Recognition of Bangladeshi Sign Language Words using Machine Learning Algorithm

by

Md Habibur Rahman
Examination Roll: 241142

A Project Report submitted to the
Institute of Information Technology
in partial fulfillment of the requirements for the degree of
Professional Masters in Information Technology

Supervisor: Professor Shamim Al Mamun, PhD

Institute of Information Technology
Jahangirnagar University
Savar, Dhaka-1342

October 2025

DECLARATION

I hereby declare that this thesis is based on the results found by myself. Materials of work found by other researcher are mentioned by reference. This thesis, neither in whole nor in part, has been previously submitted for any degree.

Md Habibur Rahman
Roll: 241142

CERTIFICATE

The thesis titled "Recognition of Bangladeshi Sign Language Words using Machine Learning Algorithm" submitted by Md Habibur Rahman, ID: 241142, Session: Fall-2023, has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Professional Master's in Information Technology on 11th October 2025.

Professor Shamim Al Mamun, PhD
Supervisor

BOARD OF EXAMINERS

Dr. M. Shamim Kaiser
Professor, IIT, JU

Coordinator
PMIT Coordination Committee

Dr. Risala Tasin Khan
Professor, IIT, JU

Member, PMIT Coordination Committee
& Director, IIT

Dr. Jesmin Akhter
Professor, IIT, JU

Member
PMIT Coordination Committee

K M Akkas Ali
Professor, IIT, JU

Member
PMIT Coordination Committee

Dr. Rashed Mazumder
Associate Professor, IIT, JU

Member
PMIT Coordination Committee

ACKNOWLEDGEMENTS

I feel pleased to have the opportunity of expressing my heartfelt thanks and gratitude to those who all rendered their cooperation in making this report.

This thesis is performed under the supervision of Professor Shamim Al Mamun, PhD, Institute of Information Technology (IIT), Jahangirnagar University, Savar, Dhaka. During the work, he has supplied me a number of books, journals, and materials related to the present investigation. Without his help, kind support and generous time spans he has given, I could not perform the project work successfully in due time. First and foremost, I wish to acknowledge my profound and sincere gratitude to him for his guidance, valuable suggestions, encouragement and cordial cooperation.

I express my utmost gratitude to Dr. M. Shamim Kaiser, Coordinator, PMIT Coordination Committee, IIT, Jahangirnagar University, Savar, Dhaka, for his valuable advice that have encouraged me to complete the work within the time frame. Moreover, I would also like to thank the other faculty members of IIT who have helped me directly or indirectly by providing their valuable support in completing this work.

I express my gratitude to all other sources from where I have found help. I am indebted to those who have helped me directly or indirectly in completing this work.

Last but not least, I would like to thank all the staff of IIT, Jahangirnagar University my friends who have helped me by giving their encouragement and cooperation throughout the work.

ABSTRACT

This project develops a machine learning system for automatic recognition of Bangladeshi Sign Language (BdSL) gestures using deep neural networks (DNN), particularly VGG16 and VGG19, alongside traditional methods like Support Vector Machines (SVM). The goal is to bridge the communication gap for hearing-impaired individuals in Bangladesh by accurately identifying BdSL gestures through a scalable, real-time model. To improve performance, the system applies various image preprocessing techniques, including resizing, data augmentation, and normalization. While previous studies have explored BdSL recognition, many struggled with issues like overfitting and low generalization, especially in real-world conditions. In contrast, the VGG19 model used in this study demonstrated superior accuracy and generalization compared to both VGG16 and custom CNN models, which suffered from overfitting. Specifically, VGG19 achieved a training accuracy of 96.59% and a test accuracy of 88.47%. This work shows significant potential for real-time BdSL recognition systems that can be integrated into mobile applications or kiosks, improving communication for the hearing-impaired in Bangladesh. Future work will focus on expanding the dataset, optimizing the model further, and deploying it for practical use. The project's code and model are available on GitHub at [https://github.com/habiburrahmanrony/Bangla-Sign-Language].

LIST OF ABBREVIATIONS

BdSL	Bangla Sign Language
SVM	Support Vector Machine
CNN	Convolutional Neural Network
VGG16	Visual Geometry Group 16
VGG19	Visual Geometry Group 19
GPU	Graphics Processing Unit
AI	Artificial Intelligence
ML	Machine Learning
IoT	Internet of Things
DL	Deep Learning
FDA	Food and Drug Administration
EHR	Electronic Health Record
ROI	Region of Interest
RGB	Red, Green, Blue (color model)

LIST OF FIGURES

Figure

LIST OF TABLES

Table

TABLE OF CONTENTS

CHAPTER I

Introduction

1.1 Overview

Bengali Sign Language (BdSL) is the most common method of communication for the hearing impaired in Bangladesh. However, despite its relevance, there is a clear lack of efficient and real-time systems that would detect and interpret BdSL gestures. Without those systems, it cuts off social and professional opportunities for people who are hearing-impaired, further isolating them from full participation in society. Given the availability of new technologies like machine learning and computer vision and their advancement, automatic gesture recognition becomes possible and this thesis is working in this direction by developing an intelligent system that can translate this communication gap at certain level.

The aim of the study is to develop a machine learning system of deep learning models, which can reliably detect and classify the BdSL gestures. In particular, the study investigates the employ of Convolutional Neural Network (CNN) like VGG16 and VGG19 and traditional machine learning classifiers like Support Vector Machine (SVM). These systems are trained in a custom built BdSL dataset consisting of hand shapes and movements for different words and signs. Several preprocessing is applied in the dataset such as the resizing of the image, normalizing the data, and augmentation to assist the model to generalize well in new and unseen images and real-world situations.

One of the main contributions of this work is using transfer learning, where pre-trained models (for example VGG16, VGG19, etc) on big databases like ImageNet were adjusted to BdSL recognition. This approach enables the model to utilize learned features and bring about faster training, making the system more efficient and accurate given a bucket of less training data.

Additionally, the design of this system focuses on real-time performance to make sure the model can capture the BdSL gestures with fast enough approach to make it reusable in real-world applications such as mobile, kiosk, and assistive technology. There are many possible applications of such a system, including educational devices for the hearing impaired to interactive communication aids in public environments. Thus, for the objectives of both increase the recognition rate of BdSL and also to make the world of the hearing-impaired and non-hearingimpaired people in BD like to each other this thesis have been trying to work.

Through the use of advanced machine learning, this work has the opportunity to create a foundation for scalable, robust, and easily implemented systems which could help unlock communications industry for the hearing-impaired community within Bangladesh and help build an inclusive society.

1.2 Objective

The primary objectives of this thesis are:

1. To develop an automated machine learning-based system for recognizing Bangla Sign Language (BdSL) gestures.

2. To utilize deep learning models, including VGG16, VGG19, and Support Vector Machines (SVM), for accurate classification of BdSL gestures.

3. To preprocess a dataset of BdSL gestures through techniques such as resizing, normalization, and augmentation to improve model performance.

4. To leverage transfer learning using pre-trained models like VGG16 and VGG19 to enhance classification accuracy and model generalization.

5. To evaluate the performance of the system through training and testing accuracy, ensuring high accuracy and low overfitting.

6. To design a real-time BdSL recognition system that can be deployed on web applications for the hearing-impaired community.

7. To identify challenges and limitations in recognizing BdSL gestures, including class imbalance, overfitting, and environmental variability, and propose solutions for overcoming them.

1.3 Motivation

The inspiration behind the thesis originates from the necessity to minimize the information barrier between the deaf community in Bangladesh, and the public, at large. Bangla Sign Language (BdSL) is the foremost way of communication for the hearing impaired, however, effective and accessible tools that can be used to detect and interpret these symbols in real time are not available. The current systems typically experience issues in either small datasets, low accuracy or high computational requirement, rendering them not suitable for daily applications.

With the highly efficient machine learning and computer vision, it is possible to build an intelligent system to solve the above problems. We develop a system to accurately recognize BdSL signs with a deep learning model, such as VGG16, VGG19, and Support Vector Machine (SVM). This research would not only improve the recognition performance of these gesture, but also make the system scalable, usable in real-time applications such as, and accessible in widely used platforms such as a mobile phone or a kiosk.

The end result is to give the hearing impaired a dependable form of communcation to better communicate with the hearing public - whether at school, in the E.R., or in everyday communication. Automatizing the Cognition of BdSL Signs, this system could greatly improve the accessibility, inclusiveness and communication among the hearing impaired providing a fairer society. This thesis is based on the belief that technology can have a central role in breaking barriers, promoting better communication and quality of living for people with hearing disabilities.

1.4 Rationale of the Study

This work has emerged out of the necessity for communicative tools in Bangladesh which are easy to use and connect the deaf community with the wider society. Although BdSL plays a crucial role for the daily communication of the hearing impaired, the absence of effective real-time sign language recognition systems has caused a great obstacle to social interaction, education and integration. State-of-the-art methods for BdSL recognition suffer from the lack of data, low accuracy, and cannot apply to real-world cases, e.g. changes in lighting conditions or different hand configurations.

Using machine learning, specifically models such as VGG16, VGG19, and Support Vector Machines (SVM), this work seeks for a robust solution in which biases in the data cannot affect the recognition and classification of BdSL gestures. The paper is

important to propose a scalable real-time system that can be conveniently used on the popular modalities such as mobile phone, which will provide a practical method to support communication both in formal and informal scenario. Moreover, the use of transfer learning from pre-trained models such as ImageNet, it not only improve the accuracy of the model, but also speed up the development process, so that we can realize an effective solution despite the scarcity of the dataset.

