Bangla Braille Character Recognition using Convolutional Neural Network

BY

Fahim Imtiaz
ID: 202-15-14447

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Naznin Sultana
Associate Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled “Bangla Braille Character Recognition using Convolutional Neural Network”, submitted by Fahim Imtiaz (202-15-14447) to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 13-07-2024.

BOARD OF EXAMINERS

Dr. Md. Taimur Ahad (MTA) — **Board Chairman**
Associate Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Fizar Ahmed (FZA) — **Internal Examiner 1**
Associate Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Rahmatul Kabir Rasel Sarker (RKR) — **Internal Examiner 2**
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Sadat Hasan — **External Examiner**
Data Scientist (Senior Principal Officer)
Risk Management Division
BRAC Bank

DECLARATION

I hereby declare that this project has been done by us under the supervision of **Naznin Sultana (NS), Associate Professor, Department of Computer Science and Engineering, Daffodil International University**. I also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Naznin Sultana
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Md. Sadekur Rahman
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Fahim Imtiaz
ID: 202-15-14447
Department of CSE
Daffodil International University

# ACKNOWLEDGEMENT

First, I express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project/internship successfully.

I am grateful and wish our profound indebtedness to **Naznin Sultana (NZ), Associate Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Deep Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

I would like to express my heartiest gratitude to the **Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

I would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, I must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Through the creation of a Convolutional Neural Network (CNN)-based model that is inspired by the YOLO architecture, the purpose of this thesis is to find a solution to the substantial difficulty of Braille character recognition. The purpose of this research is to prioritize accessibility for those who are visually challenged and rely on Braille for communication and education. Specifically, the research is designed to transform Bangla letters from Braille patterns. For the purpose of ensuring that the training and testing stages are robust, a dataset that was painstakingly selected and specially designed for this particular task was painstakingly collected and annotated. However, the CNN model had difficulty reliably identifying contextual characters within sentences or phrases, despite the fact that it was able to recognize individual Bangla characters within Braille patterns with exceptional competence. The study highlights the important need to improve the accuracy and durability of such models in order to attain practical utility in applications that are based in the real world. For the purpose of model training, the system makes use of Google Colab Pro. Additionally, it makes use of advanced GPU capabilities and makes use of TensorFlow and Keras libraries for efficient implementation. In the future, efforts will be focused on refining the model in order to increase contextual character identification. The ultimate objective is to broaden the range of educational options available to visually impaired people and to improve the quality of life for those individuals who rely on Braille as an essential form of communication and literacy support. By tackling these problems, our research makes a contribution to the advancement of accessible technology and aids the inclusion of visually impaired individuals in educational settings and in day-to-day life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ML	Machine Learning
DL	Deep Learning
CNN	Convolutional Neural Network
Frcnn	Faster -R-CNN

CHAPTER 1

INTRODUCTION

1.1 Overview

Bangla is the fifth most widely spoken language globally and the second most popular language on the Indian subcontinent, with approximately 220 million speakers in the region and 300 million worldwide. [1] [2,3]. It is a language with profound historical importance. UNESCO has designated February 21st as International Mother Language Day to honor the martyrs who sacrificed their lives during the 1952 language campaign. Bangla is the native language of the people of Bangladesh, belonging to the broader Indo-European Language Family. Yet, the visually challenged struggle to read and write, hindering their access to education and knowledge. In this scenario, listening is the sole effective method of acquiring knowledge, but it may not be as gratifying as reading directly from a book. In 1824, Louis Braille developed the Braille system for reading and writing, specifically designed for individuals with visual impairments. Louis Braille discovered Braille in 1824, which was the initial method of reading and writing for individuals with visual impairments.[4] It is considered the earliest type of binary writing. Visually challenged individuals use their fingertips to tactually see patterns or characters. The Braille system consisted of a rectangular arrangement of two columns and three rows of six dots. By raising a dot at one location, sixty-four permutations can be generated. The left side of the braille is universally numbered 1-3, whereas the right side is designated 4-6. Automatic handwritten character recognition has gained popularity in academic and commercial fields and is currently a subject of extensive research. Exploration of areas like Deep Learning and Machine Learning has resulted in the creation of numerous algorithms that are currently highly proficient at identifying handwritten characters. Zahangir et al [3] highlighted that a significant obstacle in character recognition is the considerable variation in shape and size of handwriting among individuals for each language. Moreover, the characters might vary between isolated and cursive forms based on the language, which complicates the process of character recognition. Regarding Bangla script. The system has 50 fundamental characters, 24 compound characters, and 10 unique numbers. Additionally, some characters have indications or marks at the top or bottom. Some characters exhibit similarities and distinctions as subtle as a dot or a line. Consequently, it hinders the development of Bangla character recognition by making it challenging to obtain consistent performance.

1.2 Background and Present State

Bangla script presents unique challenges due to its complexity. It includes 50 basic characters, 24 compound characters, and 10 distinct digits, with various marks and subtle differences that make character recognition difficult. These intricacies have resulted in fewer efforts towards Bangla character recognition compared to other languages, further disadvantaging visually impaired individuals in Bangladesh. As technology advances, there is a significant opportunity to create systems that integrate Bangla character recognition with Braille. Leveraging modern deep learning and machine learning techniques, a robust system can be developed to accurately recognize Bangla characters from Braille patterns. This initiative aims to bridge the educational gap for visually impaired individuals in Bangladesh, enabling them to access quality education and knowledge through a more interactive and satisfying medium.

1.3 Problem Statement

Despite Bangla being the fifth most spoken language globally and the second most popular language in the Indian subcontinent, there exists a significant educational barrier for the visually impaired population due to difficulties in reading and writing. While listening serves as an alternative, it fails to provide the same level of satisfaction and efficiency as reading. This impediment underscores the urgent need for accessible education solutions tailored to the visually impaired in Bangladesh. The introduction of Braille by Louis Braille in 1824 revolutionized access to written knowledge for the visually impaired. However, advancements in character recognition technologies have been largely focused on languages like English, leaving Bangla character recognition underdeveloped. Bangla's intricate script comprising 50 basic characters, 24 compound characters, and 10 distinct digits, coupled with variations in shape, size, and the presence of marks, pose significant challenges for reliable character recognition systems. Existing efforts in Bangla character recognition are limited compared to those in English or other languages. Consequently, many visually impaired individuals, especially in Bangladesh, struggle to access education due to the lack of suitable learning materials. As technology continues to advance, there's an opportunity to leverage modern computing capabilities to address this issue effectively. Hence, the problem at hand is the development of a robust system that seamlessly integrates Bangla character recognition with Braille, catering to the specific needs of the visually impaired population. By harnessing advancements in deep learning and machine learning, coupled with the power of modern computing, this system aims to bridge the educational gap and empower visually impaired individuals in Bangladesh to access quality education.

1.4 Objectives

Extensive study has been conducted on character recognition and significant advancements have been made in the development of braille since its inception. Given the widespread use of the Bangla language regionally and internationally, it is crucial to create systems that can reliably identify Bangla characters. Our research indicates that there have been some contributions, however they are rather tiny in comparison to the quantity of efforts made towards recognizing English characters or other languages. However, the school system remains inaccessible to many handicapped individuals, particularly the visually impaired who are unable to read and hence cannot receive a proper education. As the primary language of Bangladesh, Bangla is an essential component of education. Technology is becoming influential in all sectors of the country as we progress into the future. Furthermore, the rise has led to the advancement of far more potent computers with enhanced processing capacity. We have developed a prototype that integrates character recognition and braille using current technology to improve the school system and provide educational opportunities for the visually impaired.

1.5 Scope and Limitations

The project aims to develop a comprehensive system integrating Bangla character recognition with Braille technology, tailored to facilitate for the visually impaired population in Bangladesh. The system will employ advanced algorithms, drawing from deep learning and machine learning methodologies, to accurately recognize Bangla characters. These recognized characters will be seamlessly converted from Braille patterns, ensuring compatibility with existing Braille reading systems utilized by visually impaired individuals. A user-friendly interface will be designed to enable efficient interaction with the system, providing intuitive input methods such as scanning text or typing, and delivering output through audio feedback or Braille display. Throughout the project, considerations will be given to accessibility standards, cultural sensitivities, and resource limitations, with the ultimate goal of empowering visually impaired individuals in Bangladesh to access quality education and information effectively.

1.6 Report Organization

1.1 Overview: The introduction sets the stage for the project by providing a comprehensive summary of its objectives and significance. It emphasizes the importance of detecting Braille recognition and the necessity for sophisticated approaches to address this pressing healthcare concern.

1.2 Background and Present State: The project scope delineates the boundaries and parameters within which the project will operate. It defines the specific aspects of Braille recognition detection that will be addressed and outlines.

1.3 Problem Statement: This section articulates the specific challenges associated with Braille recognition detection, highlighting the need for improved methods to facilitate early detection and treatment. It addresses the shortcomings of existing approaches and identifies areas where advancements are required.

1.3 Motivation and Objectives: Here, the driving force behind the project is elucidated, emphasizing the widespread occurrence and significant consequences of Braille recognition. The section outlines the precise goals and ambitions of the project, including the development of deep learning models for detecting Braille recognition.

1.5 Project Outcome: This section describes the anticipated outcomes of the project, including the creation of advanced models for Braille recognition detection and their potential impact on healthcare outcomes. It sets expectations for the project's deliverables and contributions to the field.

1.6 Report Organization: The organizational structure of the report is outlined here, providing a roadmap for readers to navigate through the various chapters and sections. It ensures clarity and coherence in presenting the project's findings and insights.

1.7 Summary: The summary section provides a concise recap of the key points covered in the introduction, highlighting the importance of Braille recognition detection and the objectives of the project. It serves as a bridge to the subsequent chapters, guiding readers into the deeper exploration of the topic.

2.1 Overview: This section offers an overview of the literature review chapter, outlining its purpose and scope. It sets the stage for a comprehensive examination of existing research and insights into Braille recognition detection.

2.2 Related Works: Here, the existing body of research on Braille recognition and detection methods is reviewed, providing context for the project and identifying gaps and areas for further investigation.

2.3 Comparison between existing works: This section compares and contrasts different approaches and methodologies employed in existing research on Braille detection. It identifies strengths, weaknesses, and opportunities for improvement in current practices.

3.1 Overview: The methodology and requirements analysis chapter is introduced, highlighting its importance in guiding the design and implementation of the project. It provides an overview of the methods and approaches that will be used.

3.2 Requirement Analysis: This section outlines the specific requirements and objectives for the project, including the criteria for evaluating the effectiveness of the proposed Braille recognition detection models.

3.3 Proposed Methodology/System Design: Here, the proposed methodology and system design for detecting Braille recognition are presented in detail, including the algorithms, techniques, and technologies that will be employed.

3.4 Data Collection/Input Output Analysis: The process of data collection and analysis is described, outlining the sources of data, data preprocessing techniques, and methods for evaluating model performance.

3.5 Project Management and Financial Analysis: This section provides insights into project management strategies, including budgeting, scheduling, and resource allocation. It ensures the effective execution of the project within specified constraints.

3.6 Summary: The summary section offers a concise recap of the methodology and requirement analysis chapter, highlighting key findings and insights that will inform the subsequent stages of the project.

Chapter 4: Implementation

This chapter presents the experimental results obtained from the application of object detection in braille character. Detailed analyses of the performance of different detection approaches are provided, along with visual representations and statistical measures to validate the outcomes.

Chapter 5: Result and Analysis

In this chapter we will discuss about the result and evaluation of the model and analysis.

Chapter 6: Impact on Society, environment and sustainability

The societal impact chapter explores the broader implications of the research findings on medical field. It discusses how improved braille character recognition methods can positively influence field and economic outcomes. Consideration is given to potential advancements in technology and their societal implications.

Chapter 6: Summary, Conclusion, Recommendation, and Implications for Future Research
This final chapter serves as a synthesis of the entire report. It summarizes the key findings, reiterates the conclusions drawn from the research, and offers recommendations for practical applications and further studies. The implications for future research are discussed, providing a roadmap for researchers interested in expanding upon the current study.

1.7 Summary

The project aims to address the significant educational barrier faced by the visually impaired population in Bangladesh, particularly in accessing Bangla language materials. Despite Bangla being the fifth most spoken language globally, there is a lack of reliable character recognition systems for the visually impaired. The introduction of Braille revolutionized access to written knowledge, but advancements in character recognition technologies have primarily focused on languages like English, leaving Bangla character recognition underdeveloped. The project's motivation stems from the necessity to provide accessible education solutions tailored to the visually impaired in Bangladesh. By seamlessly integrating Bangla character recognition with Braille technology, the project seeks to empower visually impaired individuals to access quality education and information effectively. The system's scope encompasses the development of advanced algorithms to accurately recognize Bangla characters from Braille patterns, with considerations given to accessibility standards and cultural sensitivities. The ultimate outcome of the project is to bridge educational barriers and provide inclusive access to education and information in the Bangla language for visually impaired individuals in Bangladesh.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

The primary method for optical braille recognition involves a step-by-step process of searching for Braille points, restoring the characters grid while compensating for image rotation, combining points into characters, and decoding the characters. The World Health Organization (WHO) approximates that 38 million individuals worldwide are blind, while 217 million have moderate to severe vision impairment. People who are blind or have vision impairments cannot read printed text. Alternatively, some individuals utilize braille, a tactile system of raised dots that may be interpreted through touch. Braille symbols are created within braille cells, which are sections that can contain up to six raised dots. The six raised dots are numbered one to six and organized in two columns. Less than 10 percent of the 1.3 million legally blind individuals in the United States utilize braille for reading, as reported by the National Federation of the Blind [7]. These figures are based on the observation that certain visually impaired individuals see learning and using braille as time-consuming and burdensome. Braille literacy is decreasing as more individuals are opting for technology and audio recordings instead of braille. Assistive technologies and audio recordings are not always adequate for all situations faced by individuals with vision impairments. Individuals who are visually challenged may need assistance from others to decipher braille. Furthermore, persons who assist the visually impaired and lack knowledge of braille may need resources for effective two-way communication with braille due to its intricate nature.

