TRANSLATING THE SILENT LANGUAGE: SIGN-TO-CHARACTER CONVERSION IN BANGLA

BY

NUSRAT JAHAN NORY
ID: 203-15-14466
AND

AYSHA AKTER SUMI
ID:203-15-14488

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

Saiful Islam
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Md. Ferdouse Ahmed Foysal
Lecturer
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
July 2024

APPROVAL

This Project titled "**Translating the Silent Language: Sign-to-Character Conversion in Bangla**", submitted by **Nusrat Jahan Nory** and **Aysha Akter Sumi** 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 **14th July, 2024.**

BOARD OF EXAMINERS

____

Dr. S.M Aminul Haque (SMAH) **Board Chairman**
Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

____

Md. Abbas Ali Khan (AAK) **Internal Examiner 1**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

____

Md. Aynul Hasan Nahid (AHN) **Internal Examiner 2**
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

____

Dr. Abu Sayed Md. Mostafizur Rahaman (ASMR) **External Examiner**
Professor
Department of CSE
Jahangirnagar University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Saiful Islam, Assistant 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:

Saiful Islam

Assistant Professor

Department of CSE

Daffodil International University

Co-Supervised by:

Md. Ferdouse Ahmed Foysal

Lecturer

Department of CSE

Daffodil International University

Submitted by:

Nusrat Jahan Nory

ID: 203-15-14466

Department of CSE

Daffodil International University

Aysha Akter Sumi

ID: 203-15-14488

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 **Saiful Islam, Assistant Professor,** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "**Image Processing**" 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 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

The number of people who are deaf and mute on the planet is alarmingly increasing. Approximately 2.6 million people in Bangladesh are unable to communicate with others through language. This research investigates the development of a system for translating Bangla Sign Language (BdSL) into textual Bangla characters. This paper aims to bridge the communication gap between the deaf and hearing communities in Bangladesh. The system aims to more accurately recognise sign language and translate it into Bangla (Bangla characters). We have a dataset of 11,422 Bangla sign language images, encompassing 36 Bangla alphabets and 10 numerical values. Out of these, 1,906 are original images that we personally collected from various places and various people. This diverse dataset aims to ensure robust and accurate sign language recognition. This research employs five advanced pre-trained deep learning models—ResNet50, InceptionV3, Xception, VGG16, and a hybrid model that combines ResNet50 and InceptionV3. The implementation and evaluation of these models are discussed in detail. Among the tested models, the hybrid model (ResNet 50+ InceptionV3) achieved the best accuracy 96%, compared with the other models used. The Xception Model achieved an accuracy of 95%, while VGG16 achieved 94%. The InceptionV3 model performed quite poorly, with an accuracy of 93%, and the ResNet50 model achieved the lowest accuracy of 91%. Due to its superior performance, the Hybrid Model is used in the Web Application as a part of this work. This web application demonstrates the conversion of the Bangla Sign Language images to characters. This will make it easier for our people to communicate with the deaf and dumb community by utilising a variety of approaches.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

ML	Machine Learning
SVM	Support Vector Machine
CNN	Convolutional Neural Network
BdSL	Bangla Sign Language

CHAPTER 1
INTRODUCTION

1.1 Overview

This research is centred on building a Bangla sign language recognition model that indicates a mass communication milestone initiated by the deaf and dumb people in Bangladesh. We aim to create a system that can convert recognised signs into Bangla and even create more quality for people, so for individuals with speaking and hearing problems, this system can play an important role in getting them the closest life to other normal people. This work investigates various techniques to obtain faithful results, thereby helping reduce the communication barrier for hearing-impaired individuals. The study also deals with the worldwide and local evidence of hearing loss, telling an unfortunate story about how real solutions need to be.

1.2 Background and Present State

Sign language is frequently employed to address the communication divide among the deaf and nonverbal populations. It is a crucial tool for expression and communication within these communities, with a varied vocabulary and a moderate structure. Despite the fact that numerous sign languages share similarities, they are not universally comprehensible across diverse communities. Consequently, it is imperative to adopt a high-performance, automated sign language perception system in order to enhance communication and improve the quality of life for the deaf community.

Recently, significant progress has been made in the field of Bangla sign language, with several notable achievements. The development of a sign language recognition model has facilitated the communication and utilization of Bangla sign language.

Progress in deep learning methods, especially Convolutional Neural Networks (CNNs), has led to big steps forward in the area of Bangla Sign Language (BdSL) recognition. The purpose of these advancements is to overcome obstacles in communication for the deaf and mute population in Bangladesh, thereby improving their social integration and ease of access. Customised CNN architectures have been used by researchers to obtain remarkable accuracy rates in terms of recognising BdSL letters and digits [1], [15], [17], and [18]. The utilization of advanced models such as YOLOv4 and EfficientDet-D0 allows for the real-time identification and understanding of sign language [10], [11]. Some researchers implied their models on affordable hardware like the Raspberry Pi [1] and smartphone apps [3], making it feasible to recognise Belgian Sign Language (BdSL) in different environments. Zero-shot learning and transfer learning have made BdSL recognition systems more flexible and expanded their vocabulary. This lets them recognise signs they do not know by using information about the context [12]. Practical uses of BdSL recognition technology include smartphone applications that can convert voice to sign language and text to speech. Additionally, there are affordable robots that can understand sign language, showcasing efforts to incorporate BdSL recognition into everyday communication situations. There are still difficulties in accurately representing the ever-changing characteristics of sign language, dealing with cultural subtleties, and creating interfaces that are easy for users to use. In the future, researchers will try to expand the vocabulary, make the recognition algorithms more accurate, make it easier for BdSL recognition systems to understand changing parts of sign language, and make them more sensitive to different cultures. Collectively, these advances indicate a positive path towards overcoming obstacles in communication and promoting inclusiveness for the deaf population in Bangladesh.

This paper implements some different CNN-based pre-trained models for Bangla sign language recognition and a Web Application for further demonstration and brings forth the following contributions:

- This study makes a significant contribution by creating a large-scale Bangla sign language dataset. The collection has 46 distinct categories, with an average of over 50 photos per category. Photos of left, right, and bilateral hands taken in a variety of backgrounds and lighting conditions served as the dataset's source.
- As part of the implementation of CNN-based pre-trained models, four advanced deep-learning methods have been used to recognise Bangla sign language: Resnet50, InceptionV3, Xception, and VGG16. The effectiveness of these strategies has been evaluated using criteria such as average precision (AP), precision, recall, f1 score, and loss.
- A hybrid model was created by merging two pre-trained models, Resnet50 and InceptionV3.The hybrid model has demonstrated superior performance in comparison to the separate versions.
- To provide a live demonstration of the recognition system, a web application has been developed. The backend of the application utilises FastAPI, a state-of-the-art Python web framework. The frontend of the application utilises HTML, CSS, and JavaScript to develop the user interface.

The compilation of a large-scale, comprehensive dataset of Bangla sign language and the design of a highly accurate hybrid model that combines Resnet50 and InceptionV3 to greatly improve recognition performance are the innovative aspects of this study. Moreover, the incorporation of this model into a user-friendly online application demonstrates the practical execution of real-time BdSL recognition.

1.3 Problem Statement

The deficiency or complete incapacity to hear is referred to as hearing impairment or hearing loss. This condition may have an impact on one or both hearings. 1 in 6 People worldwide, or 1.3 billion people, have a significant disability, which represents 16% of the world's population. Rehabilitation is necessary to address the disabling hearing loss of approximately 430 million individuals, including 34 million children, who account for more than 5% of the global population. The World Health Organisation (WHO) predicts that by 2050, disabling hearing loss will affect more than 700 million individuals, or one in every ten individuals [32]. Their widespread prevalence emphasises the urgent need for effective solutions to help people with hearing impairments.

A substantial number of individuals in Bangladesh, a country with a population exceeding 130 million, are affected by hearing loss. According to the Bangladesh Bureau of Statistics (BBS), there are about 455,000 people who have hearing disabilities, or 0.35% of the population [31]. Many of these individuals experience severe hearing impairments that result in substantial disabilities. These statistics emphasise the urgent necessity of interventions to support the deaf and mute population in Bangladesh.

Sign language is used as an essential way of communication and self-expression for such people. However, the existence and use of sign language vary enormously across different cultures. Even though there can be similarities between different sign languages, universal understanding is not possible. It makes a case for developing complex, automated systems for identifying the signs of sign language to advance communication and raise the living standards of the deaf group. Social apathy and barriers to integration with society compound problems faced by hearing- and speech-impaired people in Bangladesh. This underlines the need for inclusive educational resources and equally complementary activities.

Various attempts at empowering and helping deaf and mute people with educational resources and inclusive programmes in society are currently underway but need better implementation and funding to make a difference. For a person with hearing loss to enhance their well-being and integrate into mainstream society, multipronged challenges have to be addressed.

1.4 Objectives

The purpose of this research is to create a model that is capable of accurately recognising Bangla Sign Language (BdSL) using pre-trained CNN-based models. The main objective is to attain a high level of precision in BdSL identification by utilising the capabilities of well-established convolutional neural network structures. Furthermore, the objective of the project is to develop a web application that presents the recognition outcomes in a visual format, enabling users to easily access and comprehend them. The project aims to enhance the development process by leveraging pre-trained models, with a specific focus on improving accuracy and usability in real-world applications. This effort is to facilitate the removal of obstacles in communication and promote inclusion for the deaf population in Bangladesh by utilising innovative technical solutions.

1.5 Scope and Limitations

By creating a user-friendly web application for visualisation and a high-accuracy model, this initiative intends to make significant strides in the recognition of Bangla Sign Language (BdSL). The aim is to improve the accuracy of existing models for exact recognition of BdSL (Bangla Sign Language) alphabets and numbers by utilising pre-trained Convolutional Neural Networks (CNNs). The project will employ robust datasets that we have gathered from surveys and investigate opportunities to expand them, thereby guaranteeing comprehensive model training. This project will prioritise the creation of an intuitive user interface design to enhance interactivity and accessibility for both those who are deaf and those who are not.

However, some problems with this project could make it harder to carry out or change its scope. Initially, the presence and sufficiency of BdSL datasets may restrict the variety and quantity of training data, thereby potentially reducing the effectiveness of models in real-life situations. Although utilising pre-trained CNN models provides a basis, attaining constant accuracy in the presence of different illumination conditions and hand placements remains a difficult task. Limitations in processing resources can also restrict the complexity and size of models that can be used on devices with limited capabilities. Moreover, the cultural and linguistic diversity in BdSL presents difficulties in guaranteeing the model's flexibility across several regional dialects. Data protection, consent, and respectful depiction of the deaf community are just a few ethical considerations that will guide the project's development and deployment phases. Notwithstanding these obstacles, the initiative aims to greatly improve BdSL identification technology, therefore fostering inclusion and accessibility for the deaf population in Bangladesh and maybe in other regions as well.

1.6 Report Organization

Chapter 1: The introduction, Background of the work and Present State of the research, what is the problem statement of the work, what are the Objectives,and how and where should be improved(Scope) and the limitations of the previous and what can be the barrier for our work, that’s all were covered in this chapter.

Chapter 2: Contains background information on this study as well as relevant research on sign language recognition in Bangla conducted by other researchers over the previous 12 years. And they are given a comparison table for easily finding which work is better and which is best. They are given a section named “open issue” which finds the problem or issues of the relevant research works.

Chapter 3: Discussed the working flow of our research, proposed methodology, how datasets were collected, which preprocessing techniques are used, which models are used, with some discussion, the software and hardware requirements of the project, project management, and financial analysis.

Chapter 4: This chapter explains how the model utilised in this work was implemented, along with a thorough flowchart that shows how it was done. The model evaluation is also covered, with curves for Training & Validation Accuracy and Loss over epochs presented. The chapter also offers a thorough explanation of how the web application functions, covering both the frontend and backend components.
Chapter 5: A thorough comparison of all the models for deciphering and identifying Bangla Sign Language can be found in Chapter 5. Confusion matrices are used for analysis, and randomly chosen images processed by each model are visualised. The goal of the chapter is to illustrate the advantages and disadvantages of each model, providing a thorough understanding of their effectiveness in practical settings and performance indicators.
Chapter 6: This section is based on the impact of this work in life, society. How can this work sustain in environment. Ethical aspect of this work are also described in this section.
Chapter 7: This chapter is the conclusion section. What will be the best outcome of this work and what is the limitations and what can be the future work for taking the work in a better position.

1.7 Summary

Finally, by creating an accurate model and a user-friendly online application, this work attempts to improve the recognition of Bangla Sign Language (BdSL). The research uses pre-trained CNN models to provide accurate identification of BdSL alphabets and numbers, allowing communication for Bangladesh's hearing-impaired. Deep learning techniques, especially convolutional neural networks (CNNs), have made it easier to create reliable BdSL recognition systems, which has made it easier for everyone to join in and use technology. The scarcity of various BdSL datasets, as well as assuring model adaptation across cultural and language variances, are significant challenges. Despite these challenges, the project aims to develop BdSL identification technology, resulting in more inclusion and a better quality of life for the deaf community in Bangladesh and elsewhere.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

Deep learning, particularly Convolutional Neural Networks (CNNs), dominates BdSL recognition research, achieving high character, numeral, and sentence recognition accuracy. Researchers actively explore real-time detection and interpretation using diverse architectures like YOLOv4 and CNNs with text-to-speech integration. Emerging zero-shot learning techniques show promise for recognising unseen signs, and expanding vocabulary potential. Accessibility efforts include smartphone apps, low-cost robots, and Raspberry Pi implementations. Future directions involve expanding vocabulary, capturing sign language dynamism, incorporating cultural awareness, and developing user-friendly interfaces. This research signifies significant progress towards breaking communication barriers and fostering social inclusion for the deaf community in Bangladesh.

2.2 Related Works

A user-friendly method for converting Bangla Sign language to text is presented in [1]. They have customised ROI segmentation, and CNN is used for sign gesture detection. They used a custom image dataset of five signs, augmented and preprocessed, for training a CNN model. The trained model achieved high accuracy based on the distance for 0.5 metres, 97.54%, and for 1 metre, it achieved 93.18%. and was implemented on a Raspberry Pi for live sign detection via webcam. This approach can be expanded with more signs and a user interface.

A sizable dataset of 30,916 images in Bangla sign language from 25 participants was created in [2]. Using this data, they subsequently trained a potent CNN model that included six convolutional layers, three pooling layers, and others. The model demonstrated remarkable accuracy after training on random folds of the data, achieving 99.83% for characters, 100% for numerals, and 99.80% overall. This approach outperforms current methods and opens the door to improved recognition of Bangla sign language!

Researchers developed a smartphone app [3] to improve communication between hearing/speaking and deaf/mute people in Bangladesh. When having a speech-to-sign conversation, CMU Sphinx recognises the speech, translates it to text, and then displays the translated images in sign language on the screen. Moreover, Google Translate manages text-to-speech for sign-to-speech, processing input before sending it and streaming back audio. Test results indicated a promising 84.71% accuracy.

The communication gap in Bangladesh between sign and non-sign users is addressed in [4]. They developed a rule-based system that converts written and spoken Bangla into animated numerals in Bangla sign language. The system recognizes individual numerals by analyzing spoken and written Bangla. Fourteen animated numeral signs—0–9, Hundred, Thousand, Lakh, and Core—have been used to train it for both text and voice inputs. The system matches parsed numerals to the appropriate animated sign using algorithms. Additionally, it evaluates the system's efficacy using the F-measure, accuracy, precision, and recall metrics. It provides excellent accuracy for both written and spoken inputs. It received 96.7% of the voice and 100% of the text input.

A learning-based method for recognizing Bangla Sign Language is presented in [5]. The study makes use of Convolutional Neural Networks (CNN) for Bangla Sign Language recognition, Hidden Markov Models (HMM) for frame segmentation, and Leap Motion Controllers (LMC) for virtual hand tracking. CNN deciphers sign expressions, HMM interprets gestures, and the LMC records fine-grained hand data. In basic sign expressions, experimental results show a 3% error rate with distortion and a 2% error rate without distortion.