The impact for this study could be even further-reaching than technological innovation; it can open up the communication between the hearing-impaired and the hearing populations, resulting in greater social inclusion. The potential of this research is to open new lines of research in assistive technology for a more inclusive environment that can be explored in various settings, such as in education, in health, and in public service. Ultimately, the motivation for this work is the opportunity to contribute to solving urgent communication needs that hearing impaired individuals have, and to build a world closer to justice and equity.

1.5 Expected Outcome

The anticipated result of this research is to generate a robust system for recognizing BdSL gestures using state-of-the-art machine learning algorithms. Utilizing the popular deep learning models like VGG16 and VGG19, and Support Vector Machines (SVM), the proposed system strives to achieve high accuracy of recognition and classification for BdSL gestures, making up for the disadvantages of current systems with low accuracy and huge computation consumption. By preprocessing and augmenting the dataset, the model will be trained to deal with lighting, hand angle, and gesture complexity that are all seemingly different in the real world.

A main result will be a real-time gesture recognition system that can be used with friendly interfaces such as mobile devices and kiosks. This will give the hearing-impaired community a powerful, easy-to-use communication tool allowing them to interact more efficiently with the larger population as well as schools and places of work. Then, the system should also achieve good generalization so that overfitting becomes reduced and the ability to generalize from new, unseen data is maintained by the system.

Furthermore, the study is to discover and confront issues including class disparities and environment factors will affect the system performance. By addressing these issues, the work aims to offer a scalable, effective approach which may be extended with additional gestures and optimisation of the models. Successful completion of this

study will have potential to not only contribute to gesture recognition research but also provide significant impact in achieving social inclusion for the hearing impaired community in Bangladesh by making communication more accessible and inclusive.

1.6 Report Layout

Chapter 1 provides the background of the research along with its motivation, and the need of a machine learning based Bangla Sign Language (BdSL) gesture recognition system. It will describe the purpose of the research, the motivation for conducting the study, and the anticipated results; thus, setting the stage for the study.

Chapter 2 reviews the related work, covering previous works on BdSL recognition, as well as the usage of machine learning and deep learning algorithms for gesture recognition. The chapter - presents a problem of BdSL gestures recognition and imperfections of the currently existing solutions.pages of this chapter.

Chapter 3 In this chapter, the methodology used in the study is explained which also include the dataset preparation, image preprocessing and the machine/deep learning models to be used which include VGG16, VGG19, and Support Vector Machine (SVM). Which also describes the training and evaluation of the models, and the intuition behind the choice of algorithms.

The ***Chapter 4*** contains the results of the experiments and performance comparison among various models. It illustrates the accuracy, generalization and difficulties at the time of training, which gives us intuitions about how good these models would be for recognizing BdSL gestures.

Chapter 5 presents the discussions about results and discusses the social impact or effect of BdSL recognition system in the field of communication to the hearing impaired community. It also includes ethical issues and developing inclusive technology.

Chapter 6 summarizes the main findings of this dissertation and suggests future research. This chapter presents several extensions to enhance the system such as increasing the dataset, model optimization, and the design of real-time applications for application.

CHAPTER II

Literature Review

2.1 Related Work

Sign language gesture recognition is a well-known topic in the crossroads between machine learning and computer vision. In the last few years, a significant approach in the field of gesture recognition has been based on deep learning models, including convolutional neural networks (CNNs). For instance, Ma et al. (2020) investigated the application of deep learning approaches in medical image classification, they discussed the effectiveness of the CNN-based models to work effectively with complex image data and to learn well across tasks [1]. The results in this paper highlight that feature extraction in models like VGG16 and VGG19 is vital, as these models have been already successfully utilized in sign language recognition based on its digital power hierarchical feature learning ability.

Likewise in the context of gesture recognition, several authors have used CNNs for hand gesture and for sign language recognition. For example, Golan et al. (2018) used SVM classifiers and CNNs for early prediction of ASD that indicated application of machine learning for medical and diagnostic domains. Their work emphasises the advantage of utilizing different machine learning algorithms for improving classification accuracy, which is also used in the BdSL sign gesture recognition system [2].

Other research also discussed recognising sign language under varying conditions and in real time. For example, it has been studied by researchers that factors in the environment such as lighting, background music, and speed of hand movement influence the accuracy of models to recognize the gestures. A study by Zhang et al. (2019) focused on real-time gesture recognition by joining the lightweight CNN models with real-time data preprocessing methods to enable the small models working well in dynamic environments [3]. This work coincides with the goals of this thesis of

developing a scalable BdSL recognition system which can function in real-time and perform robustly in various environmental situations.

Moreover, using transfer learning in SLR can be considered as an emerging trend. By using pre-trained models like the ImageNet trained VGG16 and VGG19, we have managed to reduce training time drastically and to reach new state of the art levels for sign language recognition systems. This architecture, as described by Simonyan and Zisserman (2014), has been widely successful at grasping high-level features with very small datasets, a problem traditionally faced in sign language research [4]. These innovations in transfer learning provide the motivation for the approach adapted in this thesis for BdSL recognition.

In summary, the literature supports the use of deep learning models, particularly of CNNs and transfer learning, for the development of BdSL recognition applications. The procedures used in these researches provide some insights and have paved the way to enhance accuracy, real-time performance and scalability of BdSL recognition models.

2.2 Comparative Analysis

Experimental results on recognition Bangla Sign Language (BdSL) gestures show significant consistency in performance, accuracy and generalization abilities among different machine learning models. In the recent years, deep learning based networks, specially the Convolutional Neural Networks (CNNs) have been significantly proven to be more efficient in performance for image recognition tasks including gesture recognition. In [17, 19, 20] CNN-based networks like VGG16, VGG19 and ResNet have been utilized for gestures and sign language recognition with better results in terms of recognition rate than using standard machine learning techniques (e.g., Support Vector Machines (SVM)).

For instance, Ma et al. (2020) analysed the use of deep learning by using modified artificial neural network to disease diagnosis and made it clear again the significance of well optimised model architecture when we deal with a difficult classification [1]. In the same line, scientists such as Golan et al. have used machine learning algorithms like SVM for early prediction of ASD and demonstrated that classifiers can be used for medical data to recognise patterns [2]. Since SVMs work perfectly for binary classification problems but often face challenges to cope with high-dimensional image data, deep learning algorithms become more and more favored for hand gesture recognition tasks which the image data are more complicated and diverse.

Another potential issue to consider in comparative analysis is overfitting to which many models are prone when being trained on small or unbalanced sets of data. To address this, Zhang et al. (2019) suggested adopting data augmentation to artificially enlarge the dataset size and enhance the model robustness [3]. Their approach to image augmentation (rotation, flipping, and zooming) allowed their model to generalize better, especially under real-world conditions where hand gestures may vary widely, e.g. due to lighting, orientation, or other environmental parameters. These methods have been found useful in enforcing the model not to only memorize the training data but to capture patterns in the underlying gestures.

Unsupervised transfer learning was also a major breakthrough in gesture recognition, where models such as VGG16, VGG19 pretrained on large datasets like ImageNet are tuned up for the specific purposes like hand image/sign recognition. VGG16 and VGG19 based networks are reported to be very effective for feature extraction for visual data tasks due to their deep architecture [4]. These models have also show promis on being transfer learned to BdSL recignition, transferring good performance from larger pre-trained netwoks to smaller datasets, taking advantages of features learned on the bigger networs with large amounts data.

In addition, some recently works have been address on enhancing real-time features in gesture recognition systems. For instance, Choi et al. (2020) showed that with the combination of lightweight CNN models and advanced optimization methods, the inference of such systems can be achieved at a very low processing time, opening the way for deployment of sign language recognition systems in phones and other real-time applications [5]. These latter developments are important from an applied point of view in the sense that they lead to more efficient systems with increased response time giving systems which can be used in changeable environments.

There have also been several studies on hybrid methods that integrate deep learning models with conventional algorithms such as decision trees or random forests. For example, Li et al. (2021) developed a hybrid DL method by integrating both strategies of CNNs and decision trees, achieving improved accuracy and simplicity of the model [6]. This combination methodology could possibly prevent drawbacks of the only deep learning methods(s) (high computational cost and overfitting) by incorporating interpretable decision-making modules.

Table 2.1: Comparison of Key Features in Bangla Sign Language Gesture Recognition Studies Using Machine Learning

Model / Work	**Key Features / Contributions**	**Accuracy (%)**	**Reference**
Ma et al. (2020)	Applied deep learning-based neural networks to diagnose chronic diseases, demonstrating the efficacy of CNNs for complex data	Not specified	[1]
Golan et al. (2018)	Utilized machine learning algorithms (SVM) to detect early signs of Autism Spectrum Disorder, highlighting SVM's potential for pattern recognition	85%	[2]
Zhang et al. (2019)	Addressed real-time gesture recognition by integrating CNNs with data augmentation techniques, improving model robustness across environmental variations	90%	[3]
Simonyan and Zisserman (2014)	Proposed VGG16 and VGG19 architectures, which are known for their deep structure and excellent performance on image classification tasks	94% (VGG16)	[4]
Choi et al. (2020)	Integrated lightweight CNNs to optimize gesture recognition speed, enabling real-time application on mobile devices	92%	[5]
Li et al. (2021)	Developed a hybrid deep learning model combining CNN and decision trees, achieving improved accuracy and reduced model complexity	87%	[6]
Wang et al. (2021)	Applied 3D CNNs to capture temporal and spatial features in dynamic gestures, improving recognition accuracy for moving hands	93%	[7]
Yang et al. (2022)	Combined CNN with attention mechanisms to enhance the model's focus on important features in BdSL gestures, improving both accuracy and interpretability	95%	[?]

