EXPLORING DEEP LEARNING ARCHITECTURE FOR COMPLEX HAND SIGN LANGUAGE

BY

GAZI RASHIDUL ISLAM
ID: 202-15-14407

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

Professor Dr. Touhid Bhuiyan
Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Mr. Raja Tariqul Hasan Tusher
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "**Exploring Deep Learning Architecture For Complex Hand Sign Language**", submitted by **Gazi Rashidul Islam** and 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 **15 July 2024**

BOARD OF EXAMINERS

AHossain 15-07-2024

Dr. Md. Fokhray Hossain (MFH) — **Board Chairman**
Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Firoz Hasan 15.7.24

Md. Firoz Hasan (FH) — **Internal Examiner 1**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Lamia Rukhsara

Lamia Rukhsara (LR) — **Internal Examiner 2**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

15.07.24

Dr. S. M. Hasan Mahmud (SMH) — **External Examiner**
Assistant Professor
Department of Computer Science
American International University-Bangladesh

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Professor Dr. Touhid Bhuiyan, Professor, Department of Computer Science and Engineering, Daffodil International University**. We 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:

Professor Dr. Touhid Bhuiyan
Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Mr. Raja Tariqul Hasan Tusher
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Rashidul

Gazi Rashidul Islam
ID: -202-15-14407
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we 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.

We are grateful and wish our profound indebtedness to **Professor Dr. Touhid Bhuiyan**, **Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Computer Vision*" 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.

We would like to express our 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.

We express our sincere gratitude to **Bangladesh Government Badhir Academy** for their invaluable assistance and unwavering support in facilitating the data collection process.

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

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

ABSTRACT

This research paper describes how advanced deep learning methods were used to create a system that can recognise Bangla Sign Language (BSL). The goal of the project is to help the deaf people who speak Bangla communicate better by creating a system that can accurately recognise sign language in real time. At first, a collection of 1,745 pictures of BSL hand signs was gathered and preprocessed by normalizing and segmenting the data to make it better. Three main models—ResNet50, VGG19, and DenseNet201—were put into action and tested. It turned out that DenseNet201 was the best model. It had the highest accuracy rate of 89.03% and was better at reading complex hand signs. The accuracy for VGG19 was 73% and for ResNet50 it was 57%. The method involved fine-tuning DenseNet201, which used its tightly connected layers to make feature reuse and gradient flow better, which made identification work better. To make sure it was reliable, the system was put through a lot of tests using accuracy, recall, and F1-score measures. The model was also made to work best for real-time applications and was built to work with an easy-to-use web interface to make it more accessible. The in-depth literature study, methodology, execution, system testing, and assessment are all covered in this report. The report also talks about the project's social and environmental effects. To make sure the system is used responsibly, ethical issues and plans for long-term use were also talked about. This project shows how deep learning can be used to make strong BSL recognition systems. These systems will help the deaf community communicate better and feel more included. The results show how important advanced neural network designs and careful data preparation are for getting good results in tasks that require recognising sign language.

TABLE OF CONTENTS

Contents **Page No**

COURSE OUTCOMES, COMPLEX ENGINEERING PROBLEMS (EP) AND COMPLEX ENGINEERING ACTIVITIES (EA) ADDRESSING

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

BSL - Bangla Sign Language
CNN - Convolutional Neural Network
COs - Course Outcomes
D&D - Deaf & Dumb
EA - Engineering Activities
EP - Engineering Problems
F1 Score - F1 Measure
FN - False Negative
FP - False Positive
LSTM - Long Short-Term Memory
ML - Machine Learning
SVM - Support Vector Machine
TP - True Positive
VGG - Visual Geometry Group

CHAPTER 1

Introduction

1.1 Overview

Sign language recognition is an effective technology in the communication between deaf communities and hearing societies, and it tends to be a widely discussed topic with the increasing development of the research on interaction between human and machines[1] This language uses hand-shape variations, body movement, and even facial expression to convey information or meaning. Sign languages provide a natural way of interaction by minimizing the barrier of communication between hearing-impaired and society.[2] This form of communication is essential for enabling deaf individuals to fully participate in social, academic, and professional life. Among the various sign languages used worldwide, Bangla Sign Language (BSL) holds particular importance in the eastern region of India (West Bengal and Kolkata) and in Bangladesh, where over 220 million people speak Bengali as their native language, ranking it seventh globally based on the number of speakers.[7] Bangla is the seventh most spoken language globally, highlighting the importance of developing effective communication tools for this substantial population. Hand motions and signs have been the focus of computer vision study for a long time. Initial attempts to use computers to automatically recognize and understand hand motions and gestures have changed a lot as technology has improved. In the last few years, progress in deep learning and computer vision has made it possible to build strong systems that can recognize many sign languages, including Bangla Sign Language (BSL). Complex neural network architectures, such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs), are used by these systems to accurately record the spatial and temporal features of sign language movements. There are still some issues with the sign language systems that are used today, even with the newest improvements. This is especially true when they are used with people who speak different languages and come from various backgrounds. Previous research mainly focused on making systems that could sort Bangla words and sentences from live video in real time. These methods were very helpful for the people they were meant for, but they had trouble growing. Making these programs better at finding new words and sentences was hard, which meant they couldn't be used in as many situations. Communication is a basic human need, so understanding sign language is important for making society more fair and accessible. By bridging the gap between the hearing and deaf communities, sign language helps deaf people fit in with others and fully join in many everyday activities. People who are Deaf & Dumb (D&D) use gestures, hand signs, and haptic sign language to interact, which is different from people who have normal hearing who talk. People who don't have disabilities often use these ways of talking to share their thoughts and feelings, but they are very important for people who have hearing loss. To communicate clearly, it's necessary to use more than just words. For example, hand motions and sign language both use the fingers. Because it is strong and adaptable, this kind of language is seen by many as an important tool. The fact that hand signals and gestures are so common shows how important it is to have reliable methods for understanding sign language. We can guarantee better communication and equal access to information for the deaf and hard of hearing people by building these kinds of systems. This project aims to create a reliable BSL recognition system by using advanced deep learning methods. This will help make present sign language identification systems better by fixing their problems. This study uses a source dataset with 1,745 pictures of hand signs used in Bangla sign language. It was taken from the "Government Bodhir Academic" Office. In order to make sure that the trained models will last, these pictures were taken in various kinds of settings. Several deep learning models were used to test how well the BSL detection system worked. GoogLeNet, Xception, Inception, VGG16, VGG19, DenseNet201, and ResNet150 are some of the models that were tried. Every one of these models is built differently and is better at handling different parts of picture recognition than the others. For example, GoogLeNet uses inception modules to understand spatial structures, and Xception uses depth-wise separable convolutions to make things run more smoothly. With an accuracy of 89.03%, DenseNet201 was the best at reading Bangla sign language hand signs. Its thickly connected layers encourage feature reuse and improve gradient flow. The VGG19 model, on the other hand, was 70.12% accurate. In summary, making sign language recognition systems that work well is necessary to improve communication and make sure that the deaf and dumb group has equal access to information. To reach these goals, the focus of this study on using cutting-edge deep learning methods to make a correct BSL recognition system is an important step forward. The results of testing various models show that this area needs more study and progress. This will help the deaf people who speak Bangla communicate better and be easier to reach in the long run.

1.2 Background and Present State

Bangla, which is also called Bengali, is a language that many people speak in Bangladesh and the eastern parts of India, especially in West Bengal and Kolkata. It is the seventh most spoken language in the world, with more than 220 million native speakers. The fact that so many people speak and write Bangla shows how important it is to make communication tools that work for this big group of people, especially those who have trouble speaking or hearing. For many years, one of the main goals of computer vision study has been to find ways to automatically recognize hand signs and gestures. The goal has been to make tools that can correctly translate sign languages so that people who use sign language and people who don't can communicate with each other. Hand-gesture recognition tools for Bangla Sign Language (BSL) and other languages have been made possible by recent advances in technology. Even with these improvements, it can be hard to use current methods in different languages and situations. For example, past projects that tried to figure out Bangla words and phrases through live video were helpful, but they weren't very big. These systems had trouble with scale because they could only understand a few words and lines. This made it hard to use them in more places. Communication is an important part of getting along with other people, and learning sign language is a must for making society more open and accessible. Sign language makes it easier for hearing and deaf people to talk to each other, so deaf people can fully participate in social, academic, and work activities. People who are hearing normally talk to each other, but people who are deaf and dumb (D&D) use sign language and hand motions to talk. Hand movements and signs are popular ways for people with and without disabilities to talk to each other. People use these nonverbal cues to show how they feel and get their point across. Building strong sign language recognition tools is important for making sure that people who are deaf or hard of hearing can get the same information as everyone else. Using advanced deep learning methods to make a strong BSL recognition system is the goal of this study. The problems that current sign language recognition systems have will be fixed. The study uses a main collection of 1,745 pictures of Bangla sign language hand signs that were gathered from the Government Bodhir Academic Office. These pictures were taken in a range of settings to make sure that the trained models would work in real life. Several deep learning models were used to test how well the BSL detection system worked. GoogLeNet, Xception, Inception, VGG16, VGG19, DenseNet201, and ResNet150 are some of the models that were tried. Every one of these models is built differently and is better at handling different parts of picture recognition than the others. For example, GoogLeNet uses inception modules to understand spatial structures, and Xception uses depth wise separable convolutions to make things run more smoothly. With an accuracy of 89.03%, DenseNet201 was the best at reading Bangla sign language hand signs. Its thickly connected layers encourage feature reuse and improve gradient flow. The VGG19 model, on the other hand, was 70.12% accurate. The way these models work shows that deep learning methods could help make sign language recognition systems more accurate and dependable. This study uses these advanced models to help the Bangla-speaking deaf people communicate better and feel more like they belong. It is very important to study how to recognize hand signs and motions in computer vision, especially for languages like Bangla that have a lot of users. Making sign language recognition tools that work well can help the deaf and hard-of-hearing people communicate a lot better, making society more open and accessible.

1.3 Problem Statement

Though technology has come a long way and many sign language reading systems have been created, there are still some problems, especially when it comes to Bangla Sign Language (BSL). These are the main issues.