An artificial neural network and support vector machines (SVM) system for translating Bengali Sign Language to text is presented in [6]. The system uses Microsoft Kinect as input to recognize joints and wrists, record hand signals, and classify the data using support vector machines (SVM). The support vector model processes contour features that were extracted using the convex hull method. With a five-layer neural network, the suggested model can classify images of hand gestures made by men and women with an accuracy of 84.11%. This model effectively translates sign language to text using a blend of machine learning and image processing techniques.

Using a deep transfer learning-based convolutional neural network with a random forest classifier, [7] proposes a hybrid model for automatic Bangla Sign Language recognition. 'Ishara-Bochon' and 'Ishara-Lipi' datasets are used to verify the system's performance. Due to the small size of their datasheet, they used data augmentation to create a larger dataset. Additionally, they have scaled the feature value to a smaller range of 0 to 1 rather than a larger range of 0 to 255. To remove unnecessary features from sign images, they proposed a background elimination algorithm. As for character recognition, the system achieves 91.67% accuracy, 93.64% precision, 91.67% recall, and 91.47% f1-score, and for digit recognition, 97.33%, 97.89%, 97.33%, and 97.37%.

Bengali Sign Language (BdSL) alphabet signs can now be recognized by a powerful AI system created by [8] using a VGG19 neural network as its foundation. To help the deaf community in Bangladesh communicate more effectively, they have created a diverse dataset. A comprehensive database for training was assembled by gathering more than 12,000 BdSL signals from both deaf and non-deaf people. The system demonstrated its effectiveness in real-world applications by achieving 89.6% accuracy on unseen data, using a pre-trained VGG19 network that was customized with additional layers for BdSL recognition.

Bengali Sign Language (BdSL) recognition is addressed in [9] research with a pre-trained deep learning model (VGG16). Using data augmentation and fine-tuning the model on a dataset of 1147 BdSL images, they achieved high accuracy rates (96.33% training, 84.68% validation). This technique works better than earlier strategies and could help close the communication gap between the hearing and the deaf communities.

Real-time identification of Bangla Sign Language (BdSL) on a low-cost device is tackled with deep learning in [10] work. Detectron2, EfficientDet-D0, and YOLOv7 are the three deep-learning models the researchers developed for BdSL image recognition. They achieved this by merging new and pre-existing datasets. They achieved 94% in EfficientDet-D0, 94.9% in Detectron2, and 94.2% in YOLOv7. Training time, mAP@.5, mAP@.5-.95, and detection accuracy were among the performance metrics. The Detection 2 model offers the highest accuracy, or mAP, while the YOLOv7 Tiny technique achieves the shortest training time. In the end, a webcam and Jetson Nano edge device have been equipped with the proposed Bangla sign language detection system for real-time inferencing using the YOLOv7 model.

It was proposed by [11] to detect and understand Bangla Sign Language (BdSL) in real time. The authors created a dataset of 12,500 images covering BdSL alphabets, digits, and even custom signs for sentences. They used YOLOv4 to detect objects in the dataset. A dictionary is used to map the detected signs to text using this powerful model, known for its speed. Sentences are formed by combining words and using special signs for punctuation. Finally, Google Text-to-Speech converts the text to spoken language.With an impressive 97.95% accuracy and real-time performance (0.485 seconds for sentence generation), this system holds promise for smoother communication. However, there are areas for improvement, such as including more signs, proper punctuation, and handling numbers.

Zero-shot learning (ZSL) and transfer learning are the two methodologies that [12] employed in their study. Transfer Learning is a kind of deep learning where pre-trained models are refined to recognise signals from an extensive body of data. The accuracy of this approach was an astounding 93.68%. Conversely, the Zero-Shot Learning technique forecasts invisible indicators by utilizing their attributes (such as hand position and movement) that are acquired from recognized signs. Although it is not as precise (54.34% for unseen indicators), it provides opportunities to identify novel signs that are not included in the training set. They utilize a large dataset that contains 35,149 pictures of one-handed BdSL alphabet signs.

used 3D printing to design and build a low-cost humanoid robot for sign language interpretation to lower the costs related to robot production. They used a Kaggle data set (https://tinyurl.com/rrb3zua) and a private dataset, both of which are used in [8].

Additionally, they used convolutional neural networks (CNN) and recurrent neural networks (RNN) for deep learning models. Their proposed robot can imitate sixteen alphabets from Bangla sign language and recognize ten medical symptoms. They were 98.19% accurate at identifying sign language in Bangla.

The paper by [14] presents the CNN architecture for Bangla Sign Language (BdSL) recognition and translation. Their dataset contains 4600 hand sign images that were taken with an HD Webcam. They used thresholding, resizing, and grayscale to split the preprocessed image data into training (80%) and testing (20%) groups. A CNN model was built with Keras using convolutional, pooling, and dense layers. It was then trained using the Adam optimizer and the ReLU activation. The proposed model recognizes the 36 letters and 10 numbers of BdSL. With a remarkable validation accuracy of 99.57% and 99.41% on unseen data, the system showed that it could facilitate real-world communication. Sign recognition techniques use both vision-based and sensor-based methods.

An assistive communication system based on deep learning is developed by [15] This system recognizes hand sign digits and converts them into spoken Bangla for assistive communication. Approximately 1200 images of hand signs [Online Source: (http://lttm.dei.unipd.it/downloads/gesture/)] and created 2000 sources [Dataset] are used to train a convolutional neural network (CNN) model. Images are pre-processed and the model is trained with an RMSProp optimizer and a specific cost function. The trained model recognizes hand sign digits in real time and feeds the results to a text-to-speech engine. This engine translates the digits into Bangla text and uses Google Text to Speech API for spoken output. The model achieves 92% accuracy on unseen data, demonstrating its potential for real-world use.

Support vector machines (SVM) are the main tool used in [16] work to recognize Bangla sign language. They involve two steps: first, segmenting hand signs, and then, using Gabor filters, recognizing those signs. The extracted feature vector's dimensionality is decreased using kernel principal component analysis (PCA), and the extracted features are then classified using support vector machines (SVM) classifiers. It had a 97.7% accuracy rate. 4800 photos in the dataset were used for testing. More pictures could be examined, though.

The Ishara-Lipi dataset, which included 1800 photos of 36 BdSL characters, was used in the [17] paper. There are four convolutional neural network layers in this model.It brought to light the possible effects on hearing-impaired people's lives of automatic sign language recognition. For both learning rate and model optimization, the Adam (Adaptive Moment Estimation) optimization algorithm was used. The model's 92% accuracy can be increased by making adjustments or switching to a different model. Islam et al. (2018) employed Principal Component Analysis, Skeleton Detection, and Hidden Markov Models. Furthermore, CNN was trained to identify only the Bangla numbers 0 through 9. They utilized the same dataset and attained 95% accuracy.

Focusing on the identification of Bangla numerical signs, [18] work employed a camera to collect data. To enhance the image's qualities, bandlet transformation was applied as part of the preprocessing process. A logarithm replacement method was used to lessen the impact of additional light on the images. Low-resolution image problems were addressed with the D-LBP (Local Binary Pattern) method. The dimensions were computed using preprocessed photos. Large blobs were removed from the binary images by the model after it had created a binary picture by further segmenting the skin color of the images. Convolutional Neural Networks (CNN), one of the deep learning techniques, allowed the model to achieve an astounding 99.8% accuracy rate.

Words and sentences are recognized using SVM in [19] input video. To identify the skin region within RGB images containing hand signs, the images are converted into YCbCr color space for skin color segmentation. The skin color segmentation, labeling, and filtering techniques are used by the model to distinguish the hand region from other skin regions. The hand portion is subjected to LBP (Local Binary Pattern) to extract features. These features are then used for classification, yielding an accuracy of 94.49% for sentences and 94.26% for words. To improve accuracy, a larger dataset can be used, as they did with 2400 images for training and testing.

Using the Ishara-Lipi database, [20] works employed a different CNN (Convolutional Neural Network) architecture to identify the alphabets of Bengali signs. This groundbreaking Convolutional Neural Network (CNN) achieves 99.86% accuracy with 12 convolutional layers, 3 depth-wise convolutional layers with batch normalization, and 3 pooling layers. It is anticipated that the model will advance sign language identification research and help the deaf community.

The work presented in [21] focuses on the classification of alphabets and numerals in Bangla Sign Language (BdSL) using a deep learning model.Three pre-trained CNN models (ResNet18, MobileNet_V2, and EfficientNet_B1) for classification and three CNN models (DenseNet201 FPN, UNet, and M-UNet) for semantic segmentation and background removal were used to optimize the process using stochastic gradient descent (SGD) and DenseNet201 FPN, UNet, and M-Unet. With an overall accuracy of 99.99%, ResNet18 was the most successful classifier, according to the study, while EfficientNet_B1 had the lowest accuracy, at 99.05%. This work sheds light on effective deep-learning techniques for BdSL segmentation and classification issues.

The challenge of computer vision-based sign language detection and recognition is discussed in [22] paper on normalized cross-correlation-based two-handed sign language recognition for Bangla characters. The deaf community in Bangladesh uses Bangla sign language, which is specifically covered in this paper. The recommended approach consists of two steps: recognition and refining. Potential regions for refinement are identified using a Red-Green-Blue (RGB) color model. Filtering and refining techniques are then applied. Using a statistically based template matching technique, hand sign regions are identified with 96% accuracy in the recognition step. The effectiveness of the proposed method is evaluated with various hand sign images on the same background color. Variations in the background could affect the accuracy. The paper highlights the importance of sign language recognition for improving communication with the deaf community in Bangladesh and references previous research in the field.

Convolutional neural networks (CNNs) are used in the method proposed by [23] to recognize Indian sign language movements acquired through continuous sign language films in selfie mode. A dataset containing 200 signals performed by several participants from five different viewing angles is constructed. An input layer, four convolutional layers, five rectified linear units (ReLu), two stochastic pooling layers, one dense layer, and a SoftMax output layer make up the CNN model. Three batches of distinct datasets are used to train the model. Achieving a 92.88% recognition rate in comparison to other classifier models on the same dataset is possible through increasing the number of datasets used for training. Stochastic pooling is found to be suitable for the application, and the CNN architecture shows better accuracy in recognition compared to other classifiers.

A novel method for Sign Language Recognition (SLR) in Bangla Sign Language (BdSL) using depth information is presented in [24]. They highlight the challenges posed by the complexity and large variations in hand actions in SLR. To address these challenges, the authors utilize MediaPipe, a cross-platform depth-map estimation framework, to extract depth information from RGB images. They describe the process of hand key-point detection, pre-processing, and the generation of CSV files containing normalized x, y, and z coordinate values of hand key points. SVM(RBF) was found to be the most accurate model, with 98.65% accuracy, when KNN, XGB, SVM(Linear), RFC, SVM(Polynomial), and DTC were used. Additionally, they validate the effectiveness of their approach by running MediaPipe on a benchmark American Sign Language (ASL) dataset. The paper emphasizes the affordability and efficiency of the MediaPipe framework as a low-cost alternative for extracting depth information from RGB images, making it more accessible for research and applications in sign language recognition.

Presented a novel Convolutional Neural Network (CNN) model for computer vision-based Sign Language Recognition in [23]. The dataset that the model utilizes consists of 46 images of Bangla vowels, consonants, and numerals that were taken in different lighting and background conditions. To address concerns brought up by previous research, the two datasets, "BdSL_OPSA22_STATIC1" and "BdSL_OPSA22_STATIC2," were carefully chosen. The suggested six-layer ConvNeural framework produced impressive accuracy rates by aiming for efficient feature extraction. The CNN model outperformed pre-trained models with an accuracy of 98.38% for one dataset and 92.78% for another, proving its ability to overcome challenges and progress in the field of sign language recognition.

The application of Histogram of Oriented Gradients (HOG) features to enhance single-handed Bengali sign language recognition is the main topic of a paper by [24]. Grayscale images are processed to extract HOG features, which are then used to train a K-Nearest Neighbors (KNN) classifier for the recognition of Bengali sign language motions. With a noteworthy accuracy of 91.1%, the suggested method successfully categorizes these motions, demonstrating its value for real-time applications. The authors also outline the future directions of their research, which include using concatenated feature vectors to detect motion signs in real time and expanding sign identification with movements other than those found in the alphabet.

To handle the complexity of Bangla Sign Language (BdSL) symbols, a novel architecture based on convolutional neural networks (CNNs) is presented by [25]. It combines an image network and a pose estimation network. The study emphasizes the significance of hand position and visual cues in the recognition of sign language. Future directions for development are also explored in the paper, such as training a specialized posture estimation network for BdSL hand pose estimation and curating a BdSL dataset for future research. The proposed model demonstrated impressive results, with an accuracy of 91.51%, with the aid of a CNN-based image network and a pose estimation network.

2.3 Comparison between existing works

Table 2.3.1. Comparative analysis of few works of the last 12 years regarding translating sign language into Bangla character

SL No.	Author’s Name	Used Algorithm	Best Accuracy with Algorithm
1	Khan et al.[1]	CNN	CNN–97.54% (0.5m)
2	Islam et al. [2]	CNN	CNN – 100% (numerals)
3	Shahriar et al. [3]	CMU Sphinx	CMU Sphinx– 84.71%
4	Rahaman et al. [4]	Rule-based system ● Algorithm 1: Algorithm for Bangla Numeral Parsing ● Algorithm 2: Algorithm for Voice Recognition	Algorithm 1– 100%
5	Farhad et al. [5]	CNN with HMM	CNN – 2% error (no distortion)
6	Chowdhury et al. [6]	SVM	SVM–84.11%
7	Das et al. [7]	Hybrid (CNN, Random Forest)	Hybrid – 97.33% (Digit)

8	Rafi et al. [8]	VGG19 CNN	VGG19 CNN–89.60%
9	Hossen et al. [9]	VGG16 CNN	VGG16 CNN– 84.68%
10	Siddique et al. [10]	● Detectron2 ● EfficientDet-D0 ● YOLOv7	Detectron2--94.9%,
11	Talukder & Jahara[11]	YOLOv4	YOLOv4 – 97.95%
12	Nihal et al. [12]	● Zero-shot ● transfer learning	Transfer learning– 93.68%
13	Nihal et al.[13]	● CNN ● RNN	CNN – 98.19%(Dataset-1)
14	Alam et al.[14]	CNN	CNN – 99.57%
15	Ahmed et al.[15]	CNN	CNN – 92%
16	Uddin & Chowdhury[16]	SVM	SVM – 97.70%
17	Islam et al.[17]	CNN	CNN – 92%
18	Islam et al.[18]	CNN	CNN – 94.88%
19	Shamrat et al. [19]	CNN	CNN – 99.80%
20	Santa et al. [20]	SVM	CNN – 94.49%(Sentences)
21	Hossein & Ejaz[21]	CNN	CNN – 99.86%
22	Podder et al. [22]	CNN(ResNet18, MobileNet_V2, and EfficientNet_B1)	CNN(ResNet18) – 99.99%
23	Deb et al. [23]	Template Matching with normalized cross-correlation	Template Matching – 97.5%(Detection)
24	Rao et al. [24]	CNN	CNN – 92.88%
25	Rayeed et al. [25]	● KNN ● SVM (Linear, Polynomial, RBF) ● XGB ● RFC ● DTC	SVM (RBF) – 98.65%
26	Rahaman et al. [26].	CNN	CNN-98.38%(BdSL_OPSA22_STATIC1)

27	Tabassum et al. [27]	• SVM • KNN • ANN	KNN-91.1%

2.4 Open Issues:

Sign language has no international standard. Different languages have their own sign languages. Both America and Britain use English as their official language, but their sign languages are different. The signs used in the Bangla language are inconsistent and can change at any time. Therefore, whenever changes are made to sign language signs, our work will need to be adjusted to account for these variations.

While many works already exist that use recognition of sign language in Bangla, only a small number of works also rely on any kind of interface. Since the Bangla language is quite a complex language with some ligatures, their model sometimes struggles to generate texts. Our work will concentrate on using an interface designed for real-time image processing to recognise the Bangla language. The analysis of all these papers revealed some gaps or failings.

Despite the existence of various research works on Bangla sign language generation and recognition, a critical gap remains in the limitations of available datasets. As there is no unified sign language data for Bangla, most studies rely on small datasets, hindering the models' generalizability and robustness. These datasets often lack expert opinions, participant diversity, backgrounds, lighting, and sign variations. Additionally, publicly accessible and well-annotated datasets are scarce, further limiting progress.