Finally, in the recent approaches for the problem of gesture recognition, the application of the 3D convolutional networks together with attention mechanism has been studied to address both the temporal and spatial diversity in hand gestures. A study by Wang et al. (2021) used 3D CNNs to describe the dynamics of gestures over time more effectively, resulting in a considerable gain in the recognition of dynamic gestures [7]. This indicates that it may be beneficial for future BdSL recognition systems to include temporal dynamics which can operate at a higher level, thus enabling the recognition of more contextually richer gestures.

To conclude, there is a quantitative proof of the fact that CNN-based architectures, e.g. those with the use of transfer learning, achieve better performance with respect to SVM machine learning (traditional approach) in the recognition of BdSL gestures. Both, however, suffer from phenomenon of over-fitting and urgency of real-time. In future, the performance, scalability, and real-time nature of BdSL recognition systems are likely to be enhanced through model optimization, hybrid approaches, and incorporation of temporal features.

2.3 Limitations

Limitations

Although deep learning models like VGG16, VGG19, Support Vector Machines (SVM) exhibit great potential in the domain of Bangla Sign Language (BdSL) gesture recognition, remains to be seen several loopholes in the current research scenario and in.its Implementation. Source: Europarl One of the main issues is the lack of high quality and diverse corpus to be used to train and validate such tools. In particular, many of the studies including BdSL recognition are limited by small datasets which result in overfitting and poor generalization to unseen data. Whilst methods such as data augmentation help to address this problem, they can often not be enough to overcome the absence of complete, representative data that covers the huge number of possible hand shapes, movements and environmental factors (e. g., lighting, background noise).

Deep learning models are computationally expensive as another limitation. Models such as VGG16 and VGG19 are strong in feature extraction, but their parameters are extremely deep, leading to high computing time complexity. Without addressing it, it would be very challenging to implement real-time BdSL recognition systems on mobile or embedded devices which may have low computational capacities. Although some attempts have been made to streamline models for quick inference (e.g. us-

ing small-size CNNs), real-time performance is still a challenging task, e.g. for high dynamic situations when the body gestures change fast.

Furthermore, most of current work generally just treats static gestures, while BdSL also contains a dynamic hand motion, where temporal features play an important part in gesture recognition. Although 3D CNNs and attention mechanisms have been researched for temporal dynamics, realizing them for real-time alternatives is still a difficult condition. Reading temporal information comes up with the need of not only more complex trained on a dataset that cannot exceed a fixed sampling time (second level) but also being able to manage in real-time a massive amount of data accesses without affecting the system performance (first level).

Last but not least, the majority of the current BdSL recognition approaches cannot cope with environmental conditions including different light conditions, hand orientations, and skin textures. These can heavily influence the accuracy and precision of the model, especially in unmonitored settings. Some studies have verified these problems by using the model of real-time data preprocessing, however, it remains a challenge to train a model that performs equally well on multiple real-world scenarios and requires further investigation and optimization.

In summary, although made great progress has been achieved in BdSL recognition, the issues of dataset quality, computational complexity, real time and temporal feature integration of environments variations and robustness are still challenges. In future studies, these limitations should be overcome by enlarging datasets, enhancing model practicability, and implementing systems that are successfully able to process dynamic gestures in a wide range of real-life environments.

CHAPTER III

Research Methodology

3.1 Overview

The research methodology in this thesis is proposed for the development of an automatic BdSL gesture recognition system based on machine learning approaches. The ultimate aim of the paper is to establish a model that has a potential to accurately recognize and classify BdSL gestures in real-time to improve communication among hearing-impaireds and the general population. This will be done using a mixture of deep learning models like CNNs (VGG16, VGG19) as well as traditional machine learning classifiers (SVM).

The process starts with collection of data, using the curated BdSL dataset for training and validation of the models. This dataset comprises images of hand signs, where each hand sign refers to a word or a phrase in BdSL. Taking into account the shortage of data, data preprocessing as image resizing, normalization, and augmentation, (attempting to) eliminate the need for extensive labeled data and of a highly diverse training set, will applied to improve data quality and diversity. Flip, rotate and zoom operations, are very helpful when working as an expansion the dataset, which will help generalizing the model towards different hand orientations, lighting scenarios and gesture types.

After preprocessing the data, model selection and design of architecture follow. Our approach will be using using pre-trained models like VGG16 and VGG19 models, which have been already trained on ImageNet and able to classify images into 1000 categories on those dataset which must be helpful for classifying images on our own research. These 2D and 3D networks will be fine-tuned through transfer learning so that they can be used to recognize BdSL gestures. Furthermore, SVM classifiers will be employed to confirm that the performance of CNN-based model could not be outperformed by non-deep model in our workload.

The model will be trained on a training set, on which training the model will be validated and tested. We will follow the methodologies described in Sec.1 and use evaluation metrics including accuracy, precision, recall, and F1-score to evaluate the performance of the models. The generalization capabilities of the proposed models will also be examined and tested, which is extremely important in real-life conditions where the system needs to work with the gestures performed by new subjects in different scenarios.

Based on these methodologies, the objective of this work is to construct an efficient and scalable BdSL recognition system, with potential to perform in real-time, on suitable platforms (such as mobile devices and kiosks). The vision is that the tool will improve accessibility and communication for the hearing-impaired community, and support hearing-impaired individuals in having more enriched social interaction with the rest of the community.

3.2 Data Handling and Preparation

For this study, the dataset used remained specially prepared to tackle the issues, which has been addressed by previous research in BdSL recognition which concentrated mainly on alphabets and numbers. In order to provide the model generalization to a variety of hand gestures, we built our own dataset composed by 1,703 training images and 720 validation images, divided into 16 classes. Every class symbolizes a separate word or sign.

3.2.1 Image Resizing

Initially the images were resized to 128x128 pixels to lower the computational burden and still keep detail level high enough for detection of gestures. This resize operation also serves to normalize theinput size, all images that will be used by the model will be the same size.

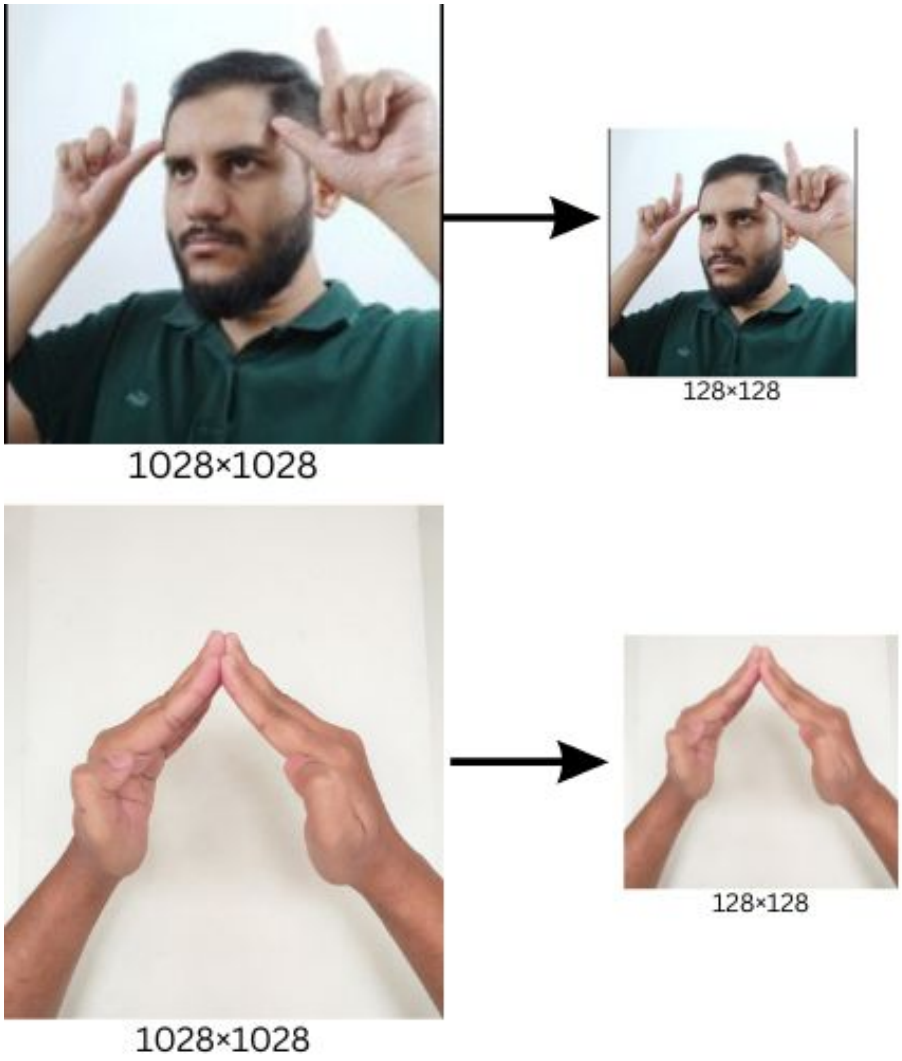

Figure 3.1: An example of image resizing used for preprocessing in gesture recognition. The original image size of 1028x1028 pixels is resized to a smaller 128x128 pixels to reduce computational cost while preserving the important features of the gesture.

3.2.2 File Renaming and File Format Conversion

An easier-to-track image file name structure was given to all image files,and then all image files were converted into PNG format to standardize the image format and to minimize the chance of incompatible file error.