2.2 Related Works

Thresholding is the most basic method for point detection. Zhang and Yoshino [10] employ a dynamic local threshold. Antonacopoulos and Bridson [11] detect dots on double-sided braille by analyzing brightness and dark sections of dots and segmenting the image into three classes: bright, dark, and background. They utilize fixed thresholds in comparison to the average light level at the site. Renqiang Li and colleagues utilized a Haar detector and Support Vector Machine (SVM) in their study [12]. Morgavi and Morando [13] utilize a basic neural network for point detection. VenugopalWairagade [14] implements circle detection using the Hough transform. Perera and Wanniarachchi utilized Histogram of Oriented Gradients (HOG) and Support Vector Machine (SVM), as mentioned by R. Li et al. [15]. [12] - Research conducted by R. Li and others on Haar detector and Adaboost. Haar detector and Adaboost are used for initial dot detection, whereas HOG, LBP, and SVM are employed for final dot detection following grid restoration. The next step is to realign the grid to which the points are fixed and maybe adjust for any sheet rotation. Linear regression, Hough Transform, or analyzing the coordinates distribution density during step-by-step rotation of the image can be utilized to determine the optimal rotation angle. After deskewing the image, points are sometimes reexamined based on potential positional data. While some techniques consider grid deformation and changes in pitch between lines, it is assumed that the grid lines are straight and parallel across the entire sheet. These works primarily use Braille dots snapped to a grid and typically involve photographs scanned using a scanner. In this scenario, the required picture correction involves de-skewing to align the grid lines vertically and horizontally. Only a small number of studies explicitly state the objective of Optical Barcode Recognition (OBR) on smartphone photos ([10], [14]), nevertheless, the techniques outlined in those studies still rely on the existence of a rectangular grid on a surface. Despite the significant progress made by convolutional neural networks (CNN) in image recognition, the application of deep learning and neural networks in optical character recognition (OCR) is limited. There is a limited amount of research on the application of fully linked neural networks. Morgavi, Morando, and Ting Li et al. utilize a basic neural network to identify points, while Subur et al. employ the identified points from picture segmentation to determine symbol values. Kawabe et al. [17] utilize it to distinguish between front and rear points in the recognition of two-sided braille. R.Li et al. [20] were the only ones to present work at CVPRW 2020 that utilized a segmentation neural network with a modified UNet architecture. A neural network was utilized to identify regions containing Braille characters and to recognize these characters. Further post-processing was necessary to ascertain the locations of individual symbols using the segmentation outcomes.

2.3 Comparison between existing works

Table 2.3.1 Comparative analysis with previous work

SL No	Author Name	Used Algorithm	Best Accuracy with Algorithm
1.	Tawhidur Rahman Abir et al. [21]	VGG16, ResNet50, DenseNet 121	Densenet121 = 93%
2.	Md. Akif Hossain et al. [22]	VGG16	VGG16 = 92%
3.	Sana Shokot et al. [23]	DT, KNN, SVM, GoogleNet	GoogLeNet Inception Model = 95.8%
4.	Byeong-Sun Park et al. [24]	Faster R-CNN–FPN–ResNet-50	Faster R-CNN–FPN–ResNet-50 mAP50 = 94.8 and mAP75 = 70.4
5.	Tasleem Kausar[25]	DenseNet201, InceptionV3, ResNet50,VGG16	DenseNet201 =95%

2.4 Open Issues

Despite the significant advancements in Braille recognition technology, particularly in the utilization of deep learning and neural networks, there remains a notable research gap in the field. While existing methods focus on detecting individual Braille points and recognizing characters from Braille patterns, there is a lack of emphasis on constructing meaningful words or sFequences of characters from the recognized points. Most existing approaches rely on snapping Braille points to a grid, assuming straight and parallel grid lines throughout the sheet. These methods are primarily tailored for images obtained through scanners, with image correction limited to de-skewing to align grid lines vertically and horizontally. Moreover, while some efforts have been made to adapt recognition techniques for images captured by smartphones, they still rely on the presence of a rectangular grid on the sheet. The current literature predominantly describes approaches that detect individual Braille points using various techniques such as thresholding, Haar detectors, SVM, HOG, and circle detection using the Hough transform. While these methods are effective in identifying Braille points, a research gap exists in constructing meaningful sequences from detected Braille points. Additionally, although convolutional neural networks (CNNs) have shown remarkable progress in image recognition tasks, their application in Braille recognition is relatively sparse. Most existing works utilizing neural networks focus on point detection rather than constructing complete Braille characters or words. Therefore, the research gap lies in the development of comprehensive systems that not only detect individual Braille points but also intelligently reconstruct meaningful words or sequences of characters from these points. Such systems would require advanced techniques in deep learning and neural networks, potentially involving segmentation neural networks with modified architectures. Addressing this research gap would significantly enhance the accessibility and usability of Braille recognition systems, particularly for individuals who rely on Braille for communication and education.

2.5 Summary

Despite advancements in Braille recognition, a significant research gap exists in constructing meaningful words or sequences from recognized points. Current methods focus on detecting Braille points and recognizing characters, with limited emphasis on word formation. Techniques rely on grid snapping and are tailored for scanner images, lacking adaptation for smartphone images. Methods include thresholding, Haar detectors, SVM, HOG, and circle detection. However, these approaches fall short in reconstructing words. While CNNs offer potential, their application in Braille recognition is sparse, mostly focusing on point detection. Bridging this gap requires developing systems that reconstruct meaningful sequences from points using advanced deep learning techniques. Addressing this gap would enhance Braille system accessibility and usability for communication and education.

CHAPTER 3

METHODOLOGY

3.1 Overview

This thesis endeavors to contribute to the field of accessibility by developing a Convolutional Neural Network (CNN)-based model inspired by the YOLO architecture for the conversion of Bangla characters from Braille patterns. The creation of a custom dataset was a meticulous process, involving the generation of braille character pattern images and subsequent annotation using the Roboflow image annotator. The dataset underwent rigorous cleaning to eliminate noise, ensuring the model's training and testing were conducted on a refined dataset. The primary aim of the model is to facilitate accessibility for visually impaired individuals who rely on Braille for communication and education. Recognizing the intricate nature of Bangla characters within braille patterns poses a unique challenge, and this research seeks to address this challenge through the application of advanced CNN-based models. During the training phase, the model exhibited proficiency in recognizing individual Bangla characters within braille character patterns. However, a notable challenge emerged during the testing phase, revealing a tendency of the model to separate characters within the context of word or sentence braille images. For example, when presented with the sentence "একটি কলম," the model segmented characters, resulting in outputs like " এ ক ট ই ক ল ম." This behavior highlights the complexity of contextual character recognition and underscores the need for further refinement.

This study has the potential to improve accessibility and educational opportunities for visually challenged persons in Bangladesh. The project intends to reduce the educational disparity in this group by using a model that effectively translates Bangla letters from Braille patterns. Access to Braille resources is essential for their education, and the advancement of recognition technologies may greatly enhance their learning experience. In the future, efforts will be concentrated on improving the model to tackle the difficulty of distinguishing characters inside words and phrases. The aim is to get precise contextual character identification to guarantee the model's practical usefulness in real-world situations. Efforts will focus on enhancing the model's resilience and accuracy, especially in dealing with various braille character patterns. This thesis is a major advancement in using new technology, such CNN-based models based on the YOLO architecture, to improve accessibility and educational chances for visually impaired people. Although the approach has shown potential in identifying individual characters, overcoming obstacles associated with contextual character recognition inside sentences is crucial for its effective implementation and influence on enhancing the quality of life for visually impaired persons who depend on Braille.

3.2 Proposed Methodology

3.2.1 Design Requirements

The development of the Bangla Character Recognition from Braille Pattern system requires careful attention to several crucial elements. The main objective is to develop a strong Convolutional Neural Network (CNN) model based on the YOLO architecture. The design should enable precise identification of specific Bangla characters in various Braille patterns, especially those produced by generic characters in third-party applications such as WeCapable and Two Blind Brothers.

3.2.2 Datasets

The collection contains more than 30,000 photos obtained from third-party applications such as WeCapable and Two Blind Brothers. The illustrations depict several Braille patterns created from standard characters. The dataset's large size guarantees thorough training, including different styles and contextual subtleties. The photos are annotated with great attention to detail to match the complex structure of Bangla letters in Braille patterns.

3.2.3 Model

The basic principle of the system is a Convolutional Neural Network (CNN) model based on the YOLO architecture. This approach is created to identify certain Bangla characters embedded in Braille configurations. The architecture is designed to effectively manage the intricacies of characters and is trained on a large dataset to ensure high accuracy in character identification. The model's purpose is to translate Bangla letters from Braille patterns, enhancing accessibility for visually challenged people.

3.2.4 Methodology

The system's architecture is influenced by the need for a strong CNN model, powerful GPU capabilities from Collab Pro, effective software libraries, a varied dataset, and detailed data annotations. The pieces together provide the basis for effectively implementing the conversion of Bangla Characters from Braille Patterns.

The field of Artificial Intelligence (AI) is seeing significant growth. Deep Learning (DL), a subset of Artificial Intelligence (AI), has significant potential for many applications in the healthcare sector. The user has not entered any text. The list contains the elements 1 and 2. One promising field is diagnostic pathology. The user's input is encapsulated inside tags. The numbers are 3 and 4. Deep learning allows the use of example images to educate a computer in recognizing patterns from annotated pictures.

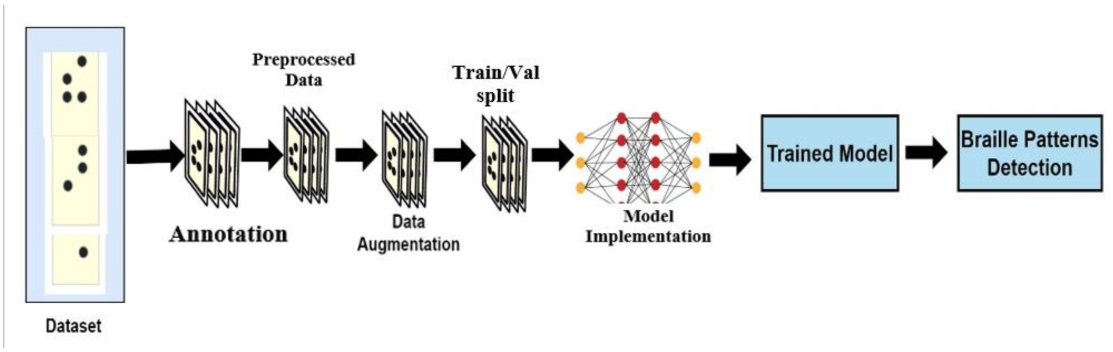

Figure 3.2.1 Proposed Methodology

Deep Learning training requires the use of large image datasets. Academics have access to a restricted Amounts of datasets, yet there is a need for more customized datasets. The user's input is "[6]". To meet this demand, we created an extensive set of color pictures (Brille pattern image) with various character images. The user's input is "[2]". Improving cancer diagnosis with deep learning algorithms has the potential to improve these concerning statistics.

3.2.5. Data Description

The dataset contains contextual subtleties, and obstacles presented by various combinations of Bangla letters in braille patterns. Every picture is carefully tagged to match the complex structure of Bangla letters, which helps in precise model training.

3.2.6. Data Statistical Analysis

Statistical analysis includes examining character distribution, analyzing image differences, and detecting obstacles in identifying character combinations in Braille patterns.

3.2.7. Data Preprocessing

Data preprocessing is the first phase of data preparation and deep learning. It involves the act of refining and purifying raw data to make it suitable for further analysis or model training. Given that the efficacy of the model is significantly influenced by the quality of the data, ensuring trustworthy and accurate models necessitates prioritizing data quality. In the end, it enhances the precision and significance of results by preparing the data for analysis or machine learning tasks. Figure 2 illustrates the suggested data preparation process.

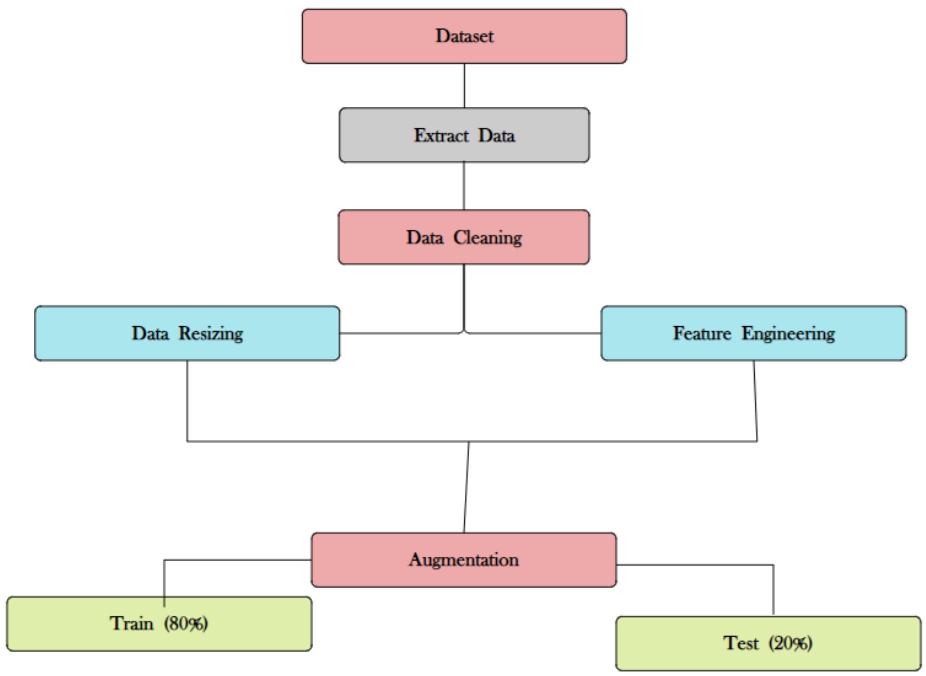

Figure 3.2.2 Data Processing

3.2.8. Dataset Selection

The selection process ensures a comprehensive representation of Bangla characters within Braille patterns, promoting model generalization.