Moving beyond datasets, tackling continuous signing and signer independence requires more attention. Combining visual and in-depth information holds promise for more robust recognition, yet it remains underexplored.

Despite progress, achieving efficient and accurate real-time recognition remains a challenge. Evaluating the usability and accessibility of these systems for the deaf community also requires more research. Overall, while advancements have been made in Bangla sign language recognition, several areas demand further exploration to develop truly robust, generalizable, and user-friendly systems that bridge the communication gap for the deaf community in Bangladesh.

2.5 Summary

The communication gap between hearing or speaking people and the deaf and mute communities in Bangladesh has been the subject of numerous studies. These studies have used diverse approaches, including deep learning, convolutional neural networks (CNN), support vector machines, hidden Markov models, and other methods.

Several researchers [1], [2], [6], [7], [8], [9], [10], [14], [15], [17], [18], [19], [20], [21], [22], [23], and [25] used CNN models to recognise Bangla Sign Language with noteworthy accuracy rates. They used many CNN architectures, with accuracy ranging from 88.11% to 99.99%. By using Support Vector Machine (SVM), on the other hand, and [22] showed good accuracy, ranging from 94.49% to 97.7%. The accuracy rates of studies [5, 12, and 24] that used CNNs and Hidden Markov Models (HMM) together, as well as transfer learning and depth information extraction, ranged from 92.88% to 98.65%. In [4], rule-based techniques were used to achieve a high accuracy of 100% for text input and 96.7% for voice input. For communication purposes, [3] and [13] tackled robot sign language interpretation and speech-to-text and text-to-speech conversion using various proposed models with accuracy ranging from 84.71% to 98.19%.

There are a variety of methods for efficiently interpreting and communicating language, and they show encouraging accuracy rates when it comes to recognising the Bangla Sign Language.

CHAPTER 3
METHODOLOGY

3.1 Overview

The objective of this study is to create a model for the recognition of Bangla sign language to facilitate communication between hearing and deaf communities. A thorough methodology that includes data collection, processed data, model development, evaluation, and refinement will be used to accomplish this paper. In Figure 3.1.1, we have tried to show our overall approach. The figure is given below:

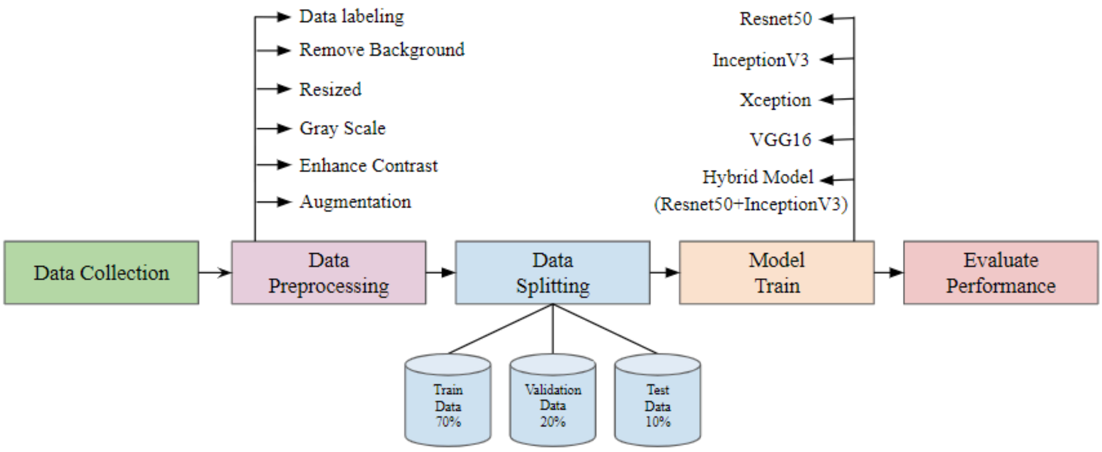
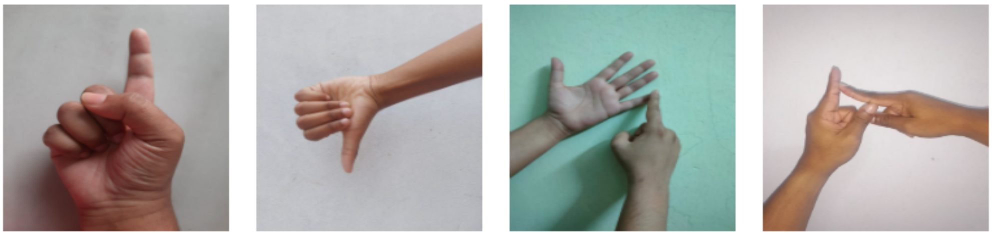


Figure 3.1.1: Working Flow of this paper

3.2 Proposed Methodology

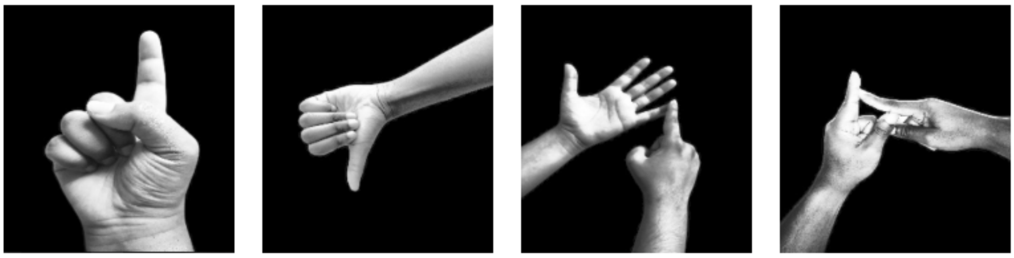
Data Collection Procedure

We collect extensive data for sign language recognition by combining surveys with image capture. We clearly state the objectives of our research, which include analysing sign language viewpoints and hand gestures as well as how they are utilised to start this process. We ensure our dataset by taking a wide range of signing images against various backgrounds in both indoor and outdoor settings. We have 11422 sign language images in the dataset, including 36 Bangla alphabets and 10 numerical values, where the original images are 1906. By involving sign language users in structured surveys or interviews, we can accomplish our study goals. We took pictures of these exchanges in sign language using cameras or smartphones from 100 volentiars. Ms. Arafat Sultana Lata who is the Sign Language Interpreter & Trainer of Society of the Deaf & Sign Language Users (SDSL) in Bangladesh Television, graded the photographs. In Figure 3.2.1, we have given an example of our Raw Data that we collected, and Figure 3.2.2 showed an example of processed data;



১ ৯ ই ছ

Figure 3.2.1: Some Raw Data from our Dataset



১ ৯ ই ছ

Figure 3.2.2: Some Processed Data from our Dataset

Here is the dataset link: **Bangla_Sign_Language_Dataset**

Table 3.2.1: Total Number of Collected Data

	Numerial	Alphabates
Raw Dataset Total: 1906	394	1512

Table 3.2.2: Total Number of Dataset

	Numerial	Alphabates
Augmented Dataset Total: 11422	2638	8784

Data Preprocessing

In data mining, data preprocessing is a crucial part. It describes the processes of integrating, cleansing, and converting data to prepare it for analysis. Enhancing the data's quality and adapting it to the particular data mining task is the aim of data preparation. In our work, we used several preprocessing techniques,

- **Labelling:** We used labelling to help our models learn to associate input data with the required outputs, like in classifying input data and predicting better outcomes. Accurate labelling is required for training the models and achieving reliable predictions. And we verified the full dataset by an expert.
- **Remove Background:** It is used especially to remove irrelevant or distorted elements from images in our dataset. This is important for focusing strictly on the main object. We gather data on different backgrounds, so if we don't remove the background, it can have a major negative impact on classifying the corrected sign, even in the model and also in the interface
- **Resize:** Resizing is the process of altering an image's or dataset's dimensions to make sure their size and form are consistent while preparing the data. To meet the respective input and ensure efficiency and standardisation, this is essential. Our data was resized to 256x256 pixels, which standardised the input size for any model.
- **Grayscale:** We used grayscale to convert the RGB images into grayscale, which reduces the complexity of image data by eliminating colour. It can enhance the speed of processing and the efficiency of certain algorithms. It requires less storage and computational resources compared to colour images.
- **Enhance Contrast:** Enhancing contrast is used in our dataset to adjust the difference between the light and dark areas, which helps to make features more distinct. There are many techniques, but we used histogram equalisation and gamma correction (Gamma = 1.5; where Gamma < 1: Darkens the image, reducing the contrast; Gamma = 1: No change; the image remains unchanged, Gamma > 1: Lightens the image, increasing the contrast) for accurate colour representation and proper brightness.
- **Data Augmentation:** We used augmentation in our dataset to create new data from existing data by horizontal flipping, height and width shift ranging, zooming, and shearing. This augmentation technique is widely used to increase the diversity and robustness of data. With the help of augmentation, the main dataset, which contains 11422 images, has 1906 images in its raw form.

Model Selection & Design

Initially, we attempted to implement five advanced deep learning pre-trained models that are based on CNN and previously trained using Imagenet. Later on, we have to build a hybrid model for better performance. The models that we have implemented in our work are:

- Resnet50
- InceptionV3
- Xception

- VGG16
- Hybrid Model (Resnet50+InceptionV3)

Convolutional Neural Networks (CNNs): CNNs are deep learning models that are made to handle visual input tasks like classifying images and finding objects. Convolutional layers are employed to process pictures, utilising adjustable filters to extract hierarchical elements such as edges and textures. Pooling techniques decrease the spatial dimensions, whereas fully linked layers are responsible for classification or regression tasks. Activation functions such as ReLU promote non-linearity, whereas dropout layers avoid overfitting by deactivating neurons during training. CNNs are very adaptable, capable of processing not just pictures but also audio and time series data. This versatility is crucial for sophisticated pattern recognition and has greatly contributed to advancements in different disciplines of machine learning. Their capacity to independently acquire meaningful characteristics from unprocessed data has transformed domains including computer vision, natural language processing, and healthcare analytics. [31]

Resnet50: In 2015, Microsoft Research developed ResNet50, a deep learning breakthrough for photo classification. ResNet50's 50 layers use residual connections, or skip connections, to train extraordinarily deep neural networks. These connections allow the network to acquire residual functions that transmit data straight from input to output. This fixes the normal deep network vanishing gradient problem. The design includes convolutional layers for feature extraction, identity and convolutional blocks for feature transformation, and fully connected layers for classification. Image recognition studies showed ResNet50, trained on the enormous ImageNet dataset, reaching human accuracy. In addition to classification, this technique is used for item identification, facial recognition, and medical image analysis. Its versatility and effect in numerous computer vision domains are evident. ResNet50's advances are still helping construct deeper and more effective neural networks that can do more than automated visual recognition. [32]

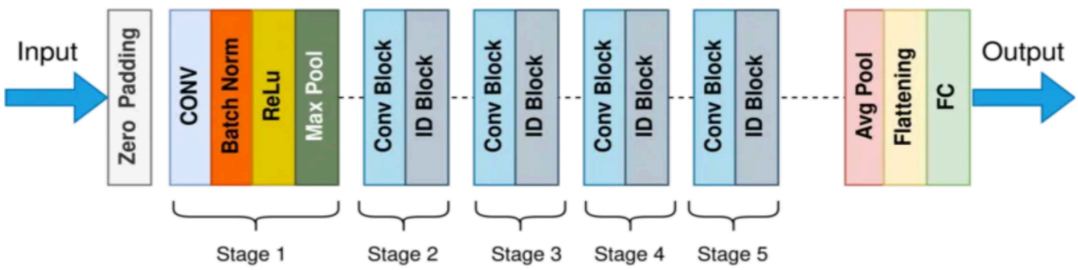


Figure 3.2.3: Resnet50 Architecture[32]

InceptionV3: Google Research created InceptionV3, a complex convolutional neural network, for the Inception model series. It was released in 2015 to improve picture identification precision and computation performance. Inception modules, which include parallel convolutional layers with different kernel sizes, are Inception V3's primary innovation. This method lets the network record properties at several dimensions and abstraction levels, enhancing its ability to learn complicated visual patterns. Inception V3 improves training stability and generalisation using factorised convolutions and aggressive regularisation methods like batch normalisation and dropout. To increase gradient flow and handle vanishing gradients, the approach uses auxiliary classifiers during training. Inception V3 has performed well in picture classification, object identification, and visual recognition benchmarks after pre-training on ImageNet. This method is popular for accurate and fast applications since it uses fewer computational resources and model size. This has made it essential to modern deep learning architectures for computer vision. [33]

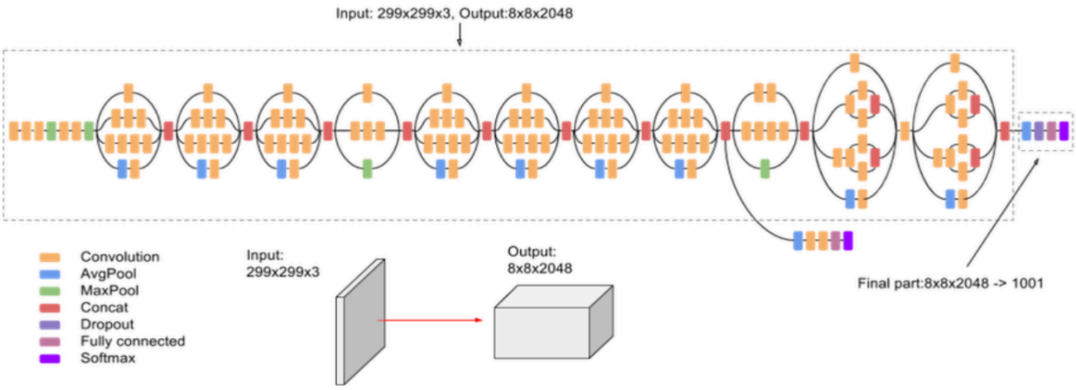


Figure 3.2.4: InceptionV3 Architecture [33]

Xception: The Xception architecture surpasses the Inception network in image classification tasks by employing depth-wise separable convolutions. In contrast to traditional CNNs, Inception V3 employs a distinct approach by utilising 1x1 convolutions to handle cross-channel correlations separately from spatial convolutions. Xception uses depthwise separable convolutions instead of Inception modules, which involve breaking down correlations into depthwise convolutions (where each input channel has its own filter) and pointwise convolutions (using 1x1 convolutions to merge outputs). By using parameters effectively, Xception achieves a tiny performance advantage over Inception V3 on ImageNet and a big advantage on a larger dataset with 350 million pictures and 17,000 classes. The key findings indicate that the inclusion of residual connections enhances accuracy and that the need for dropout decreases while working with larger datasets. The architectural modifications of Xception improve the efficiency of deep learning without adding complexity to the model. [34]

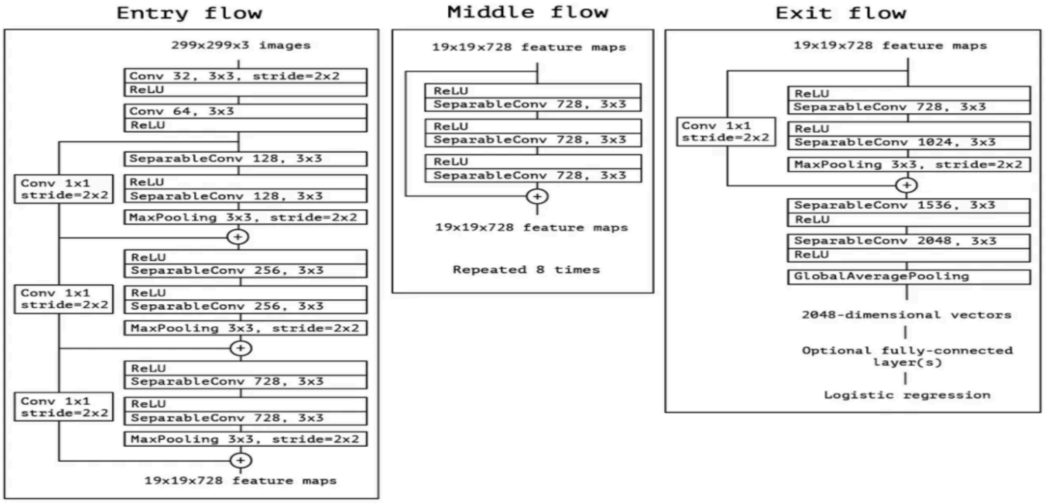


Figure 3.2.5: Xception Architecture [34]