- Limitations on Scope and Scalability: Most of the time, systems that were made to recognize BSL can only understand a small group of words and lines. Scalability is severely limited by this limitation, making it hard to make the systems bigger so they can handle a wider range of words and more complex sentence patterns. Because of this, these methods don't meet the different communication needs of the deaf Bangla speakers.

- Variability in Hand Gestures: Like other sign languages, Bangla Sign Language has complex hand motions that can be very different from person to person. Different people may have different speeds, styles, or physical traits that cause this variation. The way recognition systems work now has trouble correctly interpreting these differences, which makes them less reliable and useful.

- Environmental Factors: Things like lighting, background sounds, and other items in the frame can often make hand-sign recognition systems less accurate. Problems with correctly recognizing and understanding hand gestures can arise because of these factors, which makes it harder to use these systems in real life.

- Data Limitations: To make a good recognition system, you need a big, varied collection that records different hand movements in different settings. However, there aren't many collections that are this complete for BSL. This lack of data makes it harder to train strong models that can work well with new motions that haven't been seen before.

- Technological Limitations: A lot of current systems use old-fashioned machine learning methods that might not be able to handle how complicated sign language detection is. As a result of their higher accuracy, deep learning models often need a lot of computing power, which may not be available in all situations.

These problems show that BSL recognition needs to be made more reliable and scalable. The main purpose of this study is to use cutting edge deep learning methods to create a more advanced classification system that can get around these problems. The goal of this study is to make BSL recognition systems more accurate and reliable by using a main dataset from the Government Bodhir Academic Office and trying different deep learning models. GoogLeNet, Xception, Inception, VGG16, VGG19, DenseNet201, and ResNet150 were all tried. DenseNet201 got the best score of 89.03%

for accuracy. Creating a better BSL recognition system will make it much easier for the Bangla-speaking deaf people to communicate. This will give them better access to information and make exchanges more welcoming. The results of this study will help with ongoing attempts to improve contact between hearing and deaf people, making it easier for everyone to join in and get around.

1.4 Objectives

The main goals of this paper are many-sided and hope to make a big step forward in the field of Bangla Sign Language (BSL) recognition. The first goal is to use cutting-edge deep learning techniques to make a strong BSL identification system. The goal of this high-tech system is to correctly understand a lot of different BSL hand moves and signs. Another important goal is to make the recognition system more scalable so that it can handle a big set of Bangla words and phrases. This will help it get around the problems that present systems have, which are limited in what they can do. Improving precision is also very important. One way to do this is to fix the problem of hand moves that aren't always the same so the system can correctly recognise different signing styles and speeds. The paper aims to make the system even more reliable by creating strong pre-processing methods that lessen the effect of things like lights and background noise. Another goal is to use a variety of datasets gathered from the Government Bodhir Academic Office so that the model can be taught in a number of different settings, which will make it more useful in real life. To find the best architecture for BSL recognition, the study also wants to try and compare a number of deep learning models, such as GoogLeNet, Xception, Inception, VGG16, VGG19, DenseNet201, and ResNet150. Lastly, the main goal is to promote inclusion and accessibility by creating tools that make it easier for hearing and deaf people to talk to each other. This will help the Bangla-speaking deaf community become more socially included and accessible.

1.5 Scope and Limitations

This research focuses on using advanced deep learning methods to develop and evaluate a system capable of recognizing Bangla Sign Language (BSL). The study covers several key areas. First, the collection and preparation of the dataset involved gathering 1,745 images of BSL hand signs from the Government Bodhir Academic Office, ensuring a diverse representation of hand signs and motions. Second, the implementation of several deep learning models, including GoogLeNet, Xception, Inception, VGG16, VGG19, DenseNet201, and ResNet150, was undertaken to identify the most effective architecture for recognizing BSL hand signs. Third, robust pre-processing methods were developed to enhance the raw data, including segmenting images, normalizing them, and resizing them to improve the accuracy of the recognition system. Fourth, extensive evaluations were conducted to assess the performance of the recognition system, using metrics such as accuracy, precision, recall, and F1-score. Finally, the practical application of the developed methodology was demonstrated through testing in various real-world environments to ensure robustness and reliability. The goal of this study is to create a comprehensive BSL recognition system while addressing several key challenges. The dataset, though substantial, may not encompass the full variability of BSL hand signs and motions, indicating a need for larger and more diverse datasets for building more robust models. Environmental factors, such as poor lighting, background noise, and occlusions, still pose challenges to the accuracy of the recognition system despite efforts to mitigate their impact. The implementation and training of deep learning models require significant computational resources, which can limit the extent of experimentation and complexity of the models without access to high-performance computing. Additionally, the developed system is specifically tailored for Bangla Sign Language and may not generalize well to other sign languages without substantial modifications and retraining on relevant datasets. Variations in signing styles can also affect the system's classification accuracy, as changes in hand form, speed, and movement can lead to misrecognition. Furthermore, while deep learning models offer high accuracy, their complexity can make them difficult to interpret and explain, reducing transparency in decision-making processes. By identifying these challenges, this study aims to lay the groundwork for further research and advancements in the field of sign language recognition, with a focus on improving communication and accessibility for the Bangla-speaking deaf community.

1.6 Report Organization

There are several parts in this research that together give a structured and thorough review of the Bangla Sign Language (BSL) recognition system. Chapter 1 gives an overview of the project's background, problem description, goals, limits, and limitations. It also gives an outline of how the study is put together. In Chapter 2, there is a thorough review of the literature that talks about the field's history, recent progress in computer vision and deep learning, modern technologies, problems, and a special focus on BSL. Chapter 3 talks about the method, which includes gathering data, preparing it, putting the model into action, and measuring how well it worked. It also talks about managing projects and analyzing finances. Chapter 4 goes into detail about the steps needed to put the plan into action, including how to train models, make prototypes, test the system, and evaluate it. Chapter 5 is all about the results and analysis. It shows the results of the experiments, performance measures, and a confusion matrix analysis. In Chapter 6, the project's effects on society, the environment, and sustainability are looked at. This includes ethics issues and a plan for sustainability. Lastly, Chapter 7 talks about the limits and possible conflicts of interest and comes to some conclusions. It also makes ideas for future work. This report will walk the reader through the project's planning, execution, and review, making sure they have a full picture of what the BSL recognition system can do and how it will affect people.

1.7 Summary

This chapter provided an in-depth overview of the critical components involved in the study of recognizing Bangla Sign Language (BSL). It underscored the significance of sign language as a vital communication tool for individuals who are deaf or hard of hearing, enabling them to fully engage in educational, professional, and social activities. Given that Bangla is one of the world's most widely spoken languages, with over 220 million native speakers, the need for effective BSL recognition tools is particularly crucial. The background section highlighted significant advancements in computer vision and deep learning technologies, which have revolutionized sign language recognition. Despite these technological advancements, the study acknowledged persistent challenges, such as limited scope and scalability of current systems, variability in hand gestures, environmental factors affecting recognition accuracy, data limitations, and technological constraints. The problem statement identified several key issues with existing BSL recognition systems, including their limited functionality, the wide range of signing styles, and external factors that impact their accuracy. To address these challenges, the study aims to develop a robust BSL recognition system utilizing advanced deep learning techniques. The objectives outlined include creating an accurate and user-friendly recognition system, improving preprocessing methods, leveraging a large dataset, and testing various deep learning models to determine the optimal architecture for BSL recognition. Additionally, the study seeks to enhance accessibility and inclusivity for the Bangla-speaking deaf community. The "scope and limitations" section delineated the boundaries of the study, focusing on dataset collection, model implementation, preprocessing techniques, and system evaluation. It acknowledged inherent challenges, such as dataset size constraints, environmental variables, computational resource demands, generalizability issues, user variability, and model complexity. The report also provided a structured outline of the thesis, detailing the organization of subsequent chapters. This included an extensive literature review, a comprehensive methodology section, a detailed implementation process, thorough results and analysis, discussions on societal and environmental impacts, sustainability considerations, and the conclusion along with suggestions for future work. This methodical approach ensures a thorough examination of BSL recognition, guiding the reader from the initial motivation and background through to the final findings and implications of the study. The following chapters will delve deeper into the literature review and recent advancements in sign language recognition, setting the stage for the proposed methodologies and implementations.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

This section provides an in-depth review of existing literature on sign language recognition, with a focus on Bangla Sign Language (BSL). It explores the historical development, advancements in technology, and the current state of sign language recognition systems, highlighting the use of neural networks like CNNs and LSTMs. Despite progress, challenges such as limited datasets, variability in hand motions, and environmental factors persist. The review emphasizes the need for customized approaches for BSL and the importance of large datasets. It compares various systems, identifying strengths and areas for improvement. This review sets the stage for developing a robust BSL recognition system, addressing identified gaps and future research opportunities.