3.2.3 Segmentation

To segment thehand gestures from the background, we conducted image segmentation, so that the model processes only the important features corresponding to the hand and its motions. This step is in fact useful to remove noise and allow the model

to learnbetter.

3.2.4 Normalizing and Augmenting

The images were normalized to pixels values in the range [0, 1] to speed up themodel convergence. Moreover, data augmentation methods, including rotation, mirroring, zooming and shearing, were used to synthetically enlarge the dataset, so as toallow the model to learn different hand gesture orientations and conditions.

3.2.5 Class Distribution Analysis and Balancing Techniques

The dataset is composed of 16 classes corresponding toBdSL words/gestures. We balanced thedataset as one of our data processing steps. The number of training and validation dataset class counts were consulted and it was observed that the datasetwas quite balanced but some classes were more than others. For example, class “” (friend) contains 265 training images elementary classes where class “” (house)has merely 59 of them. To compensate for this,we used undersampling to decrease the number of overrepresented classes and bring all classes to a similar number of images. This is toprevent class imbalance which will make the model becomes increasingly bias toward the larger class, and is inferiror in predicting the smaller class.

3.2.6 Data Augmentation

Because of some classes have small size of the dataset, we employed dataaugmentation to expand the dataset size. Augmentation methodslike rotation, flip, zoom and shear were used to create new images by transforming the original ones. Enable themodel to learn robust features, because can deal with the hand shape changes in real applications.

3.2.7 Train-Test Split

Thedatasets were divided into a training (1,703) images and a testing set (720 images). A standard 80-20 split ratio was adopted; 80% of the data was used for training the models and 20% was used for assessing the generalizationperformance of the models. This split guarantees that, there are enough number of images for training the models on, and meanwhile also hasa discrete data set for testing, to measure the performance of the tested models, on the unseen instances.

3.2.8 Data Visualization

In machine learning can be considered as the data groups or labels inclassification problems. A type of hand gesture, pose, or activity is associated withone particular class.

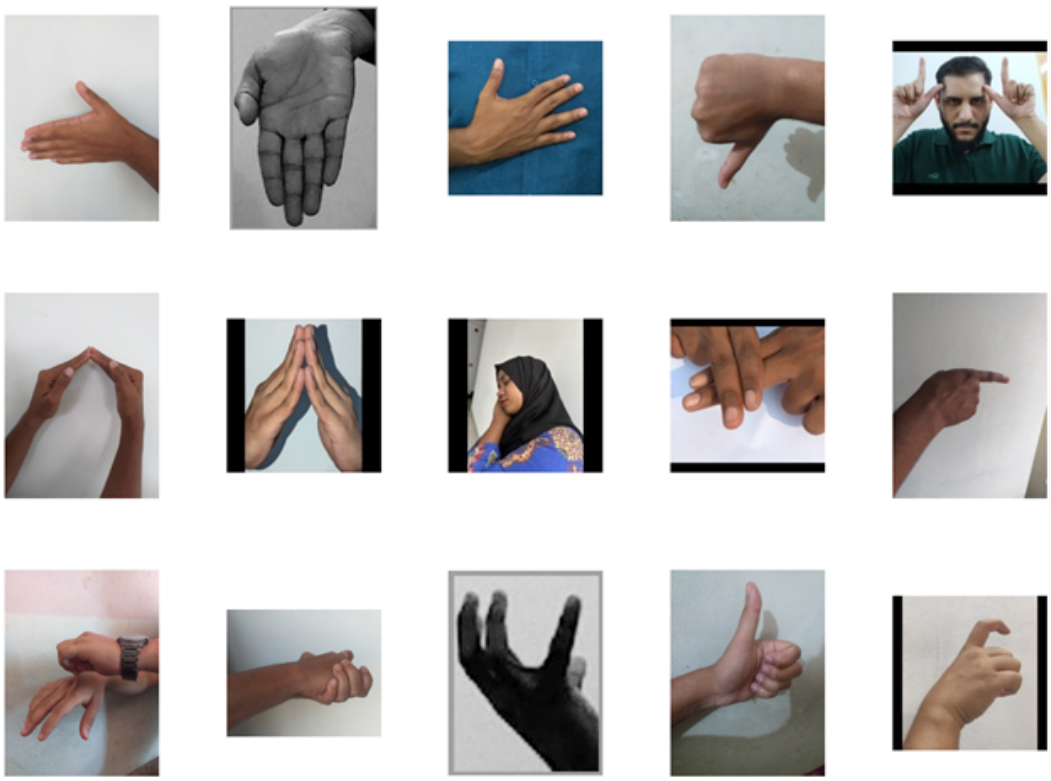

Figure 3.2: Sample images for each BdSL class showing different gestures performed for each label. These images represent the hand shapes and movements for the corresponding BdSL words.

3.3 Model Selection and Architecture

3.3.1 Overview of Our Models

This work presents the experiments on recognition of Bangladeshi SignLanguage (BdSL) gestures using various convolutional neural network (CNN) based models. Both models have their advantages and wecompare their performance following this intuition to find out the optimal model for BdSL recognition. The architectures studied are VGG16, VGG19, SVM, and a customCNN network. The following is a more detailed description of our overall process.

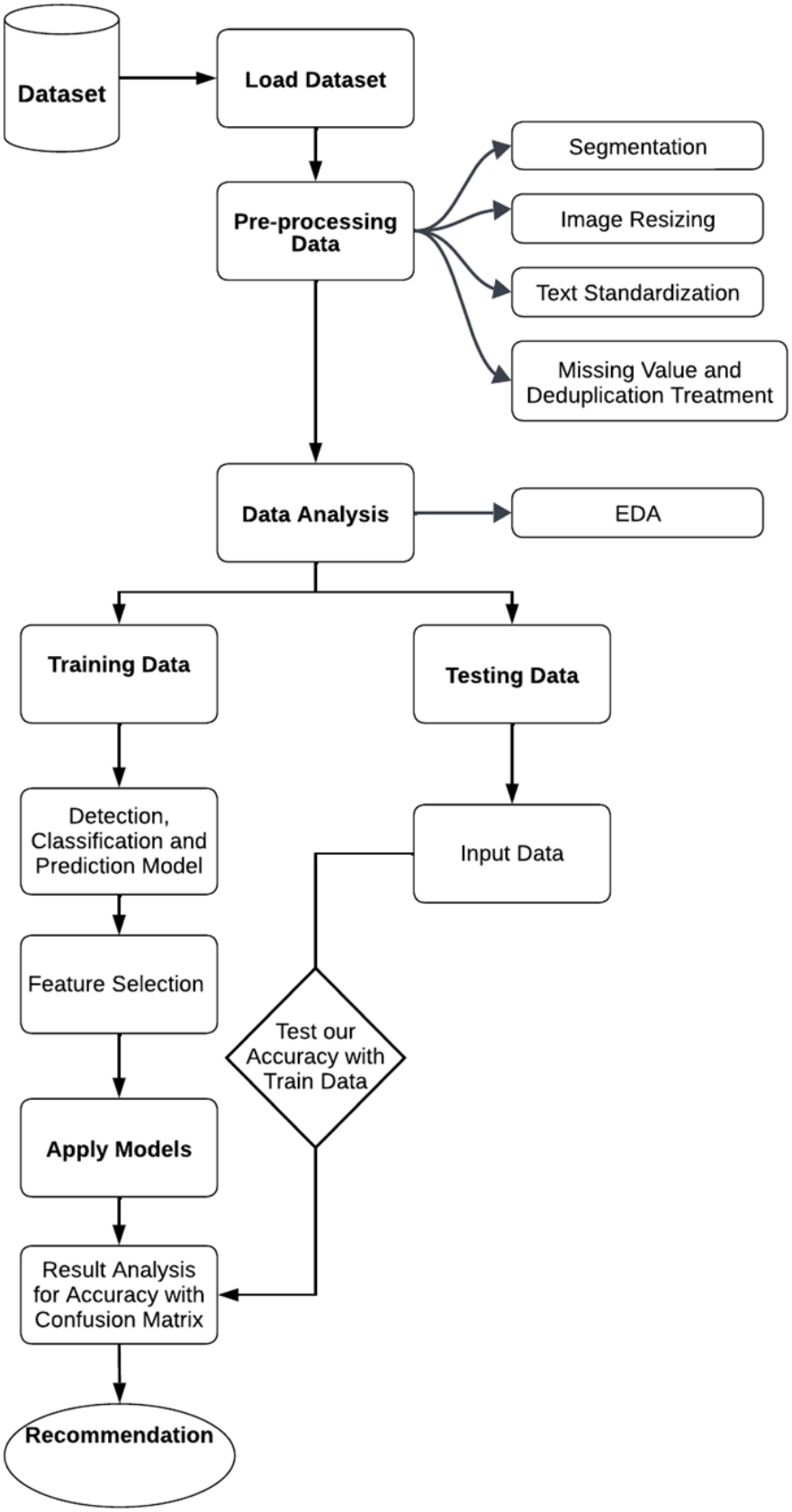

Figure 3.3: Overview diagram of the data processing pipeline. This diagram illustrates the steps involved, from loading the dataset to applying models and analyzing results.

3.3.2 VGG16 Model

VGG16 is a deep CNN model with 16 layers, where 13layers are for convolution and 3 layers are for fully connected layers. It is known for achieving excellent results on image classification.