VGG16: Oxford researchers A. Zisserman and K. Simonyan created the deep convolutional neural network VGG16 for large-scale image recognition. VGG16, introduced in "Very Deep Convolutional Networks for Large-Scale Image Recognition," has 16 layers, including convolutional and fully connected layers, with over 138 million parameters. VGG16 captures complex spatial hierarchies in 224x224 pixel pictures using tiny 3x3 convolution filters. ReLU activation algorithms introduce non-linearity to speed up training, and three fully linked layers with 4096 and 1000 channels provide robust classification. VGG16's simplicity and homogeneous architecture make it suitable for many applications despite its size. On ImageNet, which has over 14 million photos in roughly 1000 categories, it has top-5 test accuracy of 92.7%. VGG16's lengthy training durations, high model size (528 MB for learned weights), and many parameters make it susceptible to explosive gradients. VGG16 is still a popular image recognition architecture despite these issues. [35]

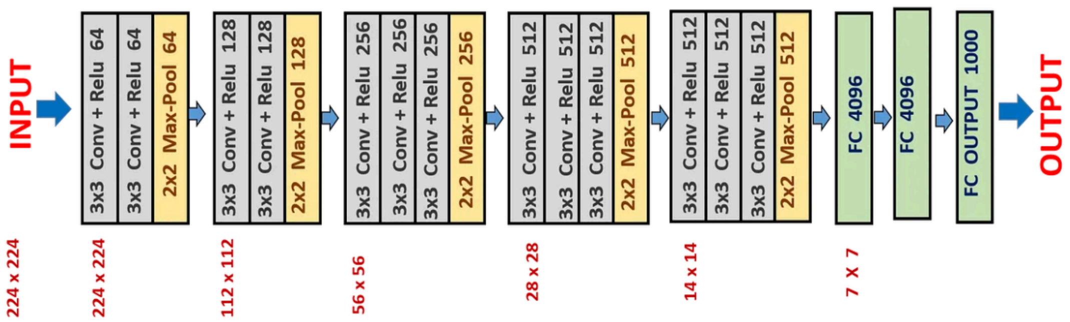


Figure 3.2.6: VGG16 Architecture [35]

Hybrid Model: In machine learning and neural networks, a "hybrid model" is a model that blends several different architectures or types of models to make the best use of each one's skills and do a better job on a given task. You can accomplish the combination by various methods; here are 3 of them:

- Architectural Integration in HM: - Integrates the architectures of numerous algorithms to create strong standalone algorithms. Examples of hybrid models are ANFIS, which combines fuzzy logic with artificial neural networks (ANN), and NBTree, which combines naïve Bayes and decision tree algorithms.
- Data Manipulation in HML:Combines data manipulation procedures with conventional ML techniques. Examples of combined techniques are FR-SVM (which combines fuzzy ranking and support vector machines) and PCA-ANN (which combines principal component analysis and artificial neural networks).
- Optimisation of Model Parameters in HML: Improves conventional optimisation techniques by including sophisticated evolutionary algorithms. Examples of combined optimisation algorithms are PSO-ANN (Particle Swarm Optimisation and Artificial Neural Network) and GA-ANFIS (Genetic Algorithm and Adaptive Neuro-Fuzzy Inference System).

Hybrid models are commonly created to tackle specific difficulties, such as enhancing accuracy, resilience, or efficiency, by using the complimentary capabilities of several models. They are extensively utilised in diverse fields like image recognition, natural language processing, and recommendation systems to attain greater performance in comparison to individual models [35].

3.3 Software and Hardware Requirements

For implementing our research work, we need high performance hardware and software requirements. Like deep learning frameworks, some libraries, and essential GPUs.

Equipment and Tools:

1. Development and Testing Servers
2. High-performance Computers for Development Team Design Software (e.g., Kaggle, Google Colab, Google Drive)
3. Framework (e.g., tensor flow)
4. Libraries (e.g., Python libraries, OpenCV libraries)
5. Collaboration Tools (e.g., Google Workplace, Google Drive, Telegram)
6. Version Control System (e.g., Git, Kaggle)
7. Plagiarism Checking Tools (e.g., Turnitin)

Software and Technologies:

1. Front-End Technologies (e.g., HTML, CSS, JavaScript)
2. Back-End Technologies (e.g., Python First API)
3. Server Hosting (e.g., Netlify)
4. Payment Gateway Integration (e.g., Bkash, Nagad, Rocket)

5. Security Software and SSL Certificates

Hardware Specifications:

1. Access to computational infrastructure with sufficient processing (e.g,memory)
2. Model development (e.g, GPU)

3.4 Project Management and Financial Analysis

Project Management:

Project Objectives:

- To develop a model for sign language recognition.
- To convert Bangla Sign into a natural language (our native language).
- To build an interface or web application that can build communication bridge between deaf and mute communities and the broader community, as well as empower them.

Project Timeline:

- Phase 1: Project Planning and Research (7 weeks)
- Phase 2: Work With Dummy Dataset (4 weeks)
- Phase 3: Raw Dataset Collection (12 weeks)
- Phase 4: Preprocessing (8 weeks)
- Phase 5: Model Implementation (7 weeks)
- Phase 6: Result Analysis and Improvement (3 weeks)
- Phase 7: Simple Software Design (5 week)

Financial Analysis:

Table 3.4.1: Estimated Cost for our Model

SN	Components	Estimated Cost (BDT)
01.	Visiting Stakeholders	1500
02.	Software and Tools	2000
03.	Data Collection and Processing	3000
04.	Documentation and Report Writing	1000
05.	Miscellaneous (e.g., licences, tools)	500
06.	Contingency (10% of total)	1000
Total Estimated Cost		9,000

3.5 Summary

The purpose of this research is to create a functional model for the identification of Bangla sign language, with the intention of enhancing communication between hearing and deaf groups. The approach involves a thorough process of gathering data through questionnaires and capturing photos, culminating in a dataset that encompass 36 Bangla alphabets and 10 numerical values. Data preparation procedures, including labelling, background removal, resizing, grayscale conversion, contrast improvement, and data augmentation, were utilised to improve data quality and prepare it for model training. After data augmentation, the dataset expanded to 11,422 images We studied and trained advanced CNN models such as ResNet50, InceptionV3, Xception, VGG16, and a hybrid ResNet50-InceptionV3 model. The hybrid model, which achieved superior classification performance by integrating the capabilities of both architectures, was exhibited. The project utilised a range of software tools and technologies for both creation and assessment, resulting in anticipated expenses amounting to 9,000 BDT.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

Several phases were involved in the implementation procedure. A heterogeneous dataset depicting Bangla sign language was initially compiled. These images were obtained from a group of 100 volunteers in different environments. The dataset expanded to 11,422 images after data augmentation, with an average of 120 images per class, encompassing 46 classes. The data underwent preprocessing procedures including labelling, background removal, downsizing to dimensions of 256x256 pixels, grayscale conversion, contrast improvement, and data augmentation to prepare it. The implementation of five advanced pre-trained deep learning models—ResNet50, InceptionV3, Xception, VGG16, and a hybrid model that combines ResNet50 and InceptionV3—was conducted. The hybrid model included characteristics from both ResNet50 and InceptionV3. These features were combined and then processed through fully connected layers using L2 regularisation and dropout to enhance the classification performance. The models were trained using the Adam optimizer, the sparse categorical cross-entropy loss function, learning rate scheduling, and early stopping to make sure they were trained well and to avoid overfitting as much as possible. Data augmentation significantly improved the ability to generalise. In Bangla sign language classification tasks, the hybrid model did better than others because it captured both specific and abstract traits more effectively.

4.2 Train Model

Resnet50: We have followed the following steps to train ResNet50:

- **Model Base:** ResNet50 is a deep convolutional neural network that has been pre-trained on ImageNet. It is configured without the top classification layer in order to preserve its ability to extract features unique to visual input.
- **Feature Extraction Layers:** To improve feature representation and minimise dimensionality, a Global Average Pooling layer is added. Two dense layers with ReLU activation are then added to the pooled output to extract high-level features.
- **Regularisation and Normalisation:** Uses a dropout layer to prevent overfitting and batch normalisation to stabilise and accelerate training by normalising inputs between layers.
- **Output Layer:** The final layer uses softmax activation to categorise photos into 46 different Bangla sign language groups. The model is optimised using the Adam optimizer with sparse categorical cross-entropy loss. It is trained for accurate gesture identification across 40 epochs on given datasets.

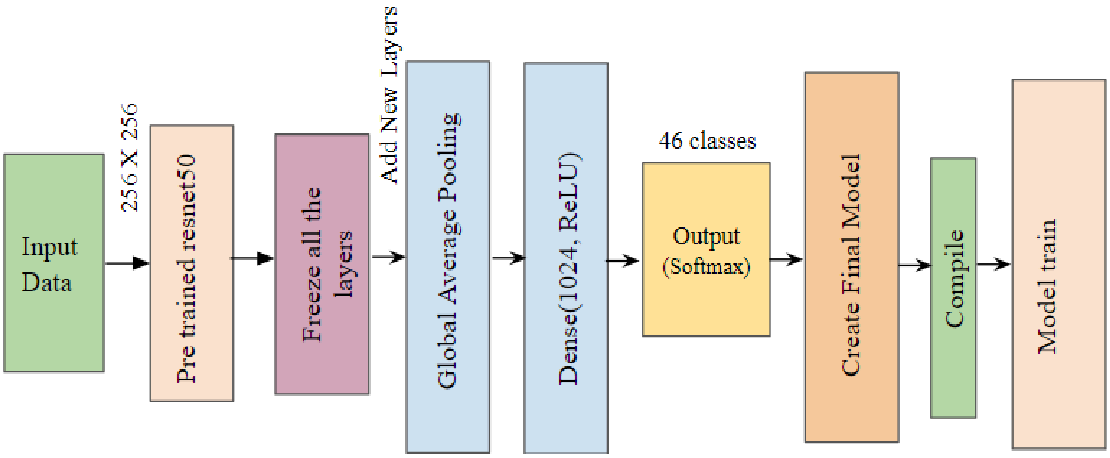


Figure 4.2.1: Working flow of Resnet50 model for our work

VGG16: The steps of the VGG16 model that we applied are as follows:

- **Model Initialization:** The VGG16 model uses pre-trained ImageNet weights where the top classification layer is not included. The input shape is defined as (256, 256, 3) to ensure compatibility with the dataset.
- **Freeze Pre-trained Model Layers:** All the layers that are stored in the pre-trained model are freezed to prevent being updated during training. And also for fixing the learned features from ImageNet.
- **Fine-tuning with New Layers:** The model used Global Average Pooling to reduce the image size (height and width) for retaining channel information. This layer includes a dense layer with 1024 neurons, ReLU activation with L2 regularisation (0.001), and batch normalization to make activations more normal.
- **Model Compilation and Training:** The Adam optimizer is used to compile the model, and the sparse categorical cross-entropy loss function, accuracy metric, and learning rate of 0.0001 are all used.
- **Training Execution:** The model is trained using the 'fit' method on the training dataset, and validation is performed on the validation dataset. The value of epochs is selected to control the duration of the training process.

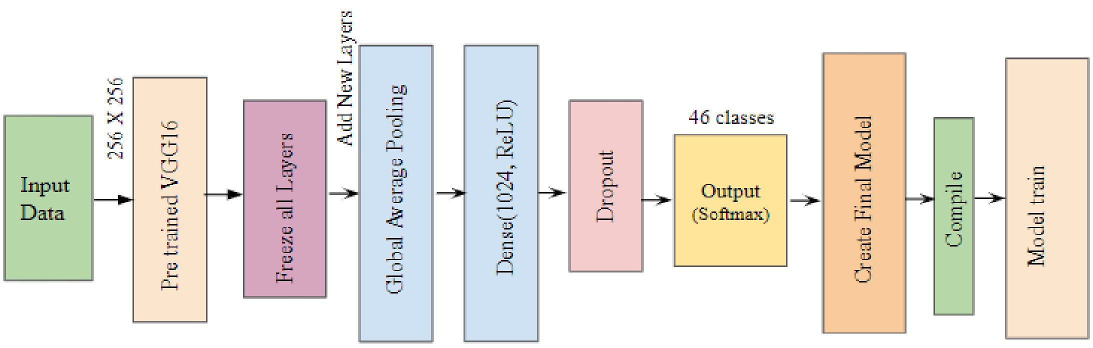


Figure 4.2.2: Working flow of VGG16 for our work

InceptionV3: Below is a concise summary of the InceptionV3 model that we have implemented:

- **Model Initialization:** An InceptionV3 model on ImageNet is initialised using pre-trained weights; the top classification layer is not included. The input shape is defined as (256, 256, 3) to ensure compatibility with the dataset.
- **Fine-tuning Layers:** To enable the model to adjust to certain properties of the new dataset, it selectively unfreezes the final 100 layers of the pre-trained InceptionV3 model.
- **Customised Top Layers:** To improve feature learning, additional layers are added to the pre-trained model. A dense layer with 2048 neurons and ReLU activation with L2 regularisation (0.001), batch normalization to make activations more normal, and global average pooling to cut down on the number of dimensions is a part of this layer.
- **Model Compilation and Training:** The Adam optimizer is used to compile the model, and the sparse categorical cross-entropy loss function, accuracy metric, and learning rate of 0.0001 are all used. The learning rate scheduling and early stopping callbacks are utilised to enhance training efficiency and mitigate overfitting.
- **Training Execution:** The model is trained using the 'fit' method on the training dataset, and validation is performed on the validation dataset. The value of epochs is selected to control the duration of the training process.

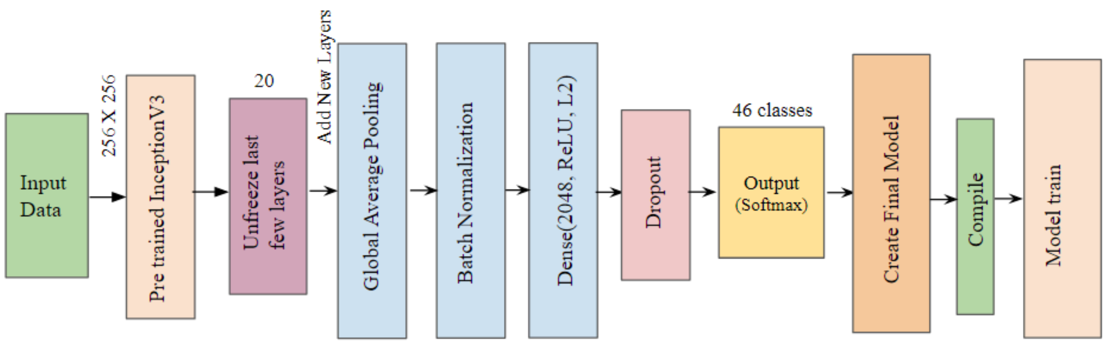


Figure 4.2.3: Working flow of InceptionV3 for our work

Xception Model: An outline of the Xception model that we have used is shown below:

- **Model Initialization:** The code initialises an Xception model with pre-trained weights from ImageNet, except for the top classification layer. The model's input shape is defined as (256, 256, 3) to ensure compatibility with the dataset.

- **Fine-tuning Layers:** This process involves adjusting the parameters of the top 50 layers of the pre-trained Xception model to better suit the unique characteristics of the Bangla sign language identification problem.
- **Model Architecture:** The Xception base model is used to extract features, and global average pooling is employed to minimise the spatial dimensions of the features.
- **Customised Top Layers:** Additional layers are appended to the pre-trained Xception foundation. These layers consist of two dense layers with 1024 and 512 neurons, respectively. Both layers utilise ReLU activation and L2 regularisation with a value of 0.01. Dropout layers, with a dropout rate of 0.5, are included after each dense layer to mitigate the risk of overfitting.
- **Model Compilation and Training:** The Adam optimizer is used to make the model, and the sparse categorical cross-entropy loss function is used along with an initial learning rate of 0.00001. The accuracy metric is used to measure performance. The learning rate scheduling callback and the early stopping callback are utilised to enhance training efficiency and mitigate overfitting.
- **Training Execution:** The model is trained by applying the 'fit' technique to the training dataset, while validation is carried out on the validation dataset. The value of epocs is determined to regulate the duration of the training process.