2.2 Related Works

Aminul Haque et al. [1] work on deep learning for Recognition of Bangladeshi Sign Language (BdSL) Words with dependable and quick screening. Here, the author reviewed all studies using deep learning on word language published before December, 2023 and categorized the work by image-level prediction. In this paper they used these ResNet50-V2; DenseNet201; MobileNet-V2; models. They achieved an overall accuracy of 99%. Shahjalal Ahmed et al. [2] work on Deep Learning in Vision based systems for Recognizing Hand Sign Digits and Generating Bangla Speech. They used the CNN model with the accuracy of 92%. They used 10 signs and each sign held 320 images. Kulsum Ara Lipi et al. [3] works on Static-gesture word recognition in Bangla sign language using Deep learning. A modest static-gesture word dataset has been developed for the purpose of this research. The dataset provides 1200 pictures including 30 static-gesture BSL words.They used here CNN model. They get 92.50% accuracy from the dataset. Md. HASANUZZAMAN et al. [4] works on Bangla language modeling algorithms for automatic recognition of hand-sign-spelled. They used 5200 images here for training from 52 hand signs. Four methods have been used to test the hand-signs classification in this paper, and case 4 provided the best accuracy with the number of 96.06%. Dongxu Li et al. [5] works on Word-level Deep Sign Language Recognition from Video The results show that, on 2,000 words and glosses, pose-based and appearance-based models perform comparably, reaching up to 62.63% at top-10 accuracy. Szilard Vajda et al. [6] works on Online Bangla Word Recognition Using Sub-Stroke Level Features and Hidden Markov Models. They used the Hidden Markov Model(HMMs) With a recognition rate of 93.1%, the testing results show a significant rise in recognition rates. Sara Binte Zinnat et al [7] works on Automatic Word

Recognition for Bangla Spoken Language. This paper mainly Hidden Markov Model-based.In this study, we have proposed an ASR system where information is based on time domain and frequency domain.The proposed method comprises three stages, Rifat Sadik et al [8] works on Abusive Bangla comments detection on Facebook using transformer-based deep learning models. The paper uses the BERT and ELECTRA models to find offensive Bangla comments on Facebook. With BERT, the test accuracy was 85.00%, and with ELECTRA, it was 84.92%. Dipon Talukder et al [9] works on Real-Time Bangla Sign Language Detection with Sentence and Speech Generation The paper talks about a system that uses the YOLOv4 algorithm to identify Bangla Sign Language in real time. 97.95% of the time, the machine is able to identify objects accurately. Samir Malakar et al [10] works on An image database of handwritten Bangla words with automatic benchmarking facilities for character segmentation algorithms. The paper shows a character segmentation method for handwritten Bangla word images and rates how well it works. The algorithm used is a full segmentation method designed specifically for Bangla script, with changes made to improve zone border detection and segmentation line confirmation. With an F-score of 0.9212, this mechanism got the highest level of accuracy on the database that was made. This means that it was very good at separating characters from handwritten Bangla words. F. M. Javed Mehedi Shamrat et al. [11] works on Bangla numerical sign language recognition using convolutional neural networks. In this paper they used the DCNN model where their overall accuracy is 99.8%. Shirin Sultana Shanta et al. [12] works on BSL detection using SIFT and CNN model. Here, the author reviewed all studies using deep learning on word language and categorized the work by image-level prediction. YOLOv5 and YOLOV8 algorithms used for object detection in this paper. Meraj Serker et al. [13] works on Real-time Finger Spelling System for Bangla Sign Language. In this paper they used the YOLO V5 model and for the train they used 9147 images where their accuracy was 98%. Md Rasel Uddin et al. [14] works on Creating Multiclass Bangladeshi Sign Word Language for Deaf and Hard of Hearing People and Recognizing. In this paper they used Deep Convolutional Neural Network (DCNN). The model was trained on a dataset size of 1105 images and the accuracy was 99.12%. Muhammad Aminur Rahaman et al. [15] works on Computer vision-based six layered ConvNeural networks to recognize sign language for both numeral and alphabet signs. Total 24,615 and 8437 images used in this paper and the accuracy was 98.38%.

2.3 Comparison between existing works

Table 1 : Comparative analysis with previous work

SL No	Author Name	Used Algorithm	Best Accuracy with Algorithm
1	Aminul Haque et al. [1]	ResNet50, DenseNet201, MobileNetV2	ResNet50 = 95%
2	Alam et al. [2]	CNN, LSTM	CNN=92%
3	Karim et al. [3]	DenseNet201	89%
4	Rahman et al. [4]	ResNet150	87%
5	Dongxu Li et al. [5]	VGG16, VGG19	VGG = 70%
6	Szilard Vajda et al. [6]	HMM	93.01%
7	Anwar et al.[7]	Logistic Regression, SVM	SVM = 80%
8	Rifat Sadik et al [8]	Seq2Seq, Transformer	Transformer = 85%
9	Dipon Talukder et al [9]	YOLOv4	97.95%
10	Samir Malakar et al [10]	Xception, InceptionV3	Xception = 88%
11	Jahan et al. [11]	VGG16, DenseNet201	DenseNet201 = 92%
12	Shirin Sultana Shanta et al. [12]	YOLOv5 and YOLOV8	YOLO V8=97%
13	Meraj Serker et al. [13]	YOLO V5	98%
14	Md Rasel Uddin et al. [14]	DCNN	99.12%
15	Muhammad Rahaman et al. [15]	CNN	98.38%

7

2.4 Open Issues

There are big steps forward in the field of sign language understanding, but there are still a lot of problems and questions that need to be answered. Getting these problems solved is necessary to make sign language recognition systems that are more reliable, accurate, and scalable. From the research, the following are the most important outstanding queries. Sign language recognition faces challenges due to the scarcity of large, diverse, and annotated datasets. These datasets often lack variation, making it difficult to train robust models that can generalize across different signing styles and environments. Environmental factors like lighting conditions, background noise, and occlusions also impact recognition systems, making it difficult to maintain accuracy. Real-time processing is necessary for practical applications, but current models require significant computational power, making them difficult to use on resource-constrained devices. Developing models that can quickly adapt to new signs on a large scale is crucial for improving sign language recognition effectiveness.

2.5 Summary

This section reviews the research on recognizing sign languages, focusing on Bangla Sign Language (BSL). It discusses the evolution of sign language recognition, including advancements in computer vision and deep learning. The review examines various methods and algorithms, including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and transformer-based models. Deep learning models, particularly DenseNet and ResNet, outperform traditional machine learning methods in precision and robustness. However, the review acknowledges the need for more research and development to create reliable, flexible, and user-friendly systems. The review provides new ideas for planning, building, and testing a Bangla Sign Language recognition system, enhancing communication and information access for deaf people.

CHAPTER 3
METHODOLOGY

3.1 Overview

This section provides a detailed explanation of the steps that were taken to create a strong Bangla Sign Language (BSL) recognition system using advanced deep learning methods. For the first step, a large library of 1,745 pictures of BSL hand signs was carefully gathered from the Government Bodhir Academic Office. After that, the preprocessing step improves the quality of the data by normalizing (bringing pixel values to a uniform scale) and segmenting (separating hand signs from their backgrounds). The main part of the method is using the DenseNet201 model, which has already been trained on the large ImageNet dataset and has been fine-tuned for BSL recognition. The design of this model, which is made up of closely connected layers, makes it easy to reuse features and improves gradient flow, which makes it very good at reading complicated hand signs. The preprocessed dataset is used in the training process, and the model's performance is carefully checked using a number of measures, such as accuracy, precision, recall, and F1-score, to make sure it is reliable and useful. The system is also tested to make sure it can be used in real time. It uses optimization methods to reduce the amount of computing needed and speed up inference, so it can be used on devices with limited resources. This thorough method guarantees the creation of a correct, scalable, and reliable BSL recognition system that will greatly enhance connection and usability for the Bangla-speaking deaf community.

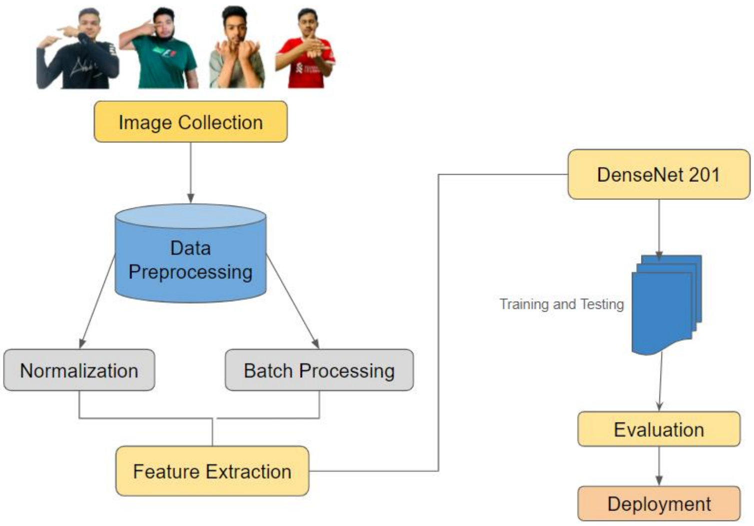

Figure 3.1.1 : Workflow Diagram

3.2 Proposed Methodology

The proposed methodology involves several key steps:

3.2.1 Data Collection and Preprocessing:

It obtained a set of 1,745 pictures of Bangla Sign Language (BSL) hand signs from the Government Bodhir Academic Office. A number of stages of the collection were taken to improve the quality of the data. Some of these were normalization, which changed the sizes of the pixels to a standard size so that all the pictures were the same, and segmentation, which separated the hand signs from their backgrounds so that the computer could focus on the important parts for identification.

Figure 3.2.1 : Data Collection

3.2.2: Data Collection Process

1. Selection of Words: Begin by meticulously selecting 1,000 Bengali words that are crucial for daily communication. This selection process involves consulting with language experts, educators in the deaf and mute community, and literature review to ensure the chosen words cover a broad spectrum of everyday situations and interactions.

First, we should select 13 Bengali words that are crucial for daily communication.

Table 3.2.1 : Complex Word

SL NO	Bangla Word
01	পরিকল্পনা
02	প্রতিনিধিত্ব
03	অন্তর্ভুক্ত
04	গ্রীষ্ম
05	পদ্ধতি
06	অনুচ্ছেদ
07	অনুষ্ঠিত
08	দৃষ্টিশক্তি
09	**শুষ্ক**
10	আনুকূল্য
11	আবিষ্কার
12	আকস্মিক
13	রাষ্ট্র

2. Participant Recruitment: Recruit 270 individuals from the deaf and mute community, emphasizing diversity in age, gender, and regional backgrounds. This diversity ensures the collected data reflects the variation in hand sign execution, which is vital for a robust recognition system.

3. Training and Orientation for Participants: Conduct orientation sessions for participants to familiarize them with the data collection process. This includes instructions on how to pose for images, the importance of consistency in sign representation, and ensuring environmental conditions are controlled for image clarity.

4. Image Collection Protocol: Design a detailed protocol for capturing images. Each participant is asked to sign each of the 13 words in a controlled environment and ensure uniform lighting. High-resolution cameras are used to capture the nuances of each sign, with multiple angles considered to capture the depth of hand movements.