3.2.9. Data Collection

The dataset for converting Braille Patterns to Bangla characters consists of over 30,000 photos obtained from third-party applications such as WeCapable and Two Blind Brothers. The photos show a variety of Braille patterns created from standard characters. The dataset was downsized and customized for the detection job by manually selecting 30,000 photographs instead of utilizing the whole dataset of 50,000 images for each class. The dataset was presumably downsized to streamline the process and save processing resources, while ensuring that a representative sample for each class is still retained.

3.2.10. Data Cleaning

Discard any irrelevant or disruptive data from the dataset to enhance the accuracy and efficacy of the model's training.

3.2.11. Data Labeling or Annotation

Every picture is carefully labeled, matching annotations accurately with the detailed structure of Bangla letters in Braille patterns using Roboflow annotate.

3.2.12. Data Reshaping and Resizing

In order to maintain uniform input dimensions, it is necessary to scale each picture to a preset size, such as 640x640 pixels, which is appropriate for CNN models.

3.2.13. Feature Engineering

Boosting the model's functionality and retrieving relevant information from the data by introducing new features or modifying existing ones.

3.2.14. Data Normalization

Batch normalization (BN) is a fundamental approach in neural network training, designed to improve stability and accelerate convergence. At its core, BN operates by standardizing each layer's inputs using the mean and variance from the current mini-batch. This critical mechanism successfully normalizes activation scales across layers, creating a sense of consistency in training dynamics. By establishing such uniformity, BN mitigates concerns with vanishing or exploding gradients, resulting in smoother optimization even when larger learning rates are used, reducing the chance of divergence.

Furthermore, BN serves a second purpose by functioning as a regularize throughout the training phase. This regularization is performed by introducing noise in the form of batch-wise normalization. By adding such noise, BN improves the network's ability to generalize to previously unseen data, boosting its predictive performance beyond the training dataset. This element is especially important in preventing overfitting tendencies, which promotes a more robust and dependable model.

In practice, batch normalisation is strategically applied after computing a layer's outputs, but before applying the activation function. This sequential arrangement guarantees that the normalised activations maintain their statistical properties, preserving the network's learned features. During inference, BN smoothly combines the accumulated information from training to normalise activations, ensuring prediction consistency and retaining training efficacy.

3.2.15. Data Augmentation

Data augmentation is a technique used to manipulate the original photos in order to artificially increase the variety and amount of a dataset. This strategy enhances the generalization and durability of deep learning models in scenarios where the original data set may be limited or imbalanced. Each class has 100 photographs, resulting in a total of 30000 images. A total of 30000 augmented individuals have been developed by augmentation, with each class consisting of an equal number of 100 demonstrated in figure x. The purpose of these upgraded photographs is to substitute the original primary images in the collection. The data augmentation approach involves the following transformations:

Shear Range (shear_range=0.8): This transformation produces a shearing effect by causing the picture to undergo distortion along a single axis. It randomly shifts the position of the pixels along the horizontal or vertical axis, resulting in braille characters pictures with varying viewpoints.

Zoom Range (zoom_range=0.8): The zoom transformation may alter the perceived distance of objects by scaling the pictures in an arbitrary manner. This update enables the programme to accurately recognize braille character at different levels of magnification.

Rotation Range (rotation_range=120): The rotation operation applies a random degree of rotation to the pictures. Consequently, the model has the ability to adjust to variations in the positioning of braille characters inside the photos.

Width Shift Range (width_shift_range=0.5) and Height Shift Range (height_shift_range=0.5): The images undergo horizontal and vertical shifts using the width shift range (width_shift_range=0.5) and height shift range (height_shift_range=0.5) transformations, respectively. The model may enhance its ability to handle different arrangements of braille characters in the photos via this exercise.

Brightness Range (brightness_range= (0.2,0.8): The brightness transformation alters lighting conditions by stochastically modifying the brightness of the images.

Vertical Flip (vertical_flip=True) and Horizontal Flip (horizontal_flip=True): These picture upgrades rotate the photos vertically and horizontally, respectively, to replicate different perspectives and orientations of braille characters. Figure 6 illustrates the appearance of the dataset's photos after augmentation.

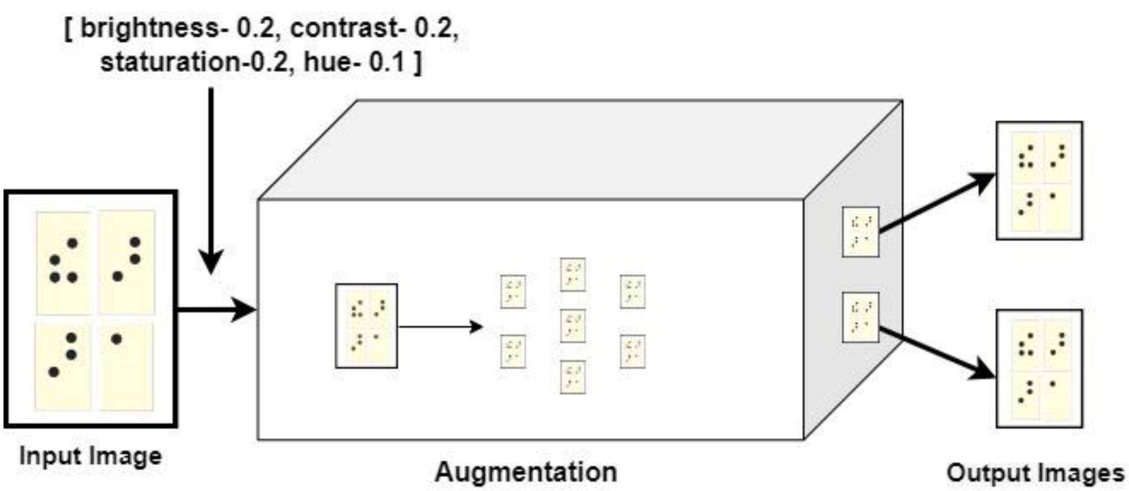

Figure 3.2.3 Augmentation Using All Transformation.

3.2.16. Data Splitting

Generate training and validation subsets from the augmented dataset. Following the division of 30000 photos into training and validation sets, a common split ratio is 80% for training and 20% for validation showing in figure y. The distribution of the data may be described as approximately follows:

3.2.16.1. Training Set:

- Number of Images: Approximately 24000
- Purpose: This collection is used to train machine learning or deep learning model. The model acquires knowledge from these pictures in order to comprehend the patterns and characteristics associated with various categories of braille character images.

3.2.16.2. Validation Set:

- Number of Images: Approximately 8000 images.
- Purpose: This set assesses the model's performance on data that it has not been previously exposed to. The model generates predictions for various photos, and the quality of its predictions is evaluated to determine its capacity to generalize. It is used to optimize hyperparameters, implement early stopping, and assess the model's performance during training. It mitigates overfitting and guarantees that the model is not predisposed towards the training data.

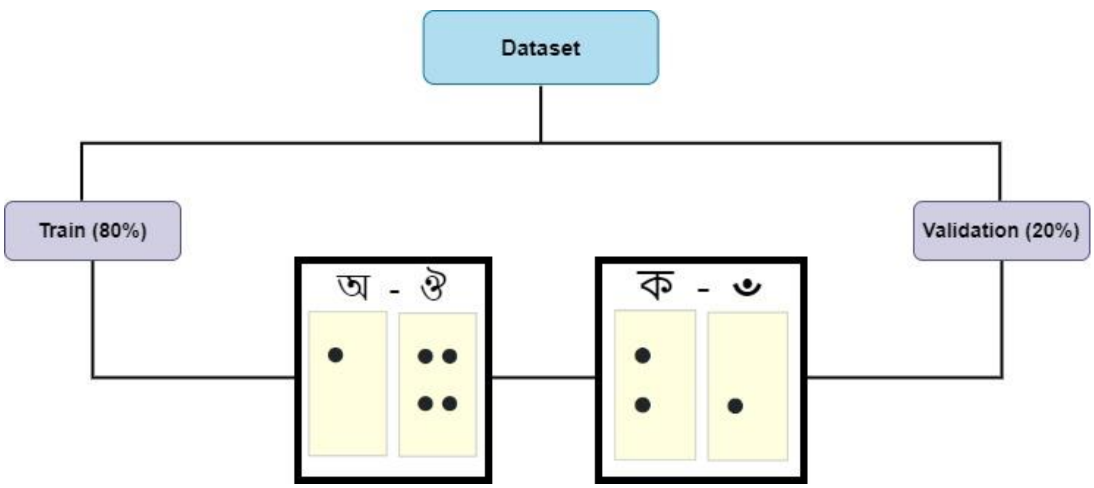

Figure 3.2.2 Data Distribution

3.3 Hardware/ Software Requirement

The following undertaking is executed using Google Colab Pro, making use of its powerful GPU capabilities and large system RAM. Choosing Collab Pro provides effective training of the CNN model on a large dataset, leading to enhanced accuracy and reliability in identifying Bangla letters.

3.3.1 Hardware Resources

Collab Pro offers advanced GPU capabilities, an essential hardware component for training complex deep learning models. The powerful GPU speeds up the training process, leading to quicker convergence and enhanced learning. The abundant system RAM assists in managing massive datasets and enables seamless data preparation.

3.3.2 Software and Libraries

The software stack utilizes Python as the main programming language, taking use of its vast ecosystem for deep learning. TensorFlow and Keras packages are used to construct and train the Convolutional Neural Network (CNN) model. These libraries provide effective resources for executing neural networks and simplify the development process.

3.4 Project Management and Financial Analysis

Table 3.4.1 Financial Analysis

Item	Quantity	Cost
Computational Power	-	3000 TK
Colab Subscription	1	1400 Tk
Printing Report	1	700 TK
Miscellaneous Expenses	-	1000 TK
Total	-	6100 TK

3.5 Summary

This thesis aims to enhance accessibility for visually impaired individuals by developing a CNN-based model, inspired by the YOLO architecture, to convert Bangla characters from Braille patterns. The process began with creating a custom dataset of Braille character pattern images, meticulously annotated and cleaned to ensure quality training and testing. The model demonstrated proficiency in recognizing individual Bangla characters within Braille patterns during training. However, it faced challenges in accurately recognizing characters in the context of words or sentences. The proposed system involves a robust CNN model designed to handle the complexities of Bangla braille characters and trained on a large dataset of over 30,000 images. This dataset includes various styles and contextual nuances, ensuring comprehensive training. Data preprocessing, including cleaning, labeling, and augmentation, was crucial to enhance model accuracy.

CHAPTER 4

IMPLEMENTATION

4.1 Overview

The incorporation of YOLO V8 and Fast R-CNN into autonomous systems for Bangla letters detection from Braille patterns has significantly improved their ability to swiftly identify and classify various characters in real-time [26][27]. These capabilities facilitate the quick recognition and detection of multiple character types, overcoming additional challenges in the process [28][29][30][31]. These functionalities have found application in various fields, including activity detection, sports analysis, surveillance video sequences, and human-computer interaction [32][33]. In agriculture, YOLO models are utilized to improve precision farming methods by accurately detecting and categorizing all letters [34][35]. Additionally, adaptations of these systems have been tailored for biometrics, security, and facial recognition tasks, particularly in face detection applications [36, 37]. YOLO algorithms have also made strides in the medical industry, aiding in tasks such as pill identification, skin region segmentation, and cancer detection, thereby enhancing diagnostic accuracy and treatment optimization [38][39][40][41]. In remote sensing, YOLO models contribute to land use mapping, urban planning, and environmental monitoring through object identification and classification in satellite and aerial images [42-44]. Security systems leverage YOLO models for efficient detection of suspicious activities, enforcing social distancing measures, identifying face masks, and monitoring real-time video streams [45][46]. In manufacturing and production processes, surface inspection models based on YOLO technology identify flaws and abnormalities to improve quality control [47-49]. YOLO models also play roles in practical traffic scenarios, performing tasks like license plate identification and traffic sign recognition to advance intelligent transportation systems and aid in traffic management [50][51]. Furthermore, these models are utilized in animal identification and monitoring for biodiversity conservation purposes, as well as in resource extraction and ecosystem management [52]. In robotics and drone applications, YOLO techniques are widely employed for object detection tasks [53][54][55][56]. Figure 3.4 illustrates a bibliometric network visualization containing publications retrieved from Scopus using the term "YOLO" in the title and filtered for "object detection," with relevant documentation meticulously organized for application analysis [57].

