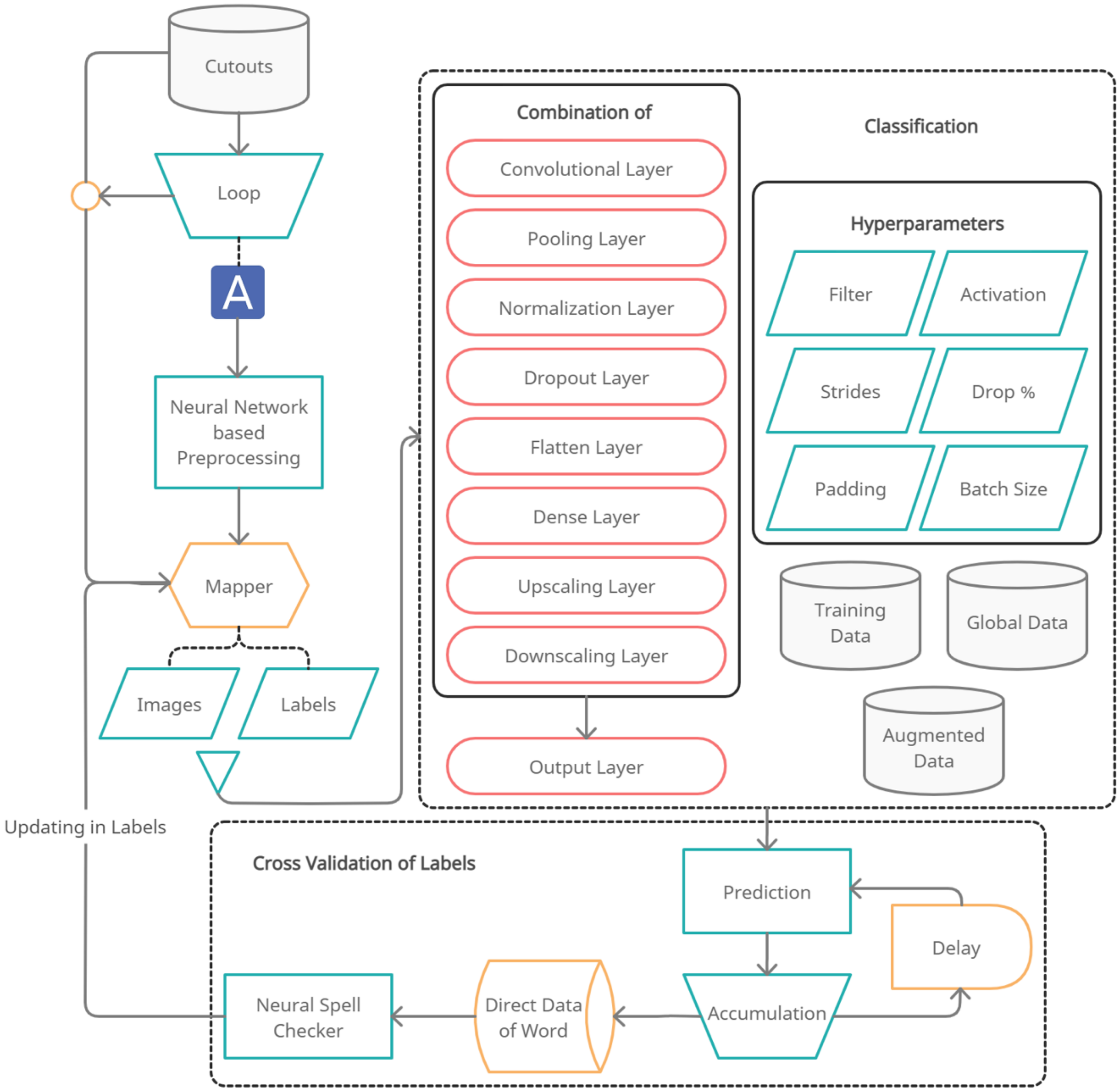


Figure 6: Pre-processing of bounding box images extracted in last stage and classification of the characters