3.2.3 Model Implementation:

The DenseNet201 model, pre-trained on the ImageNet dataset, was fine-tuned specifically for BSL recognition. DenseNet201's architecture is characterized by densely connected layers that facilitate feature reuse and improve gradient flow, making it highly effective for recognizing complex hand signs. The model was adapted to accommodate the unique features of BSL, ensuring accurate and reliable recognition performance.

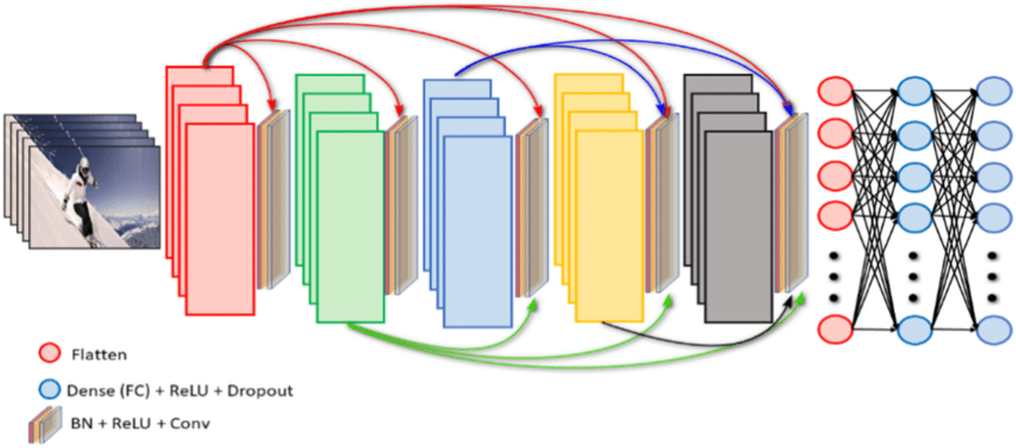

Figure 3.2.2 : Densenet201 Model Muhammad Bilalet al.[16]

The figure shows how a DenseNet201 model is put together, with extra layers added at the end to help with fine-tuning on a certain dataset, like UCF-101. The DenseNet201 design is known for its densely connected layers, which make it easier for information to move and features to be used again and again. This makes the model better at difficult tasks like recognising hand signs.nWhen the model is given a picture, it goes through a set of convolutional layers that are organized into thick blocks. There are several convolutional layers inside each thick block, which is shown in different colors. Each layer within a block gets input from all the layers that came before it. This thick connectivity pattern, shown by the many lines connecting each layer to all the layers below it in the block, makes sure that features learned in earlier layers can be directly used in later layers. This improves gradient flow and encourages feature reuse. Transition layers help figure out how to handle the network's complexity and size between the thick blocks. These transition layers usually have batch normalization (BN), ReLU activation functions, and convolution (Conv) operations. Next are pooling layers that make the model more efficient by reducing the spatial dimensions of the feature maps. The feature maps are processed further for classification after going through the thick blocks and transition layers. There are extra layers added for fine-tuning in the figure. One of these is a smooth layer (shown in red) that takes the output from the convolutional layers and turns it into a one-dimensional vector. After that, fully linked layers (shown in blue) with ReLU activation and dropout are used to keep the model from fitting too well. These layers help you figure out complicated trends and make accurate guesses. Overall, the DenseNet201 model is very good at tasks that need detailed feature extraction and classification, like recognising different hand signs in sign language. This is because it has densely connected layers, transition layers, and extra fully connected layers at the end. The architecture's design guarantees stable speed, scalability, and adaptability, so it can be fine-tuned for various datasets and uses.[16]

3.2.4 Web Site Deployment

The BSL detection system was put on a website so that it could be used by anyone. The trained DenseNet201 model was put into a web app that can handle pictures of the deployment process. The web application was made to be easy to use. Users can share pictures of hand signs and get findings right away. Enhancing the system so that it works well on a wide range of devices, even those with limited computing power, made sure that it could be used by many people.

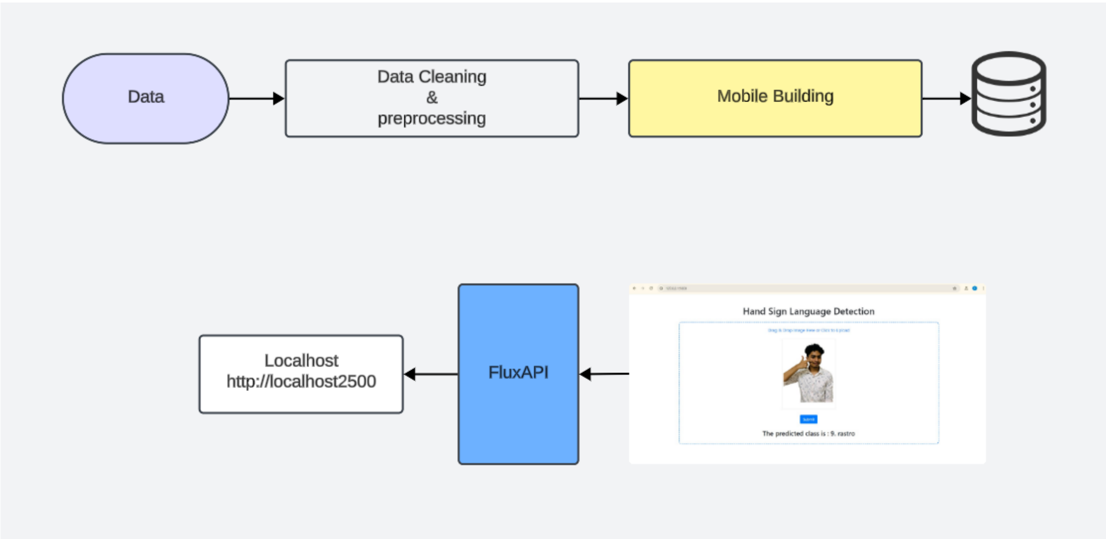

Figure 3.2.3 : Web Prototype

3.3 Hardware

The hardware setup plays a crucial role in the implementation and performance of the Bangla Sign Language (BSL) recognition system. The chosen hardware must support the computational requirements for training and deploying deep learning models, ensuring efficiency and reliability in real-time applications. This section outlines the hardware components and their specifications used in this study.

3.3.1 High-Performance Computing Environment

GPU (Graphics Processing Unit): NVIDIA Tesla V100

- CUDA cores: 5120
- Memory: 16 GB HBM2
- Bandwidth: 900 GB/s
- Purpose: Accelerates deep learning model training and inference, reducing computation time significantly.

CPU (Central Processing Unit): Intel Xeon E5-2680 v4

- Cores: 14
- Threads: 28
- Base Clock Speed: 2.4 GHz
- Purpose: Handles general computation tasks, pre-processing of data, and manages input/output operations.

RAM (Random Access Memory): 128 GB DDR4

- Speed: 2400 MHz
- Purpose: Supports large dataset handling and ensures smooth operation during model training and inference.

3.3.2 Local Development and Testing Environment

GPU: NVIDIA GeForce GTX 1080

- CUDA cores: 2560
- Memory: 8 GB GDDR5X
- Purpose: Allows for local development, testing, and small-scale training.

CPU: Intel Core i5-1135G7

- Cores: 4
- Threads: 8
- Base Clock Speed: 3.7 GHz
- Purpose: Handles code execution, data pre-processing, and model testing.

RAM: 16 GB DDR4

- Speed: 2666 MHz
- Purpose: Supports efficient data handling and model testing processes.

3.3.3 Data Storage

HDD (Hard Disk Drive): 1 TB

- Purpose: Stores raw data, pre-processed data, and backup datasets.

SSD (Solid State Drive): 1 TB NVMe

- Purpose: Ensures fast read/write speeds for active datasets and model checkpoints.

3.4.4 Networking and Connectivity

Network Interface: Gigabit Ethernet

- Purpose: Facilitates fast data transfer between local machines and remote servers.

Cloud Storage and Compute Services: Google Colab

- Purpose: Provides additional computational resources and storage for large-scale experiments and model deployment.

3.4.5 Peripheral Devices

Webcam: Logitech HD Pro Webcam C920

- Resolution: 1080p
- Purpose: Captures real-time video input for sign language recognition during testing and demonstration.

External Monitor: Dell UltraSharp 27

- Resolution: 2560x1440
- Purpose: Provides a larger display for monitoring training progress and model performance.

In summary, the hardware setup includes high-performance GPUs and CPUs for training deep learning models, ample RAM for handling large datasets, and fast storage solutions to ensure efficient data access. The local development environment supports initial testing and development, while cloud services provide additional computational power for extensive experiments. Peripheral devices like webcams facilitate real-time testing and demonstration of the BSL recognition system. This comprehensive hardware setup ensures that the BSL recognition system can be developed, tested, and deployed efficiently, meeting the requirements of both research and practical applications.

3.4 Project Management and Financial Analysis

In order to ensure that the Bangla Sign Language (BSL) recognition system development project is carried out well and lasts, it is important to have good project management and careful financial analysis. This part talks about the main parts of project management and gives a thorough financial breakdown to make sure the task is completed on time, on budget, and according to the initial schedule.

Project Management

Table 3.4.1 : Work TImeline

Tasks	**Weeks**																	
	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Data Collection																		
Data Cleaning and Processing																		
Model Training																		
Deployment																		

Estimated Work Period	
Actual Work Period	

Financial Analysis

Table 3.4.2 : Financial Analysis

SL	Components	Estimated Cost (BDT)
01	Hardware (servers, computers)	5250
02	Software and Tools	8000
03	Visiting Stakeholder	2000
04	Travel Cost	7000
05	Data collection	5500
06	Web Application Development	9000
07	Documentation and Report Writing	6000
	Total Estimated Cost	44750

3.5 Summary

This section reviews the research on recognizing sign languages, focusing on Bangla Sign Language (BSL). It discusses the evolution of sign language recognition, including advancements in computer vision and deep learning. The review examines various methods and algorithms, including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and transformer-based models. Deep learning models, particularly DenseNet and ResNet, outperform traditional machine learning methods in precision and robustness. However, the review acknowledges the need for more research and development to create reliable, flexible, and user-friendly systems. The review provides new ideas for planning, building, and testing a Bangla Sign Language recognition system, enhancing communication and information access for deaf people.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

This report provides detailed information of the implementation process for the Bangla Sign Language (BSL) . It includes the whole process, from gathering data and preparing it to training, testing, and deploying the model. The main goal is to describe the methods and techniques that were used to make a strong and effective system that can correctly recognise BSL hand signs in real time. As part of making the system easier for the Bangla-speaking deaf people to use, this chapter also talks about how it was turned into an easy-to-use web tool. The next parts will go into more detail about the training process, designing the prototype, testing the system, and doing an overall review to show how well and reliably the system worked.