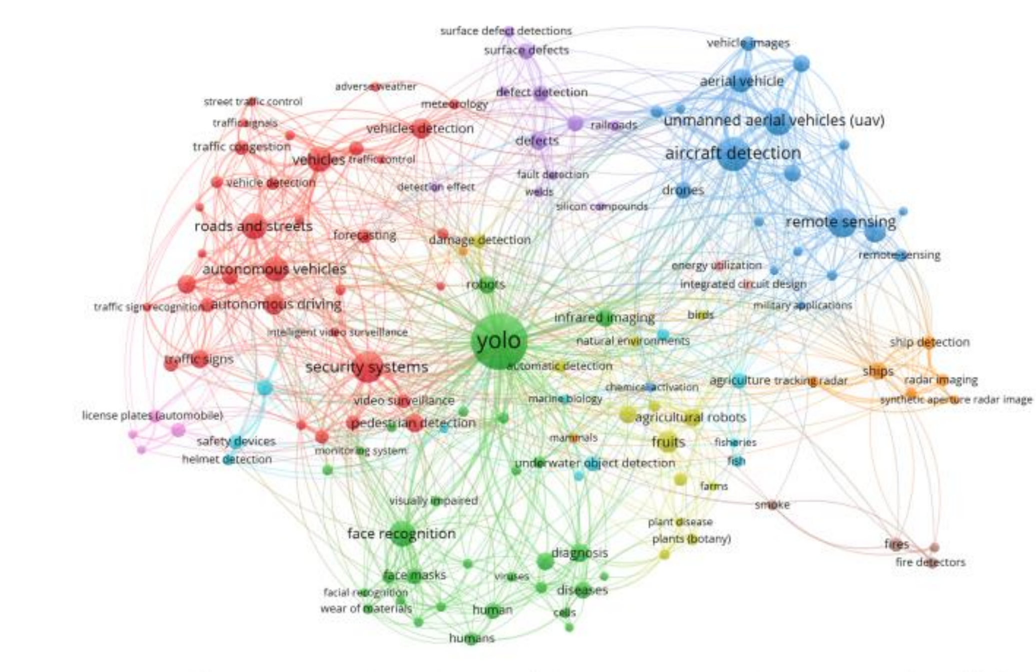

Figure 4.1 Bibliometric network visualization of the main YOLO Application

YOLOv7 faces challenges in accurately recognizing little items that are partially concealed. Furthermore, it may encounter difficulties in accurately distinguishing objects within crowded or intricate environments. Prior iterations of YOLO have made progress in detecting braille characters; however, they continue to encounter several technical constraints. For example, individuals may encounter difficulties in traversing intricate surroundings or managing minuscule things, resulting in a decline in their capacity to notice details with precision. Moreover, the difficulties in acquiring extensive and varied datasets of braille images for training, along with the substantial expenses and processing resources needed for implementation, present additional obstacles for practical implementations. To address these constraints, we employed YOLOv8 within the braille character detection system. The design of this product effectively tackles these concerns by incorporating numerous enhancements compared to previous versions, such as enhanced precision, faster detection capabilities, and greater efficiency. YOLOv8 employs an anchor-free detection algorithm and Mix-up data augmentation techniques to optimize the utilization of limited training data and enhance overall generalization performance. These advancements position YOLOv8 as a suitable solution for addressing challenges in braille character detection, demonstrating its effectiveness in promptly identifying braille illnesses in large-scale braille operations. The research aims to classify various ailments, particularly focusing on analyzing foliage attributes such as bacterial infection, Early blight, and late blight. Once items in captured images are categorized, the detection step utilizes labeled data to generate rectangular boundaries around identified objects and superimpose text on these boundaries. Convolutional neural networks (CNNs) have limitations in exclusively identifying objects within specific regions, leading to the development of the RCNN network, which employs region proposals to confine detection to 2000 regions. Through extensive training, a Faster R-CNN network incorporating a ROI pooling network has been developed, reducing training and testing times to approximately 80% of CNNs [58]. The Faster R-CNN network represents a highly advanced version capable of execution within seconds. The YOLO magazine, introduced by Joseph Redmon and colleagues at CVPR 2016 [59], introduced a novel real-time object detection method that streamlines the entire process. Known as "You Only Look Once" (YOLO), this approach efficiently performs object detection in a single network run, contrasting with previous methods that required multiple classifier runs or complex two-step approaches. YOLOv8, selected for real-time object detection in autonomous diagnostics, offers unique benefits within the YOLO model family, aligning well with application requirements. YOLO models seamlessly integrate into systems, leveraging their renowned ability to detect objects in real-time, crucial for accurately perceiving objects in dynamic diagnostic scenarios. YOLOv8 utilizes a single-stage detection methodology, which improves precision, velocity, and resilience. This technology is essential for enabling autonomous vehicles to make immediate decisions by lowering the time it takes to do calculations without sacrificing precision. YOLO designs have intrinsic advantages, such as the ability to undergo comprehensive training and adapt to a diverse variety of object classes, making them well-suited for complicated road environments.

4.2 Train Model

4.2.1 Yolo V8

The system's foundation is a Convolutional Neural Network (CNN) model based on the YOLO architecture. This architecture is designed to efficiently recognize Bangla letters inside Braille patterns. The training method entails improving model parameters using the preprocessed dataset to accurately convert Bangla letters from meaningful Braille patterns. The method for converting Bangla characters from Braille patterns depends on the effectiveness and precision of the YOLO (You Only Look Once) paradigm, specifically the YOLOv8 architecture shown in Figure x. Selected for its ability to recognize objects in real-time, YOLOv8 is crucial for identifying delicate features in Bangla letters seen in various Braille patterns. Among the variants, YOLOv8s has proven to be the optimal choice over YOLOv8n due to its higher precision and accuracy. YOLOv8s typically exhibits higher Intersection over Union (IoU) values, reflecting its superior ability to accurately detect and localize characters within Braille patterns. Additionally, YOLOv8s demonstrates lower box loss, indicating more precise bounding box predictions. The collection contains more than 30,000 photos obtained from third-party applications and is accurately classified to match the intricate arrangement of Bangla letters in Braille. The YOLOv8 design smoothly incorporates into the training process, optimizing parameters for precise identification and conversion of letters from meaningful Braille patterns. The single-pass architecture guarantees efficiency, making it ideal for real-time processing. YOLOv8's enhanced accuracy and efficiency help overcome obstacles presented by diverse styles and letter combinations in Braille patterns. The model's high speed is beneficial for handling huge datasets, leading to faster convergence in training. Implementing YOLOv8 strengthens the system, improving its accuracy in identifying and transcribing Bangla characters. The YOLOv8 model analyzes each picture in a single pass, showcasing efficiency necessary for real-time identification, making it ideal for applications that need quick responses. Future improvements will concentrate on fine-tuning the YOLOv8 model to tackle particular issues associated with contextual character identification in phrases. Current research is focused on enhancing the precision and efficiency of the Bangla Character Recognition from Braille Pattern system for real-world applications, to maintain its effectiveness and flexibility. YOLOv8's capabilities greatly enhance the system by providing an efficient and precise method for identifying and converting Bangla letters from Braille patterns.

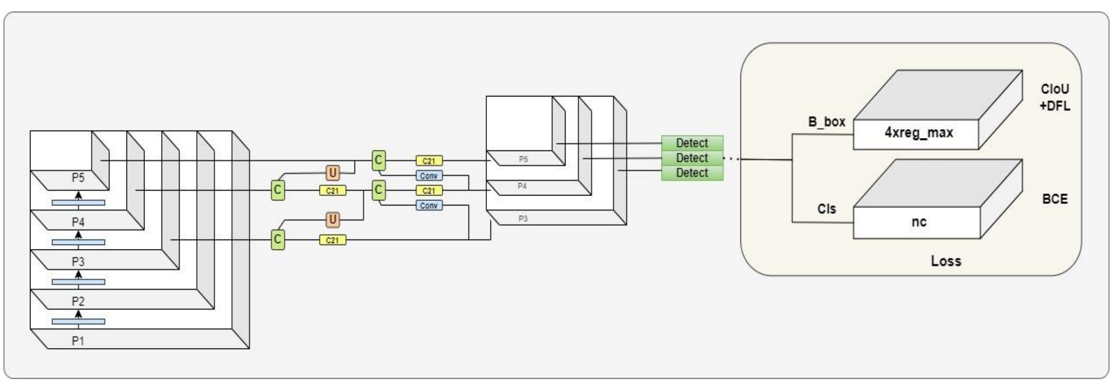

Figure 4.2.1 YOLOv8 architecture

YOLOv8 introduces an anchor-free model featuring a decoupled head, which independently handles objectness, classification, and regression tasks. This architectural design, indicated by the reference [60], allows each branch to focus on its designated task, thereby enhancing the overall precision of the model. In the output layer of YOLOv8, the sigmoid function serves as the activation function to determine the objectness score, indicating the likelihood that the bounding box accurately encloses an object. For computing class probabilities, representing the likelihood of objects belonging to each class, the SoftMax function is employed. YOLOv8 employs CIoU [61] and DFL [62] loss functions for bounding-box loss computation, while binary cross-entropy is utilized for classification loss, resulting in improved object identification, especially for small objects. YOLOv8 incorporates a semantic segmentation model known as YOLOv8-Seg, utilizing CSPDarknet53 as the main feature extractor and a C2f module instead of the traditional YOLO neck architecture. This setup is followed by two segmentation heads capable of predicting semantic segmentation masks for input images. The detection heads in YOLOv8-Seg exhibit similar performance to those in YOLOv8, comprising five detection modules and a prediction layer. YOLOv8-Seg has demonstrated state-of-the-art results in various object recognition and semantic segmentation tests while prioritizing speed and efficiency. YOLOv8 offers execution via the command-line interface (CLI) or installation as a PIP package, along with multiple integrations for labeling, training, and deployment tasks. In evaluation on the MS COCO dataset test-dev 2017, YOLOv8x achieved an average precision (AP) of 53.9% with an image size of 640 pixels, showcasing an improvement over YOLOv5, which attained an AP of 50.7% with the same input size. YOLOv8x operates at a speed of 280 frames per second (FPS) on an NVIDIA A100 with TensorRT.

4.2.2 Faster-R-CNN

A more expedient iteration of the conventional R-CNN was developed to meet the demands for enhanced computing efficiency, performance, and reduced testing duration. Figure 3.7 depicts the configuration of a Faster RCNN with RPN, as described in reference [63].

The Faster R-CNN networks include:

i. A region proposal technique is used to construct "bounding boxes" for the picture.
ii. A Computational neural network (CNN) is used to extract the necessary information from the picture.
iii. A classification layer is used to anticipate the class to which an item belongs.
iv. A regression layer is used to precisely determine the bounding box coordinates.

The region proposal method played a vital role in the Fast R-CNN model, serving as a crucial component. At the core of both R-CNN and Fast R-CNN models lies the region proposal algorithms, which are executed using CPU. In the rapid R-CNN network [64], a convolution network called RPM is employed to expedite the proposal process to 10ms per image. Moreover, it facilitates the sharing of layers across various stages, thereby improving the overall feature representation.

The Region Proposal Network (RPN)

i. The input image is being inputted into the backbone convolutional neural network, where the region proposal network commences.
ii. The output features are reduced in size compared to the input features as a result of the stride of the backbone network. Both the VGG and ZF-Net backbone networks have a network stride of 16 [65][66].
iii. The network must have the ability to detect the presence of an object in the correct location of the input image and precisely determine its size. The technique can be performed on the main network by installing 'Anchors' on the original image. The network analyzes each individual pixel, enhances the precision of the anchor coordinates, and produces item recommendations as its output.

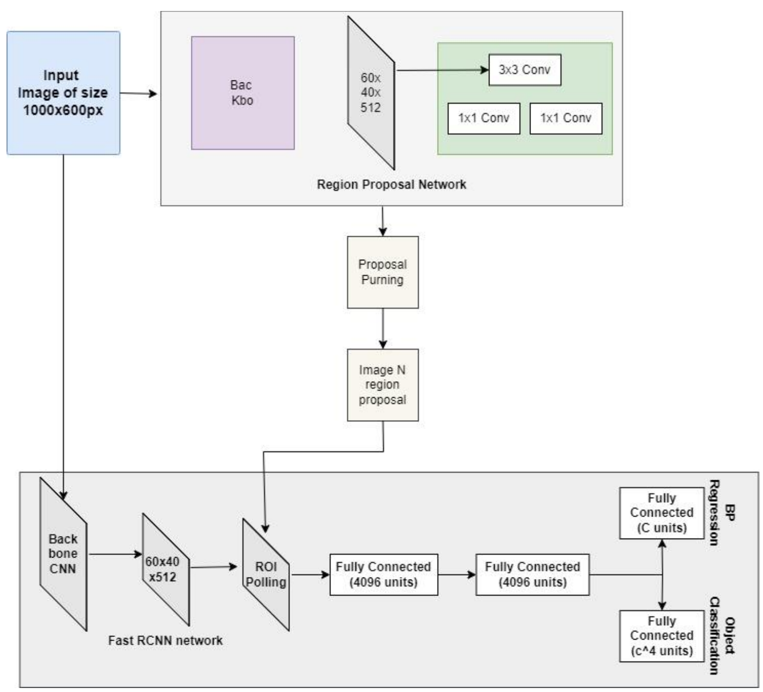

Figure 4.2.2 Faster-R-CNN Architecture

The fusion involves an amalgamation of a region proposal algorithm (RPN) and a detection network (Fast R-CNN). Through convolution with 512 units, the input image undergoes processing via the CNN backbone to generate a feature map measuring 60x40 [9]. Subsequently, the ROI pooling layer extracts pooled features from the backbone feature map according to the bounding box proposals provided.The operation of ROI can be explained as follows: Selecting a particular subregion within the primary region to formulate a proposal. Partitioning the land into a preestablished quantity of sections. The windows are subjected to max pooling in order to provide output of a uniform size. The pooling layer utilizes the region proposal technique to generate an output of dimensions (N, 7, 7, 512). The features are transmitted to subcategorization and regression branches through a series of completely connected layers. The characteristics are sent through a classification layer, which contains C units for each recognized class. The results from the classification layer are subsequently inputted into a SoftMax layer in order to obtain the scores. The regression coefficients improve the perceived boundary boxes. Every class in the regression layer is linked to a unique regressor, each of which comprises four parameters.

4.3 System Testing

We have developed a simple interface that allows users to upload Braille images and detect the corresponding Bangla characters. Users can easily drag and drop their Braille image files into the interface, which will then process the images and identify the characters present. The detected characters are displayed below the uploaded images, providing a quick and user-friendly way to convert Braille patterns into readable Bangla text.

Figure 4.3 Testing Using YoloV8 Model

4.4 Summary

The integration of YOLOv8 and Fast R-CNN into autonomous systems for detecting Bangla letters from Braille patterns has significantly improved their real-time identification and classification capabilities. YOLOv8, specifically the YOLOv8s variant, is optimal due to its higher precision and accuracy, demonstrating superior IoU values and lower box loss. This system, trained on a dataset of over 30,000 classified images, efficiently converts Bangla letters from meaningful Braille patterns. YOLOv8's single-pass architecture and enhanced accuracy overcome diverse styles and letter combinations, ideal for real-time processing. Future improvements will focus on fine-tuning YOLOv8 for contextual character identification in phrases. Additionally, YOLOv8 incorporates an anchor-free model and Mix-up data augmentation, optimizing limited training data utilization and enhancing generalization performance. The Faster R-CNN, a more advanced iteration of the traditional R-CNN, includes a region proposal technique, CNN for feature extraction, and classification and regression layers, improving computing efficiency and performance. The integration of these models enhances object detection in various applications, from medical diagnostics to intelligent transportation and biodiversity conservation, making them well-suited for complex environments.