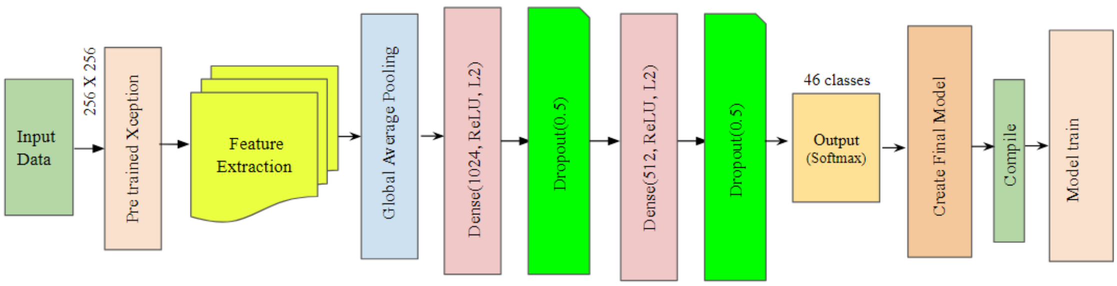


Figure 4.2.4: Working flow of Xception Model for our work

Hybrid Model: Here is the breakdown of our hybrid model

- **Model Initialization:** For the purpose of recognising Bangla sign language, the hybrid model integrates the ResNet50 and InceptionV3 architectures. The pre-trained weights from both models are imported, and the top 50 layers are adjusted to better suit the new dataset.
- **Standardised Input:** To guarantee consistency throughout the dataset and enable consistent feature extraction from ResNet50 and InceptionV3, images are standardised to 256x256 pixels.

- **Feature Extraction and Pooling:** To minimise spatial dimensions while maintaining significant features, features are retrieved simultaneously from both models and then pooled globally.
- **Fully Connected Layers:** To make generalisation better and avoid overfitting during training, pooled features that have been joined together are sent through fully connected layers that regularise and drop out L2.
- **Compiling and Training the Model:** The hybrid model is trained with sparse categorical cross-entropy loss and put together using the Adam optimizer with 0.00001 as the starting learning rate. The implementation of learning rate scheduling and early stopping techniques aims to optimise training efficiency and mitigate overfitting.
- **Training Execution:** The model is trained using 'fit' into the training dataset, and validation is performed on the validation dataset. The value of epochs is selected to control the duration of the training process.

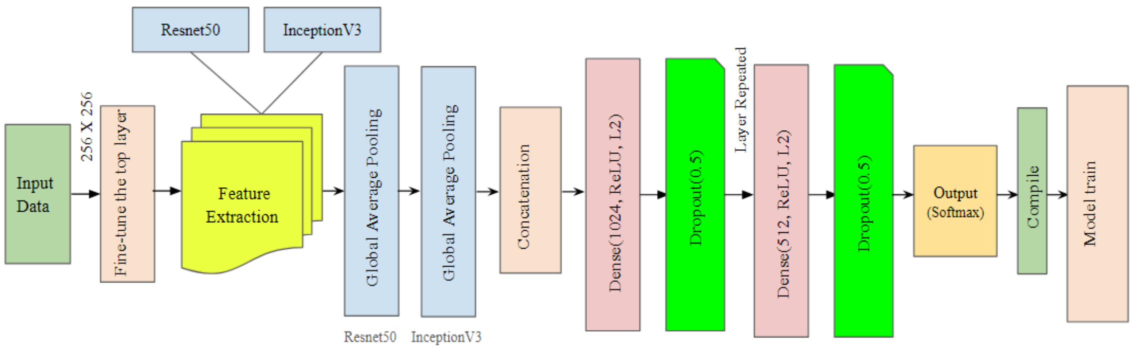


Figure 4.2.5: Working flow of the Hybrid model (Resnet50+InceptionV3) for our work

Web Application: We made a web application for building the bridge between sign language users and Bangla readers by translating Bangla sign language captured images into Bangla characters. The application consists of a client-server architecture, with a frontend developed using HTML, CSS, and JavaScript and a backend implemented in Python (FirstAPI). Users can upload hand signs through the web interface, which are then processed using background removal, resizing, gray scaling, and contrust enhancement techniques. The translation process is based on our Hybrid Model [Figure 4.2.5, which describes the working flow of Hybrid model]. Figure 4.2.6 shows the part of web application which has a basic interface. There are 3 sections- image upload, processing the images by preprocessing techniques that we used in this research without augmentation, and lastly, using the hybrid model, which gives the output of the respected character.

Figure 4.2.6: Interface of our Web Application

4.3 Model Evaluation

Our dataset was utilised to evaluate the trained models against five pre-trained deep learning models: Resnet50, VGG16, InceptionV3, Xception, and the Hybrid Model (Resnet50+InceptionV3). The models showed excellent accuracy during both training and validation. But there may be a problem with overfitting because the training accuracy was marginally greater than the validation accuracy. Our models have all gotten more than 90%, which means they were trained well on the dataset. The training and validation accuracy of these deep learning models is presented in Table 4.3.1.

Table 4.3.1: The Training and Validation Accuracy of Used Model

Model	Training Accuracy	Validation Accuracy
Resnet50	0.9994	0.9406
VGG16	0.9836	0.9547
Inception	0.9995	0.9454
Xception	1.0000	0.9489
Hybrid Model	1.0000	0.9639

Training & Validation Accuracy and loss of ResNet50:

Represent the training and validation accuracy of the Resnet50 model, with the number of epochs on the x-axis and the accuracy and loss presented on the y-axis. And the data are not overfitted.

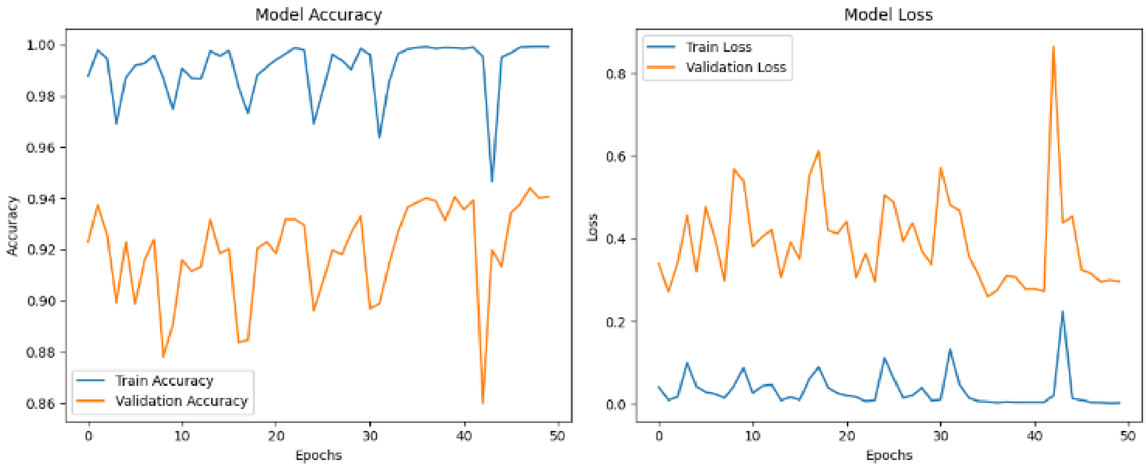


Figure 4.3.1: Training & Validation Accuracy and Loss of ResNet50 over the epochs

Training & Validation Accuracy and Loss of VGG16:

Figure 4.3.2 represents the training and validation accuracy of the VGG16 model, with the number of epochs on the x-axis and the accuracy and loss presented on the y-axis.

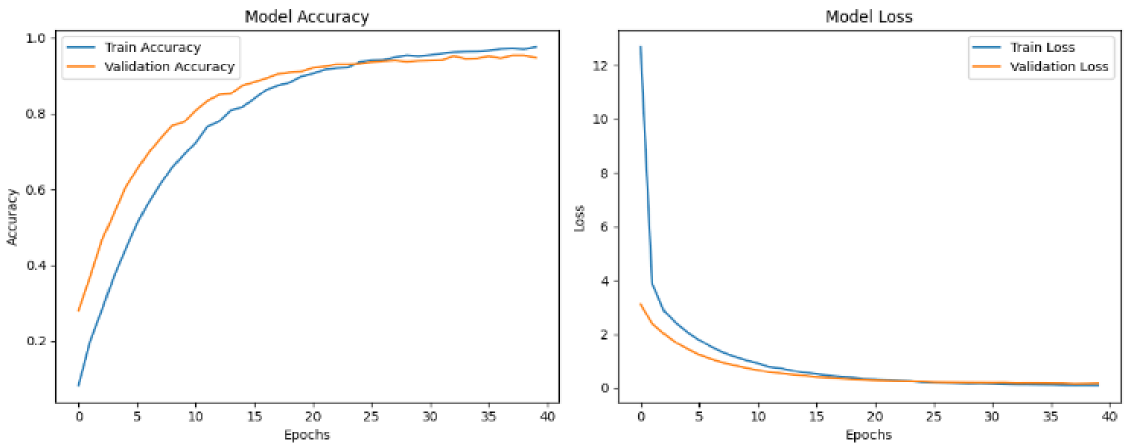


Figure 4.3.2: Training & Validation Accuracy and Loss of VGG16 over the epochs

Training & Validation Accuracy and Loss of Inception V3:

Figure 4.3.3 represents the training and validation accuracy of the InceptionV3 model, with the number of epochs on the x-axis and the accuracy and loss presented on the y-axis.

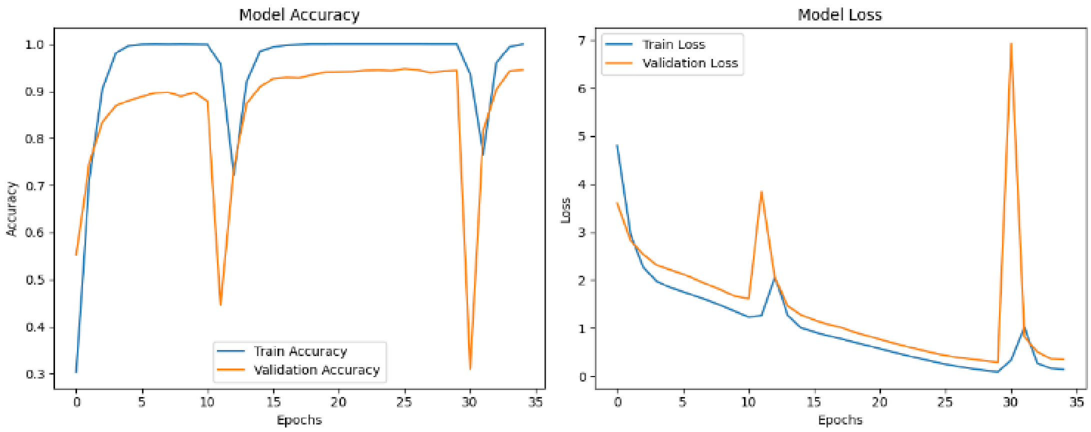


Figure 4.3.3: Training & Validation Accuracy and Loss of InceptionV3 over the Epochs

Training & Validation Accuracy and Loss of Xception:

Figure 4.3.4 represents the training and validation accuracy of the Xception model, with the number of epochs on the x-axis and the accuracy and loss presented on the y-axis.

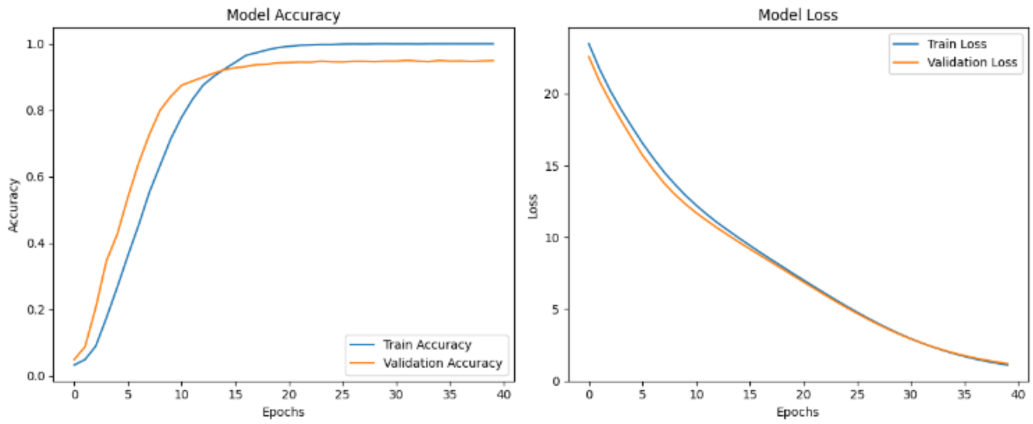


Figure 4.3.4: Training & Validation Accuracy and Loss of the Xception Model over the Epochs

Training & Validation Accuracy and Loss of the Hybrid Model (Resnet50+InceptionV3):

Figure 4.3.5 represents the training and validation accuracy of the InceptionV3 model, with the number of epochs on the x-axis and the accuracy and loss presented on the y-axis.

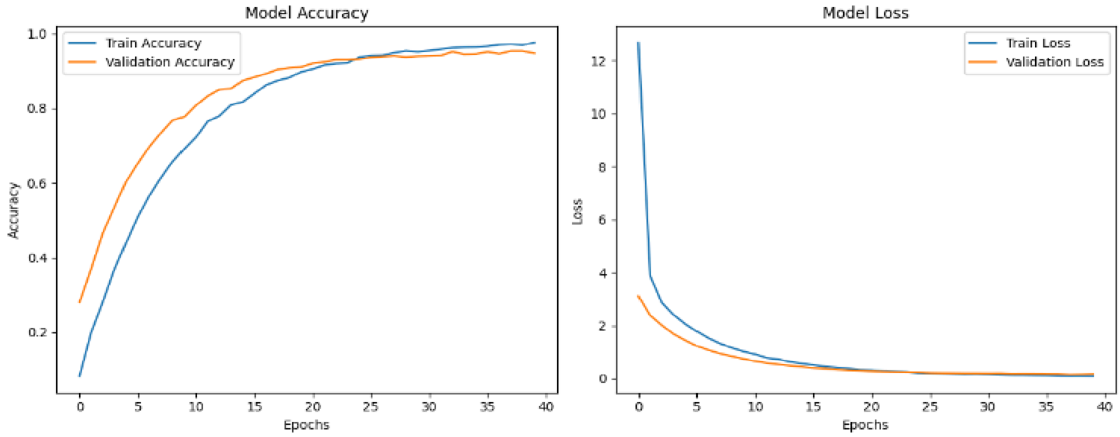


Figure 4.3.5: Training & Validation Accuracy and Loss of Hybrid Model over the Epochs

4.4 Summary

This study was carried out using five highly trained CNN architecture deep learning models: ResNet50, InceptionV3, Xception, VGG16, and a hybrid model that combines ResNet50 and InceptionV3. All five models demonstrated high performance during the experiments, the hybrid model achieved a training accuracy of 100% and a validation accuracy of approximately 96%. Similarly, the Xception model reached a training accuracy of 100% and a validation accuracy of approximately 95%. InceptionV3 and ResNet50 both achieved approximately 99% training accuracy, with validation accuracies of approximately 94%, while VGG16 reached 98% training accuracy with a validation accuracy of approximately 95%. The accompanying graph shows the training and validation accuracy curves for each of the five models, providing an illustration of their performance. The x-axis represents the number of epochs, while the y-axis represents accuracy and loss. Moreover, overfitted data does not exist in any model.

CHAPTER 5
RESULT AND ANALYSIS

5.1 Overview

This paper used several pre-trained models to find the best result. A heterogeneous dataset of 1906 photos depicting Bangla sign language was initially compiled. The dataset expanded to 11,422 images after data augmentation, with an average of 120 images per class, encompassing 46 classes. The implementation of five advanced pre-trained deep learning models—ResNet50, InceptionV3, Xception, VGG16, and a hybrid model that combines ResNet50 and InceptionV3—was conducted. Mainly, CNN base methods are evaluated using many machine learning classification model performance metrics. With the help of many aspects—accuracy, precision, recall, F1 score, and confusion matrix the best model can be found.