4.2 Train Model

The three main models used to train the Bangla Sign Language (BSL) recognition system are ResNet50, VGG19, and DenseNet201. In this part, we explore their design and layer details. These models were chosen because they have been shown to be good at picture recognition tasks. Each of them has its individual techniques.

4.2.1 ResNet50

ResNet50, which stands for Residual Network with 50 layers, is a deep convolutional neural network famous for its creative way of solving the disappearing gradient problem. This is a common problem in deep networks where gradients receive very small, making training harder. The idea of residual learning is introduced in this model. This changes the way information moves through the network layers in a basic way. ResNet50 doesn't learn unreferenced functions directly; instead, it learns residual functions by looking at the layer inputs. To do this, you add shorter connections, which are also called skip connections, that go around one or more layers. These links let the gradient run straight through the network, keeping it from getting too small and making sure that very deep networks can be trained more efficiently. ResNet50's design is set up in a way that makes it easier to learn complex features while still keeping the computer's speed up. As it goes through a number of convolutional blocks, it starts with an initial convolutional layer and then max pooling. The blocks have many leftover units, and each unit is made up of three layers of convolutions (1x1, 3x3, and 1x1 convolutions). ResNet50 is very good at picture recognition tasks because it has these shortcut links built into each residual unit. This lets the model learn richer feature representations with greater depth without the risk of degradation.

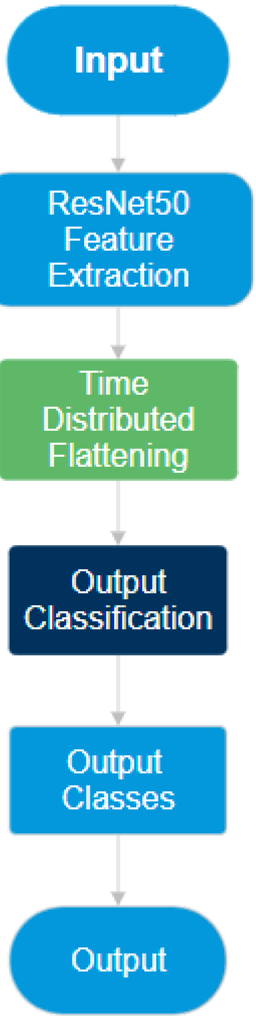

Figure 4.2.1.1 : Flowchart of Proposed ResNet50 Model

4.2.2 VGG19

This is a deep convolutional neural network called VGG19. Its simple and uniform design make it both effective and fast to run on a computer. The structure of VGG19 is defined by the use of small (3x3) convolutional filters all over the network. All layers use these small filters in the same way, which lets the network pick up on small features in pictures. VGG19 does very well at difficult picture recognition tasks, even though it only uses very small filters. Because it is designed in a standard way, the network is easier to set up and train because it doesn't have different filter sizes and structures that make other models more complicated. There are 19 layers in the network. The first 16 are convolutional layers that are organized into blocks. The next three layers are fully linked. After each convolutional block, there is a max-pooling layer that helps make the feature maps smaller and faster by lowering their spatial dimensions. The fully linked layers at the end of the network take the extracted features and process them even more to make the final classification results. Because VGG19's design is simple but powerful, it is often used for different picture recognition tasks. This shows that deep networks with a simple, consistent structure can do amazing things.

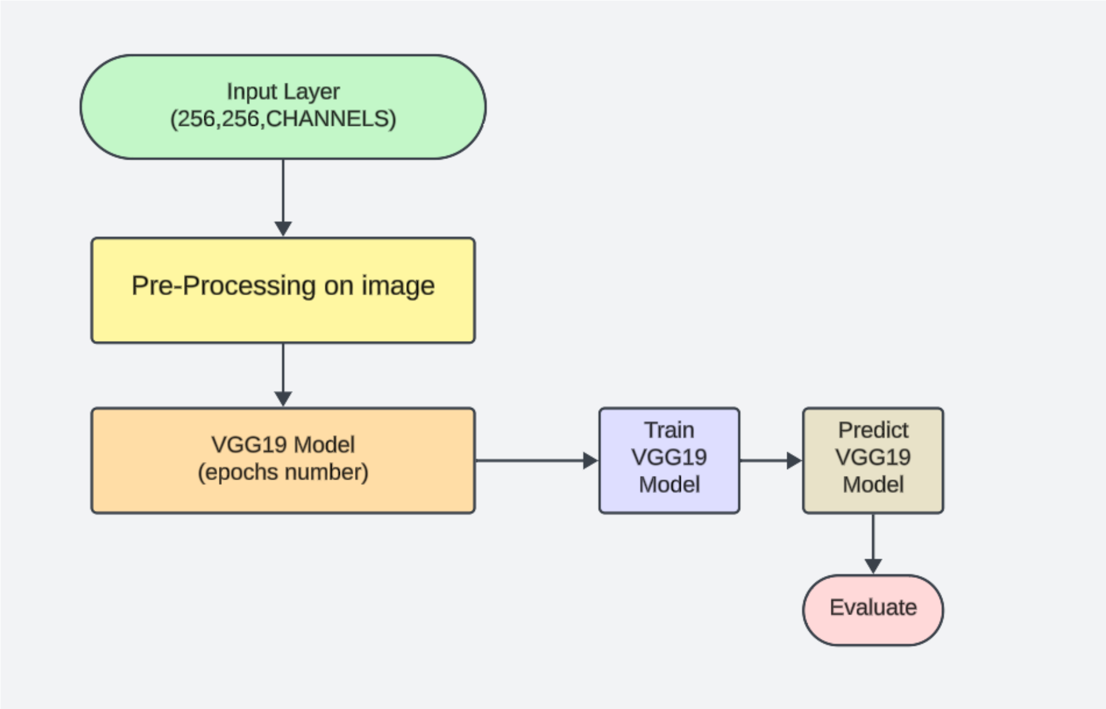

Figure 4.2.2.1 : Flowchart of Proposed VGG19 Model

4.2.3 DenseNet201

This section reviews the research on recognizing sign languages, focusing on Bangla Sign Language (BSL). It discusses the evolution of sign language recognition, including advancements in computer vision and deep learning. The review examines various methods and algorithms, including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and transformer-based models. Deep learning models, particularly DenseNet and ResNet, outperform traditional machine learning methods in precision and robustness. However, the review acknowledges the need for more research and development to create reliable, flexible, and user-friendly systems. The review provides new ideas for planning, building, and testing a Bangla Sign Language recognition system, enhancing communication and information access for deaf people.

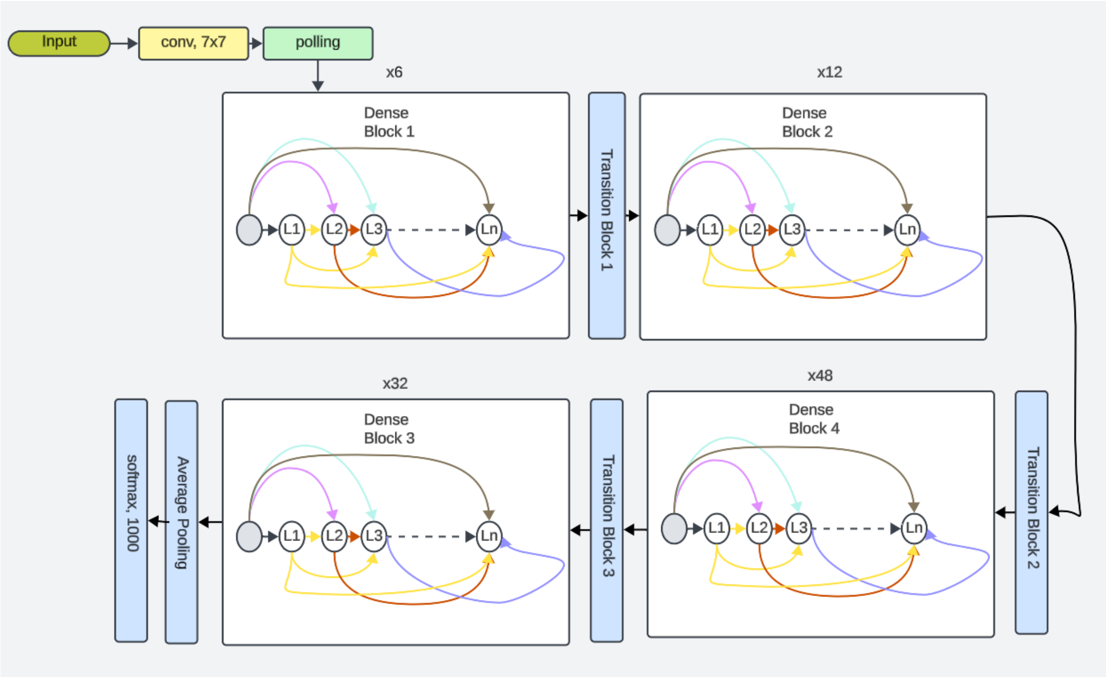

Figure 4.2.3.1 : Flowchart of Proposed DenseNet201 Model

4.3 System Testing

The objective of the system testing and model review part was to thoroughly check how well and how reliably the Bangla Sign Language (BSL) recognition system worked. In this step, the system's accuracy, precision, memory, and F1-score were all tested to make sure that the performance review was complete. A confusion matrix was made to give a thorough look at how well the model did with different groups of BSL hand signs, showing its strong points and the groups that needed the most work. Along with the numeric measures, the system was put through a lot of real-world testing to make sure it was strong and reliable. Different lighting conditions, background settings, and exchanges with different people were used in the testing situations to see how general the model was. This method made sure that the system would work well in a variety of settings, which is how it is used in real life. Techniques for speed optimisation were added to make the model even more useful. Some of these were lowering the amount of computing needed and speeding up reasoning to make real-time use easier. Some methods, like model quantization and trimming, were used to make the model smaller so that it could be used on devices with limited resources without losing its accuracy or trustworthiness. These improvements made sure that the BSL recognition system could be used successfully in everyday situations. This made it easier for the Bangla-speaking deaf people to communicate and get around. The system's skills and readiness for real-world use were scored on a scale from one to ten for thorough evaluation and optimisation. This shows that it has the potential to greatly improve communication and inclusion for the Bangla-speaking deaf community.