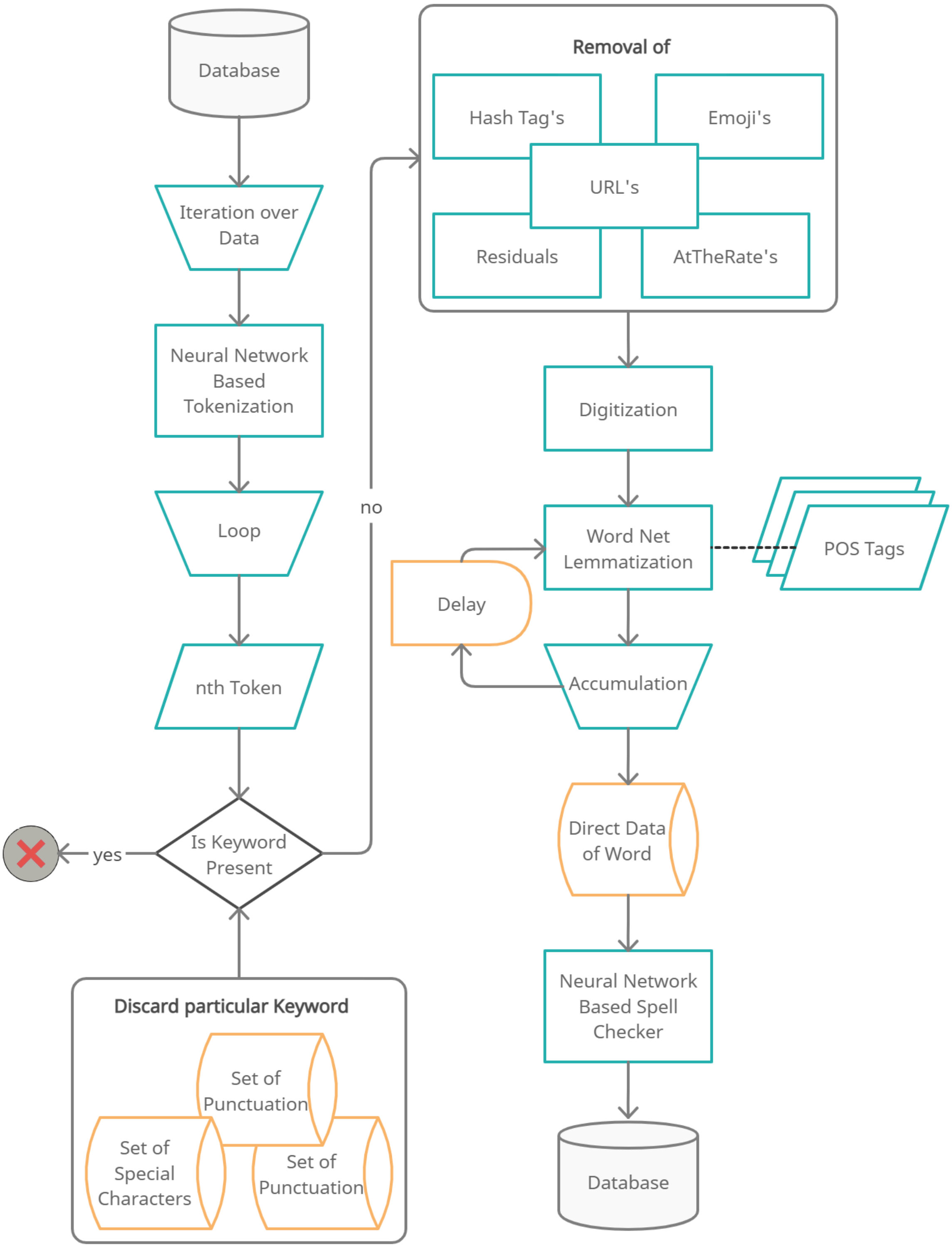


Figure 7: Database conversion to text and usage of NLP

5.7. UML Diagrams/Agile Framework

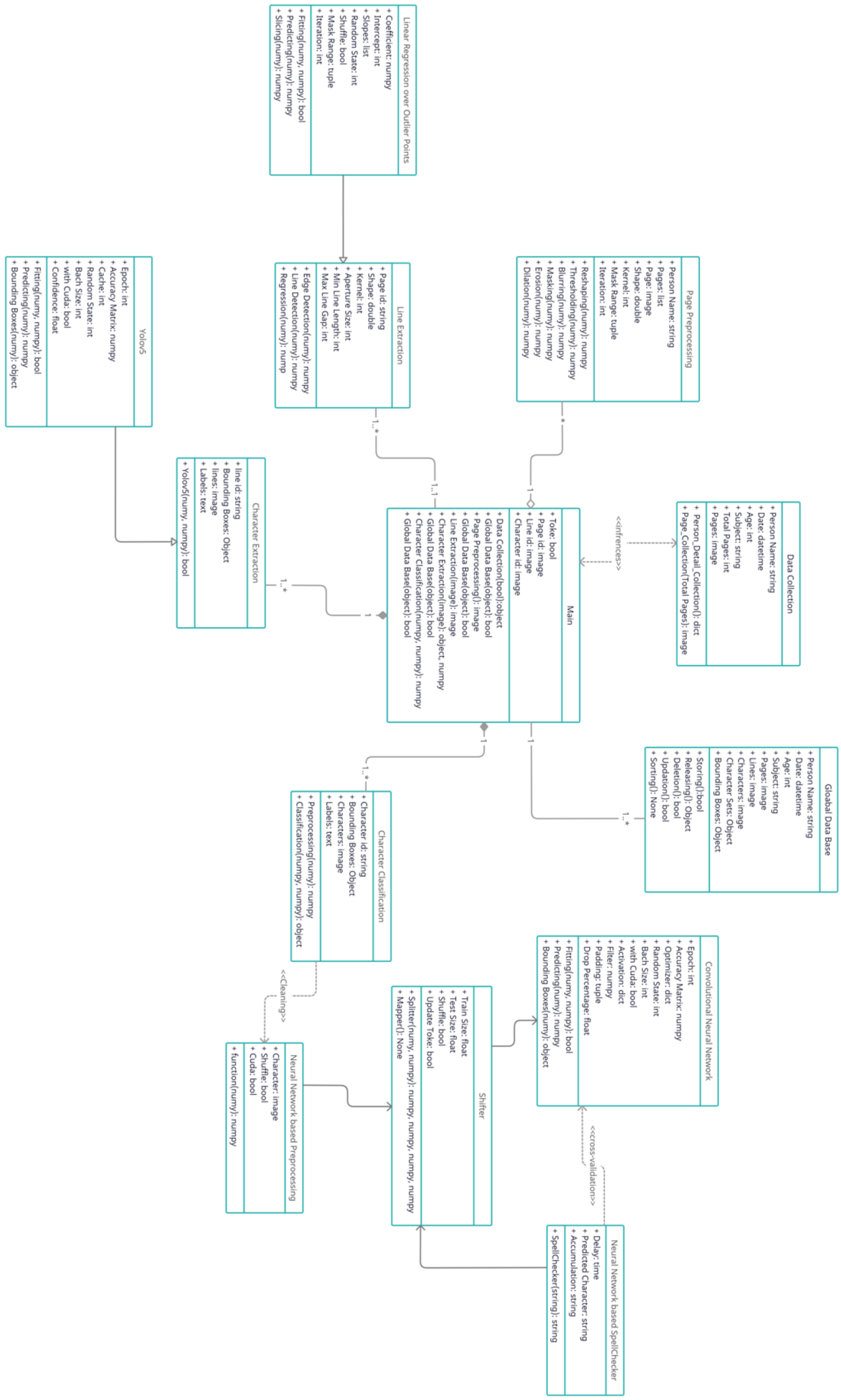


Figure 8: Class Diagram

CHAPTER 6

PROJECT PLAN

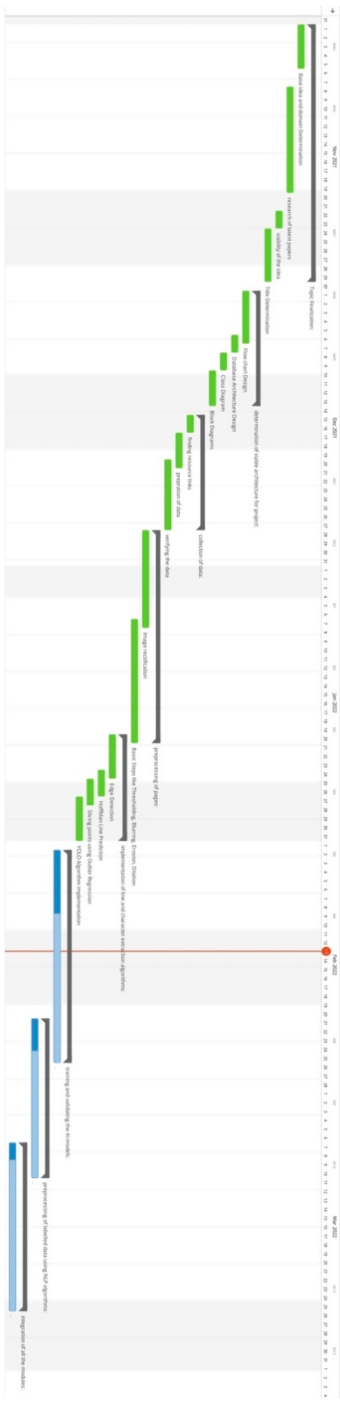


Figure 9: Gantt Chart explaining the project timeline

CHAPTER 7
IMPLEMENTATION

7.1. Dataset Description

In order to move forward with the modelling of our system, we required a dataset of handwritten papers that not only included text written by a number of writers but also included a diversity of handwriting styles used by individuals. Because we needed a sufficient number of handwritten samples from writers for both training and testing purposes for this research, we decided to go ahead and generate our own dataset rather than use a dataset that was already available to the public. This decision was made because we needed to use this research. It was requested of the writers that they write down their everyday thoughts as well as everything else they felt like. These pages that were collected were subsequently used for training. These pages contained the English alphabet, numbers, and words that had been struck out. These words that were struck out were also provided as an input to the model so that it can learn to recognise them early on in the process and train itself to do so because it is necessary to deal with them. By augmenting the data, we were also able to enhance the size of the dataset. When data is enhanced, it is beneficial since it improves the performance of models and the outcomes they produce by producing fresh and diverse instances for train datasets. This is an example of how augmenting data can be advantageous.

7.2. Methodology

There are three main steps to building a complete handwritten recognition module. The first step is to create a dataset for training the model. In the second step, the model is trained on the dataset. In the third step, the user uploads new handwritten pages to train the model on. The first stage is a comprehensive step comprising data generation, pre-processing for each output created via the phases, image improvement. The next step proceed with training the pictures on the processed photos. We then utilise this trained model to test on real-time input photos. The technique we follow for handing the image or page incorporates page pre-processing, page enhancement and then extraction of words from these pages. The output words are again pre-processed and upgraded to provide a better input for the following model. The characters are then retrieved from the processed words. Algorithms for strengthening these character cut-outs is applied to it before submitting the result to segmentation model. The segmented characters are then provided for classification. The process chain communicates to each other but independently in a way where the output of one process if erroneous or unenhanced will be addressed in the following process.

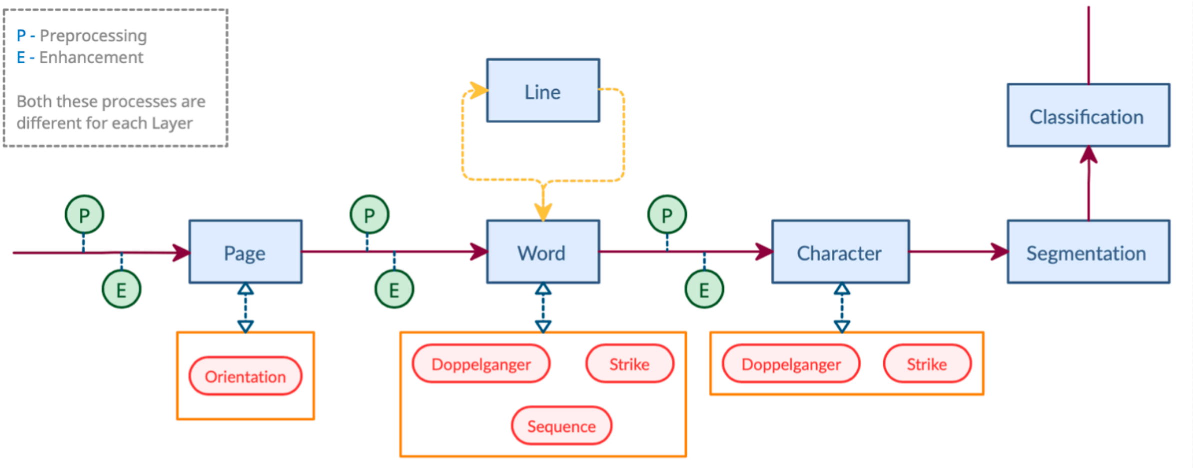


Figure 10: Process of our system

7.2.1. Page Pre-processing

The pre-processing of an image plays a very critical function in the betterment of the accuracy of any model. This is due to the fact that the environment in which an image is captured may differ from the conditions that exist in real time. The user may provide a page with a variety of backgrounds, all of which need to be managed. Additionally, different people write with a range of pen and ink colours. Another problem that needs to be fixed is determining whether or not an input image has been skewed in any way. We need to make the necessary adjustments to the image if it is skewed; otherwise, the model won't be able to recognise the correct words or characters. Finding the required angle that the image is tilted at in order to fix this problem is the first step. The majority of this stage was accomplished by applying several pre-processing techniques to the image using OpenCV. These techniques included blurring the image, adjusting the scale of the image, scaling the image, and many more.

7.2.2. Page Enhancement

It is expected that the size of the image will decrease as a result of the model's continued extraction of words and then characters ahead. When we finally get to the character extraction model, the number of pixels will have already significantly dropped. Convolutional neural networks achieve the best results when applied to images of higher resolution. Image enhancement through the use of deep learning models is the answer to this issue, which can be found by following the chain of reasoning that this leads to. In order to solve this problem, we can make use of a number of deep learning-based models, which enable us to scale the image by a factor of 2, 4, or 8. The model that we decided to move forward with is called RealESRGAN, and it is predicated on the notion that a combination of low resolution (noisy) sequence of photographs of a scene may be utilised to generate a high resolution image or image sequence. RealESRGAN is based on the idea that a high resolution image or image sequence can be generated from a combination of low resolution (noisy) images. As a result, it makes an effort to recreate the original scene image with a higher resolution given a collection of photos with a lower resolution that have been observed.

7.2.3. Word Extraction

The process of extracting words from text is one of the most crucial steps in an OCR system. There is a wide variety of approaches that can be used to accomplish the same goal. We have trained images on YOLOV5 Medium in order to facilitate the localisation of text included inside a page. The YOLOv5 platform supports a variety of architecture types, including nano, small, medium, big, and x-large. We proceeded with the experimentation of Nano, small, and medium sizes because Large and X-Large demand a significant amount of computer power. The small network achieved an acceptable level of accuracy but was unable to solve the problem of recognising cursive letters. The medium network offered a solution to this problem, making it the most effective option.