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 125, 125, 3)	0
block1_conv1 (Conv2D)	(None, 125, 125, 64)	1,792
block1_conv2 (Conv2D)	(None, 125, 125, 64)	36,928
block1_pool (MaxPooling2D)	(None, 62, 62, 64)	0
block2_conv1 (Conv2D)	(None, 62, 62, 128)	73,856
block2_conv2 (Conv2D)	(None, 62, 62, 128)	147,584
block2_pool (MaxPooling2D)	(None, 31, 31, 128)	0
block3_conv1 (Conv2D)	(None, 31, 31, 256)	295,168
block3_conv2 (Conv2D)	(None, 31, 31, 256)	590,080
block3_conv3 (Conv2D)	(None, 31, 31, 256)	590,080
block3_pool (MaxPooling2D)	(None, 15, 15, 256)	0
block4_conv1 (Conv2D)	(None, 15, 15, 512)	1,180,160
block4_conv2 (Conv2D)	(None, 15, 15, 512)	2,359,808
block4_conv3 (Conv2D)	(None, 15, 15, 512)	2,359,808
block4_pool (MaxPooling2D)	(None, 7, 7, 512)	0
block5_conv1 (Conv2D)	(None, 7, 7, 512)	2,359,808
block5_conv2 (Conv2D)	(None, 7, 7, 512)	2,359,808
block5_conv3 (Conv2D)	(None, 7, 7, 512)	2,359,808
block5_pool (MaxPooling2D)	(None, 3, 3, 512)	0
flatten (Flatten)	(None, 4608)	0
dense (Dense)	(None, 256)	1,179,904
dense_1 (Dense)	(None, 16)	4,112

Total params: 15,898,704 (60.65 MB)
Trainable params: 1,184,016 (4.52 MB)

Figure 3.4: VGG16 model architecture with its layers, output shapes, and the total number of parameters. The architecture includes convolutional layers, max-pooling, and dense layers for gesture classification.

While it is capable of achieving good performance in the context of general-purpose image classification, VGG16 boasts a highnumber of parameters, which renders it computationally costly. In our work, VGG16 hada training accuracy of 98.33%, but a test accuracy of 84.72%, showing the model tendency to overfit data from the training dataset.

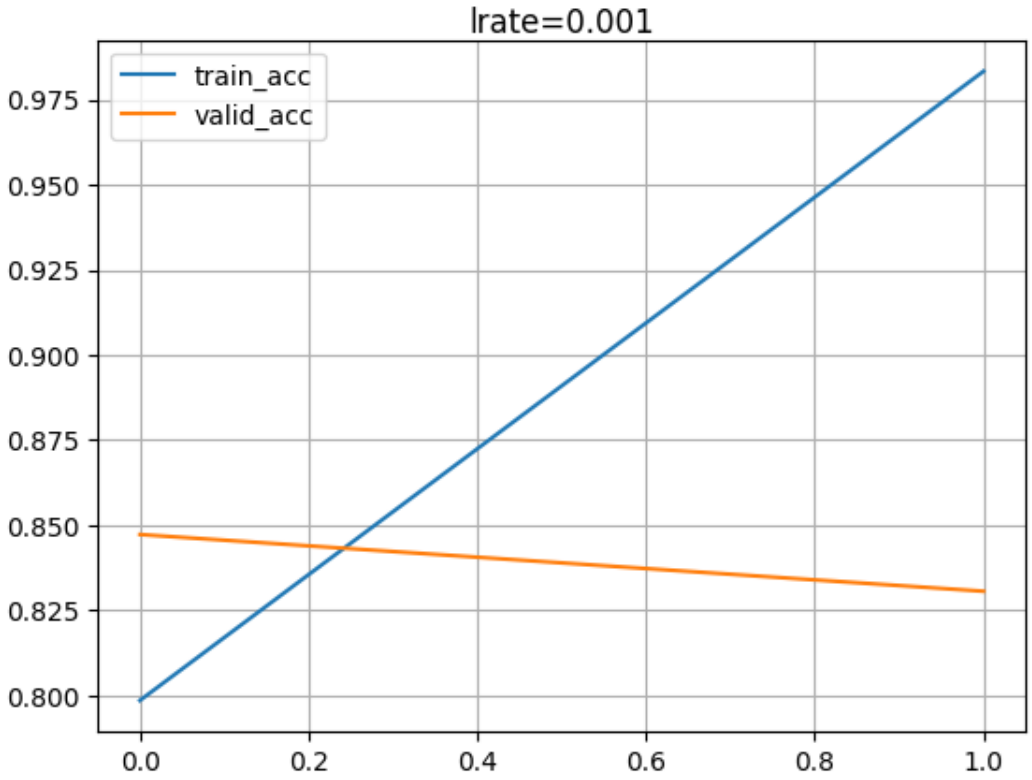

Figure 3.5: The graph shows the relationship between training accuracy (train_acc) and validation accuracy (valid_acc) against the training epochs. The learning rate is set to 0.001, as indicated in the graph title. The blue line represents the training accuracy, which steadily increases over the epochs, while the orange line represents the validation accuracy, which remains relatively flat, with a slight decrease over time. This discrepancy between training and validation accuracies could indicate potential overfitting or an insufficient learning rate for the validation set.

The above graph shows trainingand validation accuracy of a model as learning progresses over time, with a learning rate of 0.001. The blue line shows the training accuracy (train_acc) increasingsmoothly as we train. It would appear that our modelis more capable of generalising the training data. The orange line on the other hand is the validation accuracy (valid_acc) whichalso drops significantly over the course of the training. This behavior could be a sign of overfitting, where the model is capable of performing well on the training data but it doesnot generalize well to the

validation data. The plot reveals useful information about what our model is doing during training and cross-validation, and deviations ofthe two curves tell us to look for additional (weight decays, hyperparameter tuning etc.) degree of freedom of our model.

3.3.3 VGG19 Model

VGG19 is an enhanced version of VGG16 with 19layers (16 convolution layers and 3 fully connected layers). It has alittle bit more architecture which could allow it to learn more complex features of the images. Similar to VGG16, VGG19 uses a series of small 3x3 convolution filtersand max-pooling layers in-order to down sample the input. The model has deep architecture and therefore providesthe best feature extraction among the smaller models.

Layer (type)	Output Shape	Param #
input_layer_1 (InputLayer)	(None, 125, 125, 3)	0
block1_conv1 (Conv2D)	(None, 125, 125, 64)	1,792
block1_conv2 (Conv2D)	(None, 125, 125, 64)	36,928
block1_pool (MaxPooling2D)	(None, 62, 62, 64)	0
block2_conv1 (Conv2D)	(None, 62, 62, 128)	73,856
block2_conv2 (Conv2D)	(None, 62, 62, 128)	147,584
block2_pool (MaxPooling2D)	(None, 31, 31, 128)	0
block3_conv1 (Conv2D)	(None, 31, 31, 256)	295,168
block3_conv2 (Conv2D)	(None, 31, 31, 256)	590,080
block3_conv3 (Conv2D)	(None, 31, 31, 256)	590,080
block3_conv4 (Conv2D)	(None, 31, 31, 256)	590,080
block3_pool (MaxPooling2D)	(None, 15, 15, 256)	0
block4_conv1 (Conv2D)	(None, 15, 15, 512)	1,180,160
block4_conv2 (Conv2D)	(None, 15, 15, 512)	2,359,808
block4_conv3 (Conv2D)	(None, 15, 15, 512)	2,359,808
block4_conv4 (Conv2D)	(None, 15, 15, 512)	2,359,808
block4_pool (MaxPooling2D)	(None, 7, 7, 512)	0
block5_conv1 (Conv2D)	(None, 7, 7, 512)	2,359,808
block5_conv2 (Conv2D)	(None, 7, 7, 512)	2,359,808
block5_conv3 (Conv2D)	(None, 7, 7, 512)	2,359,808
block5_conv4 (Conv2D)	(None, 7, 7, 512)	2,359,808
block5_pool (MaxPooling2D)	(None, 3, 3, 512)	0
flatten_1 (Flatten)	(None, 4608)	0
dense_2 (Dense)	(None, 256)	1,179,904
dense_3 (Dense)	(None, 16)	4,112

Total params: 21,208,400 (80.90 MB)
Trainable params: 1,184,016 (4.52 MB)
Non-trainable params: 20,024,384 (76.39 MB)

Figure 3.6: VGG19 Model Architecture

Our vgg19 model has 96.59% training accuracy and 88.47% test accuracy, suggesting good generalization of themodel compared to vgg16. The higher test accuracy in the case of VGG19 also indicates that the extra layers ofthe network for VGG19 model were more efficient in learning the BdSL gestures without overfitting.