5.2 Experimental Results

This section compares the five models used - ResNet50, InceptionV3, Xception, VGG16, and a hybrid model that combines ResNet50 and InceptionV3. The models showed excellent accuracy during both training and testing. But there may be a problem with overfitting because the training accuracy was marginally greater than the validation accuracy. Our models have all gotten more than 90%, which means they were trained well on the dataset. The training and test accuracy of these deep learning models are presented in Table 5.2.1.

Table 5.2.1: Training and Test Accuracy of Used Models

Model	Training Accuracy	Test Accuracy
ResNet50	99%	91%
VGG16	97%	94%
InceptionV3	99%	93%
Xception	100%	95%
Hybrid Model (ResNet50+InceptionV3)	100%	96%

The comparative analysis of several pre-trained deep Convolutional Neural Network (CNN) architectures, namely ResNet50, InceptionV3, Xception, VGG16, and a hybrid model that combines ResNet50 and InceptionV3, for translating the Balga Sign Language into native Bangla Language. Notably, the findings showcase that the Hybrid model (ResNet50+InceptionV3) emerged as the top-performing model, attaining the highest accuracy of 96%. Compared with the top hybrid model, VGG16 and Xception are quite similar in performance, respectively at 94% and 95%, with the trained accuracy at 97% and 99%. InceptionV3 performed well, with an accuracy rate of 93% and a train accuracy of 99%. And ResNet50 indicates a marginally lower performance than the other four models, which achieved 91% with a trained accuracy of 99%. This nuanced differentiation in accuracy underscores the importance of carefully selecting the appropriate CNN architecture in the context of translating Bangla Sign Language to native Bangla Language. The detailed accuracy metrics provided in Table 5.2.1 offer a valuable reference point for understanding the comparative strengths and weaknesses of these models.

5.3 Performance/ Comparative Analysis:

CNN base methods are evaluated using many machine learning classification model performance metrics. The following table 5.4.1 is a confusion matrix that is used for the classification problem:

Table 5.4.2: Classification Matrix

		Predicted Class	
Actual Class		YES	NO
	YES	True Positive(TP)	False Negative(FN)
	NO	False Positive(FP)	True Negative(TN)

The following are the fundamentals for evaluating our model that we have used for translating Bangla Sign Language to Bangla (native language) characters:

Accuracy: Accuracy is the model's predicted percentage. It is the percentage of properly classified data in total samples [37] . Equation;

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision: Precision shows how many accurately anticipated instances were positive. Pr is helpful when False Positive (FP) matters more than False Negative (FP) [37]. It's this equation;

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

Recall: Recall, or sensitivity, measures how many positive expected occurrences were classified accurately. Recall is a helpful statistic when False negative (FN) prevails over False positive (FP) [37]. The equation below is:

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

F1 Score: The F1 score is a recall-precision metric of categorization accuracy. As the F1-score is the average of precision and recall, it gives a complete picture. Precision and recall are best when they are equal [37]. Equation;

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

The following tables are used to evaluate performance of our models:

Table 5.3.3: Performance Matrix for Resnet50

ResNet50			
	Precision	Recall	F1-Score
Accuracy			0.91
Macro Average	0.91	0.91	0.91
Weighted Average	0.92	0.91	0.91

Table 5.3.4: Performance Matrix for VGG16

VGG16			
	Precision	Recall	F1-Score
Accuracy			0.94
Macro Average	0.95	0.94	0.94
Weighted Average	0.95	0.94	0.94

Table 5.3.5: Performance Matrix for Inception V3

Inception V3			
	Precision	Recall	F1-Score
Accuracy			0.93
Macro Average	0.93	0.93	0.93
Weighted Average	0.93	0.93	0.93

Table 5.3.6: Performance Matrix for Xception

Xception			
	Precision	Recall	F1-Score
Accuracy			0.95
Macro Average	0.94	0.94	0.94
Weighted Average	0.95	0.95	0.95

Table 5.3.7: Performance Matrix for Hybrid Model

Hybrid(ResNet50+InceptionV3)			
	Precision	Recall	F1-Score
Accuracy			0.96
Macro Average	0.96	0.96	0.96
Weighted Average	0.96	0.96	0.96

Confusion matrix: A confusion matrix is a visualisation and summarization of the performance of a classification algorithm. It is mainly a table that is used to define the performance of a classification algorithm. Figures 5.3.1 to 5.3.5 show all the confusion matrixes of our used model. Accuracy can be misleading if used with imbalanced datasets, and therefore, a confusion matrix can be useful for evaluating performance.

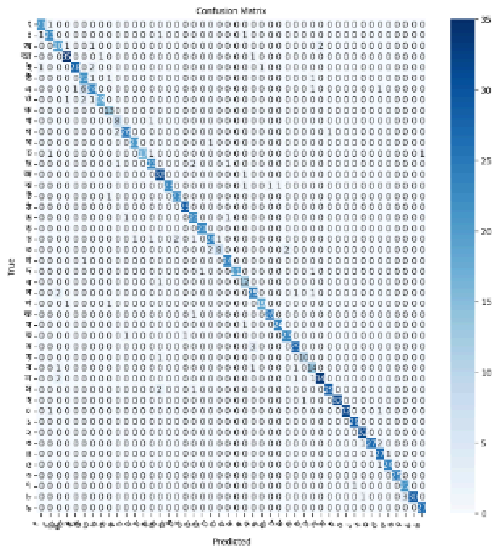


Figure 5.3.1: Confusion Matrix of ResNet50

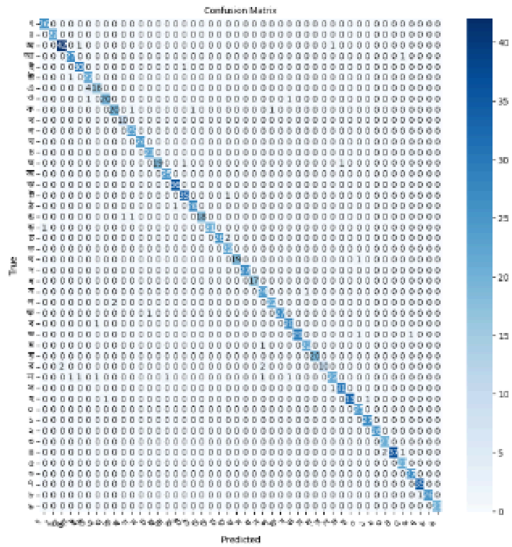


Figure 5.3.2: Confusion Matrix of VGG16

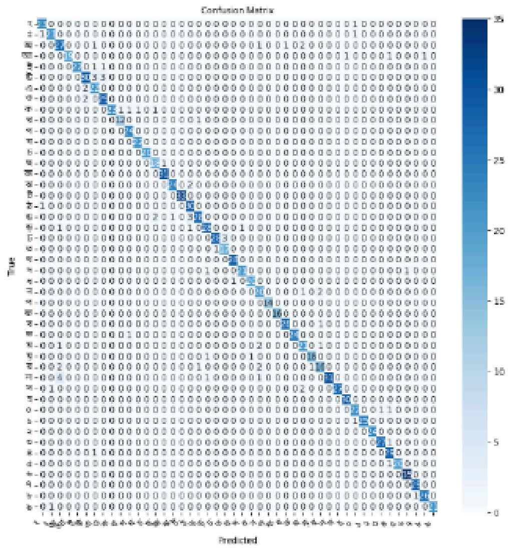


Figure 5.3.3: Confusion Matrix of InceptionV3

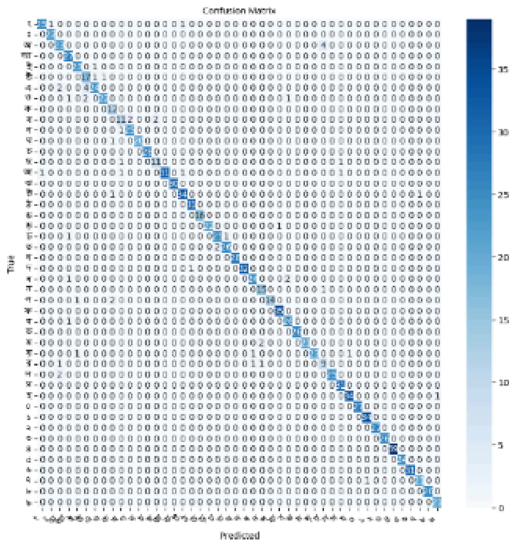


Figure 5.3.4:: Confusion Matrix of Xception

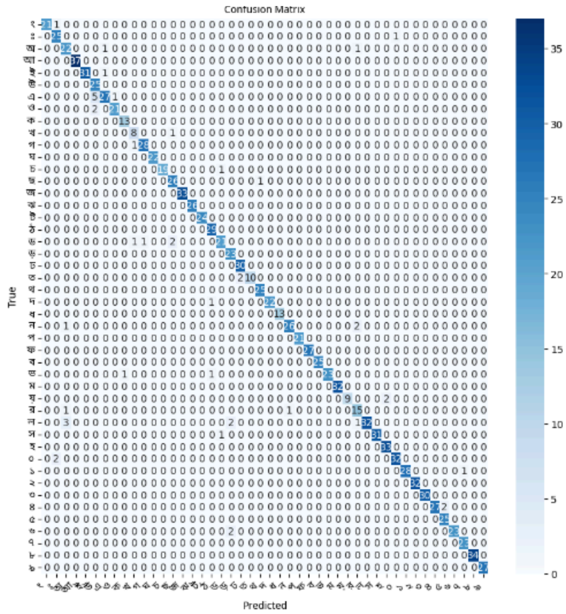


Figure 5.3.5: Confusion matrix of Hybrid Model

Visualizations: In image classification, visualization involves displaying the input pictures next to their actual and expected labels. This facilitates the comprehension of model performance and the identification of areas that require improvement. The visualization process includes displaying real-time predictions and highlighting misclassification. The following are some visualisation examples of our models that we have implemented. Though all of our models have achieved more than 90% accuracy, some of them do an excellent job of recognising sign language; some of them (e.g., Resnet50) have predicted at most 1 or 2 wrong signs.