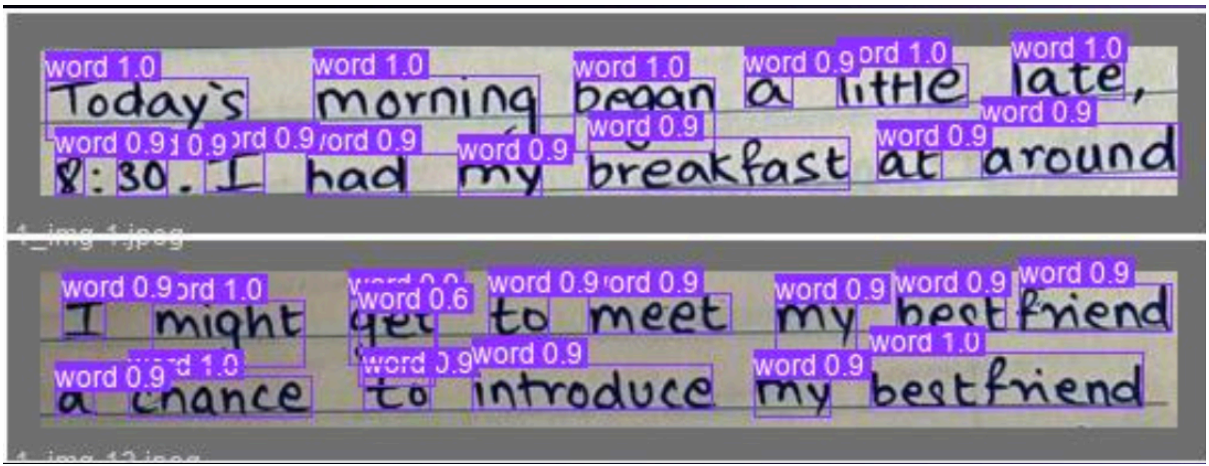


Figure 11: Word Extraction Output

YOLO generates a bounding box around the item and stores the object's height and breadth in addition to the object's x1, y1, and x2, y2 coordinates. YOLO required approximately 13 hours of training time until it was accurate 95 percent of the time for our dataset that was full of words. As a result, we received a really great confidence percentage.

When we were trying to extract words from a page, we encountered a lot of challenges. The most significant challenges that needed to be solved were the doppelganger, the sequence, and the strike problems. During YOLO testing, we noticed that it considered one-word multiple times with multiple bounding boxes overlapping it. This is the first challenged named as doppelganger. We solved this by calculating area of all bounding boxes given as an output by YOLO. It goes back for prediction to find the best suitable coordinates. It filters out the coordinates on the basis of area to find the optimum coordinate.

Figure 12: Example of a doppelganger

The next challenge we encountered was unpredictability of sequence. When we predict a bounding box, it saves all the coordinates of the same. The words are not predicted in order, and preserving the sequence is a crucial part of OCR. The basic solution we can think of for this problem is clustering and finding the clusters representing lines. But this solution has a flaw of random words being clustered together because their distance is less from the cluster but in reality the word is part of the next line. We first selected the origin, after tuning it, the centre point of the page was selected so that there is no directional bias. We computed the distance for each x1,y1 from the origin, using an equation(Eq. 1) we came up with to solve one issue, i.e., the value of x increases rapidly whereas the value of y increases gradually. Hence we needed a distance formula which helped in resolving this mathematically. The distances are then sorted to find the one with the shortest distance. This point is considered as a reference coordinate. We then iterate through the lines we get by multiplying the height of reference line with a bottom reference and top reference. Each iteration calculates the line passing the bottom and top. This line is considered as reference and the height of each x1,y1 is calculated from this line. These are either positive or negative on the basis of which we can decide where they lie. Until the last iteration we try to find the maximum number of boxes with a negative height. The points with negative height belong to line one. After this we can easily achieve the sequence of the words in the line by sorting x1 coordinates. We reset the origin and loop until all words are not deleted.

$$distance = \sqrt[2]{[\log(x2 - x1)]^2 + (y2 - y1)^2} \tag{4}$$

Eq 1: Distance Formula

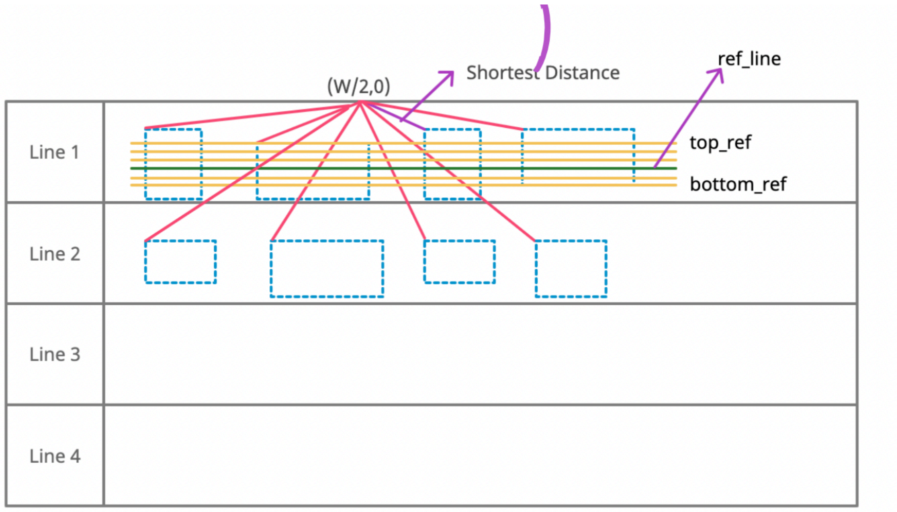


Figure 13: Diagrammatic representation of sequence algorithm

The modifications that were made to YOLO were carried out in such a way that they prevented the recognition of words that had been crossed out. However, under particular circumstances, including but not limited to instances in which it failed to detect the strikes in a word, a CNN model was built to eliminate the words that were struck out with a hundred percent degree of precision. Before settling on the architecture of the model and the parameters, we conducted experiments in which we combined hyper parameters in various ways. The combinations that produced the greatest results were then preserved, and we used those results to train our model. During the training phase, various authors' pages are combed through to find words that are then extracted and categorised as either normal or struck out. These samples were provided as input to our CNN – based binary classification model.

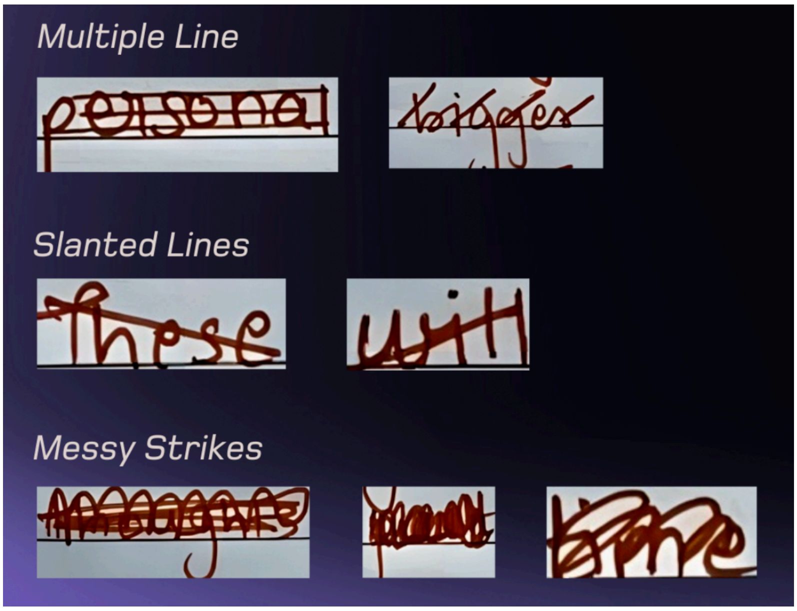


Figure 14: Types of struck out words

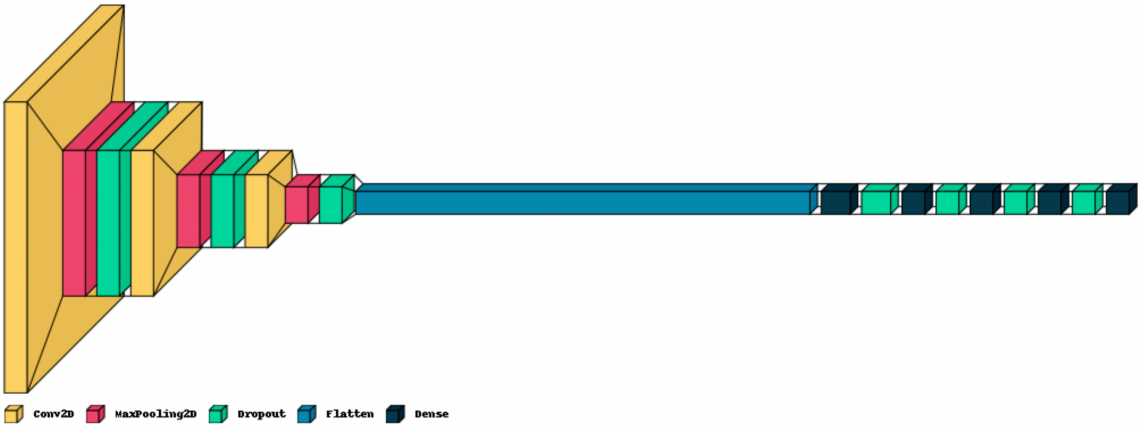


Figure 15: CNN Model employed for removal of strike words

7.2.4. Character Extraction

Because of the computational limits that were discussed earlier, we started developing our model for character extraction by employing the YOLOV5 medium. Even though we had strong recall and precision in an earlier epoch, we needed to iterate through 250 epochs before we could get stable confidence in our results. The risk of encountering a doppelganger was significantly reduced once the subject's confidence was established. The removal of characters that have been crossed out is likewise handled for those characters that are present in words. When extracting characters, our model is also responsible for removing the background from the page.