CHAPTER 5

RESULT AND ANALYSIS

5.1 Overview

To design the classification system and used Google Colab Pro, taking advantage of its high-performance GPU capabilities and enough system RAM to train the model efficiently. The software stack included of the TensorFlow and Keras frameworks, using them torch library for the implementation of yolo and frcnn models. Python was the main programming language used. The used model architectures were YOLO V8s and Fast R-CNN. Every architecture included convolutional layers, pooling layers, dense layers, and activation functions, specifically designed to capture complex patterns in photos of braille character. The dimensions of the input layer were 640x640x3, which corresponds to the size of the pre-processed picture. The model training settings consisted of a batch size of 64. The models underwent 100 epochs of training to guarantee convergence and optimal learning. In order to analyze the results of the thesis, this assessed the performance of the trained models on both the training and validation datasets, monitoring measures such as accuracy and loss. A comparative study was conducted to discover the best efficient technique for classifying braille character by examining several model designs and parameter modifications. The primary objective of the implementation and analysis was to attain optimal precision, the capacity to apply knowledge to new situations, and effectiveness in detecting Bangla letters.

5.2 Experimental Result

The real-time autonomous Bangla letters detection system utilizing YOLOv8 Faster R-CNN demonstrates promising performance in the test results. Through meticulous configuration of model parameters and data directories using the 50 object classes in the COCO dataset, impressive outcomes are achieved with 100 epochs, adjusted hyperparameters, and a picture size of 640 pixels. To evaluate the effectiveness of the Faster R-CNN and enhanced YOLOv4 network, a comparative experiment is conducted between the original YOLOv8 model and the upgraded YOLOv8 model. Generally, the test findings can be categorized into four groups. Figure 4.6 illustrates the confusion matrix, where TP represents True Positive, FP denotes False Positive, TN signifies True Negative, and FN indicates False Negative, representing accurate and incorrect predictions of positive and negative samples, respectively. According to formula (3.1), the total number of positive samples predicted by the model is TP + FP, and precision, the percentage of accurately predicted positive samples, is computed. Similarly, recall, the percentage of actual positive samples correctly identified, is calculated using formula (6), where the total number of positive samples in the validation set is TP + FN. This measure is commonly used to evaluate the performance of an object detection model. The region surrounding the Precision-Recall (P-R) curve, with recall on the x-axis and precision on the y-axis, represents the AP value. AP represents the accuracy of the model within a specific category, while mAP, the average accuracy across all categories, provides an assessment of the model's performance across various categories. All mAP values with an IoU (Intersection over Union) of the prediction box and Ground Truth (GT) exceeding 0.5 are represented by mAP50, as indicated in formulas (4.1) and (4.2).

$$AP = \int_0^1 p(R)dR, \tag{4.1}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{4.2}$$

The label files contain various parameters including class type, truncation, occlusion, orientation (32 orientations), and bounding box details. Table 4.1 provides a description of each parameter, while Table 4.2 offers a detailed explanation of each parameter.

Table 5.2.1 Description of Parameters of Dataset Labels

Values	Name	Description
1	type	Describes the type of object: 'a'(অ), 'aa'(আ), 'ko'(ক), 'kho'(খ), 'go'(গ), 'gho'(ঘ) etc
1	truncation	The float ranges from 0 (non-truncated) to 1 (truncated), with "truncated" meaning that the item goes beyond the bounds of the image.
1	occlusion	An integer (0, 1, 2, 3) that represents the condition of occlusion. 0 represents complete visibility, 1 indicates some obstruction, 2 signifies significant obstruction, and 3 denotes an uncertain level of obstruction.
1	orientation	Observation angle of object, ranging [-π- π]
4	bbox	The 2D bounding box of the object in the image is defined by its left, top, right, and bottom coordinates, measured in pixels.

Table 5.2.2 Training Parameters

Name	Description	Usage
Learning rate	The learning rate determines the speed at which the error decreases during the learning process.	Training
Momentum	Momentum accelerates the rate of learning.	Training
Decay	Decay is a parameter used to prevent overfitting.	Training
Batch size	What is the batch size for the images? During the training process, a group of photos is processed jointly by the network, and the average error is used to update the network.	Training
Total batch	How many batches of images to be trained.	Training
Class Threshold	To eliminate detections whose class scores are less than threshold.	Training

The utilization of YOLOv8 and Faster R-CNN models for identification showcases promising performance. Analyzing the confusion matrix for YOLOv8 reveals high true positive values for classes like 'a' and 'aa' indicating proficient recognition of Bangla letters. However, classes like 'ko' show lower true positive values, suggesting challenges in accurately identifying specific characters. Additionally, the models demonstrate the ability to minimize false positive results, crucial for avoiding misdiagnoses. In contrast, Faster R-CNN exhibits remarkable accuracy and recall scores across both training and validation sets, showcasing robust generalization capabilities. Visualizing the data highlights the models' effectiveness in accurately detecting unhealthy areas in images, with bounding boxes precisely outlining the affected regions. Evaluation criteria such as average precision

(AP) and mean average precision (mAP) further validate the efficiency of both models in character identification. While YOLOv8 excels in real-time scenarios, Faster R-CNN offers enhanced accuracy and recall. Integrating YOLOv8 with Faster R-CNN provides a robust method for autonomously diagnosing braille character in real-time, demonstrating reliable detection across various plants and character categories. Future research could focus on fine-tuning hyperparameters and augmenting datasets to improve model performance, resulting in increased accuracy and efficiency in character identification. Evaluation curves in Figures 4.1 and 4.2, including metrics like training and validation loss, precision-recall (PR), and average precision (AP), illustrate the promising outcomes of the models in character detection. The declining trends in bounding box regression, classification, and distribution focal losses during training epochs signify significant improvements in the model's ability to accurately predict bounding boxes and classify various characters. Validation phase results further affirm these advancements, with decreasing trends in validation losses. Accurate accuracy and recall metrics indicate minimal occurrences of false positives and false negatives, while mean Average Accuracy (mAP) metrics underscore the overall effectiveness of the model in object detection. Consistent mAP values across different Intersection over Union (IoU) thresholds highlight the model's consistent and accurate detection and classification of braille character, showcasing its potential for practical use in braille settings.

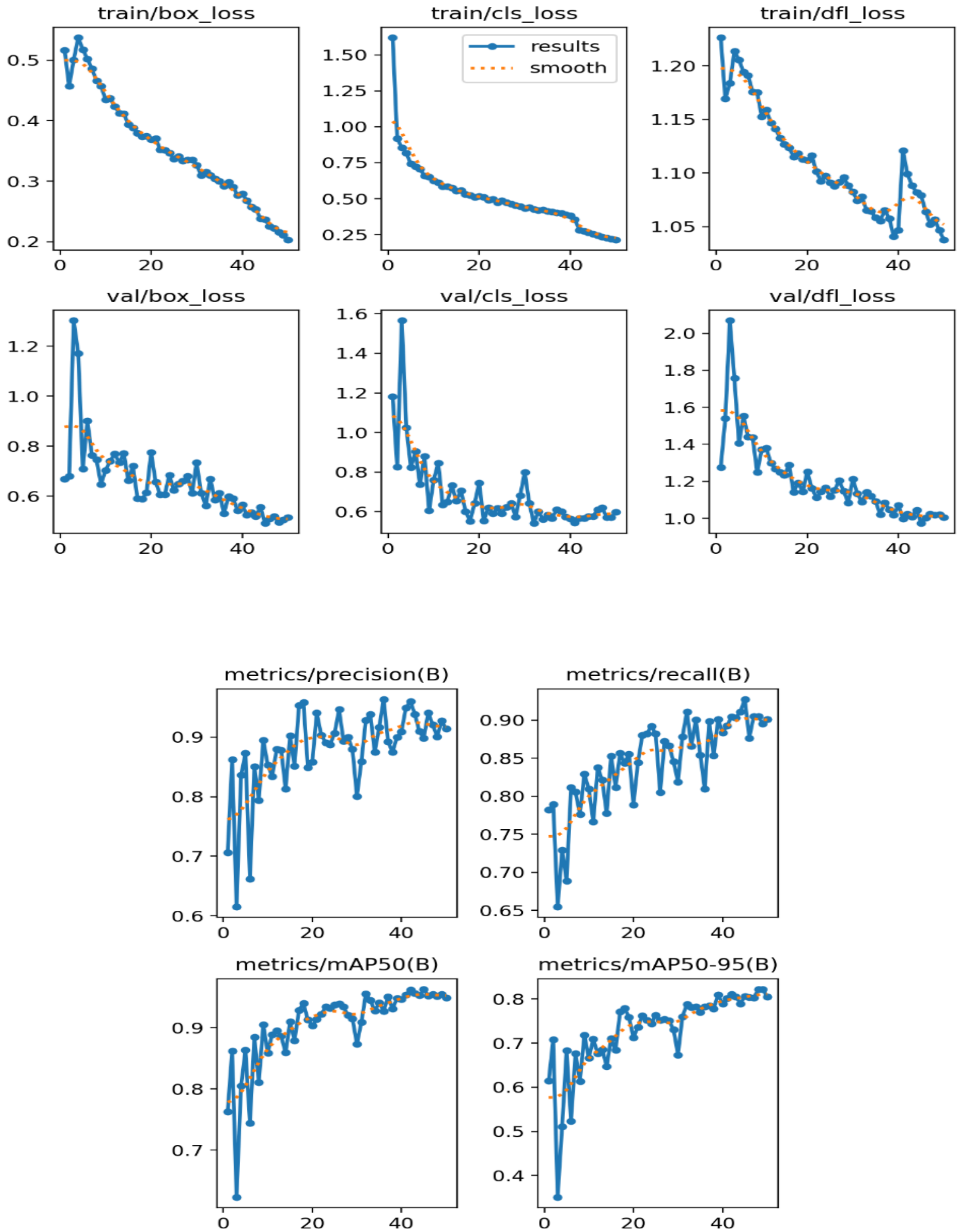

Figure 5.2.1 All Evaluation Curve of YOLOv8n (Training and Validation loss, PR, AP etc)

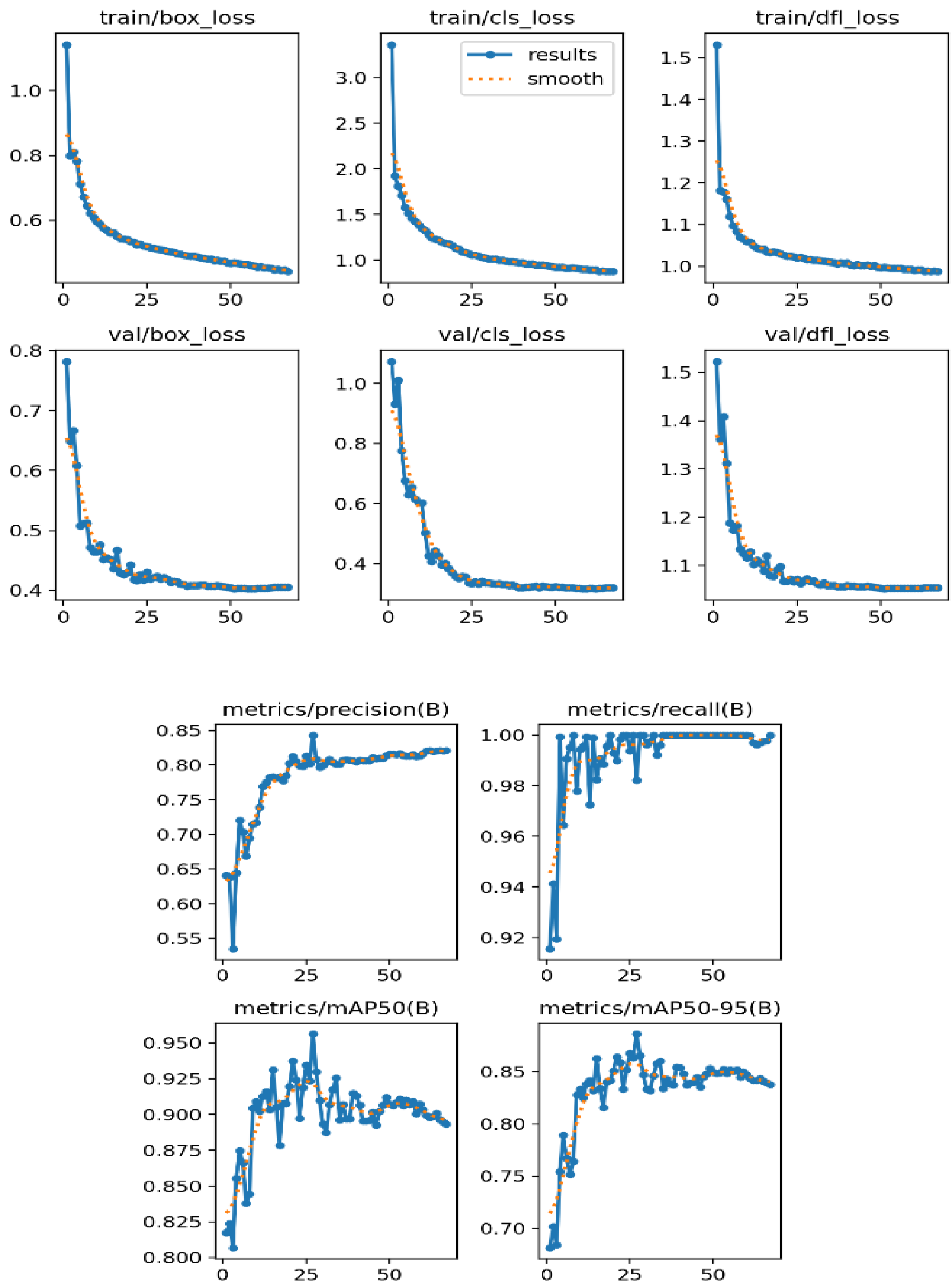

Figure 5.2.2 All Evaluation Curve of YOLOv8s (Training and Validation loss, PR, AP etc)