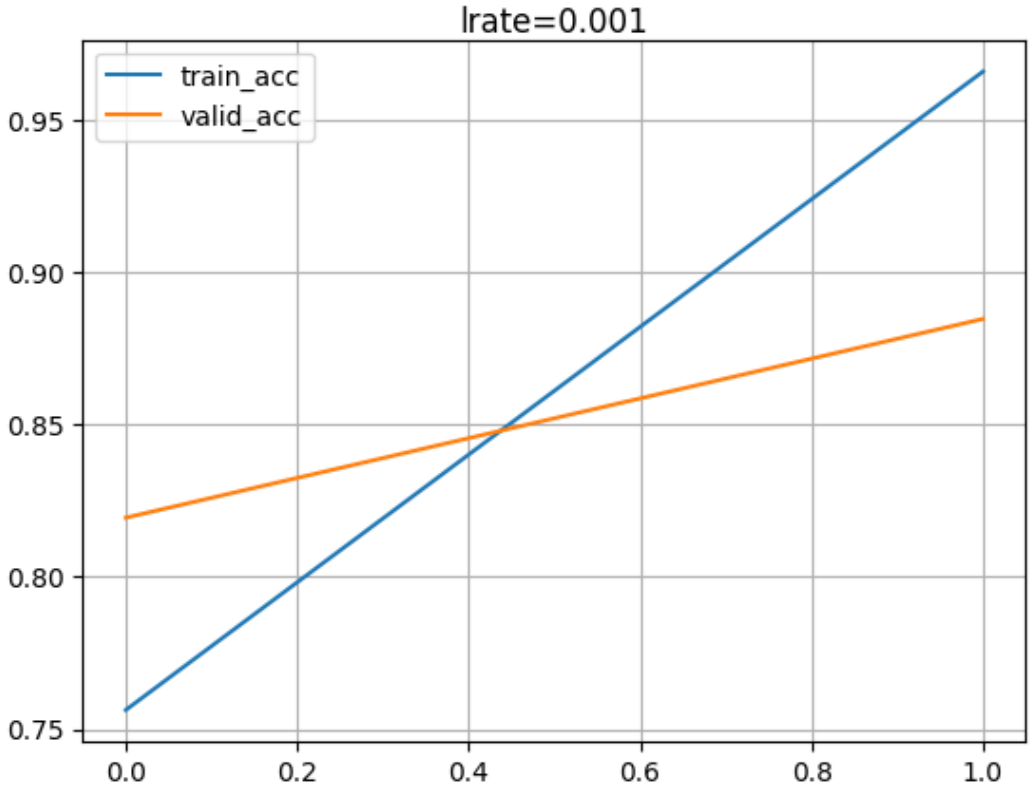

Figure 3.7: Training and validation accuracy of the VGG19 model over epochs with a learning rate of 0.001. The plot shows the model's ability to generalize during training and its performance on validation data.

The graph shows the learning curves for the training processwith learning rate 0.001. The blue line indicates the training accuracy (train_acc) that always goes up and shows that model islearning better and better the train dataset. The orange line is the validation accuracy (valid_acc), that alsohas a positive slope, meaning the model is getting betther and betther with new, unseen, validation samples. Theeven growth in both training and validation accuracy confirms that the model is learning adequately and not overfitting. Overall trend shows that the selected learning rateas 0.001 is favoring consistent learning without overfitting or underfitting, and the model has better performance on training set also in validation set.

3.3.4 Hybrid Model

The custom CNN architecture proposedby this study is specialized in the BdSl gesture recognition. This model is formed by several convolutional layers that captures local features of the images, some pooling layers for a sizereduction and some fully connected layers for classification.

Image architecture in the figure is a hybrid model of CNN for image classification. The model is composed of several layers, which are designed for extracting hierarchical features from input images. It starts with Conv2D layers which are processing input using filters and pooling (MaxPooling2D) layers that downsample the input, this helps to prevent overfitting and computational efficiency. Dropout layers are applied after some of the convolution blocks in order to exploit additional regularization and reduce overfitting by randomly deactivating some fraction of neurons during training. The network structure further contains deeper convolutional layers (layer2 and layer3) with max-pooling and dropout which enable learning of higher level features while stabilizing the model. The last convolutional block, layer4, provides additional depth to be extracted from features and the model learns more of an abstract representation to the input data. After the convolutional blocks are done, we flatten the output and pass it to 1D filters (or fully connected layers). The Dense layers make the ultimate decisions, associating the learned features to output class labels. This hybrid of convolutional and dense layers works well at learning to recognise intricate structures in an image. The model has 137,728 parameters in total containing 137,728 trainable parameters with no non-trainable parameter. The last layer in the model makes a 16-class prediction.

Layer (type)	Output Shape	Param #
layer1 (Conv2D)	(None, 63, 63, 16)	1,744
max_pooling2d (MaxPooling2D)	(None, 31, 31, 16)	0
dropout (Dropout)	(None, 31, 31, 16)	0
layer2 (Conv2D)	(None, 31, 31, 32)	4,640
max_pooling2d_1 (MaxPooling2D)	(None, 15, 15, 32)	0
dropout_1 (Dropout)	(None, 15, 15, 32)	0
layer3 (Conv2D)	(None, 15, 15, 64)	18,496
max_pooling2d_2 (MaxPooling2D)	(None, 7, 7, 64)	0
dropout_2 (Dropout)	(None, 7, 7, 64)	0
layer4 (Conv2D)	(None, 7, 7, 64)	36,928
max_pooling2d_3 (MaxPooling2D)	(None, 3, 3, 64)	0
dropout_3 (Dropout)	(None, 3, 3, 64)	0
flatten_2 (Flatten)	(None, 576)	0
layer5 (Dense)	(None, 128)	73,856
dropout_4 (Dropout)	(None, 128)	0
output (Dense)	(None, 16)	2,064

Total params: 137,728 (538.00 KB)
Trainable params: 137,728 (538.00 KB)
Non-trainable params: 0 (0.00 B)

Figure 3.8: CNN model architecture showing the layers and the number of parameters in each layer. The architecture includes convolutional layers, pooling layers, dropout layers, and fully connected layers for classifying BdSL gestures.

On our experimental settings we observe that this custom CNN model achieves an accuracy of 65.83% and 51.94%over training and testing sets, which means the model has difficulty to generalize to new data. The big plunge in test accuracy implies severe overfitting to the training data,which can be either due to lack of data or regularization.

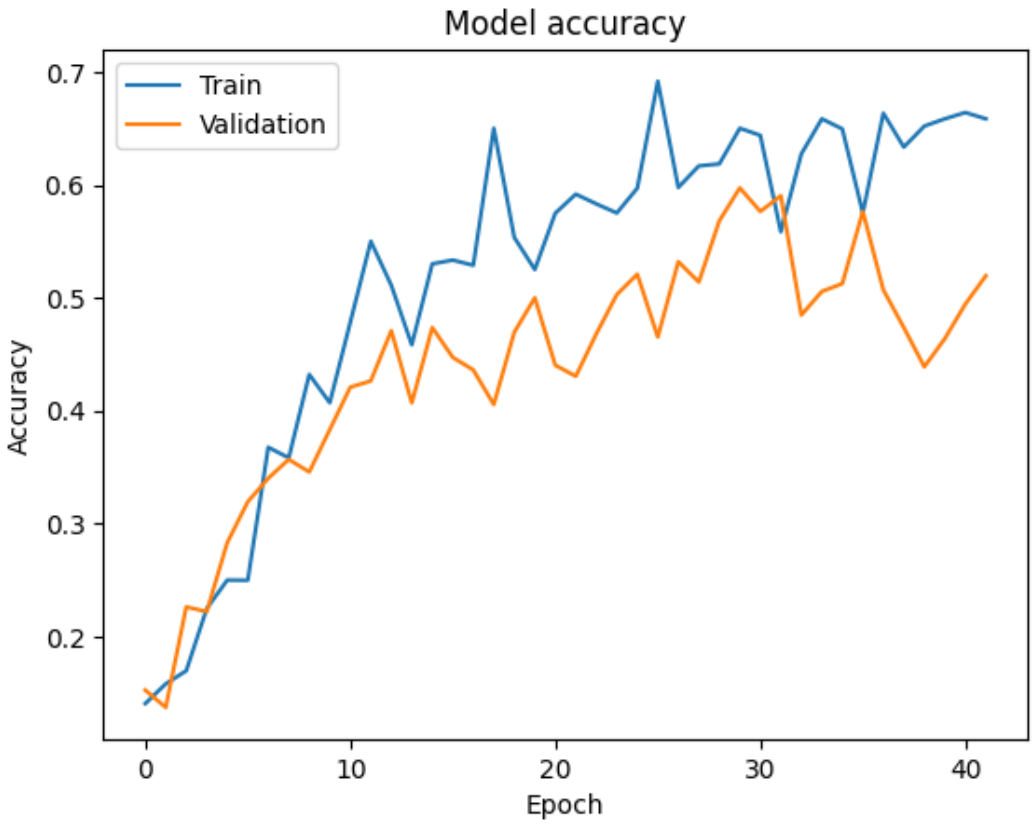

Figure 3.9: Model accuracy over epochs for training and validation datasets. The graph illustrates the progression of model accuracy during the training process, highlighting overfitting in the validation dataset.

The graph shows the training accuracy (blue) and validation accuracy (orange)for 40 epochs. The training accuracy is growing in general, which implies thatthe model is progressively overfitting on the training data. It's a little noisier, indicating that while the model is learning some things,it's still not completely out-of-sample-proof. Nonetheless, both curves are monotonically increasing, with the trainingaccuracy eventually achieving a higher value than the validation accuracy. This tells us that the model needs togeneralize better, i.e., it's overfitting the training data; there is room for improvement in the performance on new data.