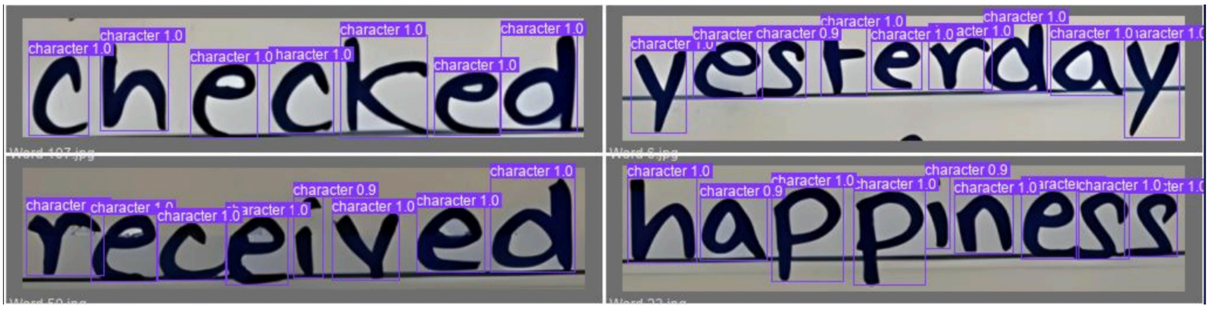


Figure 16: Characters extracted from words

7.2.5. Segmentation

Image segmentation is a method in which a digital image is broken down into several subgroups called Image segments. This helps in minimising the complexity of the image, which in turn makes it easier to perform further processing or analysis on the image. Image segmentation is a subfield of digital picture processing that concentrates on breaking up an image into its component pieces on the basis of the characteristics and qualities of those parts. The basic objective of image segmentation is to simplify the image so that it may be analysed more quickly and effectively. In image segmentation, you take an image and break it up into multiple pieces that each have their own unique characteristics.

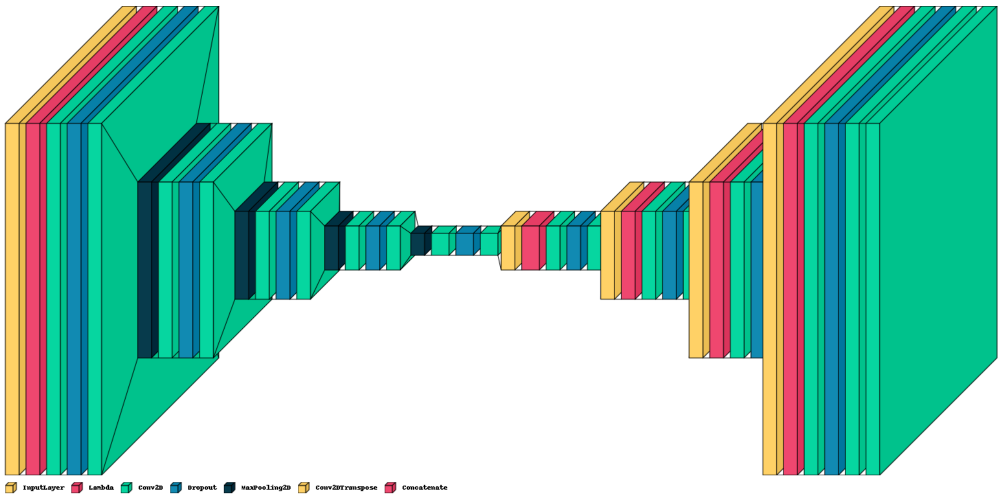


Figure 17: Residual U-Net

U-net is an architecture for neural networks that was primarily developed for the purpose of picture segmentation. The architecture of a U-net consists, at its most fundamental level, of two different ways of communication. The first path is the contraction path, also known as the encoder or the analysis path. This path, which gives classification information and is analogous to a standard convolution network, is also known as the analysis path. The second type of path is an expansion path, which is sometimes referred to as a decoder or a synthesis path. This type of road is made up of up-convolutions and concatenations with features from the contracting path. Because of this extension, the network is able to get locally relevant classification information. Additionally, the expansion path also boosts the resolution of the output, which can then be fed into a final convolutional layer to produce a fully segmented image. This is possible because of the way the expansion path works. The resulting network is nearly symmetrical, which gives it the appearance of being shaped like a u. The primary canonical job that the majority of convolutional networks are tasked with completing is to label the entire image with a single category.

Taking inspiration from U-net we designed a different model called ELIPNET. The model was designed for handling image segmentation and to improve the results than already existing models. We applied functional API to the model with lambda layers so that it will be able to handle variable input size. The output of the model is a grayscale image.

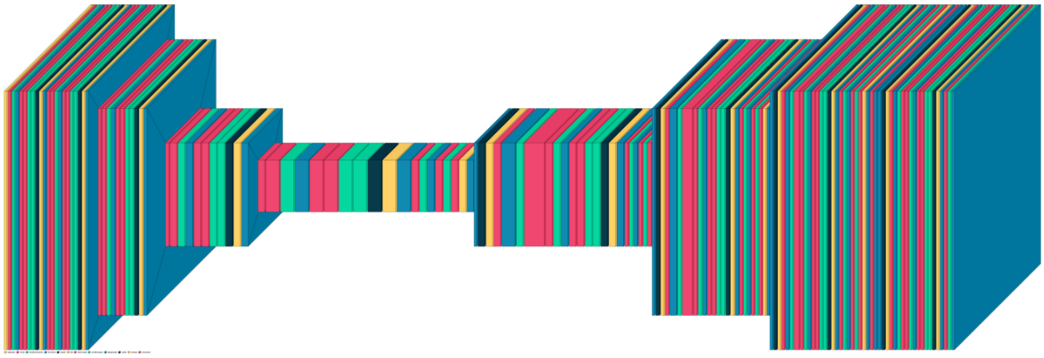


Figure 18: Attention Residual Elipnet

7.2.6. Classification

The subsequent step is the final link in the process chain, since it will produce an output of the utmost significance to an OCR model. The outputs of the segmentation model are included in the classification model as input. The ensemble technique was utilised so that we might improve the overall performance of our model. These models received two different kinds of input from us: the first kind was the output of segmentation, and the second kind was the cut-outs of characters that were produced by our YOLO model in the output of character extraction.

Before being used as an input in our ensemble models, these characters went through an initial processing stage first. An ensemble approach is essentially a collection of networks working together. After that, each model is utilised to generate a prediction, and the real prediction is calculated as the average of the predictions made by the several models.

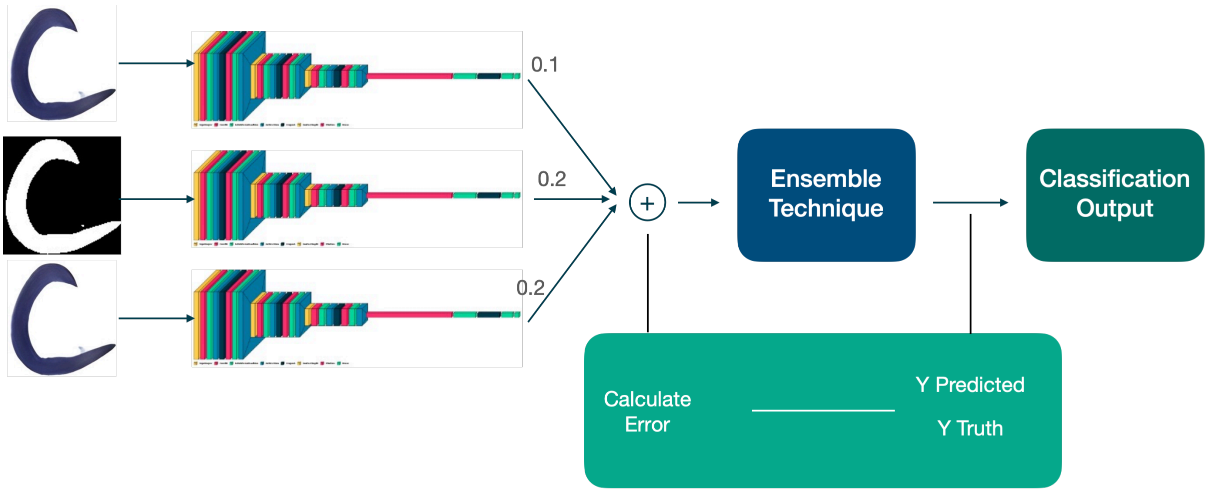


Figure 19: Classification Models using Ensemble Technique

We made use of a collection of three models, with two of the models receiving the output from character extraction, and one of the models receiving the output from segmentation. The fact that every model makes the same contribution to the overall forecast given by the ensemble is one of the drawbacks of using this methodology. Despite the fact that some models are known to perform much better or significantly worse than others, it is required that all members of the ensemble exhibit competence in comparison to the random chance that could occur. As a result, we made use of the weighted technique, in which each of the first models was given a weight of 0.1, and each of the two models that followed it each received a weight of 0.2. Because we have a limited amount of resources, we have created a model that recognises lowercase alphabets and gives them the label 26 for their class. The output of the classification model is given to the spell checker so that it may verify whether or not the output that was generated is correct, and if it isn't, it can make the necessary spelling corrections.