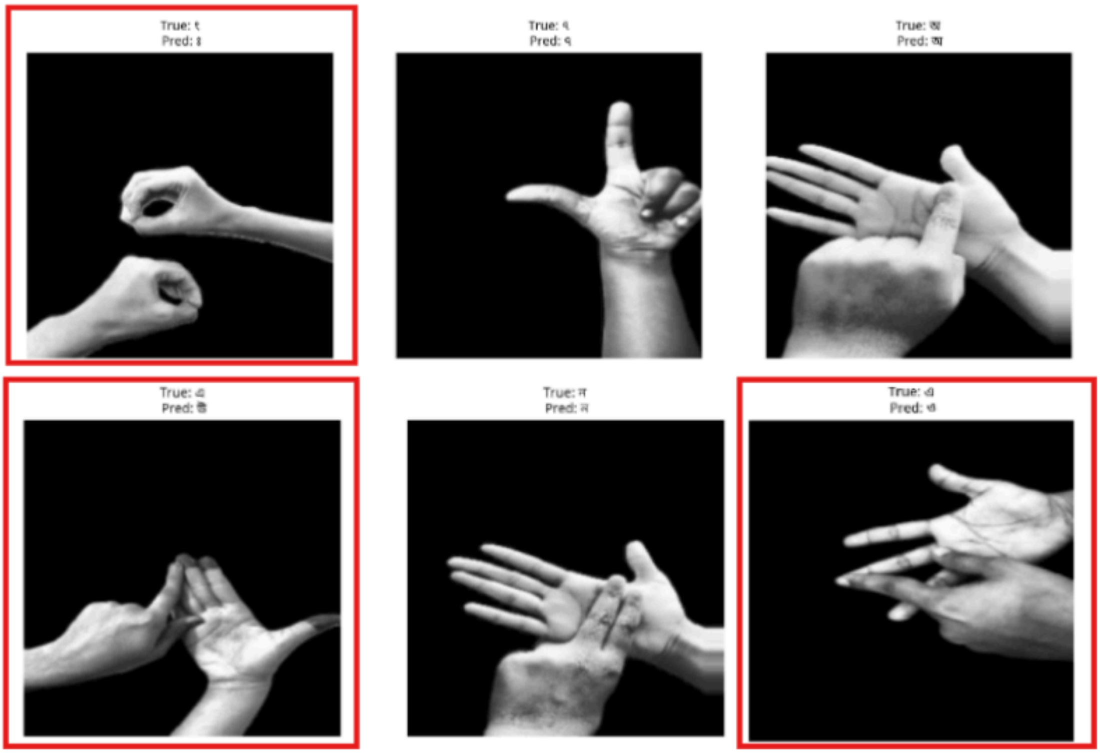


Figure 5.3.6: Visualization of Random Images using ResNet50

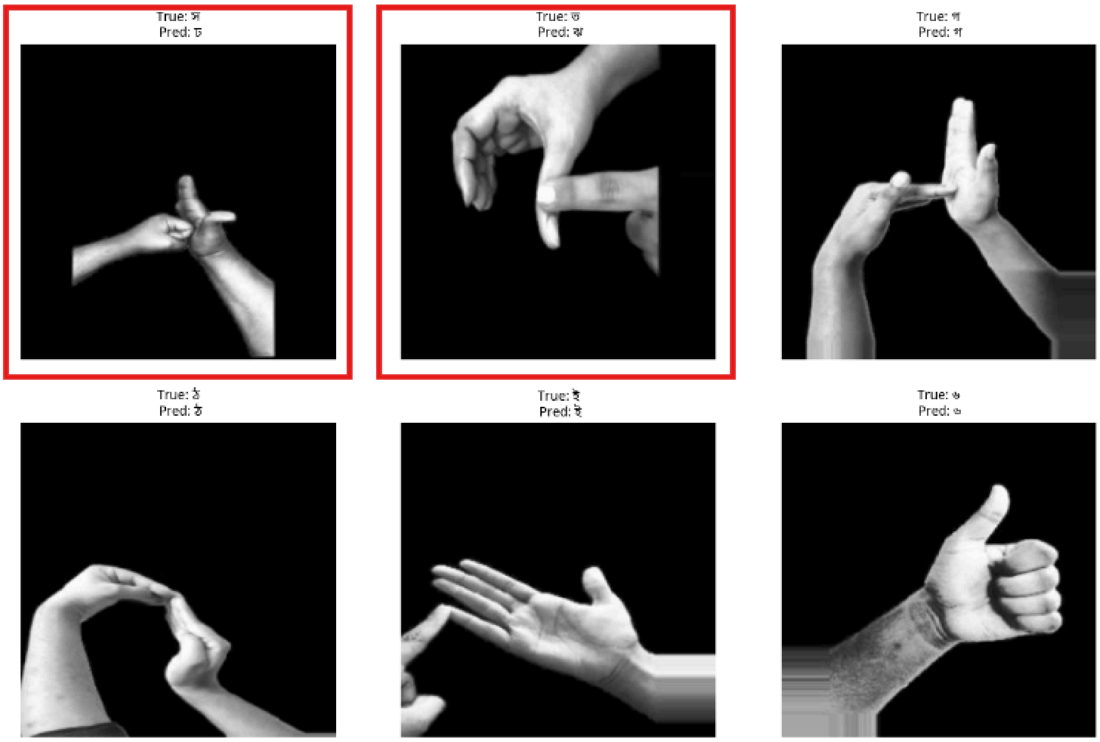


Figure 5.3.7: Visualization of Random Images using VGG16

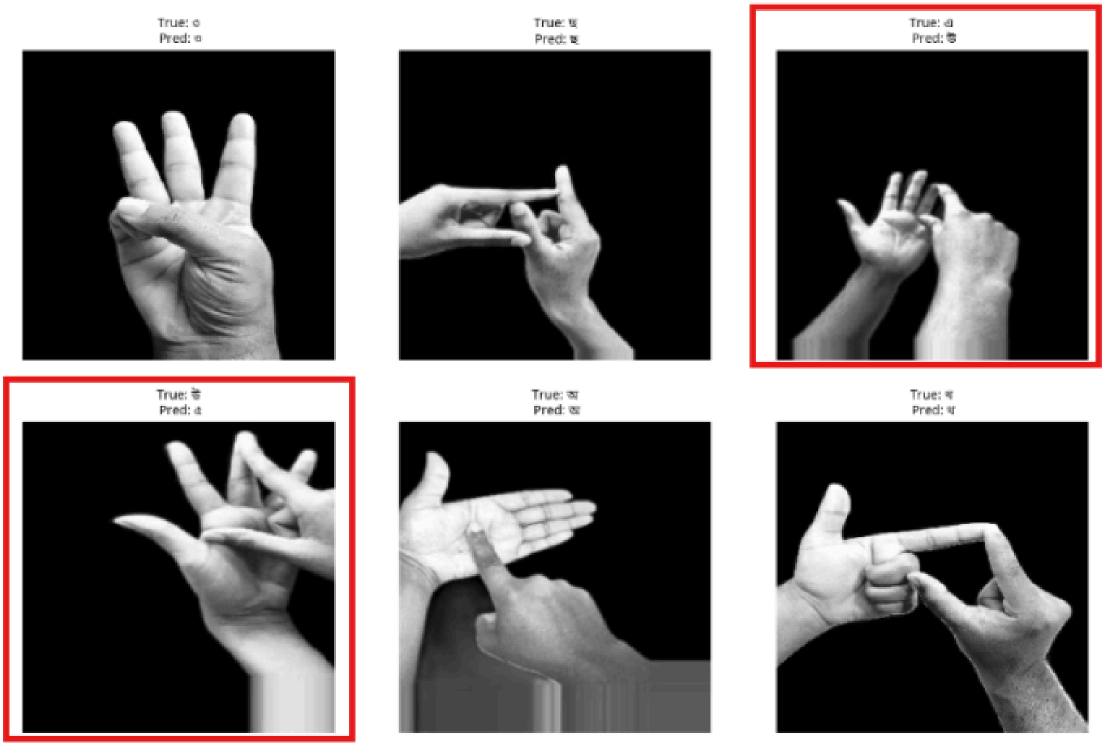


Figure 5.3.8: Visualization of Random Images using InceptionV3

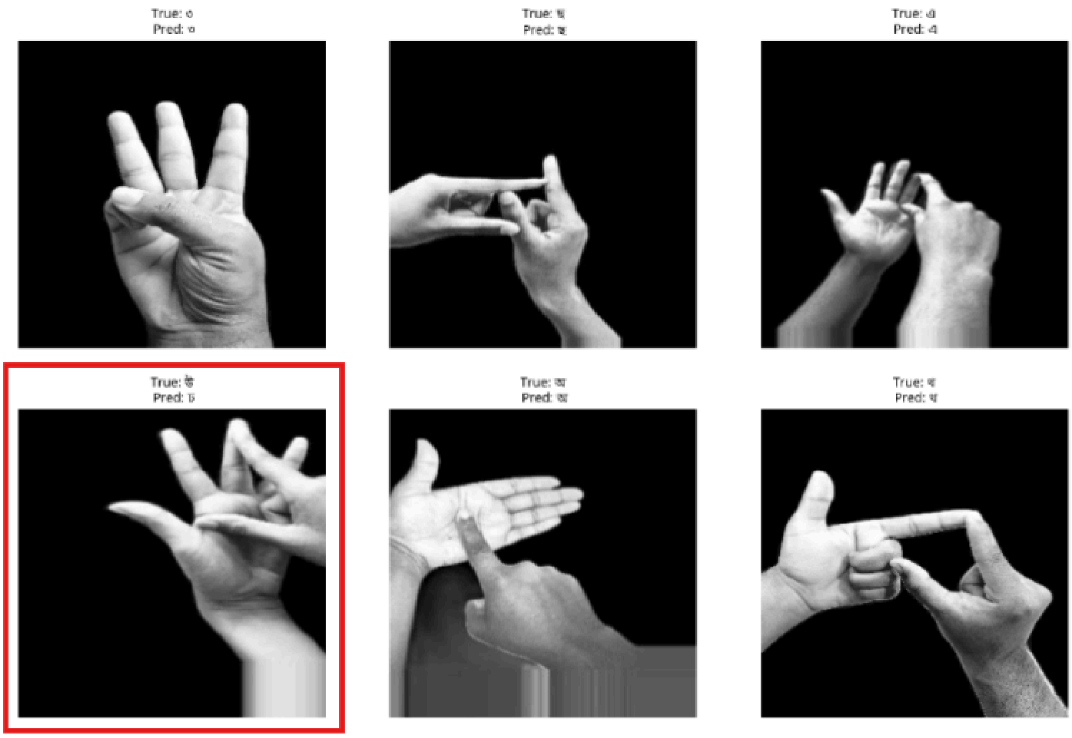


Figure 5.3.9: Visualization of Random Images using Xception

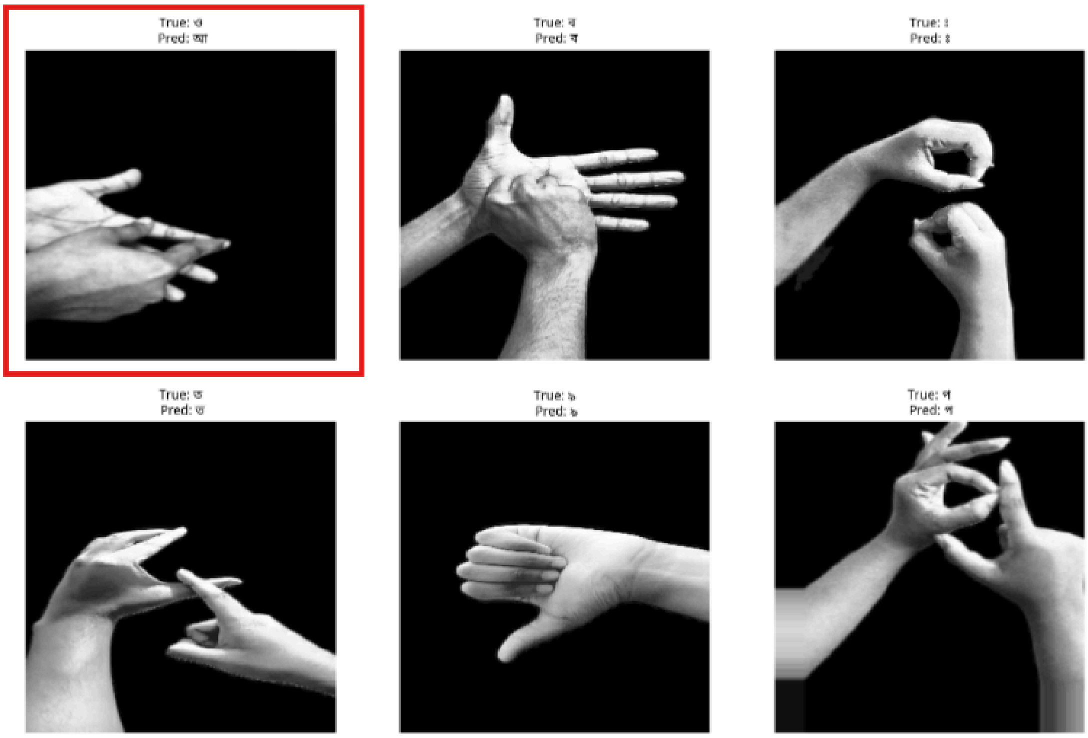


Figure 5.3.10: Visualization of Random Images using Hybrid Model

Web Application: Our web application consists of a client-server architecture, with a frontend developed using HTML, CSS, and JavaScript and a backend implemented in Python (FirstAPI). Figure 4.2.6 shows the web application. There is 3 sections- Image upload, Processed the images by preprocessing techniques which we used in this research without augmentation, and the lastly by using Hybrid model which gives the output of the respected character.

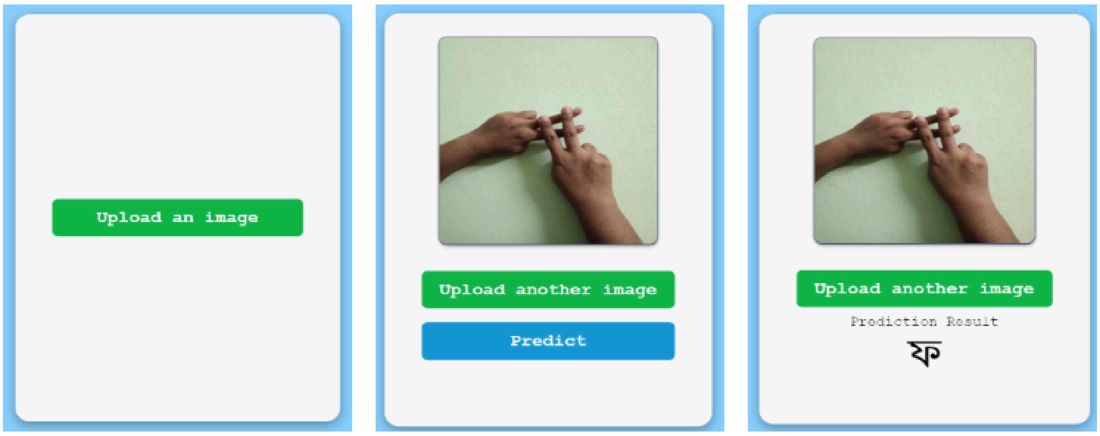


Figure 5.3.11: Visualization of Web Application

5.4 Summary

This study of “Translating Bangla Sign Language to Character” was carried out using five highly trained CNN architecture deep learning models: ResNet50, InceptionV3, Xception, VGG16, and a Hybrid Model that combines ResNet50 and InceptionV3. The findings showcase that the Hybrid model (ResNet50+InceptionV3) emerged as the top-performing model, attaining the highest accuracy of 96%. The confusion matrix summarises the performance of all algorithms (Figures 5.3.1 to 5.3.5). And with the help of visualisation, we display the input pictures next to their actual and expected labels. While every model attained an accuracy rate of more than 90%, some—like ResNet50—showed remarkable results with very few errors in sign identification.

CHAPTER 6
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

There is a considerable improvement in the quality of life for people in Bangladesh who have hearing and speech impairments as a result of the creation of a Bangla sign language recognition model. These persons are given the ability to express themselves more effectively and participate more completely in social, educational, and professional activities as a result of the model's provision of a dependable method of communication through the interpretation of sign language. An improvement in their overall quality of life is brought about by this technological innovation, which makes it easier for them to gain access to important services, educational possibilities, and work chances that were previously restricted due to communication obstacles.

6.2 Impact on Society & Environment

The approach is critical to empowering people with hearing and speech impairments by offering a dependable form of communication that is consistent with their language and cultural choices. This empowerment fosters their self-assurance and motivates them to assert their rights and capabilities within society.

The model enables hearing individuals to engage more meaningfully with the deaf and mute community by facilitating effective communication. This interaction fosters empathy, mutual respect, and comprehension, resulting in a more cohesive and harmonious social structure that embraces and appreciates diversity. The Bangla sign language recognition model's implementation in a variety of sectors, including healthcare and education, guarantees that all members of society have access to essential services. This inclusivity not only enhances the quality of life for individuals with disabilities but also contributes to the general well-being and prosperity of the local community.

In general, the Bangla sign language recognition model's societal influence surpasses mere technological advancement. It catalyses a cultural transformation towards social justice, equality, and inclusivity, thereby establishing the groundwork for a more compassionate and resilient society in which each individual's voice is acknowledged and respected.

6.3 Ethical Aspects

The development and deployment of the Bangla sign language recognition model prioritise respect for the rights of individuals with disabilities, inclusivity, and cultural diversity from an ethical perspective. The Bangla sign language community's active participation in the design and implementation phases is essential to this approach, as it guarantees that the technology faithfully reflects their cultural preferences and linguistic nuances. This collaborative process not only reinforces regard for the community's heritage and identity but also improves the model's effectiveness.

Privacy and data protection are important social issues that need strict means to protect user data and make sure data is used responsibly. In order to establish trust and maintain confidentiality, it is imperative to implement transparent data management practices and robust security protocols. Additionally, the model promotes user autonomy and independence among the hearing and speech-impaired population in Bangladesh, with the intention of empowering rather than exploiting them. It endeavours to ensure that its benefits are accessible to all, thereby reducing socio-economic disparities and promoting social inclusion.

In short, the Bangla sign language identification model is based on a set of moral principles that encourage a technologically welcoming environment that values diversity, safeguards privacy, and gives disabled people more power. By following these rules, the model helps make society more fair and polite, where everyone can take part and do well.

6.4 Sustainability Plan

Maintaining the Bangla sign language detection model requires ongoing work. This involves continuing research and development to increase accuracy and expand the model's vocabulary coverage; also, future NLP models can be included for correct sign-to-text translation. In order to increase awareness, adoption, and integration of the technology into a variety of sectors, including education, healthcare, and public services, it is essential to collaborate with local communities, educational institutions, and governmental organisations. Maintaining, updating, and expanding the model over time will be easier with long-term funding sources and relationships with stakeholders. This will make sure that it stays useful and effective over time.

6.5 Summary

In summary, the creation of a Bangla sign language recognition model is a critical technological advancement that is designed to improve the quality of life and communication of individuals with hearing and speech impairments in Bangladesh. This innovation not only addresses the communication barriers encountered by the deaf and mute communities, but also encourages social inclusion, cultural diversity, and ethical responsibility. The model promotes a more inclusive society by eliminating communication barriers, ensuring that all individuals have equal opportunities to contribute, flourish, and participate. Its sustainable implementation guarantees the community's ongoing benefits and reaffirms Bangladesh's dedication to inclusivity and accessibility for all.

CHAPTER 7
CONCLUSION AND FUTURE WORK

7.1 Conclusions

The Bangla sign language recognition model offers a significant technological innovation aimed at improving communication and inclusion for people with hearing and speech impairments in Bangladesh. The concept facilitates social integration, empowers users, and promotes a more equal society by accurately interpreting sign language and bridging the communication gap. The development approach has prioritised ethical issues, including respect for cultural diversity and privacy protection, to ensure that the model caters to the particular requirements and preferences of the Bangla sign language community.

7.2 Further Suggested Works

Moving forward, more research and development are needed to improve the model's accuracy and capabilities, notably in recognising a wider variety of vocabulary and identifying dialectical variances within Bangla sign language. Collaboration with linguists, educators, and technology specialists will be essential in improving the model and incorporating it into educational curriculum, healthcare systems, and public services in a more comprehensive manner.

In addition, investigating the possibilities of mobile applications and wearable technologies might greatly enhance the model's accessibility and usefulness in everyday situations, enabling users to converse effortlessly in many locations. In the future, the incorporation of Natural Language Processing (NLP) models shows potential for directly converting sign language motions into written text. This technological breakthrough has the potential to completely transform communication for those who have hearing impairments, allowing them to engage more extensively in digital and online relationships. These technologies improve accessibility and inclusion by connecting sign language users with the wider community.

The Bangla sign language recognition model has the potential to transform into a versatile instrument that enhances the communication and interaction capabilities of users, thereby empowering them and enriching their daily lives, by conducting research and development in these areas.

7.3 Limitations/ Conflict of Interests

The development of a Bangla sign language recognition model is confronted with numerous substantial challenges, despite its probable advantages. The lack of standardisation in Bangla sign language, which causes differences in signals across different areas and communities, is one of the main challenges. Continuous updates to the model are necessary to correctly understand and respond to changes in sign language usage over time, due to its unpredictability. A crucial problem is the limited availability of a comprehensive, standardised dataset for Bangla sign language, which is required for efficiently training the model. The scarcity of available data hampers the model's capacity to accurately convert signs into characters, thereby limiting its practical usefulness for everyday communication requirements.

Furthermore, it is technically difficult to record the finer details and subtleties of sign language movements. Continual research and development efforts are necessary to fine-tune the model to accurately recognise subtle motions and facial expressions that are crucial for sign language communication. Also, the fact that real-time apps need strong internet connectivity makes them harder to reach, especially in places with bad infrastructure.