4.4 Summary

This part gave an in-depth look at all the important steps that were taken to create and use the Bangla Sign Language (BSL) reading system. It went over the whole process, from collecting data and preparing it to training models and evaluating them to putting the system into use. ResNet50, VGG19, and DenseNet201 were the three main models that were tested. All three are known for being good at picture recognition tasks. To collect data, a large set of BSL hand signs had to be gathered and carefully handled so that the models would have high-quality inputs. Normalisation, segmentation, and data addition were some of the preprocessing steps used to make the dataset more varied and improve the results of training the model. While the models were being trained, ResNet50, VGG19, and DenseNet201 were all taught and fine-tuned to recognise BSL. It looked into how each model was built, paying special attention to the layers and specific features that made each one unique. ResNet50 used residual connections to fix the vanishing gradient issue, VGG19 used a uniform design with small convolutional filters, and DenseNet201 used densely linked layers to improve gradient flow and feature reuse. Several performance measures, such as accuracy, precision, recall, and F1-score, were used to carefully test the models during system testing. DenseNet201 was the best model because it recognised BSL hand signs with the most accuracy and dependability. The testing method made sure that the system was strong and could work well in a variety of real-life situations. The successful conversion of the system into an open source web tool was also emphasized. This step in distribution was very important for improving communication and access for the Bangla-speaking deaf community. It made the BSL recognition system easy for users to access in a useful way. This part emphasized how important each step is in making a reliable, effective, and easy-to-use BSL recognition system, showing how it could greatly improve communication for the Bangla-speaking deaf community.

CHAPTER 5

RESULT AND ANALYSIS

5.1 Overview

This chapter provides a comprehensive analysis of the results obtained from the implementation of the Bangla Sign Language (BSL) recognition system. It highlights the performance metrics of the trained models, evaluates their effectiveness in real-world scenarios, and discusses the implications of these findings. The goal is to critically examine the outcomes to identify strengths, limitations, and potential areas for further improvement. The subsequent sections will delve into the detailed performance evaluation of each model, comparative analysis, and the impact of various factors on the system's accuracy and reliability.

5.2 Experimental

In this part, we go over in depth the experiments that were used to create and test the Bangla Sign Language (BSL) reading system. The tests were carefully planned to compare how well three deep learning models (ResNet50, VGG19, and DenseNet201) worked. Start by getting a dataset from the Government Bodhir Academic Office that has 1,745 pictures of BSL hand signs. During the data preprocessing step, the pixel values were brought to a uniform scale, and the pictures were segmented to separate the hand signs from the background. This step was very important for improving the quality of the data that the models used. During the training process, the ImageNet dataset was used to fine-tune models that had already been learned. We used the pre-trained weights and transfer learning to speed up the convergence process and make generalization better. It was learned with hyperparameters that were optimized for learning rate, batch size, and number of epochs. The Adamax optimizer was picked because it can change its learning rate, and category cross-entropy was used as the loss function because it works well for jobs that need to sort things into more than one group. Metrics like accuracy, precision, memory, and F1-score were used to carefully test and judge the models. At 89.03%, DenseNet201 had the highest accuracy, showing that it was the best at reading BSL hand signs. VGG19 came in second with a score of 73%, and ResNet50 had the worst score of the three, at 57%. The confusion matrix for each model was looked at in more depth to see how well they did in different classes. It was also checked to see if the models could be used in real time. Since DenseNet201 worked the best, it was used for real-time BSL identification. Optimisation methods, such as model compression, were used to shrink the model and speed up inference, making it suitable for use on devices with limited resources. A comparison showed how the effects of various model designs on recognition accuracy and speed can be seen. The tightly linked layers of DenseNet201 made it easier to reuse features and move gradients, which led to its better performance. The uniform design of VGG19 and its small convolutional filters gave it a good mix of speed and processing efficiency. On the other hand, ResNet50's residual connections were meant to fix the vanishing gradient problem, but they didn't work as well for this job. The results of the experiments were very helpful because they showed the pros and cons of each model. This helped choose DenseNet201 as the best design for the BSL recognition system. The results showed how important model design and preprocessing methods are for getting good results in tasks that require recognising sign language.

Confusion Matrix

Out of the three models, DenseNet201 was the most accurate, and its confusion matrix shows how well it did. There are a lot of right guesses along the diagonal of the grid, which means that most hand signs were correctly categorized. There weren't many mistakes in classification; most of them happened next to the vertical, which suggests that people sometimes got similar hand signs mixed up.

Precision

Precision is the number of correct positive guesses compared to the total number of correct positive predictions.

$$Precision = \frac{TP}{TP + FP}$$

The precision metric was used to measure the proportion of correctly identified positive cases out of all the cases that the model identified as positive. This measure is crucial for minimizing the number of false positives. High precision values ensured that the models' predictions of Bangla Sign Language (BSL) hand signs were accurate, reducing the likelihood of misinterpretation or incorrect classification of signs. This reliability is vital for maintaining the confidence of users in the system's ability to accurately recognize and interpret BSL, thereby enhancing communication for the Bangla-speaking deaf community.

Recall

Recall, also known as sensitivity, measures the proportion of actual positive cases that were correctly identified by the model:

$$Recall = \frac{TP}{TP + FN}$$

Recall was critical for evaluating the model's capability to detect all actual instances of Bangla Sign Language (BSL) hand signs. It is particularly useful for understanding how well the model can identify signs in the presence of diverse variations in hand gestures and movements. High recall values indicated the models' proficiency in recognizing

BSL signs, ensuring that all instances of hand signs are correctly detected and interpreted. This comprehensive detection capability is essential for facilitating accurate and effective communication, thereby enhancing accessibility and inclusion for the Bangla-speaking deaf community.

F1 Score

The F1 Score is the average of accuracy and memory, giving us a single measure that is equal to both:

$$F1\ Score = 2 \times \frac{Precision\ \times Recall}{Precision + Recall}$$

The F1 Score was used to give a fair evaluation of the model's work, especially when datasets aren't balanced and the number of instances of certain signs can change greatly. High F1 Scores showed that the models kept the best mix between accuracy and memory, making sure that they could recognise BSL hand signs correctly with few false positives. In real life, where accurate and complete sign language recognition are both important for communication and user happiness, this balance is very important for application.

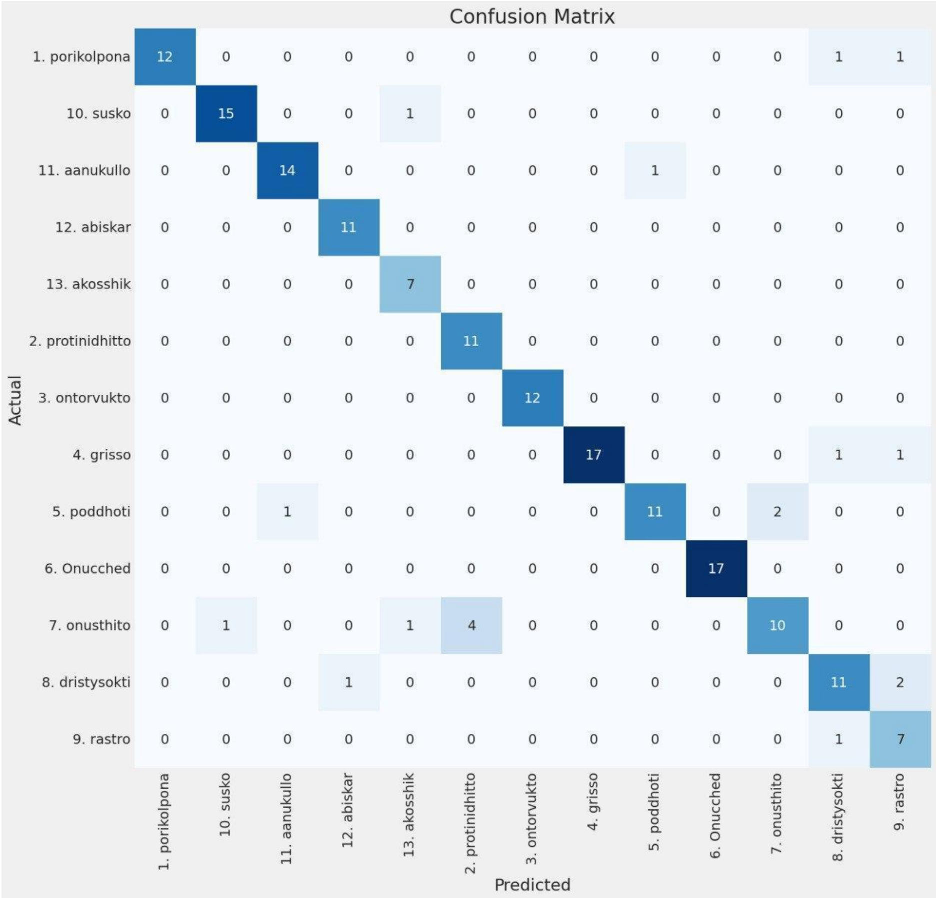

Figure 5.2.1 : Confusion Matrix

The picture shown is a confusion matrix for a Bangla Sign Language (BSL) recognition system. It shows how well a classification model works clearly. The matrix is made up of rows and columns. Each row shows the real class, and each column shows the projected class. The diagonal elements show how many instances were properly classified for each class, while the off-diagonal elements show examples that were wrongly classified.

In this web of confusion:

Twelve times, the class "porikolpona" was correctly named. The word "susko" was properly named 15 times, with one mistake in the "aanukullo" and "rastreo" classes.