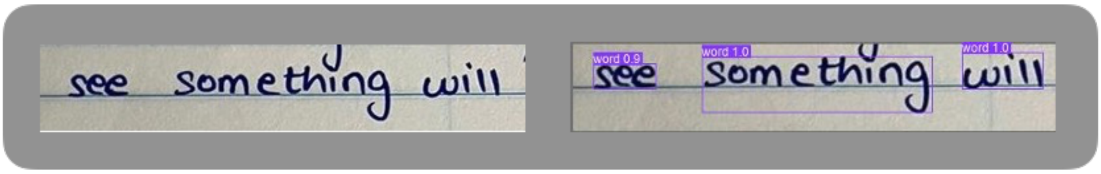
CHAPTER 8

PERFORMANCE EVALUATION AND TESTING

8.1. Testing

Test Case No.	Description	Input	Desired Output	Result Of test Case
1	Word Extraction	Page	Correctly detected words	Successful
2	Word Extraction	Page	Correctly detected words	Successful
3	Character Extraction	Word	Extracted characters	Successful
4	Character Extraction	Word	Correctly detected words	Successful

8.2. Testing Screenshots



(a) (b)

Figure 20: Test Case 1 (a) Input (b) Output

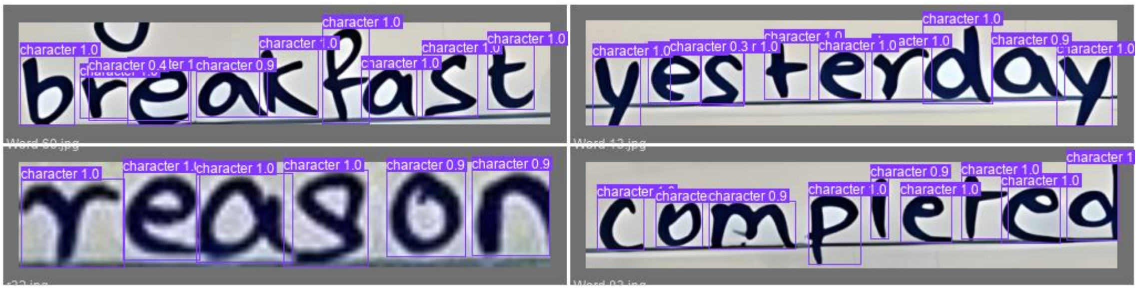


Figure 21: Testing for character extraction

CHAPTER 9

RESULTS AND ANALYSIS

Within our implementation we have multiple modules designed handle each and every part of the process important to make a robust OCR system to recognize different handwritten documents. These modules have further sub modules, trained and tested on dataset to gain an optimal model. With the initial step being pre-processing pages and enhancing the pages we got an efficient output. Our model using RealESRGAN was able to enhance the smaller resolution images successfully(Fig. 22).

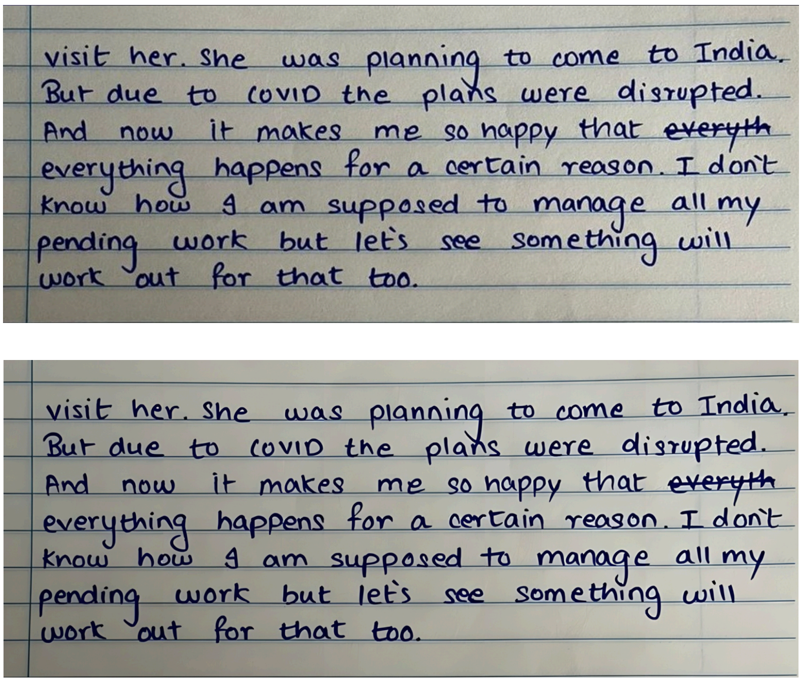
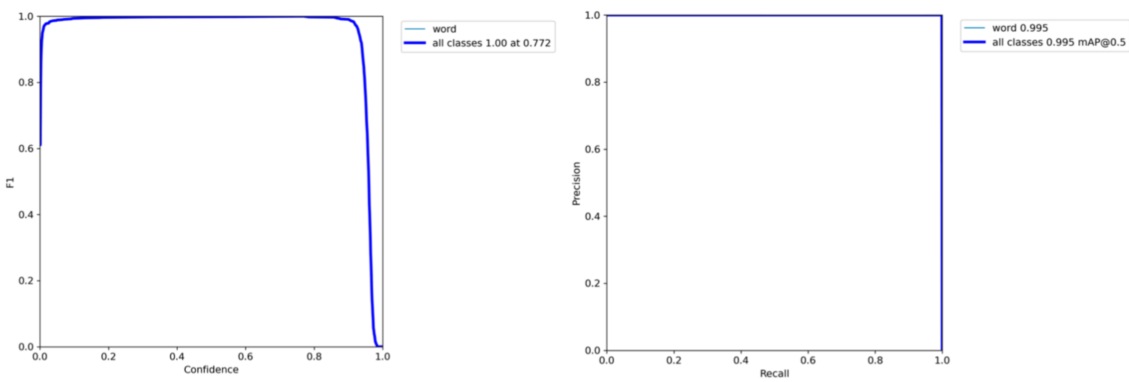
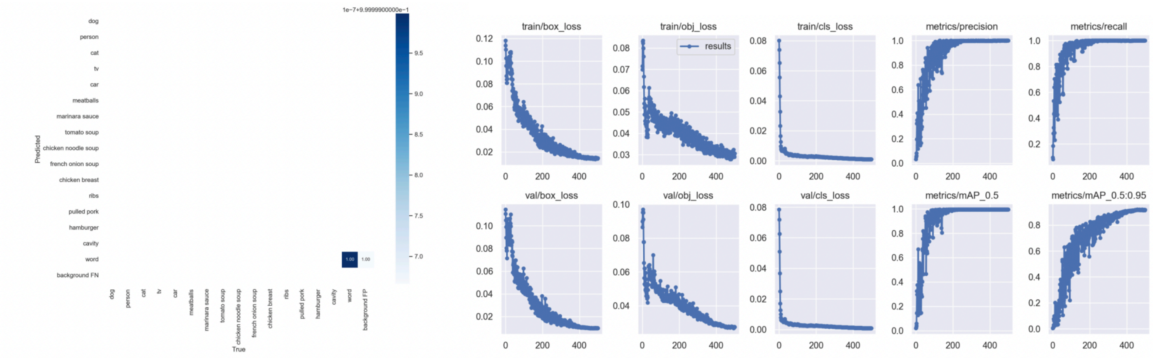


Figure 22: Page Enhancement Output



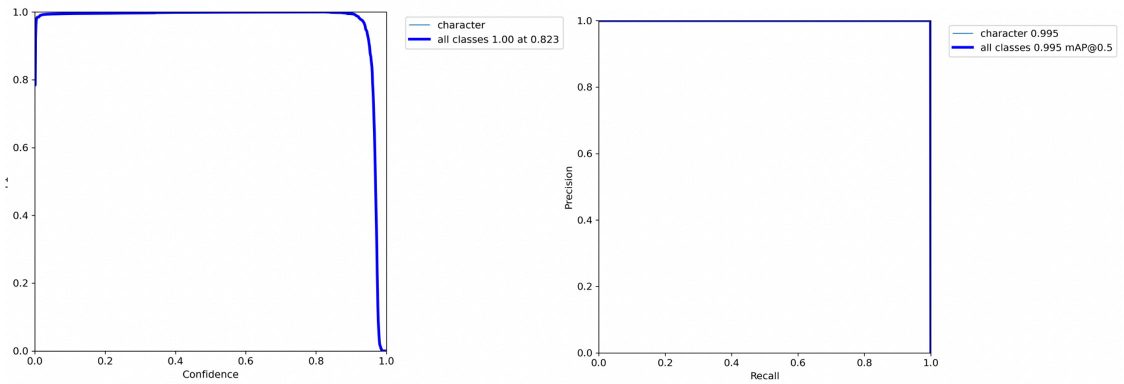
(a) (b)