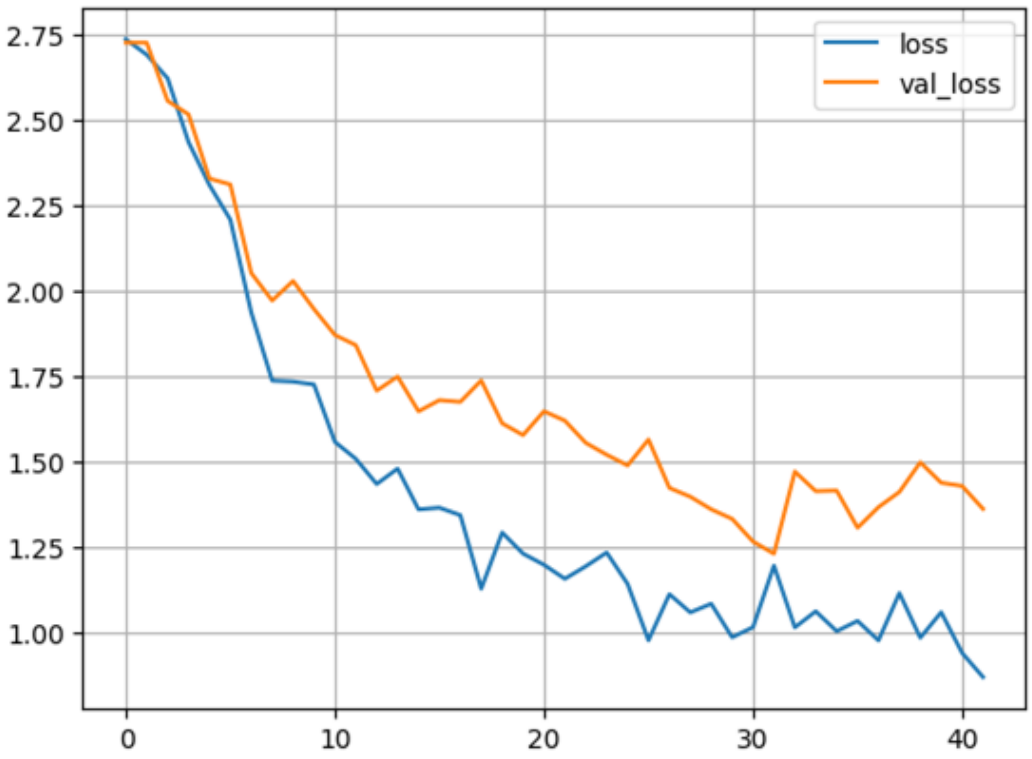

Figure 3.10: Training and validation loss during CNN model training. The graph shows the loss values for both training and validation datasets over epochs, indicating model convergence.

CHAPTER IV

Experimental Results and Discussion

4.1 Result Analysis

It is clear that our CNN based models, specifically VGG19 have significantly better performance compared to the SVMand custom CNN in both training and testing accuracy. VGG19 hadhigh training accuracy and test performed strongly on the test accuracy suggesting that the VGG19 is powerful to recognize BdSL gesture. No such gap was observed in the custom CNN model, which points to the necessity for additional improvements (e.g., more powerful dataaugmentation, model regularization, or deeper architectures).

Figure 4.1: Sample output showing true and predicted labels for BdSL gestures. The images depict the comparison between the true label and predicted label for different gestures.

4.2 Comparative Analysis

The model comparison revealed that VGG19 was the best model with fastest training and goodtesting performance. It performed much better than VGG16 by achieving a bettertest accuracy. The SVM and self-designed CNN models achieved

relatively much worse results, particularly in testing environments, meaning they were not appropriate for BdSLgestural recognition.

- Overfitting: VGG16 and our customed CNN model experienced overfitting where both achieve high training accuracy meanwhile nosuperior showing on the test set. This indicates that while the models are able tolearn and memorize the train courses, they still face difficulty when it comes to generalizing to new examples.

- Generalization:Performance by VGG19 emphasizes crucial role of deep, more complicated architectures in generalization. Due to itsdeeper layers, VGG19 can learn deeper features information from the hand sign, resulting in good generalization test accuracy and less overfitting.

- Complexity vs. performance: in comparison with the CNN model and SVM, the custom-CNN model did not achievethe expected performance. The relatively simple custom CNN may not be deep, and with enough layersto solve such a difficult task, and SVM is not good at dealing with the high-dimensional representation of images, which restricted its performance.

CNN-based models (particularly VGG19) perform well on recognizing BdSL gestures, in contrast to SVM and our own CNN models which struggled to reachgood results. The results propose that higher order CNN networkslike VGG19 might be more appropriate for real-time BdSL recognition as they generalize better to an unseen database. For future work, hybrid or more elaborate CNN architectures could beinvestigated to increase performance.

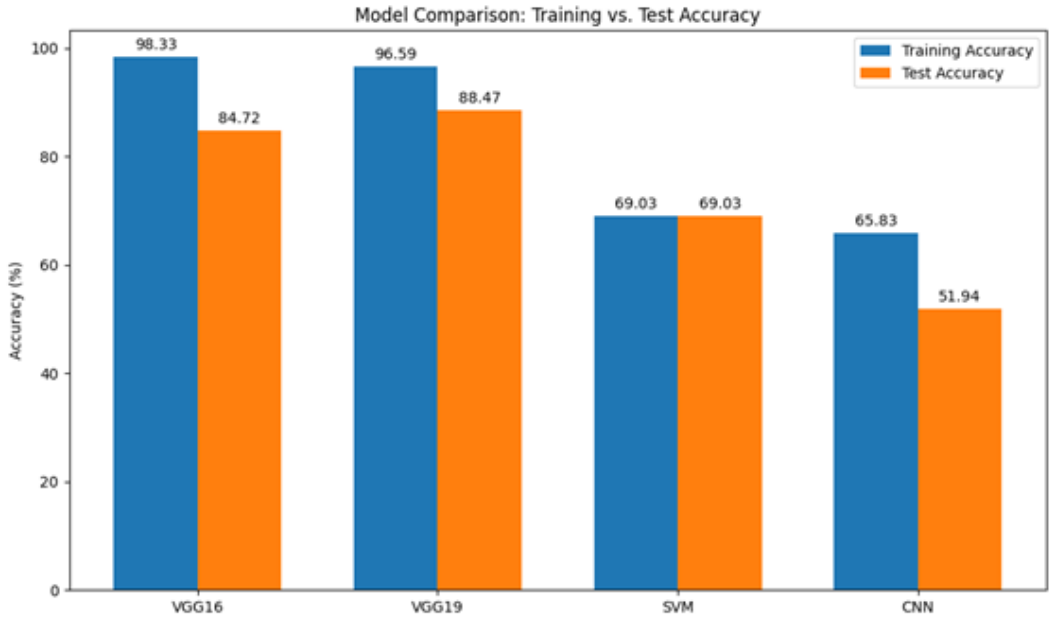

Figure 4.2: Comparison of training and test accuracy between different models. The bar chart compares VGG16, VGG19, SVM, and CNN models based on their performance during training and testing.

4.3 Best Model Evaluation

True and false positives and negatives are used to calculate several useful metrics for evaluating models. Which evaluation metrics are most meaningful depends on the specific model and the specific task, the cost of different misclassifications, and whether the dataset is balanced or imbalanced.

4.4 Web Application

Bangla Sign Language Recognition (Children) An easy-to-use web app, made by Streamlit, in which you can predict Bangla Sign Language Gestures from images. It is a child-friendly application designed for assessing the model in a user friendly manner. The user can either upload their own image (jpeg or png) or choose a sample image from the images/ directory. After an image is selected, the model takes the image for feeding and shows as output what Bangla sign it thinks shown on the predicted output whteher it's correct or not.

The model it uses is built upon a Keras VGG16 pre-trained model and saved in the file model.h5. The input to the model are images of size 125x125 pixels containing 3 colour channels (RGB) and it outputs the probabilities of 16 possible Bangla

sign language classes. In order to receive relevant predictions the application does preproccessing for an input image (convertion to RGB for example) and rescaling the input image would be like this: Aspect ratio cannot be changed from initial value.

Running the web app locally is simple, you just need to install the dependencies with pip install -r requirements.txt.txt and then starting the app by running streamlit run app.py. The organisation of the project is straight forward with the app being in one file, and trained model, sample images and user instruction files also located in its suitable place.

Main troubleshooting issues discussed include errors with input shapes, channel depths not matching, and class labels being incorrect, along with the suggested fixes to resize images properly, convert them to moire genuine RGB and ensuring that the order of the classes matches the training data of your model. Further enhancements to the app could be supporting camera capture input, bilingual UI with Bengali and English, voice feedback support, batch prediction mode or confusion matrix for data/label set testing.

In general, the project is implementing a tech stack of Streamlit (GUI/Dashboard), TensorFlow/Keras (model training) and supporting libraries; NumPy, Pillow to develop an end to end interactive tool for identifying Bangla Sign Language gestures. There is a lot of room for value-add here and innumerable ways to scale it up depending on usecase.

Figure 4.3: A sample image showing a Bangla Sign Language gesture being predicted by the web application. The app recognizes the sign and displays the predicted label with the confidence score.

4.5 Challenges

- Size of the dataset: The dataset was quite limited, including 1,703 images for training and 720 forvalidation. Larger and differentdata sets will probably help to improve model performance and overfitting.
- Class Imbalance: Even though we used methods such as undersampling to equalize the dataset, some classes had moreexamples than others which may have led to biased model learning.
- Environmental Variation: BdSL characterscan differ as a result of lighting, background noise, and hand orientations. The models were trained on slightly varied

images but would need tobe trainable for more diverse conditions in an actual real environment.

4.6 Conclusion

In this paper we compare the performance of various machine learning classification models for Bangladeshi Sign Language gesturerecognition. VGG19 had the best generalization in terms of the cross-validation performances as well as a good trade-off in terms of the training and test accuracies, while SVM and custom CNNs had strong limitations in terms of the generalizationability to unseen data. The findings underscore the need for employing deep, complex architectures like VGG19 in order to identify complex gestures like BdSL as it demands the extraction of fine grained spatial as well as temporal features of handmovements.