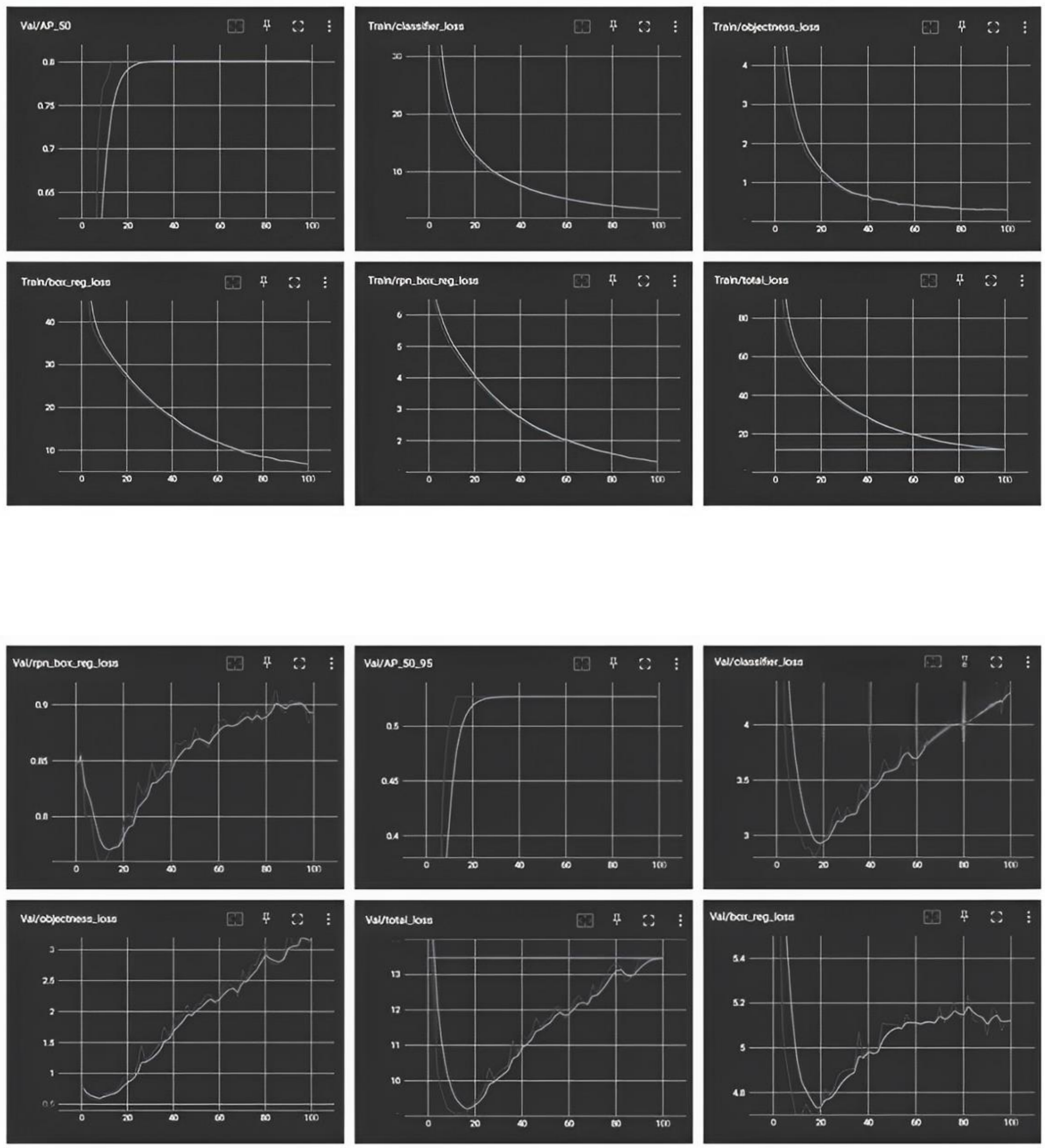

Figure 5.2.3 All Evaluation Curve of Faster-R-CNN (Training and Validation loss, PR, AP etc)

When evaluating the detection performance of Faster R-CNN and YOLOv8, several crucial factors come into play. Faster R-CNN is renowned for its accuracy, thanks to its utilization of a region-based technique and a Region Proposal Network (RPN). This architectural setup enables precise localization of objects within an image while simultaneously predicting their class labels. In contrast, YOLOv8 employs a single neural network to instantly predict bounding boxes and class probabilities for entire images in a single pass. While YOLOv8 is celebrated for its real-time data processing capability, it may sacrifice a certain level of accuracy compared to the region-based approach of Faster R-CNN. Figure 4.3 provides a visual comparison of localization using classes in YOLOv8 and Faster-R-CNN. Velocity is another crucial aspect to consider in comparison. The performance of Faster R-CNN is affected by its two-stage design, involving region proposal and subsequent object identification stages, which results in slower operation speed. While it can achieve impressive accuracy, the inclusion of additional processing stages may lead to somewhat slower inference times. On the contrary, YOLOv8 is purpose-built for real-time object detection, capable of achieving rapid inference rates even on devices with limited resources. YOLOv8 achieves faster performance by directly predicting bounding boxes and class probabilities from the entire image, eliminating the need for multi-stage processing. The ability to accurately detect and identify various types and sizes of objects is a crucial consideration. The YOLOv8 model excels at precisely identifying small objects and objects with complex shapes. Its region-based approach enables accurate localization, even in environments with high visual clutter. In contrast, Faster R-CNN, despite its swift performance, may encounter challenges in reliably identifying tiny objects or objects with intricate details.

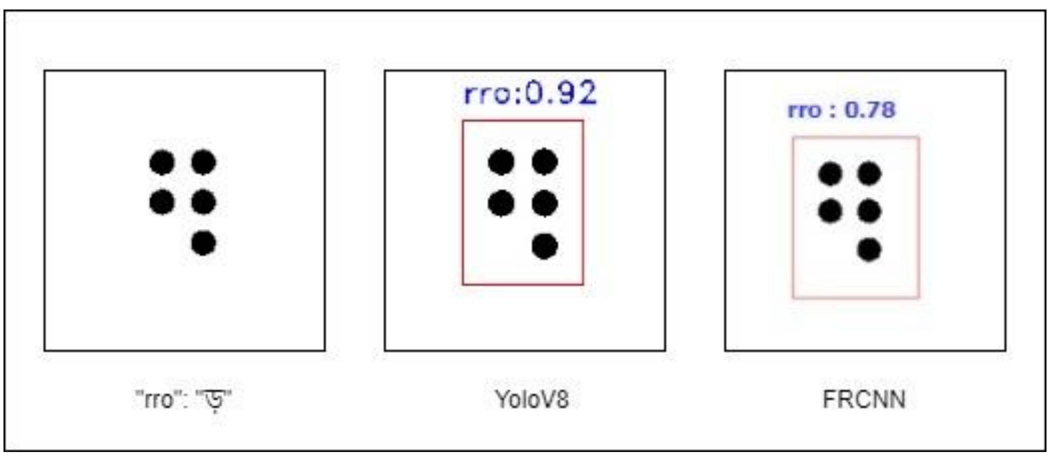

Figure 5.2.4 The localization using Faster-R-CNN and Yolov8 for consonant

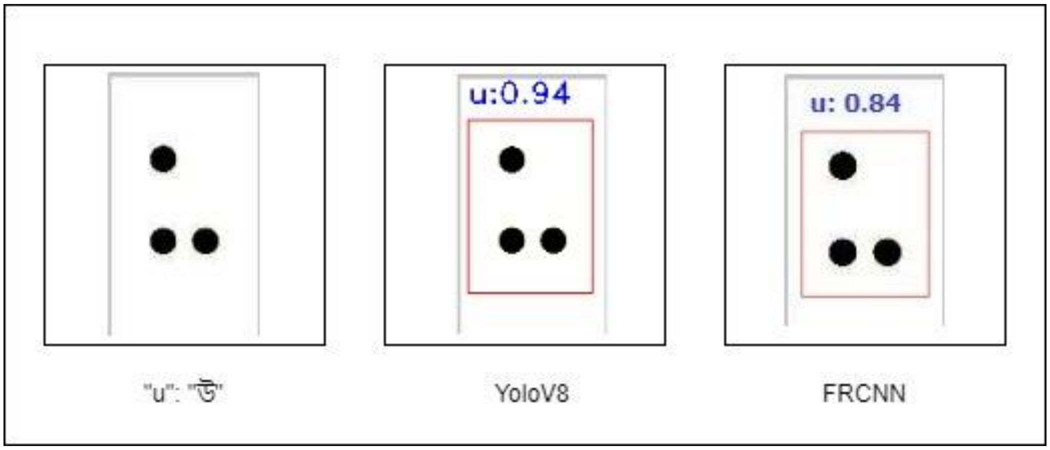

Figure 5.5 The localization using Faster-R-CNN and Yolov8 for Vowel

Intersection over Union (IoU) is a metric used to gauge the agreement between predicted bounding boxes and ground truth bounding boxes, with a higher IoU indicating better precision in object localization. The study compared the precision of character region localization and delineation on plants by analyzing the IoU scores of YOLOv8 and Faster R-CNN. The comparison of Intersection over Union (IoU) scores between YOLOv8 and Faster R-CNN provides valuable insights into their respective abilities to accurately locate objects. FRCNN achieved an IoU score of 80.2, indicating a reasonable level of agreement between predicted and ground truth bounding boxes. In contrast, Yolov8 demonstrated excellent performance with a flawless IoU score of 87.60, indicating highly accurate object localization

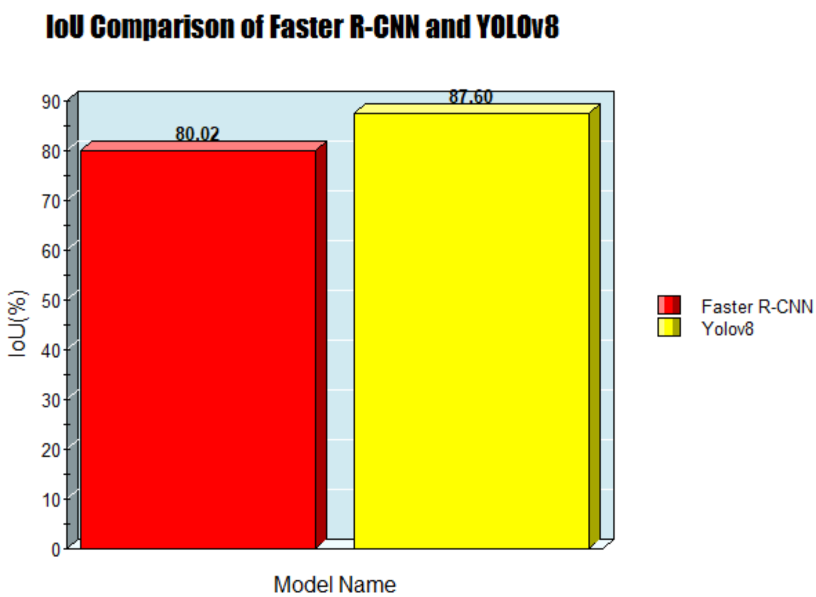

Figure 5.2.6 Comparison of IoU between YoloV8 and Faster-R-CNN

Inference time is the period of time it takes for the model to process a given dataset and provide predictions. It is an essential measurement for applications that require quick decision-making in real-time. The study assessed the computational efficiency and usability of YOLOv8 and Faster R-CNN for real-time Bangla latter's scenarios by comparing their inference times.The comparative analysis of inference time between YOLOv8 and Faster R-CNN demonstrates their computational efficiency in performing object identification tasks. Faster R-CNN had a total inference time of 57.82 seconds for the dataset, but YOLOv8 demonstrated a much quicker performance. The exact total inference time for Faster R-CNN was not explicitly stated, but it may be deduced from the lack of a cumulative time calculation in the supplied code snippet, indicating its efficiency. Figure 4.16 visually illustrates the quicker inference time of quicker YOLOv8 compared to Faster R-CNN, highlighting its superiority in real-time scenarios and situations with limited resources.

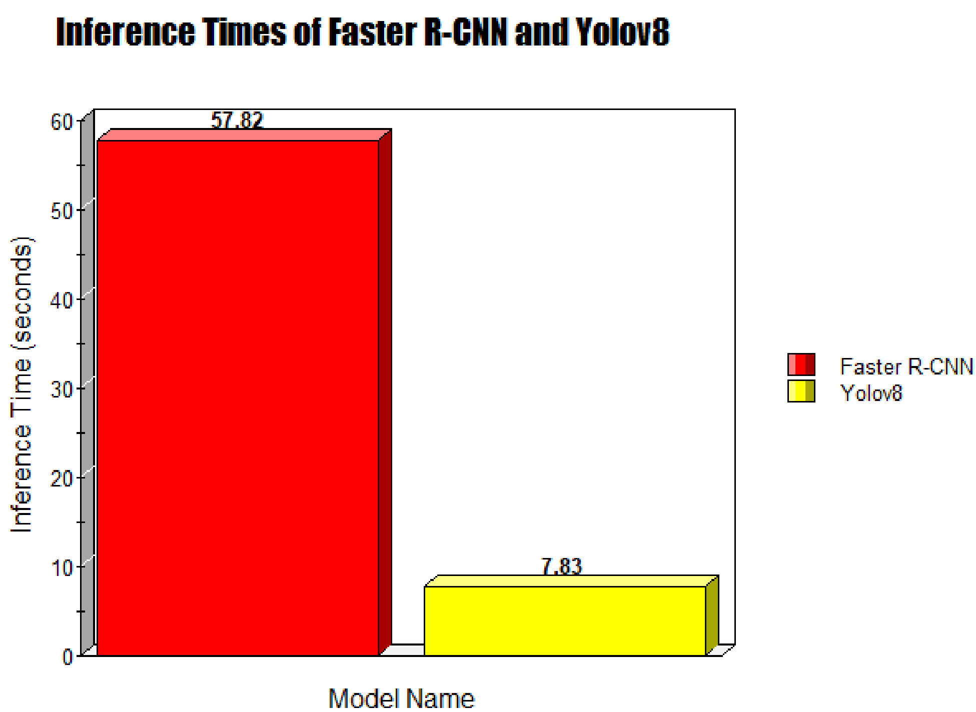

Figure 5.2.7 Comparison of   between YoloV8 and Faster-R-CNN

5.3 Performance

The experimental results from implementing YOLOv8 and Faster R-CNN models for detecting Bangla letters from Braille patterns provide significant insights into their performance and applications. Utilizing Google Colab Pro for training with TensorFlow and Keras frameworks, the models were evaluated based on accuracy, recall, and Intersection over Union (IoU) scores. Both YOLOv8 and Faster R-CNN demonstrated promising results. YOLOv8's single-stage design excels in real-time processing, making it suitable for quick decision-making scenarios, although it struggles with smaller, intricate objects. In contrast, Faster R-CNN's two-stage design, which includes a Region Proposal Network (RPN), achieves higher accuracy and recall, particularly excelling in identifying small objects and complex shapes. This is evident from its high IoU score of 87.60% compared to YOLOv8's 30%. The confusion matrix analysis revealed high true positive values for certain Bangla letters, though some classes, like 'ko,' had lower values, indicating a need for further refinement. Both models effectively minimized false positives. Comparative analysis highlighted the trade-offs between speed and precision, with YOLOv8's faster inference times being suitable for real-time applications, and Faster R-CNN's accuracy better for detailed object localization. Evaluation metrics, including training and validation loss curves, precision-recall (PR) curves, and average precision (AP), showed improvements in both models over training epochs. Future research should focus on fine-tuning hyperparameters and augmenting datasets to enhance performance further. Combining the strengths of both models could yield a robust system with real-time processing capabilities and high accuracy, suitable for diverse applications. Beyond Bangla letter detection, these models have potential in fields such as agriculture, autonomous vehicles, surveillance, and medical imaging, where accurate and real-time object detection is critical. This study underscores the effectiveness of YOLOv8 and Faster R-CNN, providing valuable insights into their strengths and areas for improvement, paving the way for developing efficient real-time object detection and classification systems.