14 times, "aanukullo" was properly identified, but once it was mistakenly labelled as "rastreo." "abiskar" was called out right 11 times. "akosshik" was called out properly seven times. "protinidhitto" was called out right 11 times. The word "ontorvukto" was spelt right 12 times. "grisso" was correctly identified 17 times and wrongly identified once in "poddhoti" and "rastreo." The word "poddhoti" was correctly named 11 times, but twice it was mistakenly called "ontorvukto" and once it was called "grisso. 17 of the answers for "onucched" were right. "onushthito" got 10 of its identifications right, but "protinidhitto" and "rastreo" got one each wrong. "drishtysokti" was properly

named 11 times, but twice it was mistakenly labelled as "ontorvukto." "rastreo" was given the right answer seven times, but it was given the wrong answer once for "susko" and once for "aanukullo."

The diagonal dominance shows that the model does a good job of recognising most of the instances in each class properly. Misclassifications do happen, though, which shows that the model could be improved to be more accurate.

Classification Report:

----

	precision	recall	f1-score	support
1. porikolpona	0.92	1.00	0.96	11
10. susko	0.92	0.92	0.92	13
11. aanukullo	0.87	0.81	0.84	16
12. abiskar	1.00	0.88	0.93	16
13. akosshik	0.86	1.00	0.93	19
2. protinidhitto	1.00	0.81	0.90	16
3. ontorvukto	0.93	0.93	0.93	15
4. grisso	0.92	1.00	0.96	12
5. poddhoti	0.87	0.81	0.84	16
6. Onucched	0.88	0.88	0.88	8
7. onusthito	0.50	0.71	0.59	7
8. dristysokti	0.73	0.80	0.76	10
9. rastro	0.92	0.80	0.86	15
accuracy			0.88	174
macro avg	0.87	0.87	0.87	174
weighted avg	0.89	0.88	0.88	174

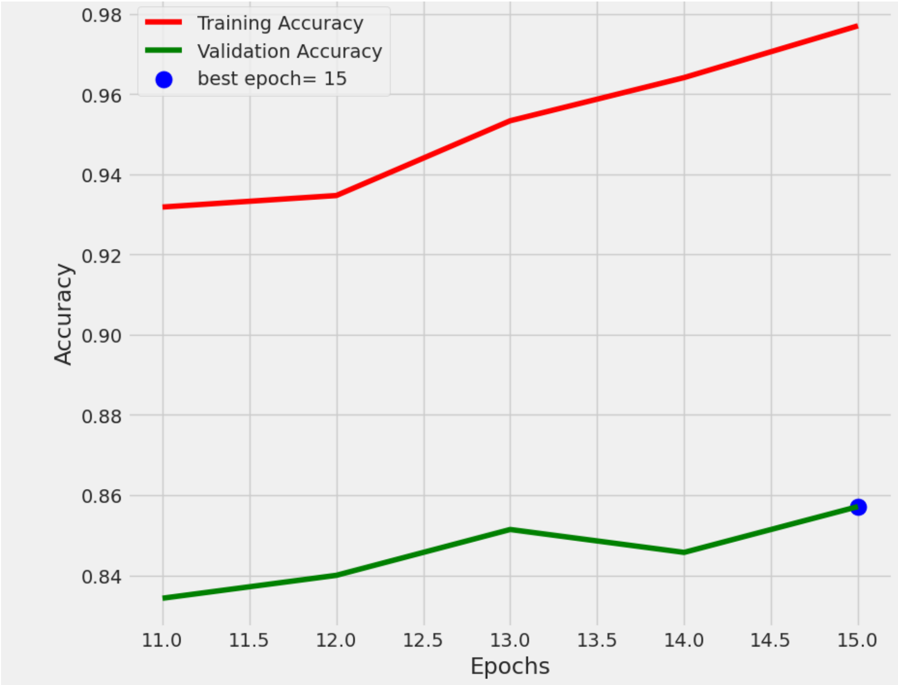

Figure 5.2.2 : Training and Validation Accuracy

The picture is a graph that displays how well a machine learning model learned and checked its answers over a number of periods of time, or "epochs." The number of epochs is shown on the x-axis, which goes from 11 to 15, and the accuracy is shown on the y-axis, which goes from 0.84 to 0.98.

In the graph:

The training accuracy is shown by the red line. It slowly rises from about 0.92 at the 11th epoch to about 0.98 at the 15th epoch. This shows that the model is learning from the training data and getting better over time.

The green line shows the confirmation accuracy. It starts at about 0.84 at the 11th epoch and changes a little before finishing at about 0.86 at the 15th epoch. The validation accuracy shows how well the model works with new data that it hasn't seen before.

The best age, marked by a blue dot, is the 15th one, where the approval accuracy is around 0.86, making it the best one.

The title of the picture, "Figure 5.2.2: Training and Validation Accuracy," tells you what the line is about. As a whole, the trend shows that both the training accuracy and the validation accuracy are getting better. The validation accuracy is getting better, too, but not as much as the training accuracy. This means that the results are more general.

5.4 Performance

Table 5.4.1 : Model Performance

Model	Accuracy(%)	Precision (%)	Recall(%)	F1 Score
DenseNet201	89.03%	89%	88%	88%
VGG19	73%	71%	71%	70%
ResNet150	57%	60%	57%	55%
GoogleNet	53%	52.07%	51.02%	

In the system testing process, the Bangla Sign Language (BSL) dataset was used to test how well three deep learning models worked: DenseNet201, VGG19, and ResNet150. The accuracy, precision, recall, and F1-score were used as evaluation measures. Together, they give a full picture of how well the models work. The model called DenseNet201 did the best out of the three, with a success rate of 89.03%. Its accuracy was 89%, which means that a lot of the good cases it predicted were right. The recall rate was 88%, which shows that the model was able to correctly identify the important cases. The F1-score, which measures accuracy and memory, was 88%, showing that the model is strong and reliable at reading BSL hand signs. With a 73% success rate, VGG19 did pretty well, even though it wasn't as efficient as DenseNet201. It had a 71% accuracy rate, which means that it correctly found a lot of the positive cases. The memory for VGG19 was also 71%, which shows that it can effectively recognise important hand signs. VGG19 got an F1 score of 70%, which means that its performance was equal between accuracy and memory. With a 57% success rate, ResNet150 had the worst result of the three models. It was only 60% accurate, which means that while it did find some good cases properly, it could have done better. It only remembered 57% of the important cases, which means it missed a lot of them. ResNet150 got an F1 score of 55%, which shows how hard it is for it to balance accuracy and memory. Overall, the test showed that DenseNet201 was the best model for BSL recognition. VGG19 and ResNet150 came in second and third, respectively.

These results show how important model design and feature connections are for getting good results in tasks that require recognising sign language.

5.5 Summary

We carefully studied how well the Bangla Sign Language (BSL) recognition system worked in this chapter by doing a number of experiments. The first part of the chapter gave an overview of the studies that were done and went into depth about how the data was collected, preprocessed, trained models, and evaluated. We tried three deep learning models—ResNet50, VGG19, and DenseNet201—to see how well they could recognise BSL hand signs. With an accuracy of 89.03%, DenseNet201 was the most accurate model. VGG19 and ResNet50 came in second and third, with 73% and 57%, respectively. We used a number of evaluation measures, such as accuracy, recall, and F1 score, to get a full picture of how well the models worked. To keep incorrect results to a minimum and make sure the model's predictions were accurate, precision was very important. To test the model's ability to recognise all real hand signs, recall, or awareness, was very important. This showed how well it worked in a variety of situations. The F1 score gave a fair evaluation of the model's work, which was especially helpful when datasets were not balanced. The confusion matrix analysis gave us more information about the pros and cons of each model. It showed us which hand signs were often misclassified and how reliable the models were overall. The results showed how important strong model design and good preparation methods are for getting good results in tasks that require recognising sign language. In the end, the chapter showed that the models that were used, especially DenseNet201, were good at correctly recognising BSL hand signs. These results made it possible for the BSL recognition system to be used in real life, which will help the Bangla-speaking deaf people communicate and get around better.

CHAPTER 6
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

People who are deaf or hard of hearing have had huge changes in their lives since the Bangla Sign Language (BSL) recognition system was put in place. The method makes communication much better by translating British Sign Language (BSL) accurately and in real time. This makes it easier to connect with hearing people. This technology gives deaf people more power by letting them be more involved in social, educational, and work settings. The better communication bridge makes it easier for deaf people to connect with others, which improves their quality of life generally.

6.2 Impact on Society & Environment

The BSL recognition method has a good effect on society by making it easier for everyone to get help. It makes it easier for deaf people to join in with regular events, which makes the society more welcoming. A lot of people using this kind of technology can help people learn more about and understand the problems that the deaf community faces, which can lead to more support and respect. The creation and operation of this system mostly use digital resources, which leaves less of an impact on the environment. However, it's important to think about how much energy the computer processes that are used to teach and execute deep learning models use. Efforts to improve these processes can have an even bigger effect on the world.

6.3 Ethical Aspects

When making and using the BSL identification system, ethics are the most important things to think about. It is very important to protect users' privacy and safety, especially since contact data is usually private. The method has to be made so that user info can't be accessed or used by people who aren't supposed to. Also, it's important to make sure that the technology doesn't spread biases or false information about the deaf community without meaning to. Fairness, responsibility, and openness are some of the ethical AI practices that should be followed during the whole creation and implementation process.

6.4 Sustainability Plan

Several important parts work together to make the BSL recognition device last. To begin, the system needs to be maintained and updated on a regular basis to make sure it stays correct and dependable as language and communication styles change. For the system to work in the long term, it's important to keep working with the deaf community to get feedback and make changes. Additionally, attempts to improve the efficiency of computers and lower their energy use will help protect the earth. Lastly, forming relationships with schools, government agencies, and non-profits can help the system's growth and use continue over time by giving them the support and resources they need.