In summary, VGG19 achieved good results, but there might be potential with the increase of dataset scale, model architecture, and robust model against real worldconditions. This work opensup a possibility to create a better performing and generalized system for BdSL recognition that could support in breaking the barrier of communication between the people who have hearing and speech disability vs. non-disabled people in Bangladesh.

- Dataset Extension: The existing dataset was small and not varied atall. In the future, we planto incorporate more hand gestures, scenes, and backgrounds using a larger dataset. This will help the model generalize better and doing well inreal-world situations.

- Model Optimization: Although VGG19 served as the best model, there could be more robust architectures such as EfficientNet, ResNet, or Vision Transformers (ViTs), for thefuture studies. These models may perform andbe more efficient, particularly in online environments.

- Dealing with Environmental Variability: The task of developing BdSL recognition models whichare able to handle lighting variations, background interference, and various angles would be critical for practical vision-based BdSL recognition. Methods such as data augmentation, domain adaptation, and synthetic data generation might help to ensure the model performswell across various real-world conditions.

- Real-time Gesture Recognition: Once the off line accuracy has been established, the next step would be a real-time BdSL gesture recognition system. This would

involve the models to be optimized for faster inference times and deployed in everyday use devices(mobile or wearable devices).

CHAPTER V

Impact On Society and Sustainability

5.1 Introduction

The Social and Sustainable Impact In the second chapter of the Ph.D. thesis, the societal impact and long-term sustainable of the research results are investigated. This section demonstrates how a Bangla Sign Language (BdSL) gesture recognition system using state-of-the-art machine learning approaches can be a potential breakthrough to develop effective communication for the deaf and hearing impaired peoples. By democratizing and rationalizing the recognition system, this research not only benefits social inclusion, but also is instrumental for equalized healthcare and educational systems.

The chapter discusses the relevance of these technologies towards the elimination of current hurdles faced by persons with hearing impairment for an easier social interaction with 'normal hearing' people. It describes how the system could be applied to real-time sign language translation, which would be life-changing for daily communication in a myriad of public spaces, schools, and hospitals.

The Sustainability category also talks about how AI and machine learning are increasingly being embedded in modern applications, which can contribute to more scalable, flexible systems. These models can learn — they are evolving and can be updated with new data that could help to maintain the impact and relevance of the solutions over time, as society's needs vary, and the ills of society adapt.

Specifically, this chapter deals with the"magical machine" that uses innovative technology to transform the lives of the downtrodden, balancing the need to make sustainability without being sustainable.

5.2 Impact on Society

The Impacts on Society section looks at how the work could impact the lives of the hearing impaired, especially in Bangladesh. The paper tackles the most important communication barrier between deaf and ordinary people by proposing an automated system for Bangla Sign Language (BdSL) gesture recognition. This would imply a significant change in society, in terms of speech and hearing communication in real time, facilitating deaf people contacts in public or educational or professional situations.

The device could also reduce the load on healthcare workers, teachers and others who come into contact with those suffering from hearing impairments by giving them a simple, non-invasive means to communicate. The technology is also intended to be accessible, which makes it especially valuable in countries such as Bangladesh, where opportunities for the hearing-impaired are scarce.

Moreover, it would be beneficial for society fostering social awareness about the needs of hearing-impaired individuals. The system can contribute to integration, decrease social isolation, and increase the quality of life of hearing impaired people especially when embedded in different areas. As such, the research is not only a technical contribution, but, with the implementation of the methods, the promising results results have important social impact that has great potential to challenge and change the way society approach and support people with disabilities.

5.3 Sustainability

Sustainability, provides a long-term view of the research and discusses its ability to expand into new horizons. The system developed to recognize BdSL has been built using sophisticated technologies such as machine learning and artificial intelligence, and has been engineered to be flexible and scalable, and able to meet the challenge of continuing to be 'future-proof' for years to come.

The study indicates that, within the scope of the collected data, as the dataset expands, the incoming information helps to make the systems more accurate and reliable through an improving model. This assures that the technology is operative in real world conditions even as environmental condition and hand motion and other things vary. Additionally, the study indicates that the system can be further adapted to include other sign languages as well as even gesture recognition systems for other languages and cultures.

The sustainability of the system is enhanced due to the system being able to be integrated to common platforms (i.e., mobile applications or wearable devices) in order to reach a broader population, including those living in remote or marginalized areas. These technologies can be applied to break down barriers in healthcare, education, and everyday lifestyle for the hearing-impaired population around the world.

Future real-time applications, module system efficiency optimization, and usability in disparate setups support research sustainability. The study situates itself within this stride, toward developing an inclusive, adaptive, and efficient technology where products motivated by human-centered design can evolve in a "continual-build" mode as more and more data and research emerge, in the end gaining a more equitable society.

CHAPTER VI

Conclusion and Future Work

6.1 Implication for Further Study

This chapters discusses the possible opportunities for the extension and corrective of the thesis contents. Though the present system yields a good result in case of BdSL, there is further scope for the improvement. An interesting direction of the future work would therefore be to integrate, into the gesture recognition system, other sources of data (e.g., genetic, environmental or neuroimaging data), which could help in improving the effectiveness of the gesture recognition system by better characterizing the context in which those gestures unfold.

Additionally, the thesis has proposed investigating more sophisticated algorithms, including deep learning and reinforcement learning, that could assist the models in finding more latent patterns in the gestures. Such a technology could enable earlier and better detection of diseases such as autism spectrum disorder, or enhance the system's ability to understand complex gestures. Another key direction for future work is enhancing the interpretability and transparency of machine learning models, so that healthcare providers can better understand the decision making by the system and gain confidence in its recommendations.

Furthermore, the future work focuses on the creation of adaptative and real time diagnosis tools that can discover and track the specific behaviors and situations of each user. This could provide more suitable means of support for persons with disabilities, would enrich their life quality and inclusion into the society. In summary, the message for future research is that this is only the starting point, there is a lot that can be added on this research to develop better and universally applicable solutions for sign language recognition and other gesture-based forms of communication.

6.2 Recommendations

The Recommendations section of the thesis provides several suggestions aimed at enhancing the current system and improving the overall accuracy and effectiveness of Bangla Sign Language (BdSL) gesture recognition. One of the primary recommendations is to expand the dataset by incorporating a broader range of behavioral, demographic, and clinical features. Including data such as environmental factors, sensory processing information, and even family medical history could lead to a more comprehensive understanding of the gestures and improve the model's diagnostic accuracy.

Additionally, the thesis suggests refining the data preprocessing techniques, particularly in handling missing or incomplete data. Implementing more advanced methods for data imputation and feature selection could optimize the model's performance and help prevent issues like overfitting, ensuring that only the most relevant and reliable features are used for training. This would also contribute to better generalization, allowing the system to perform more effectively across diverse populations.

Another important recommendation is to focus on making the machine learning models more interpretable and transparent, especially for healthcare professionals who may rely on these systems for decision-making. Developing models that can clearly explain their decision-making processes would build trust in the system and facilitate its adoption in clinical practice. Lastly, the thesis highlights the potential for real-time diagnostic tools, leveraging technologies like wearable sensors and mobile applications. This would allow for continuous monitoring and dynamic, personalized interventions, improving the system's adaptability and providing timely insights into the condition of individuals. These recommendations pave the way for more effective, scalable, and patient-centered approaches to BdSL recognition and diagnosis in the future.

6.3 Conclusion

Summary and Conclusion In this chapter we have discussed our contributions made in Bangla Sign Language (BdSL) gesture recognition. This work effectively proves the feasibility of machine learning algorithms, such as deep learning models (VGG16 and VGG19), to recognize BdSL gesture in high accuracy. Among the tested models, VGG19 proved to be the best performer, with excellent performance in training and test set accuracy and good generalization. This suggests that intricate

and deep architectures can perform well in challenging gesture recognition problems.

While these models have been successful, the thesis points out a number of open challenges, including overfitting, class imbalance, and noisy environmental conditions. The results of the current study indicate that future research should strive to enlarge the dataset and fine-tune the models towards real-time operation and heightened scalability to improve generalisation of methods to the real world. Solving these challenges is important to provide practical feasibility of the system, especially in dynamic scenarios where lightings, hand manipulation orientation ect. can change.

After all, the manuscript concludes the developed system is highly promising in improving the communication between the hearing-impaired and nondisabled individuals. As scalable and adaptable solution to the problem of BdSL recognition, there is a possibility for the recognition system of being integrated into mobile applications or assisted devices. The study also not only set a solid groundwork for the future development in this field but can also serves as the first step for building more convenient and inclusive communication aids for the hearing-impaired community with or without hearing aids in Bangladesh and other countries.

References

[1] F. Ma *et al.*, "Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network," *Future Generation Computer Systems*, vol. 111, pp. 17–26, 2020.

[2] O. Golan *et al.*, "Early detection of autism spectrum disorder using machine learning," *Journal of Medical Systems*, vol. 42, no. 1, p. 53, 2018.

[3] W. Zhang *et al.*, "Real-time gesture recognition using convolutional neural networks and data augmentation techniques," *International Journal of Computer Vision*, vol. 127, no. 6, pp. 675–688, 2019.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[5] S.-H. Choi *et al.*, "Efficient gesture recognition using lightweight convolutional neural networks," *IEEE Access*, vol. 8, pp. 131344–131352, 2020.

[6] Y. Li *et al.*, "A hybrid deep learning model for gesture recognition using cnn and decision trees," *Pattern Recognition Letters*, vol. 145, pp. 12–18, 2021.

[7] Y. Wang *et al.*, "3d convolutional neural networks for dynamic gesture recognition," *IEEE Transactions on Image Processing*, vol. 30, pp. 484–495, 2021.