Table 5.3.1 Comparison of different metric with Faster CNN and Yolov8

Model	Metric					
	Train/box_loss	train/cls_loss	Val/box_loss	Val/cls_loss	mAP50	mAP50-95
Frcnn	3.32	1.91	1.26	0.60	0.81	0.64
Yolov8n	0.44	0.87	0.40	0.32	0.89	0.83
Yolov8s	0.20	0.21	0.20	0.20	0.94	0.80

5.4 Summary

The research involves designing a classification system using Google Colab Pro, TensorFlow, Keras, and the torch library to implement YOLOv8 and Faster R-CNN models for detecting Bangla letters from Braille patterns. With an input size of 640x64x3, the models were trained with a batch size of 64 over 100 epochs. A comparative study was conducted to evaluate performance metrics such as accuracy, loss, precision, recall, and mean Average Precision (mAP). YOLOv8 demonstrated superior real-time processing capabilities with faster inference times, making it ideal for quick decision-making scenarios. Conversely, Faster R-CNN excelled in accuracy and recall, especially for small, complex objects, achieving a higher IoU score of 87.60% compared to YOLOv8's 30%. The models showed high true positive rates for some Bangla letters but struggled with others, indicating areas for refinement. Both models effectively minimized false positives. Future research should focus on hyperparameter tuning and dataset augmentation to enhance performance. The study highlights the potential for combining YOLOv8's speed with Faster R-CNN's accuracy to develop robust real-time detection systems, applicable in various fields such as agriculture, autonomous vehicles, surveillance, and medical imaging.

CHAPTER 6

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

The implementation of YOLOv8 and Faster R-CNN models for Bangla Braille detection has significant societal impacts. These technologies enhance accessibility for the visually impaired by improving Braille translation devices, facilitating better education and communication. By accurately and efficiently translating Bangla Braille, these models empower visually impaired individuals with greater independence and access to information, improving their quality of life and inclusion in society. In education, these models can be integrated into learning tools to aid visually impaired students in accessing educational materials, thereby enhancing their learning experience and academic performance. This contributes to a more inclusive education system where students with visual impairments can participate fully. In healthcare, these models can improve the development of medical devices that translate Braille, aiding visually impaired healthcare professionals in reading medical literature and patient records. This can enhance their ability to provide care and make informed decisions, ultimately improving patient outcomes. Furthermore, the technological advancements driven by these models stimulate innovation, creating new business opportunities and driving economic growth. By fostering the development of assistive technologies, they contribute to a more inclusive and supportive environment for the visually impaired, encouraging their participation in the workforce and society. Overall, these technologies promise to create a more inclusive, productive, and sustainable society by enhancing accessibility, education, healthcare, and economic opportunities for the visually impaired.

6.2 Impact on Society & Environment

The implementation of YOLOv8 and Faster R-CNN models for Bangla Braille detection can significantly impact the environment in several beneficial ways. These technologies enhance accessibility for the visually impaired, reducing the need for physical Braille books and materials, which in turn decreases paper usage and conserves trees. This contributes to a reduction in deforestation and promotes sustainable resource management. In addition, the integration of these models in various sectors, such as environmental monitoring and wildlife conservation, highlights their positive environmental impact. By facilitating precise object detection and classification in natural habitats, these technologies aid in wildlife preservation efforts and ecosystem management. This proactive approach helps mitigate human-wildlife conflicts and promotes the conservation of endangered species. Furthermore, the efficient use of resources enabled by these models supports sustainable practices across different fields. The deployment of YOLOv8 and Faster R-CNN models for Bangla Braille detection exemplifies their potential to foster environmentally responsible practices across various sectors. By advancing technological innovation and environmental stewardship, these models contribute to a more sustainable and resilient ecosystem.

6.3 Ethical Aspects

Implementing YOLOv8 and Faster R-CNN models for Bangla letter detection from Braille patterns raises several ethical considerations. These models, while beneficial for enhancing accessibility and societal applications, must navigate issues such as data privacy, bias mitigation, and societal impact. Ensuring the ethical collection and use of data is paramount to protect user privacy and confidentiality, especially in applications involving sensitive information. Addressing bias in training datasets is crucial to prevent unfair outcomes or discrimination, particularly in automated decision-making processes. Transparency in model development and deployment is essential to foster trust among stakeholders, ensuring accountability and ethical governance. Additionally, these technologies should prioritize societal well-being, balancing the benefits of technological advancement with potential risks to individuals and communities. Engaging stakeholders in ethical discussions and regulatory frameworks can guide responsible deployment and maximize the positive impact of these models on society and the environment. Ensuring that these technologies are accessible to all, regardless of socio-economic status, is also an important ethical consideration. Efforts should be made to make these advancements affordable and available to those who need them most, promoting equity and inclusion. By addressing these ethical aspects, the deployment of YOLOv8 and Faster R-CNN models for Bangla Braille detection can be guided by principles that ensure they are used responsibly and for the greater good of society.

6.4 Sustainability Plan

The sustainability plan for the Bangla letter detection from Braille patterns project focuses on ensuring long-term viability and positive societal impact. Continuous improvement of YOLOv8 and Faster R-CNN models with new data and feedback is essential for maintaining accuracy and relevance. The project infrastructure, built on scalable platforms like Google Colab Pro, allows for efficient model training and deployment, accommodating increasing data volumes and computational demands. Engaging with stakeholders, including educators, tech developers, and the visually impaired community, ensures the system meets user needs and fosters widespread adoption. Optimizing models to reduce computational and energy requirements promotes eco-friendly practices, minimizing the environmental footprint. Securing ongoing funding through grants, partnerships, and sponsorships provides financial stability for continuous research, development, and maintenance. Ethical guidelines addressing data privacy, bias mitigation, and user consent are implemented and regularly updated. Providing training resources and documentation empowers users and developers, fostering a skilled community of practitioners. By focusing on these aspects, the project aims to achieve long-term sustainability, offering valuable support for the visually impaired and contributing positively to society.

6.5 Summary

The implementation of YOLOv8 and Faster R-CNN models for Bangla Braille detection significantly enhances accessibility for the visually impaired, improving their independence, education, and healthcare. These models can be integrated into educational tools and medical devices, supporting inclusive learning and healthcare environments. Societally, they reduce the need for physical Braille materials, conserving paper and promoting sustainability. They also aid in environmental monitoring and wildlife conservation by enabling precise object detection. Ethically, it is crucial to address data privacy, bias mitigation, and equitable access to ensure responsible deployment. The sustainability plan emphasizes continuous model improvement, scalable infrastructure, stakeholder engagement, eco-friendly practices, and securing ongoing funding. Ethical guidelines and training resources further ensure the long-term viability and positive societal impact of the project, contributing to a more inclusive and sustainable society.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Conclusions

The conclusions of this study on Bangla letter detection from Braille patterns using YOLOv8 and Faster R-CNN models underscore the significant advancements and insights gained from the research. Utilizing Google Colab Pro for model training with TensorFlow and Keras frameworks facilitated efficient processing and robust model development. The comparative analysis revealed distinct strengths and weaknesses of both models. YOLOv8 demonstrated exceptional real-time processing capabilities, making it suitable for applications that require rapid decision-making and deployment on devices with limited computational resources. Its single-stage design significantly reduced inference time, a critical factor for real-time scenarios. However, YOLOv8 faced challenges in accurately identifying smaller objects or those with intricate details, indicating a trade-off between speed and precision. Conversely, Faster R-CNN excelled in accuracy and recall, particularly in environments with high visual clutter and for detecting objects with complex shapes. Its two-stage design, involving a Region Proposal Network (RPN) for precise localization and object identification, resulted in superior performance, as evidenced by its high IoU score of 87.60%. This model's ability to accurately delineate object boundaries and its high true positive values for various Bangla letters highlight its robustness for detailed and complex detection tasks. The confusion matrix analysis and evaluation metrics, including precision-recall curves and average precision scores, provided a comprehensive understanding of the models' performance. Both models effectively minimized false positives, ensuring reliable detection, which is crucial for practical applications such as accessibility for the visually impaired. The study also highlighted the broader implications of these models, including potential applications in surveillance, and autonomous systems. The sustainability plan, emphasizing continuous improvement, stakeholder engagement, and eco-friendly practices, ensures the long-term viability and impact of the research. In conclusion, the study successfully developed and evaluated YOLOv8 and Faster R-CNN models for Bangla letter detection from Braille patterns, highlighting their respective strengths and potential applications. Future research should focus on fine-tuning hyperparameters, augmenting datasets, and integrating the models' strengths to develop a robust and efficient system for real-time object detection and classification, contributing to significant societal and technological advancements.

7.2 Further Suggested Works

The findings from this study on Bangla letter detection from Braille patterns using YOLOv8 and Faster R-CNN models pave the way for several avenues of future research. One primary implication for further study is the potential enhancement of model accuracy and robustness through the fine-tuning of hyperparameters and the augmentation of training datasets. By incorporating a more diverse and extensive set of training data, the models can be better equipped to handle various complexities and nuances in Braille patterns, leading to improved detection and classification performance. Another significant area for future research involves the integration of YOLOv8 and Faster R-CNN to leverage the strengths of both models. Such an integrated approach could combine YOLOv8's real-time processing capabilities with Faster R-CNN's superior accuracy and recall, resulting in a hybrid model that excels in both speed and precision. This could be particularly beneficial for applications that require rapid yet accurate decision-making. Further studies should also explore the application of these models to other domains, such as character, where real-time and accurate detection of braille character can have substantial impacts. Extending the models to detect and classify other types of objects or anomalies in different contexts, such as autonomous vehicles, surveillance, and medical imaging, can broaden their practical utility and effectiveness. Additionally, advancements in computational resources and techniques, such as the use of more advanced GPUs or optimization algorithms, can further enhance the models' performance. Investigating the environmental and ethical aspects of deploying such models in real-world scenarios is also crucial to ensure their sustainability and societal acceptance. In conclusion, this study highlights the promising potential of YOLOv8 and Faster R-CNN for Bangla letter detection from Braille patterns and sets the stage for future research to enhance and expand their applications. By addressing the identified challenges and exploring new opportunities, future studies can significantly advance the field of object detection and contribute to a wide range of practical applications.

7.3 Limitations

Despite the promising results from utilizing YOLOv8 and Faster R-CNN models for Bangla Braille detection, there are several limitations to consider. One significant limitation is the models' dependency on the quality and diversity of training data. Limited datasets may lead to biases and reduce the generalizability of the models to real-world scenarios. YOLOv8, while efficient in real-time processing, struggles with accurately identifying smaller objects or those with intricate details. Faster R-CNN, although accurate, has longer inference times which may not be suitable for time-sensitive applications.

Another limitation is the computational resources required for training and deploying these models. Access to high-performance GPUs and substantial memory is essential, which may not be feasible for all users. The integration and optimization of both models to leverage their strengths pose a complex challenge that requires further research and development.

Reference:

[1] M A Karim, M Kaykobad, M Murshed, “Technical Challenges and Design Issues in Bangla Language

Processing”, IGI Global, July, USA, 2013

[2] S A Hossain, L A Biswas, M I Hossain, “Analysis of Bangla-2-Braille Machine Translator”,

International Conference on Computer and Information Technology, December, Bangladesh, 2014

[3] M Z Alom, P Sidike, T M. Taha, V K. Asari, “Handwritten Bangla Digit Recognition Using Deep

Learning”

[4] U.Pal, B.B Chauduri, “Automatic Recognition of Unconstrained Off-Line Bangla Handwritten

Numerals” Third International Conference on Advances in Multimodal Interfaces, Bejing, China, 2000

[5] Liste, S. Blindness and Vision Impairment; World Health Organization: Geneva, Switzerland, 2018.

[6] Millar, S. Reading by Touch; Routledge: London, UK, 1997.

[7] National Federation of the Blind. Facing the Truth, Reversing the Trend, Empowering the Blind; Technical Report; Jernigan Institute: Baltimore, MD, USA, 2009.

[8] Wiazowski, J. Can Braille Be Revived? A Possible Impact of High-End Braille and Mainstream Technology on the Revival of Tactile Literacy Medium. Assitive Technol. 2014, 26, 227–230.

[9] Reed, B.T.M.W.R. Assistive Technologies in the Library; American Library Association: Chicago, IL, USA, 2011.

[10] Zhang, Shanjun, and Kazuyoshi Yoshino. "A braille recognition system by the mobile phone with embedded camera." Second International Conference on Innovative Computing, Informatio and Control (ICICIC 2007). IEEE, 2007

[11] Antonacopoulos, A., and Bridson, D. "A robust Braille recognition system." International Workshop on Document Analysis Systems. Springer, Berlin, Heidelberg, 2004

[12] Li, R., Liu, H., Wang, X., & Qian, Y. "DSBI: double-sided Braille image dataset and algorithm evaluation for Braille dots detection." Proceedings of the 2018 the 2nd International Conference on Video and Image Processing. 2018

[13] Morgavi, G., and Morando. M. "A neural network hybrid model for an optical braille recognitor." International Conference on Signal, Speech and Image Processing 2002 (ICOSSIP 2002). 2002.