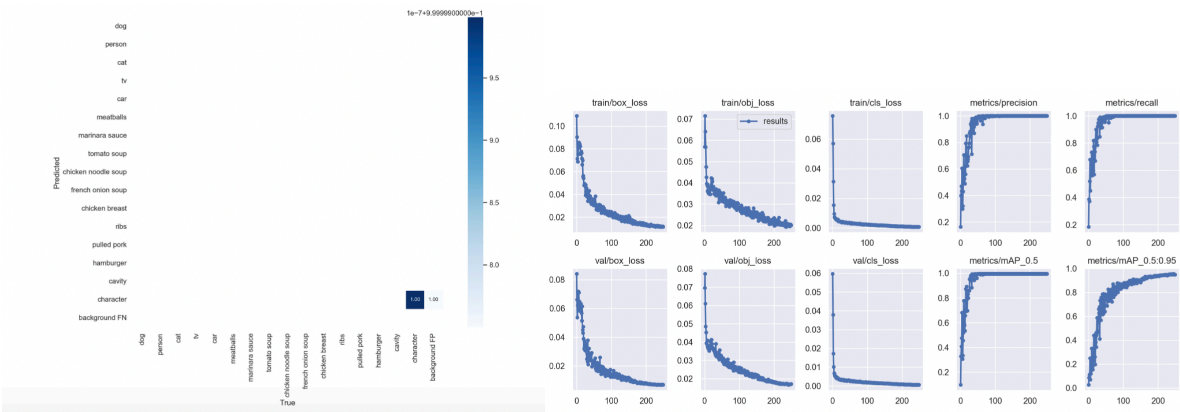
(c) (d)

Figure 23: Word Extraction Results.(a) F1-Curve, (b) PR-Curve, (c) Confusion Matrix, (d) Training/Validation results per epoch

Our text extraction model was based on YOLOV5m which takes care of localizing texts. Initially a YOLOV5m model is employed to extract words. The results for these YOLO models were very significant and after a lot of tuning and experimentation we got a good confidence for word and character as the object to be detected(Fig. 23,24).



(a) (b)



(c) (d)

Figure 24: Character Extraction Results.(a) F1-Curve, (b) PR-Curve, (c) Confusion Matrix, (d) Training/Validation results per epoch

Once we extract characters, these characters are sent to our segmentation and classification model to predict what these characters labels are. The output of process chain till the step of character extraction is as follows :

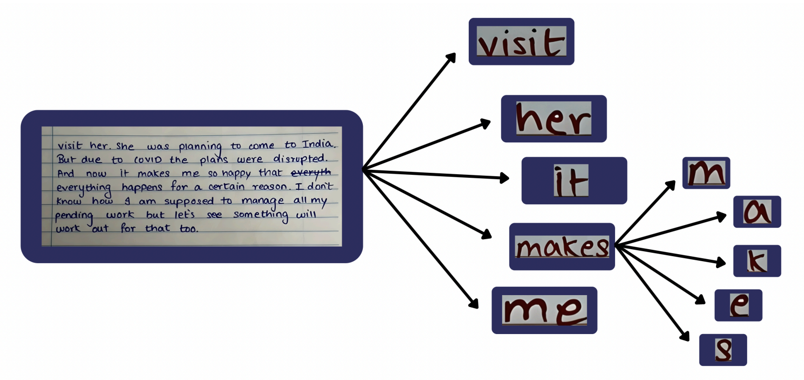


Figure 25: Process output till character extraction

Figure 26: Segmentation Output

With the help of ensemble technique we built a collection of three CNN models which collectively gave an accuracy of 98%