Maintaining user trust and ensuring responsible deployment also require careful consideration of ethical considerations around data privacy and the possible monetization of the technology. Lastly, the field is missing a sufficient amount of expert collaboration and opinion, which is essential for the effective refining of the model and the consideration of the unique linguistic and cultural aspects of Bangla sign language. The deaf and mute population in Bangladesh can only be fully supported and empowered by the Bangla sign language recognition model if these obstacles are addressed via cooperative efforts, creative research, and moral governance.

REFERENCES

[1] Khan, S. A., Joy, A. D., Asaduzzaman, S. M., & Hossain, M., "An efficient sign language translator device using convolutional neural network and customized ROI segmentation". In *2019 2nd International Conference on Communication Engineering and Technology (ICCET)* (pp. 152-156). IEEE, April 2019.

[2] Islalm, M. S., Rahman, M. M., Rahman, M. H., Arifuzzaman, M., Sassi, R., & Aktaruzzaman, M. "Recognition bangla sign language using convolutional neural network". In *2019 international conference on innovation and intelligence for informatics, computing, and technologies (3ICT)* (pp. 1-6). IEEE, September 2019.

[3] Shahriar, R., Zaman, A. G. M., Ahmed, T., Khan, S. M., & Maruf, H. M. "A communication platform between bangla and sign language". In *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)* (pp. 1-4). IEEE, December 2017.

[4] Rahaman, M. A., Hossain, M. P., Rana, M. M., Rahman, M. A., & Akter, T. "A rule based system for bangla voice and text to bangla sign language interpretation". In *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)* (pp. 1-6). IEEE, December 2020..

[5] Farhad, Y., Prasad, P. W. C., Abeer, A., Elchouemi, A., & Sasikumaran, S. "Bangla Sign Language Recognition using Convolutional Neural Network". In *International Conference on Intelligent Computing, Instrumentation and Control Technologies*, 2017.

[6] Chowdhury, A. R., Biswas, A., Hasan, S. F., Rahman, T. M., & Uddin, J. "Bengali sign language to text conversion using artificial neural network and support vector machine". In *2017 3rd International Conference on Electrical Information and Communication Technology (EICT)* (pp. 1-4). IEEE, December 2017.

[7] Das, S., Imtiaz, M. S., Neom, N. H., Siddique, N., & Wang, H. "A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier" . *Expert Systems with Applications*, *213*, 118914, 2023.

[8] Rafi, A. M., Nawal, N., Bayev, N. S. N., Nima, L., Shahnaz, C., & Fattah, S. A. "Image-based bengali sign language alphabet recognition for deaf and dumb community". In *2019 IEEE global humanitarian technology conference (GHTC)* (pp. 1-7). IEEE, October 2019.

[9] Hossen, M. A., Govindaiah, A., Sultana, S., & Bhuiyan, A. "Bengali sign language recognition using deep convolutional neural network" . In *2018 joint 7th international conference on informatics, electronics & vision (iciev) and 2018 2nd international conference on imaging, vision & pattern recognition (icIVPR)* (pp. 369-373) IEEE,June 2018 .

[10] Siddique, S., Islam, S., Neon, E. E., Sabbir, T., Naheen, I. T., & Khan, R. “Deep Learning-based Bangla Sign Language Detection with an Edge Device”. *Intelligent Systems with Applications*, *18*, 200224, 2023.

[11] Talukder, D., & Jahara, F. “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, December 2020.

[12] Nihal, R. A., Rahman, S., Broti, N. M., & Deowan, S. A. “Bangla sign alphabet recognition with zero-shot and transfer learning”. *Pattern Recognition Letters*, *150*, 84-93, 2021.

[13] Nihal, R. A., Broti, N. M., Deowan, S. A., & Rahman, S. “Design and development of a humanoid robot for sign language interpretation”. *SN Computer Science*, *2*, 1-17, 2021.

[14] Alam, M. S., Tanvir, M., Saha, D. K., & Das, S. K. “Two dimensional convolutional neural network approach for real-time bangla sign language characters recognition and translation”. *SN Computer Science*, *2*, 1-13,2021.

[15] Ahmed, S., Islam, M. R., Hassan, J., Ahmed, M. U., Ferdosi, B. J., Saha, S., & Shopon, M. “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,* 2019.

[16] M. A. Uddin and S. A. Chowdhury, "Hand sign language recognition for Bangla alphabet using Support Vector Machine," *2016 International Conference on Innovations in Science, Engineering and Technology (ICISET)*, Dhaka, Bangladesh, pp. 1-4, 2016.

[17] Islam, M. S., Mousumi, S. S. S., Rabby, A. S. A., & Hossain, S. A. (2019). A simple and mighty arrowhead detection technique of Bangla sign language characters with CNN. In *Recent Trends in Image Processing and Pattern Recognition: Second International Conference, RTIP2R 2018, Solapur, India, December 2018, Revised Selected Papers, Part I 2* (pp. 429-437). Springer Singapore.

[18] Islam Md.S., Mousumi S.S.S., Azad Rabby A.K.M.S., Hossain S.A., Abujar S. “A Potent Model to Recognize Bangla Sign Language Digits Using Convolutional Neural Network”,*Procedia Computer Science*, 143 , pp. 611-618,*2016* .

[19] Shamrat, F M & Chakraborty, Sovon & Billah, Md & Kabir, Moumita & Shadin, Nazmus & Sanjana, Silvia. “Bangla numerical sign language recognition using convolutional neural networks. Indonesian Journal of Electrical Engineering and Computer Science”. 23. 405-413. 10.11591/ijeecs.v23.i1.pp405-413, 2021.

[20] U. Santa, F. Tazreen and S. A. Chowdhury, "Bangladeshi hand sign language recognition from video" ,*2017 20th International Conference of Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, 2017, pp. 1-4, 2017.

[21] Hossein, M. J., & Ejaz, M. S.Recognition of Bengali sign language using novel deep convolutional neural network. In *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)* (pp. 1-5). IEEE, December 2020.

[22] Podder, K.K.; Chowdhury, M.E.H.; Tahir, A.M.; Mahbub, Z.B.; Khandakar, A.; Hossain, M.S.; Kadir, M.A. “Bangla Sign Language (BdSL) Alphabets and Numerals Classification Using a Deep Learning Mode” l. *Sensors 22*, 574,2022.

[23] Deb, K., Khan, M.I., Mony, H.P., & Chowdhury, S., “Two-Handed Sign Language Recognition for Bangla Character Using Normalized Cross Correlation.”, *Global journal of computer science and technology, 12*, April 2022.

[24] G. A. Rao, K. Syamala, P. V. V. Kishore and A. S. C. S. Sastry, "Deep convolutional neural networks for sign language recognition," *2018 Conference on Signal Processing And Communication Engineering Systems (SPACES)*, Vijayawada, India, pp. 194-197, January 2018.

[25] S. M. Rayeed, Gazi Wasif Akram, Sidratul Tamzida Tuba, Golam Sadman Zilani, Hasan Mahmud , Md. Kamrul Hasan, "Bangla sign digits recognition using depth information," Proc. SPIE 12084, Fourteenth International Conference on Machine Vision (ICMV 2021), 120840P , March 2022.

[26] 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, December 2023.

[27] Tabassum, T., Mahmud, I., Uddin, M. D. P., Emran, A. L. I., Afjal, M. I. B. N., & Nitu, A. M., “Enhancement of single-handed bengali sign language recognition based on hog features," Journal of theoretical and applied information technology, 98, 743 - 756, March 2020.

[28] Bangladesh Bureau of Statistics, available at <<https://bbs.gov.bd/>>, last accessed on 28-06-2024 at 12.30AM.

[29] World Health Organization , available at << https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss >>, last accessed on 28-06-2024 at 12.35 AM.

[30] He, K., Zhang, X., Ren, S., & Sun, J. “Deep residual learning for image recognition” . In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778), 2016.

[31] Medium , available at << https://medium.com/@draj0718/convolutional-neural-networks-cnn-architectures-explained-716fb197b243 >>, last accessed on 28-06-2024 at 12.50 AM.

[32] Medium , available at << https://medium.com/@nitishkundu1993/exploring-resnet50-an-in-depth-look-at-the-model-architecture-and-code-implementation-d8d8fa67e46f >>, last accessed on

28-06-2024 at 01.00 AM.
[33] Google Cloud, available at << https://cloud.google.com/tpu/docs/inception-v3-advanced >>, last accessed on 29-06-2024 at 01.00 AM.
[34] Medium , available at << https://ninad98.medium.com/xception-deep-learning-with-depthwise-separable-convolutions-b8eb6fcaee58 >>, last accessed on 29-06-2024 at 01.00 PM.
[35] Medium , available at << https://medium.com/@siddheshb008/vgg-net-architecture-explained-71179310050f >>, last accessed on 29-06-2024 at 02.00 PM.
[36] JPT , available at << https://jpt.spe.org/hybrid-machine-learning-explained-nontechnical-terms >>, last accessed on 29-06-2024 at 02.30 PM.
[37] Medium , available at << https://medium.com/analytics-vidhya/accuracy-vs-f1-score-6258237beca2 >>, last accessed on 29-06-2024 at 09.00 PM.

Appendix A

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

Title: Translating the Silent Language: Sign-to-Character Conversion in Bangla
Students ID: 203-15-14466, 203-15-14488

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a **Bangla Sign Language** translation 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 goals 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 modelling while adhering to relevant constraints in this FYDP	**PO5**

CO7	Analyse 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 endeavours 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 endeavours, 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 endeavours.	**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, CO3, CO5, CO6, CO7 and CO8	**K1 (Natural Science Fundamentals)** means a systematic, theory-based understanding of the natural sciences applicable to the discipline, which information we used to build our knowledge on different techniques and proposed some preprocessing and model in our 3.2 Proposed Methodology section as well as for final defense project design. **K2 (Mathematics, Numerical analysis, Statistics Fundamentals)** means conceptually based on mathematics, numerical analysis, statistics, and formal aspects of computer and information science to support analysis and modeling applicable to the discipline which we used for our table 2.3.1 for comparison between existing works and in table 3.4.1 for estimating the cost for our project in 3.5 Project Management and Financial Analysis. In chapter 4 Table 4.3.1 describe the training and Validation accuracy of used model and also 4.3.1 to 4.3.5 gives the curves of these models(	**Page no:** [26-30] **Section:** [3.2] **Page no:** [1-7] **Section:** [2.3, 4.3,5.3]

				ResNet50, Vgg16, Inception V3, Xception, Hybrid model(ResNet50+Inception).The 5.3 contain the performacne matrix of implemented models Table 5.3.3 to 5.3.7.	
				K3 (Engineering Fundamentals) means a systematic, theory-based formulation of engineering fundamentals required in the engineering discipline which information we used for our data analysis and statistical comparison in our 2.3 comparisons between previous works in the 2.3.1 table, 3.3 software and hardware requirementst analysis and 3.4 Project Management and Financial Analysis in chapter 3. **K4 (Specialist Knowledge)** means engineering specialist knowledge that provides theoretical frameworks and bodies of knowledge for the accepted practice areas in the engineering discipline; much is at the forefront of the discipline, which we have mentioned slightly in our 3.2 proposed methodology which are data collection, data preprocessing,model selection and design. The project applies **K6 (engineering practice & design)** by the figure of process of used models in Chapter 3.It also contain the web application design of our work.	**Page no:** [30] **Section:** [2.3, 3.3,3.4] **Page no:** [33] **Section:** [3.2] **Page no:** [33] **Section:**

					[3.2]
				This project ensures to **K8 (Research Literature)** by synthesizing insights from recent studies, to advance techniques to recognizedBangla SIgn Language using deep learning, showcasing a comprehensive understanding of current methodologies.	**Page no:** [5-7] **Section:** [1.2]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project addresses **EP-2 (Challenges you faced/ you may face during the solution problem)** by recognizing the Bangla Signs, including the limitations. We gathered data from various people with various backgrounds, which are provided in Chapter 3. One of the biggest problems we encountered was that people were reluctant to contribute pictures of signs. And the main of this work is the signs are used in Bangla language can cahnge at any time because they are inconsistent. Many papers used CNN to obtain the desired results, leaving us wondering which model to employ. Through comparative analysis, it confronts challenges in understanding spatial distributions, offering insights for refining diagnostic methodologies.	**Page no:** [23-24] **Section:** [2.4, 3.2]

3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project addresses **EP-3 (Choose one significant solution among several)** by meticulously comparing experimental outcomes, highlighting Transfer Learning as the chosen significant solution for enhanced recognition of Bangla Sign all are given in chapter-5.	**Page no:** [36-57] **Section:** [5.2]
4.	**EP4: Familiarity of Issues**	Yes	CO8	**EP-4 (You are not usually facing this kind of domain earlier. Your project has an impact on another domain not only in CSE)** We did a detailed review of the literature that covered academic papers and other online resources published over the past few previous 12 years to gain an understanding of the scope of our study. To further improve our knowledge, we also went to a school working with the Deaf and hearing-impaired community. Our analysis found a common pattern: earlier research employed small or similar datasets. Our research documentation's Chapter 2 goes into detail about these findings.	**Page no:** [16-24] **Section:** [2.2, 2.4]
5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A

6.	**EP6: Extends of stakeholders involved and conflicting requirements**	Yes	CO8	**The CEP, EP-6 (Involve diverse groups of stakeholders with widely varying needs)** As for the stakeholders, our proposed method will directly help the deaf and dumb while also allowing them to create a channel of communication with the public. Regarding the conflicting requirement, there is no government-certified or broadly applicable sign language for Bengali, which will lead to some issues that are discussed.	**Page no:** [16-24] **Section:** [1.1, 1.4, 2.4]
7.	**EP7: Interdependence**	Yes	CO5	**The CEP, EP-7 (Are high-level problems including many parts or sub-problems)** As we mention the signs used in the Bangla language are inconsistent and subject to change at any moment. Therefore our work will need to be modified to account for any changes made to sign language signs.	**Page no:** [26-34] **Section:** [2.4]

8.	**-**	Yes		**The CO4 (Your project has a budget to evaluate and estimate the cost required for our FYDP.)** The number of workers involved, the technique used to collect data, the cost of the hardware and software, the cost of travel, and other relevant factors are some of the variables that affect our project's budget and expenses. As of right now, the project's finalised cost estimate and budget are given in chapter - 3 Project Management and Financial Analysis.	**Page no:** [26-34] **Section:** [3.4]

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 the best outcome in recognising or in translating Bangla Signs to Bangla characters.	**Page no:** [33-35] **Section:** [3.3]

2.	**EA2: Level of interaction**	Yes		Bangla has no unified sign language, hence, the sign language can be vary according to place or time. As for now we have most commonly used by researchers and label them according to their rules. This sorts of conflicting issues are mentioned in 1.5, 2.4, and 7.3	**Page no:** [4, 14, 42-43] **Section:** [1.5, 2.4, 7.3]
3.	**EA3: Innovation**	Yes		For better performance, we have utilised 2 pre-trained models(ResNet50 and InceptionV3), and combined them to build a hybrid model. Not only that, we have also build a web application for live demonstrations of our work. Both of them are mentioned in Chapter 3 and 4.	**Page No:** [16-22, 24-31] **Section:** [3.2, 4.2, 4.3]
4.	**EA4: Consequences for society and the environment**	Yes		This project contributes to society by improving the communiaction bridge between deaf or mute person with normal people through translating Bangla signs to Bangla character, while also promoting environmental sustainability by employing efficient computational resources. The chaper 6 is all about this EA4.	**Page no:** [58-61] **Section:** [6.1, 6.2, 6.3, 6.4]

5.	**EA-5: Familiarity**	Yes		This project expands upon existing research by examining a novel approach to translating sign language through transfer learning, demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field.	**Page no:** [7-10] **Section:** [4.2]

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:** [13] **Section:** [3.4]
2	CO9	Yes	The project demonstrates adherence to ethical principles by prioritizing privacy, and transparently documenting the research process, ensuring responsible knowledge dissemination and social well-being through the ethical web application which comply with **CO9**.	**Page no:** [60] **Section:** [6.3]
3	CO10	No	N/A	N/A

4	CO12	Yes	The project's dedication to continuous learning **(CO12)** data collection, statistical analysis, development of models and web applications, and thorough experimental results and analysis, showcases a commitment to staying updated and refining techniques to address modern challenges.	**Page no:** [26-34, 37-55] **Section:** [4.2, 4.3, 5.2, 5.3, 7.2]