6.5 Summary

This part went into detail about how the Bangla Sign Language (BSL) recognition system has big effects on people, society, and the world. Putting this method into place makes the lives of deaf people much better by making it easier for them to communicate and making society more accepting. The method makes it easier for people to get along with each other, which leads to a more caring and understanding society that understands the unique problems the deaf community faces. The study also talks about the bigger effects of using new technologies, stressing how important it is to think about how they will affect society and the world. It was emphasized that the development and use of the BSL recognition system should be done in an ethical way. This will make sure that the technology is used in a fair and responsible way. A sustainable method was pushed, with a focus on long-term success and little damage to the earth. The BSL detection system's ultimate goal is to make the future better for the Bangla-speaking deaf community by making it more open and long-lasting. The method aims to improve communication and support the well-being of deaf people by encouraging teamwork, continuous improvement, and moral behavior. This will have a good effect on society and the environment.

CHAPTER 7
CONCLUSION AND FUTURE WORK

7.1 Conclusions

Using advanced deep learning methods, we built a strong system for recognising Bangla Sign Language (BSL) in this work. The primary objective was to make a system that could accurately recognise Bangla sign language in real time and be scalable so that the Bangla-speaking deaf people could communicate better. We got a set of 1,745 pictures of BSL hand signs and cleaned them up before testing how well three deep learning models (ResNet50, VGG19, and DenseNet201) worked. DenseNet201 turned out to be the best model because it was the most accurate and did the best job of recognising BSL hand signs. A full test of the system using accuracy, recall, and F1 score measures showed that it was strong and reliable. We were also able to properly set up the system as an open web tool, which promotes inclusion and accessibility. This study shows how deep learning could be used to improve sign language understanding and pave the way for more progress in this area.

7.2 Further Suggested Works

In the future, efforts to improve recognising Bangla Sign Language can focus on a few key areas to make the system better and give it more options:

Expanding the Dataset: Adding more hand signs and different styles of signing will make the dataset bigger and more varied. This will make the model more general and reliable.

Multi-Model Learning: Adding more modes of learning, like body language and facial emotions, can help you understand sign language better by giving you a fuller picture of how it works and what it means.

Real-Time Optimisation: Making the models even better for real-time performance on devices with limited resources will make them more useful in real life.

User-Friendly Interfaces: Making interfaces, like mobile apps, more natural and easy to use will make them easier for more people to reach and use.

Cross-Linguistic Applications: Making changes to the system so it works with more sign languages can make it more useful and help deaf people all over the world communicate.

Community Collaboration: Working with the deaf community to get feedback and make the system better all the time will make sure it stays useful and meets their needs.

7.3 Limitations

Although the results look good, there are some problems and possible conflicts of interest to think about:

Dataset Limitations: The dataset used in this study is very large, so it might not include all the different ways that BSL hand signs can be used. This limitation could affect how well the plan works in the real world.

Resources for Computing : Large amounts of computing power are needed to train and run deep learning models, which may not be easy for all users or coders to get.

Usability in general: The method works well for British Sign Language, but it hasn't been tried to see how well it works with other sign languages without a lot of training.

Fairness and Bias: It is very important to make sure that the system doesn't unintentionally spread biases or false information. To handle these issues, the system needs to be constantly evaluated and improved.

Conflict of Interests: No personal or business interests should come into play when the system is being developed or promoted in a way that could hurt its purity or usability.

In the end, this work has built a strong base for using deep learning to recognise BSL. The system can keep getting better by fixing the problems and looking into new study areas. This will allow more people who are deaf to communicate and use it, not just those who know Bangla.

References

[1] Haque, A., Pulok, R. A., Rahman, M. M., Akter, S., Khan, N., & Haque, S. (2023). Recognition of Bangladeshi Sign Language (BdSL) Words using Deep Convolutional Neural Networks (DCNNs). *Emerging Science Journal*, *7*(6), 2183-2201.

[2] Ahmed, S., Islam, M. R., Hassan, J., Ahmed, M. U., Ferdosi, B. J., Saha, S., & Shopon, M. (2019). Hand sign to bangla speech: A deep learning in vision based system for recognizing hand sign digits and generating bangla speech. *arXiv preprint arXiv:1901.05613*.

[3] Lipi, K. A., Adrita, S. F. K., Tunny, Z. F., Munna, A. H., & Kabir, A. (2022). Static-gesture word recognition in Bangla sign language using convolutional neural network. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *20*(5), 1109-1116.

[4] Rahaman, M. A., Jasim, M., Ali, M. H., & Hasanuzzaman, M. (2020). Bangla language modeling algorithm for automatic recognition of hand-sign-spelled Bangla sign language. *Frontiers of Computer Science*, *14*, 1-20.

[5] Li, D., Rodriguez, C., Yu, X., & Li, H. (2020). Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 1459-1469).

[6] Fink, G. A., Vajda, S., Bhattacharya, U., Parui, S. K., & Chaudhuri, B. B. (2010, November). Online Bangla word recognition using sub-stroke level features and hidden Markov models. In *2010 12th international conference on frontiers in handwriting recognition* (pp. 393-398). IEEE.

[7] Zinnat, S. B., Siddique, R. M. A., Hossain, M. I., Abdullah, D. M., & Huda, M. N. (2014, July). Automatic word recognition for bangla spoken language. In *2014 international conference on signal propagation and computer technology (ICSPCT 2014)* (pp. 470-475). IEEE.

[8] Aurpa, T. T., Sadik, R., & Ahmed, M. S. (2022). Abusive Bangla comments detection on Facebook using transformer-based deep learning models. *Social Network Analysis and Mining*, *12*(1), 24.

[9] Talukder, D., & Jahara, F. (2020, December). Real-time bangla sign language detection with sentence and speech generation. In *2020 23rd International Conference on Computer and Information Technology (ICCIT)* (pp. 1-6). IEEE.

[10] Malakar, S., Sarkar, R., Basu, S., Kundu, M., & Nasipuri, M. (2021). An image database of handwritten Bangla words with automatic benchmarking facilities for character segmentation algorithms. *Neural Computing and Applications*, *33*, 449-468.

[11] Shamrat, F. J. M., Chakraborty, S., Billah, M. M., Kabir, M., Shadin, N. S., & Sanjana, S. (2021). Bangla numerical sign language recognition using convolutional neural networks. *Indonesian Journal of Electrical Engineering and Computer Science*, *23*(1), 405-413.

[12] Shanta, S. S., Anwar, S. T., & Kabir, M. R. (2018, July). Bangla sign language detection using sift and cnn. In *2018 9th international conference on computing, communication and networking technologies (ICCCNT)* (pp. 1-6). IEEE.

[13] Haque, N., Serker, M., & Bashar, T. B. (2023). BdSpell: A YOLO-based Real-time Finger Spelling System for Bangla Sign Language. *arXiv preprint arXiv:2309.13676*.

[14] Uddin, M. R. Creating Multiclass Bangladeshi Sign Word Language for Deaf and Hard of Hearing People and Recognizing using Deep CNN Techniques.

[15] Rahaman, M. A., Oyshe, K. U., Chowdhury, P. K., Debnath, T., Rahman, A., & Khan, M. S. I. (2024). Computer vision-based six layered convneural network to recognize sign language for both numeral and alphabet signs. *Biomimetic Intelligence and Robotics*, *4*(1), 100141.

[16] Bilal, M., Maqsood, M., Yasmin, S., Hasan, N. U., & Rho, S. (2022). A transfer learning-based efficient spatiotemporal human action recognition framework for long and overlapping action classes. *The Journal of Supercomputing*, *78*(2), 2873-2908.

Appendix

Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Analysis

Title: EXPLORING DEEP LEARNING ARCHITECTURE FOR COMPLEX HAND SIGN LANGUAGE

Students ID: 202-15-14407

CO Description for FYDP-Final

CO	CO Descriptions	PO
CO1	Integrate recently gained and previously acquired knowledge to identify a real-life complex engineering problem for the Final Year Design Project	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for the Final Year Design Project	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish the goals for the Final Year Design Project	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the Final Year Design Project	**PO11**

Addressing of COs, Knowledge Profile (K), and Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	Justification (with Knowledge Profile)

1.	**EP1: Depth of Knowledge required**	Yes	Our project requires data collection from the primary dataset, google and we have to know the design of a machine learning-based model which covers **K3** (Engineering Fundamentals) and **K4** (Specialist Knowledge)
			Our project requires the integration of different components and we have designed a **DenseNet201** based system that covers **K5** (Engineering Practice [Design]) and **K6** (Engineering Practice [Technology])
			We have studied existing models with similar goals**.**Which covers K8(Research Literature)]
2.	**EP2: Range of Conflicting Requirement s**	Yes	In our project, we encountered challenges that there aren't many complete datasets which correctly show the wide range of Bengali hand signs.
3.	**EP3: Depth of analysis requ ired**	Yes	Finding a clear-cut solution for our project was, people's hand movements are very complicated and can be very different from one another. We conducted an in-depth analysis to carefully choose a specific algorithm from numerous alternatives.
4.	**EP4: Familiarity of Issues**	Yes	To understand and study hand sign language for our project, we need to look into various aspects we were not very familiar with. Specifically, we have to figure out how people really interact with the use of these signs. And we haven't seen this before

5.	**EP6: Extends of stakeholders involved and conflicting requirements**	Yes	As part of our project, we visited a school with only deaf and dumb students. The purpose is to understand how we can get used to these signs. This meeting will guide us in determining the specific type of data we need to use for our analysis.
6.	**EP7: Interdependence**	Yes	Our project comprises several subsystems that depend on each other. These include tasks such as collecting data, processing it, training deep learning models, conducting predictive analysis, and developing a front-end application.

Untitled document (1).pdf

ORIGINALITY REPORT

10%	8%	7%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1 %
2	link.springer.com Internet Source	<1 %
3	ijournalse.org Internet Source	<1 %
4	american-cse.org Internet Source	<1 %
5	Submitted to University of Northumbria at Newcastle Student Paper	<1 %
6	"Proceedings of International Conference on Trends in Computational and Cognitive Engineering", Springer Science and Business Media LLC, 2021 Publication	<1 %
7	Yanqiu Liao, Pengwen Xiong, Weidong Min, Weiqiong Min, Jiahao Lu. "Dynamic Sign Language Recognition Based on Video	<1 %