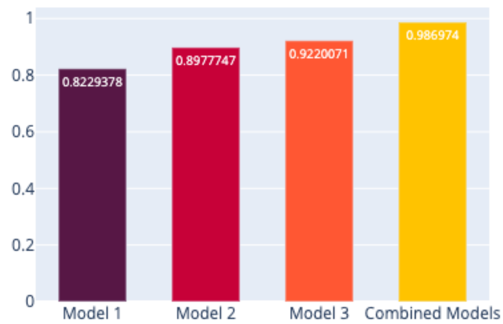


Figure 27: Classification accuracy using Ensemble Technique

CHAPTER 10

GUI Screenshots

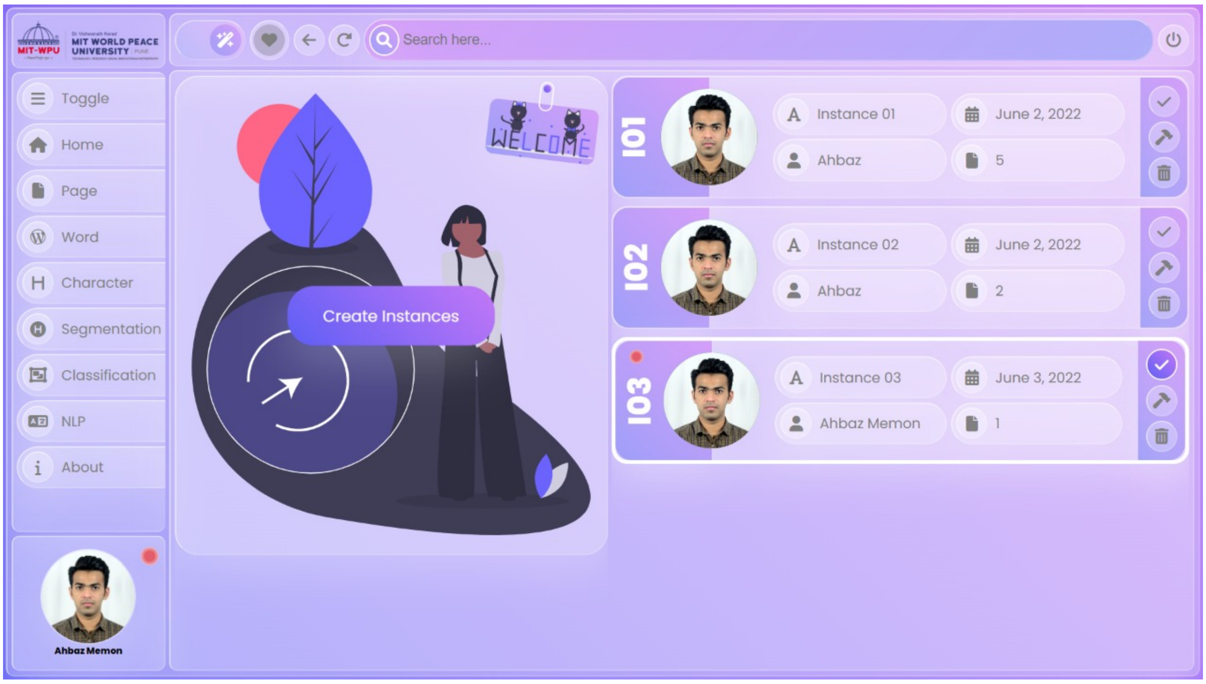


Figure 10: GUI Snapshot of the Landing Dashboard Page

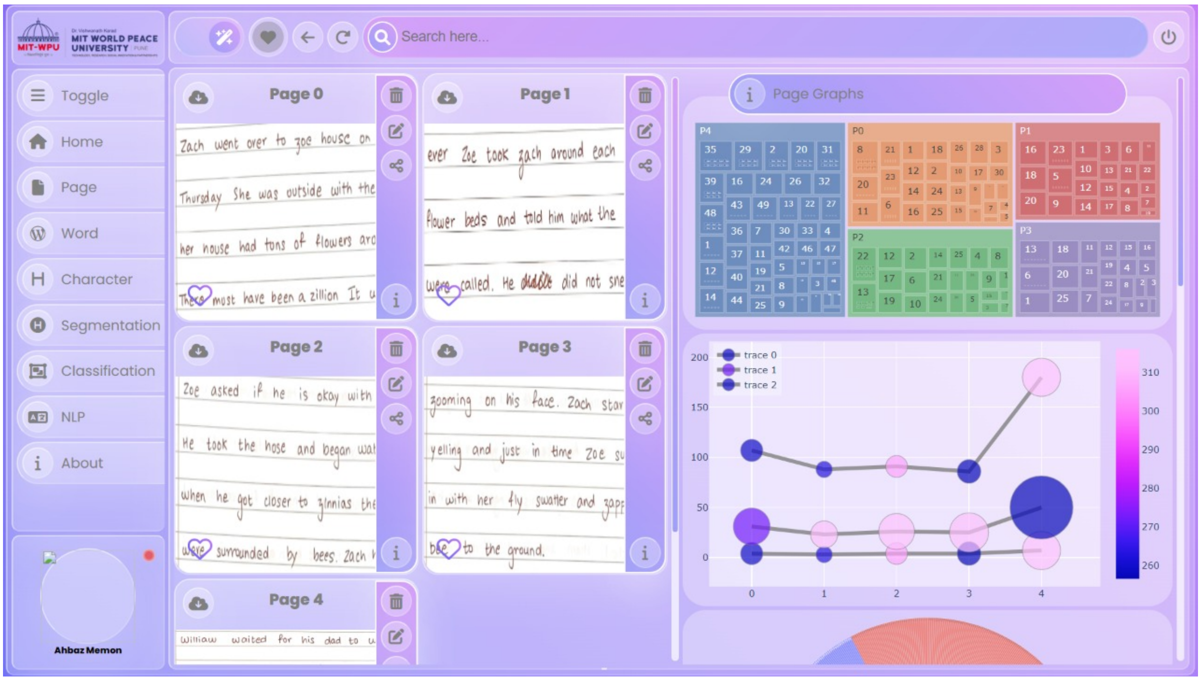


Figure 11: GUI Screenshot of Page Preprocessing

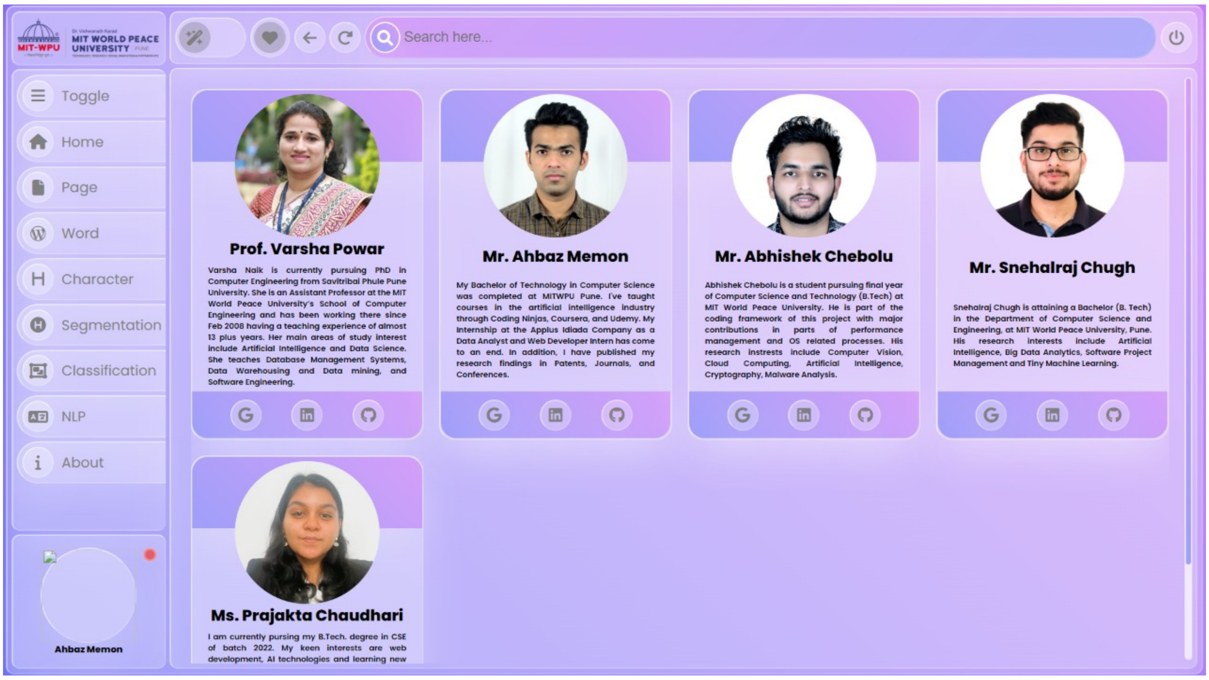


Figure 12: GUI Screenshot About Section

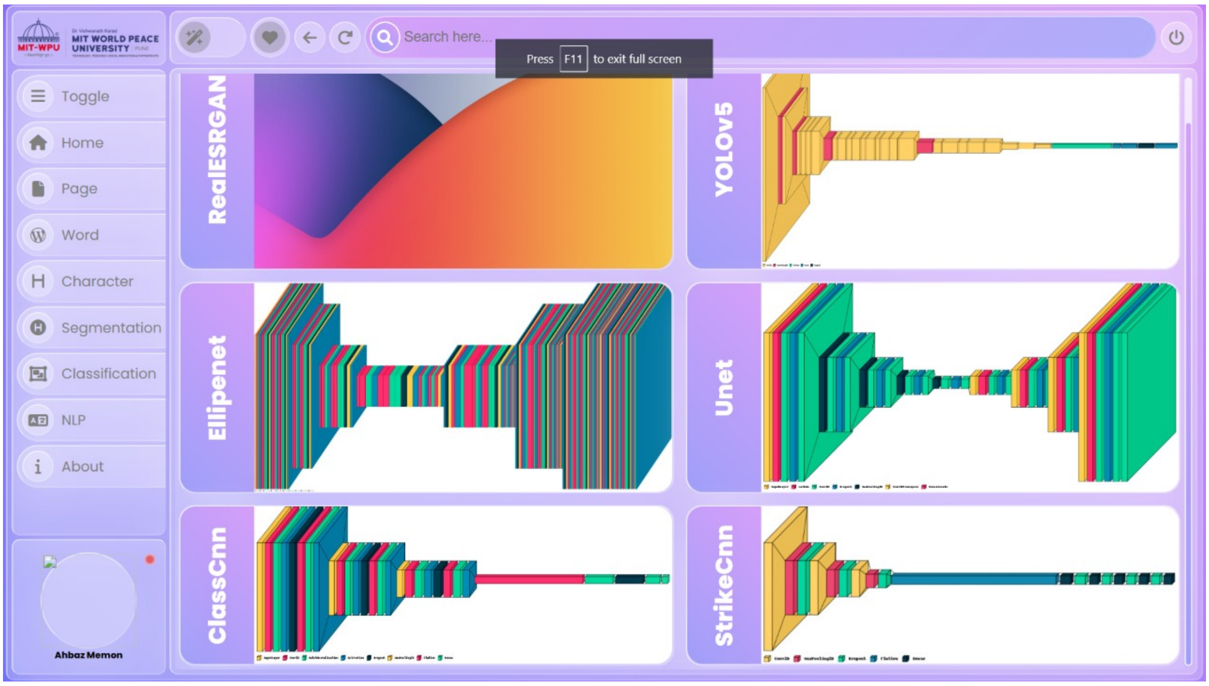


Figure 13: GUI Screenshot Models Section

CHAPTER 10
APPLICATIONS

The process of storing documents is experiencing a positive transition as a direct consequence of the increasing implementation of optical character recognition technology across a wide range of industries. When combined with this method, digitised texts can be transformed into documents that, in addition to merely being image files, also contain text information that machines can recognise and search for. This is an improvement over the previous method, which only used image files. This technology can decrease the need for individuals to manually input critical documents, which saves time and effort while still allowing for the retrieval of internet databases in their physical form. On the other hand, textual character recognition will automatically pull the required data and incorporate it into the document on its own initiative.

We are able to synthesise voice with the assistance of the text that can be digitised with the help of OCR. There are a great number of important applications for speech processing, and these applications can be found in a wide variety of settings. These devices aid people with impairments in overcoming a number of difficulties that are contextual in nature. People with cognitive difficulties, such as those associated with vision impairment, autism, or the fact that they are pre-literate or uneducated, amongst other things, may show to be useful. In addition, members of general public make extensive use of degrees of contact with speech outputs.

In circumstances in which the transcript that is generated is essential to the successful completion of the interaction and the capabilities of the technology, the dependability of the character recognition sector is absolutely essential. The vast majority of programmes receive great deal of praise when they have been made available on a portable device such as a phone. Because these devices have such low material requirements, the versions that will eventually be put into production will need to make greater use of the available resources .

Diverse comics include written typography; the use of synthesising enables the look to be retained even when the comic is translated into a different language. Furthermore, the penned writing lends itself to a variety of imaginative applications, such as customised eBooks. The emphasis of the programme would be on gaining a knowledge of the person's feelings and ideas at the moment the notes were made, and this would be a vital feature. If the writer wrote in a hurry, the angle of his or her strokes may indicate that they were writing more quickly than usual.

CHAPTER 12
CONCLUSION

Our study has examined and analysed many approaches for extracting textual content from scene photos, with the goal of improving their effectiveness. We've gone through the fundamentals of character recognition using photos. In which we examined several image processing approaches in a specific sequence for character recognition using digitised images, as well as the deployment of the textual recognition system, we also covered the use of text recognition software.

Handwritten typographic components were identified with high accuracy and efficiency using our in-house letter recognition technology; scribbled numerals will be identified with high speed and accuracy using CNN, which will be utilised to provide accurate output.

When we trained the networks, we utilised a different dataset, and when we achieved the appropriate accuracy after retraining, we used various datasets to evaluate the learned model's ability to convert data into letters.

In contrast to traditional and manual methods, our invention will assist authors in automating the process of creating letters in their own style using machines in a quicker and simpler manner, resulting in time and labour savings for them.

In the future, the suggested study might be applied to a variety of other language-producing techniques with a high degree of accuracy and precision.

CHAPTER 11
FUTURE SCOPE

The output of our model is a stream of handwritten characters, words, and phrases that are generated from a written page. This output can be further analysed to generate useful insights on the personality and character qualities of an individual. Character strokes can reveal information about a person's patterns, personality traits, and other aspects of their life. Graphology is the name given to the academic discipline that examines the individual strokes that make up a person's handwriting. The next step in the development of our model could involve the detection of various types of strokes within the characters in order to gain a deeper understanding of an individual. This could also assist us in providing answers to questions like what the person was feeling at the time they were writing the handwritten documents.

In light of the fact that we are planning a research on sentiment analysis, the digitised text that we produce as our output can also serve as an input to NLP models, which can provide more information regarding the author's emotions. Employing speech synthesis based on the generated text is one method that can be used in conjunction with NLP models to broaden the application of our model. Text-to-speech conversion modules can help to broaden the user base of optical character recognition (OCR) by making it possible for visually impaired people to benefit from these systems.