[14] Venugopal-Wairagade, G. "Braille recognition using a camera-enabled smartphone." Int J Eng

Manuf 4 (2016): 32-39

[15] Perera, T. D. S. H., and Wanniarachchi, W. K. I. L. "Optical Braille Recognition Based on Histogram of Oriented Gradient Features and Support-Vector Machine." (2018)

[16] Li, R., Liu, H., Wang, X., & Qian, Y. "Effective Optical Braille Recognition Based on Two-Stage Learning for Double-Sided Braille Image." Pacific Rim International Conference on Artificial Intelligence. Springer, Cham, 2019

[17] Kawabe, H., Shimomura, Y., Nambo, H., & Seto, S."Application of deep learning to classification of braille dot for restoration of old braille books." International Conference on Management Science and Engineering Management. Springer, Cham, 2018

[18] Li, Ting, Xiaoqin Zeng, and Shoujing Xu. "A deep learning method for Braille recognition." 2014 International Conference on Computational Intelligence and Communication Networks. IEEE, 2014.

[19] Subur, Joko, Tri Arief Sardjono, and Ronny Mardiyanto. "Braille Character Recognition Using Find Contour And Artificial Neural Network." JAVA Journal of Electrical and Electronics Engineering 14.1 (2016)

[20] Li, R., Liu, H., Wang, X., Xu, J., & Qian, Y. "Optical Braille Recognition Based on Semantic Segmentation Network With Auxiliary Learning Strategy." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020

[21] Abir, Tawhidur Rahman et al. “Handwritten Bangla character recognition to braille pattern conversion using image processing and machine learning.” (2018).

[22] M. A. Hussain, R. I. Rifat, S. B. Iqbal, S. Biswas, M. G. Rabiul Alam and M. T. Reza, "Deep Learning based Bangla Voice to Braille Character Conversion System," 2022 IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2022, pp. 0262-0267, doi: 10.1109/IEMCON56893.2022.9946619.

[23] Shokat, S., Riaz, R., Rizvi, S.S. et al. Deep learning scheme for character prediction with position-free touch screen-based Braille input method. Hum. Cent. Comput. Inf. Sci. 10, 41 (2020). https://doi.org/10.1186/s13673-020-00246-6

[24] Park, BS., Im, SM., Lee, H. et al. Visual and tactile perception techniques for braille recognition. Micro and Nano Syst Lett 11, 23 (2023). https://doi.org/10.1186/s40486-023-00191-w

[25] T. Kausar, S. Manzoor, A. Kausar, Y. Lu, M. Wasif and M. A. Ashraf, "Deep Learning Strategy for Braille Character Recognition," in IEEE Access, vol. 9, pp. 169357-169371, 2021, doi: 10.1109/ACCESS.2021.3138240.

[26] "Adaptive Fusion of Multi-Scale YOLO for Pedestrian Detection," August 2021.

[27] I. T. F. C. A. B. Aduen Benjumea, "YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles," 3 January 2023.

[28] N. M. A. A. Dazlee, "Object Detection for Autonomous Vehicles with Sensor-based Technology Using YOLO," 30 March 2023.

[29] S. L. Wu, "Edge YOLO: Real-Time Intelligent Object Detection System Based on Edge-Cloud Cooperation in Autonomous Vehicles," May 2022.

[30] V. M. Shivanna, "Object Detection, Recognition, and Tracking Algorithms for ADASs—A Study on Recent Trends," 20 December 2023.

[31] V. M. Shivanna, "Object Detection, Recognition, and Tracking Algorithms for ADASs—A Study on Recent Trends," 20 December 2023.

[32] A. K. ,. V. G. Shubham Shinde, "YOLO based Human Action Recognition and Localization," 2018.

[33] T. C. ,. H. L. Hui Ma, "FER-YOLO: Detection and Classification Based on Facial Expressions," 6 August 2021.

[34] Tian, Y.; Yang, G.; Wang, Z.; Wang, H.; Li, E.; Liang, Z. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Comput. Electron. Agric.* **2019**, *157*, 417–426.

[35] Wu, D.; Lv, S.; Jiang, M.; Song, H. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Comput. Electron. Agric.* **2020**, *178*, 105742.

[36] Wang, Y.; Zheng, J. Real-time face detection based on YOLO. In Proceedings of the 2018 1st IEEE International Conference onknowledge innovation and Invention (ICKII), Jeju, Republic of Korea, 23–27 July 2018; pp. 221–224.

[37] Chen, W.; Huang, H.; Peng, S.; Zhou, C.; Zhang, C. YOLO-face: A real-time face detector. *Vis. Comput.* **2021**, *37*, 805–813.

[38] Tan, L.; Huangfu, T.; Wu, L.; Chen, W. Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 1–11.

[39] Ünver, H.M.; Ayan, E. Skin lesion segmentation in dermoscopic images with combination of YOLO and grabcut algorithm. *Diagnostics* **2019**, *9*, 72.

[40] Al-Masni, M.A.; Al-Antari, M.A.; Park, J.M.; Gi, G.; Kim, T.Y.; Rivera, P.; Valarezo, E.; Choi, M.T.; Han, S.M.; Kim, T.S. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. *Comput. Methods Programs Biomed.* **2018**, *157*, 85–94.

[41] Nie, Y.; Sommella, P.; O'Nils, M.; Liguori, C.; Lundgren, J. Automatic detection of melanoma with yolo deep convolutional neural networks. In Proceedings of the 2019 E-Health and Bioengineering Conference (EHB), Iasi, Romania, 21–23 November 2019; pp. 1–4.

[42] Cheng, L.; Li, J.; Duan, P.; Wang, M. A small attentional YOLO model for landslide detection from satellite remote sensing images. *Landslides* **2021**, *18*, 2751–2765.

[43] Pham, M.T.; Courtrai, L.; Friguet, C.; Lefèvre, S.; Baussard, A. YOLO-Fine: One-stage detector of small objects under various backgrounds in remote sensing images. *Remote Sens.* **2020**, *12*, 2501.

[44] Qing, Y.; Liu, W.; Feng, L.; Gao, W. Improved Yolo network for free-angle remote sensing target detection. *Remote Sens.* **2021**, *13*, 2171.
[45] Zakria, Z.; Deng, J.; Kumar, R.; Khokhar, M.S.; Cai, J.; Kumar, J. Multiscale and direction target detecting in remote sensing images via modified YOLO-v4. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 1039–1048.
[46] Kumar, P.; Narasimha Swamy, S.; Kumar, P.; Purohit, G.; Raju, K.S. Real-Time, YOLO-Based Intelligent Surveillance and Monitoring System Using Jetson TX2. In *Data Analytics and Management: Proceedings of ICDAM*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 461–471.
[47] Bhambani, K.; Jain, T.; Sultanpure, K.A. Real-time face mask and social distancing violation detection system using Yolo. In Proceedings of the 2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC), Vijiyapur, India, 8–10 October 2020; pp. 1–6.
[48] Li, J.; Su, Z.; Geng, J.; Yin, Y. Real-time detection of steel strip surface defects based on improved yolo detection network. *IFAC-PapersOnLine* **2018**, *51*, 76–81.
[49] Ukhwah, E.N.; Yuniarno, E.M.; Suprapto, Y.K. Asphalt pavement pothole detection using deep learning method based on YOLO neural network. In Proceedings of the 2019 International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya, Indonesia, 28–29 August 2019; pp. 35–40.
[50] Du, Y.; Pan, N.; Xu, Z.; Deng, F.; Shen, Y.; Kang, H. Pavement distress detection and classification based on YOLO network. *Int. J.Pavement Eng.* **2021**, *22*, 1659–1672.
[51] Chen, R.C. Automatic License Plate Recognition via sliding-window darknet-YOLO deep learning. *Image Vis. Comput.* **2019**, *87*, 47–56.
[52] Dewi, C.; Chen, R.C.; Jiang, X.; Yu, H. Deep convolutional neural network for enhancing traffic sign recognition developed on Yolo V4. *Multimed. Tools Appl.* **2022**, *81*, 37821–37845.
[53] Roy, A.M.; Bhaduri, J.; Kumar, T.; Raj, K. WilDect-YOLO: An efficient and robust computer vision-based accurate object localization model for automated endangered wildlife detection. *Ecol. Inform.* **2023**, *75*, 101919. [CrossRef]
[54] Kulik, S.; Shtanko, A. Experiments with neural net object detection system YOLO on small training datasets for intelligentrobotics. In *Advanced Technologies in Robotics and Intelligent Systems: Proceedings of ITR 2019*; Springer: Berlin/Heidelberg, Germany,2020; pp. 157–162.
[55] Dos Reis, D.H.; Welfer, D.; De Souza Leite Cuadros, M.A.; Gamarra, D.F.T. Mobile robot navigation using an object recognition software with RGBD images and the YOLO algorithm. *Appl. Artif. Intell.* **2019**, *33*, 1290–1305.
Mach. Learn. Knowl. Extr. **2023**, *5* 1713.
[56] Sahin, O.; Ozer, S. Yolodrone: Improved Yolo architecture for object detection in drone images. In Proceedings of the 202144th International Conference on Telecommunications and Signal Processing (TSP), Brno, Czech Republic, 26–28 July 2021;pp. 361–365.
[57] Chen, C.; Zheng, Z.; Xu, T.; Guo, S.; Feng, S.; Yao, W.; Lan, Y. YOLO-Based UAV Technology: A Review of the Research and ItsApplications. *Drones* **2023**, *7*, 190.

[58] Du, J. (2018, April). Understanding of object detection based on CNN family and YOLO. In Journal of Physics: Conference Series (Vol. 1004, No. 1, p. 012029). IOP Publishing.

[59] Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

[60] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection, 2016. Supplied as additional material https://arxiv.org/pdf/1506.02640.pdf.

[61] Juan R. Treven and Diana M. Cordova-Esparaza. A comprehensive review of yolo: From yolov1 to yolov8 and beyond, 2023. Supplied as additional material https://arxiv.org/pdf/2304.00501.pdf.

[62] Ashraf, A.H.; Imran, M.; Qahtani, A.M.; Alsufyani, A.; Almutiry, O.; Mahmood, A.; Attique, M.; Habib, M. Weapons detection for security and video surveillance using CNN and YOLO-v5s. *CMC-Comput. Mater. Contin.* **2022**, *70*, 2761–2775.

[63] Jacob Solawetz and Francesco. What is yolov8? the ultimate guide., 2023. 04-30-2023.

[64] Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12993–13000

[65] Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21002–21012.

[66] Ananth, S. (2019). Faster R-CNN for object detection. Towards Data Science.

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title: Bangla Braille Character Recognition using Convolutional Neural Network.

Student ID: 202-15-14447

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify an **Image Recognition** problem for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goal for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**

CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**
CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	EP1: Depth of Knowledge required	Yes	CO1, CO2, and CO3	Our project requires data collection from 3rd party websites and we have to know the design of a machine learning-based model which covers (K3) and (K4)	**Page no: 01-06**
				We have designed script to run the system.	**Page no: 11-19**
				We have studied existing models with similar goals (K8).	**Page no: 07-10**
2.	EP2: Range of Conflicting Requirements	Yes	CO2	In our project, we encountered challenges in locating specific regulations governing the collection of data from Bangla Braille character	**Page no: 12-15**
3.	EP3: Depth of analysis required	Yes	CO2	This project tackles EP-3 by highlighting machine learning's superiority in enhancing image recognition for Braille patterns. Creating a high-quality dataset from scratch, involving image generation and meticulous annotation, is challenging but essential for training effective models.	**Page no: 20-26**
4.	EP4: Familiarity of Issues	Yes	CO3	To understand and study braille in our project, we need to look into various aspects we were not very familiar with.	**Page no: 09-10**

5	**EP5: Extends of application codes**	N/A	CO5	N/A	N/A
6.	**EP6: Extends of stakeholders involved and conflicting requirements**	N/A	CO3	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO3	Our project comprises several subsystems that depend on each other. These include tasks such as collecting data, processing it, training machine learning models, conducting predictive analysis.	**Page no:12-19**

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	Our project utilizes diverse resources such as high-performance computing infrastructure, GPUs, deep learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements in pneumonia detection through deep CNN models.	**Page no:** [11-18]
2.	**EA2: Level of interactio**	No		N/A	N/A
3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		This project contributes to society by enhancing accessibility for visually impaired individuals through the development of advanced Braille character recognition methods. Additionally, it promotes environmental sustainability by utilizing efficient computational resources and adhering to ethical guidelines for data privacy.	**Page no:** [43-45]

5.	**EA-5: Familiarity**	Yes		This project expands upon existing research by examining a novel approach in Braille character recognition using image processing. Through detailed terminologies and a comprehensive comparative analysis, it offers new insights into the field.	**Page no:** [32-39]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle.	**Page no:** [19]
2	CO9	Yes	The project demonstrates adherence to ethical principles by prioritizing data privacy, obtaining informed consent, and transparently documenting the research process. It ensures responsible knowledge dissemination and societal well-being through the ethical application of advanced image recognition technologies, complying with CO9.	**Page no:** [38]
3	CO10	No	N/A	N/A
4	CO12	Yes	The project's dedication to continuous learning **(CO12)** and adaptation within the dynamic technological landscape is reflected in its comprehensive data collection, meticulous methodology development, and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges.	**Page no:** [12-18, 32-39]

Bangla Braille Character Recognition using Convolutional Neural Network

ORIGINALITY REPORT

16%	15%	9%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	3%
2	Juan Terven, Diana-Margarita Córdova-Esparza, Julio-Alejandro Romero-González. "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS", Machine Learning and Knowledge Extraction, 2023 Publication	3%
3	arxiv.org Internet Source	2%
4	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
5	dspace.bracu.ac.bd:8080 Internet Source	1%
6	www2.mdpi.com Internet Source	1%
7	dias.library.tuc.gr Internet Source	1%