The topic of software applications that can produce images of handwritten text automatically is currently receiving a lot of attention. The models that are employed for this purpose make an effort to emulate human handwriting in order to achieve the desired effect of making the text appear more human-like as opposed to computer-generated. It does it by utilising a dataset that is already in existence and which comprises the names of characters as well as the words. The classification result can be used by our model to generate text in the writer's handwriting. This text would be machine generated, but it would be essentially the same as the writer's handwriting

REFERENCES

1	Revathi, A. S., & Modi, N. A. (2021, March). **Comparative Analysis of Text Extraction from Color Images using Tesseract and OpenCV.** In *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*(pp. 931-936). IEEE.
2	Ajmal, M., Ahmad, F., Martinez-Enriquez, A. M., Naseer, M., Muhammad, A., & Ashraf, M. (2018, December). **Image to Multilingual Text Conversion for Literacy Education.** In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 1328-1332). IEEE.
3	Palekar, R. R., Parab, S. U., Parikh, D. P., & Kamble, V. N. (2017, April). **Real time license plate detection using openCV and tesseract.** In *2017 international conference on communication and signal processing (ICCSP)* (pp. 2111-2115). IEEE.
4	Kalyani, D., & Vijay Kumar, P. (2021). **Text-Based Handwritten Recognition Through an Image Using Recurrent Neural Network.** In *ICT Analysis and Applications* (pp. 475-486). Springer, Singapore.
5	Narayan, A., & Muthalagu, R. (2021, March). **Image Character Recognition using Convolutional Neural Networks.** In *2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII)* (pp. 1-5). IEEE.
6	Dome, S., & Sathe, A. P. (2021, March). **Optical charater recognition using tesseract and classification.** In *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)* (pp. 153-158). IEEE.
7	Parthiban, R., Ezhilarasi, R., & Saravanan, D. (2020, July). **Optical character recognition for English handwritten text using recurrent neural network.** In *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*(pp. 1-5). IEEE.
8	Ansari, G. J., Shah, J. H., Yasmin, M., Sharif, M., & Fernandes, S. L. (2018). **A novel machine learning approach for scene text extraction.** *Future Generation Computer Systems*, *87*, 328-340.
9	Sandyal, K. S., & Kiran, Y. C. (2019, September). **Analysis on Preprocessing Techniques for Offline Handwritten Recognition.** In *International Conference on Intelligent Data Communication Technologies and Internet of Things* (pp. 546-553). Springer, Cham.
10	Latha Kalyampudi, P. S., Srinivasa Rao, P., & Swapna, D. (2019, January). **An efficient digit recognition system with an improved preprocessing technique.** In *International Conference on Intelligent Computing and Communication Technologies* (pp. 312-321). Springer, Singapore.
11	

	Shah, M., Mehta, S., Mody, P., Roy, A. S., & Khachane, S. P. (2018). **Handwriting Recognition of Diverse Languages.** *International Journal of Computer Science and Mobile Computing*, *7*(4), 109-114.
12	Chaudhari, K., & Thakkar, A. (2019). **Survey on handwriting-based personality trait identification.** *Expert Systems with Applications*, *124*, 282-308.
13	Zheng, F., Luo, S., Song, K., Yan, C. W., & Wang, M. C. (2018). **Improved lane line detection algorithm based on Hough transform.** *Pattern Recognition and Image Analysis*, *28*(2), 254-260.
14	Wang, Z., Yang, D., & Tong, Q. (2020, February). **A Method of Workpiece Coherent Line Detection Based on Progressive Probabilistic Hough Transform.** In *Proceedings of the 2020 9th International Conference on Software and Computer Applications* (pp. 141-146).
15	Haifeng, D., & Siqi, H. (2020, September). **Natural scene text detection based on YOLO V2 network model.** In *Journal of Physics: Conference Series* (Vol. 1634, No. 1, p. 012013). IOP Publishing.

PLAGIARISM REPORT

Dupli Checker

PLAGIARISM SCAN REPORT

0% Plagiarised	100% Unique	Date	2022-02-13
		Words	998
		Characters	6299

Content Checked For Plagiarism

We want to develop a personal module to assist and better analyse an individual through their handwriting using artificial intelligence, than what has been done so far which is quite generic and limiting. We are combining four domains, .i.e, Deep Learning, Computer Vision, NLP and Psychology for our research.

* Our focus is to invent new methods for letter segmentations in a sentence
* Creation of customized personal database for every different handwriting (classified entity)
* Psychological Analysis of an individual upon the style of writing (stores, size, angle of letters)
* A module will be integrated to help deaf and blind people understand emotions and the feeling behind the hand-written message at the time it was written.
* The machine can understand the cancelled-out/scribbled words by itself as to analyse the page more human-like.
* We will be particularly analysing the finding for every sentence which are:
* Analysing the person's nature/personality
* Understanding emotions at the time of writing
* Sentiment analysis of sentence

The number of ways individuals can take photos using modern technology has increased dramatically in the contemporary age. It is conceivable that these photos include the essential written type of data that have to be edited or preserved electronically.

It's indeed possible to browse for and identify information in photographs, as well as papers, with aid of said Handwriting Analysis Architecture. Using a computer application, a written message is produced by interpreting the letters of an image or digitized document.

Recurrent learned algorithms and faster processors have made it possible for a much more comprehensive writing identification system that really can translate photos of a few segmental letters or indeed a phrase.

The method has a wide range of potential uses. Capabilities include analyzing financial cheques, identifying information from cards, assisting the sighted with identifying information on writings, as well as portraying emotions associated with the research process. For the moment, our program solely recognises text in English. Other languages could be implemented in the long term.

Dupli Checker

PLAGIARISM SCAN REPORT

0% Plagiarised	100% Unique	Date	2022-02-13
		Words	836
		Characters	5572

Content Checked For Plagiarism

2
Automation within Handwriting Analysis

Application of project

Optical character recognition technology is being employed in a variety of sectors and is transforming the document storage process in a positive way. When used in conjunction with this technique, digitised texts can be transformed into docs that have text information that machines can identify and hunt for rather than just picture files. While retrieving online databases physically, this technology can reduce the need for people to input critical documents, which saves effort and time. Conversely, textual character recognition pulls the necessary data and integrates it into the document on its own initiative. The accuracy of the result achieved in this manner is high, as is the effective waiting period, which is much longer than one minute.

There are many major uses for speech processing, and these uses are found in a broad range of situations. Such technologies assist impaired people in overcoming a variety of contextual obstacles. People who have cognitive challenges owing to vision impairment, autism, or who are pre-literate or uneducated, among other factors, may prove to be valuable. Degrees of interaction with speech outputs are also widely used by members of the general public in their daily lives.

The reliability of the character recognition sector is vital in situations where the transcript generated is critical to the finished interaction and the competence of the technology. Most programmes are greatly admired when they are made accessible on a portable device such as a phone. These gadgets feature minimal material requirements, and as a result, the models that will be implemented must be more resource-intensive as a result. It is more pleasant to use a hand gesture recognition system that requires less convergence speed, particularly when the entire programme includes handwritten recognition software as a subsystem with such a limited resource architecture.

2

Dupli Checker

PLAGIARISM SCAN REPORT

0% Plagiarised	100% Unique	Date	2022-02-13
		Words	996
		Characters	7198

Content Checked For Plagiarism

2
Automation within Handwriting Analysis

3.1. Project Assumptions

With this tool, past studies' flaws, such as their inflexible mathematical formalism and contradictory theoretical premises when coping with inaccurate inputs, are no longer present. Despite the sophisticated mathematical analysis, the most common objections may be divided into two categories:

1) Constraints imposed by the computational mathematics assumption that were chosen. Inside the theory, a few of these hypotheses is the mathematical distributions of the condition lifetime, which is one of the hypotheses. As a consequence, several fundamental characteristics of the architecture of communication are not well represented in the systems that are now accessible.

2) The rigid numerical framework, which really is incompatible with the imperfect quality of writing and spoken data.

The reality that our approach is built solely on information concerning pen motions results in the limitation of pen-hand paradigm that they have created. Furthermore, the theories are predicated on the notion of a fixed location on the printed sheet, upon which the hand-pen mechanism is supposed to rotate as a finite string of letters is created by the pens. Clearly, this hypothesis might be validated by research that makes use of the 3D motion monitoring system.

We feel that if characteristics of the real personal notes and the feelings of the author are accounted for, a much more realistic version than that which has been described in several studies might be constructed. In this regard, the strategy that will be discussed in this section may quickly prove to be beneficial.

For the most part, 3D motion monitoring will be valuable in graphemic studies in either evaluation studies where hypotheses about the link between physiological major axes in calligraphy and aspects of handwritten imprints need to be tested or handwriting imprints themselves.

Dupli Checker

PLAGIARISM SCAN REPORT

0% Plagiarised	100% Unique	Date	2022-02-13
		Words	217
		Characters	1601

Content Checked For Plagiarism

2
Automation within Handwriting Analysis

CHAPTER 7

CONCLUSION

Our study has examined and analysed many approaches for extracting textual content from scene photos, with the goal of improving their effectiveness. We've gone through the fundamentals of character recognition using photos. In which we examined several image processing approaches in a specific sequence for character recognition using digitised images, as well as the deployment of the textual recognition system, we also covered the use of text recognition software.

Handwritten typographic components were identified with high accuracy and efficiency using our in-house letter recognition technology; scribbled numerals will be identified with high speed and accuracy using CNN, which will be utilised to provide accurate output.

When we trained the networks, we utilised a different dataset, and when we achieved the appropriate accuracy after retraining, we used various datasets to evaluate the learned model's ability to convert data into letters.

In contrast to traditional and manual methods, our invention will assist authors in automating the process of creating letters in their own style using machines in a quicker and simpler manner, resulting in time and labour savings for them.

In the future, the suggested study might be applied to a variety of other language-producing techniques with a high degree of accuracy and precision.