STATIC SIGN LANGUAGE RECOGNITION USING IMAGE PROCESSING AND DEEP LEARNING TECHNIQUES

A DISSERTATION

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF DEGREE
OF
MASTER OF TECHNOLOGY
IN
COMPUTER SCIENCE & ENGINEERING

Submitted by:

PRANAV
2K20/CSE/16

Under the supervision of
Dr. Rahul Katarya
(Professor)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

JUNE, 2022

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi - 110042

CANDIDATE'S DECLARATION

I, Pranav, Roll No. 2K20/CSE/16 student of M. Tech (Computer Science and Engineering), hereby declare that the project Dissertation titled **"**Static Sign Language Recognition using Image Processing and Deep Learning Techniques**"** which is submitted by me to the Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of and Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi	Pranav
Date:	2K20/CSE/16

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi - 110042

CERTIFICATE

I hereby certify that the Project Dissertation titled **“Static Sign Language Recognition using Image Processing and Deep Learning Techniques”** which is submitted by Pranav, Roll No. 2K20/CSE/16, Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge, this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi	Dr. Rahul Katarya
Date:	Professor
	Department of CSE

ACKNOWLEDGMENT

The success of this project depends on the help and contribution of a large number of people as well as the organization. I am grateful to everyone who contributed to the project's success.

I'd want to convey my gratitude to Dr. Rahul Katarya, my project guide, for allowing me to work on this research under his supervision. His unwavering support and encouragement have taught me that the process of learning is more important than the ultimate result. Throughout all of the progress reviews, I am appreciative to the panel faculty for their assistance, ongoing monitoring, and motivation to complete my project. They assisted me with fresh ideas, gave crucial information, and motivated me to finish the task.

PRANAV

2K20/CSE/16

ABSTRACT

The deaf and mute community has a great difficulty articulating their thoughts and opinions to everyone else; sign language is their most eloquent means of communication, but the general public is unaware of sign language, making it difficult for the mute and deaf to communicate with others. To address this communication gap, a system that can accurately convert sign language gestures to speech and likewise in real-time is required.

This work proposes Effi-CNN, an image Sign Language Recognition (SLR) system. Our system uses transfer learning with EfficientNetB2 as the basic model to transform sign gesture photographs to words. We've also created a system that translates hand movements into text instantaneously.

We evaluated our on eight publically available datasets, including the Massey University gesture dataset, ArSL2018 dataset, MNIST-ASL dataset, and others. Comparing our results to state-of-the art algorithms, the experimental findings show that our technique is more successful. The results show that our Effi-CNN surpasses most of current existing solutions, and it has the ability to categorise a large number of gestures with a low rate of error.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

1. SL: Sign Language
2. ISL: Indian Sign Language
3. ArSL: Arabic Sign Language
4. ASL: American Sign Language
5. RADAR: Radio Detection and Ranging
6. EEG: Electro-encephalogram
7. RF: Radio Frequency
8. SLR: Sign Language Recognition
9. VBT: Vision Based Techniques
10. AI: Artificial Intelligence
11. RGB-D: Red Green Blue Depth
12. SBT: Sensor Based Techniques
13. CNN: Convolutional Neural Network
14. HSV: Hue Saturation Value
15. Fps: frames per second
16. RoI: Region of Interest
17. WRN: Wide Residual Network
18. SVM: Support Vector Machine
19. GBM: Gradient Boosting Machine
20. RCNN: Region-based Convolutional Neural Network
21. HOG: Histogram of Oriented Gradients
22. LBP: Local Binary Pattern
23. PCA: Principal Component Analysis
24. MLP: Multi-Layer Perceptron
25. ViT: Vision Transformer
26. BW: Black and White
27. CSV: Comma Separated Values

CHAPTER 1

INTRODUCTION

Because not everyone can hear or speak, such persons rely on unique communication techniques rather than their voices to engage with society. People with hearing and speech problems utilise hand gestures and accompanying actions, called sign language (SL) to communicate their intended ideas[1]. SL is a set of movements organised by Structure, grammar, semantics, pragmatics, and morphology. SL is the natural mode of communication for those who are deaf or mute [2].

There is no common SL, and SLs have developed organically as different groups of people interact. Different countries and regions use different SLs, and almost every country has an official SL [3]; examples include the India SL (ISL), Arabic SL (ArSL), American SL (ASL), and so on. Even in nations where the spoken language is the same, the sign languages utilised differ, for example, British SL and ASL.

Figures 1.1 and figure 1.2 show how the ASL alphabet 'Q' appears similar to the Indian SL alphabet 'U.' (ISL) [4], [5]. Most linguistic characteristics, such as grammatical structure, change between SLs, just as they do between spoken languages [6]. Even the simplest alphabetical representation as illustrated in figure 1 might change from one SL to the next, including the shape of the sign and how it is executed [3]. The datasets for this investigation were picked from three distinct SLs.

Over 430 million people (43.2 crore adults and 3.4 crore children), or 5% of the world's population, require rehabilitation to overcome their "disabling" hearing loss. By the year 250, every tenth individual may be at danger of hearing loss [7]. In underdeveloped countries, children with hearing loss and deafness are seldom educated, while adults with hearing loss have a higher chance of being unemployed or being in lower paying jobs.

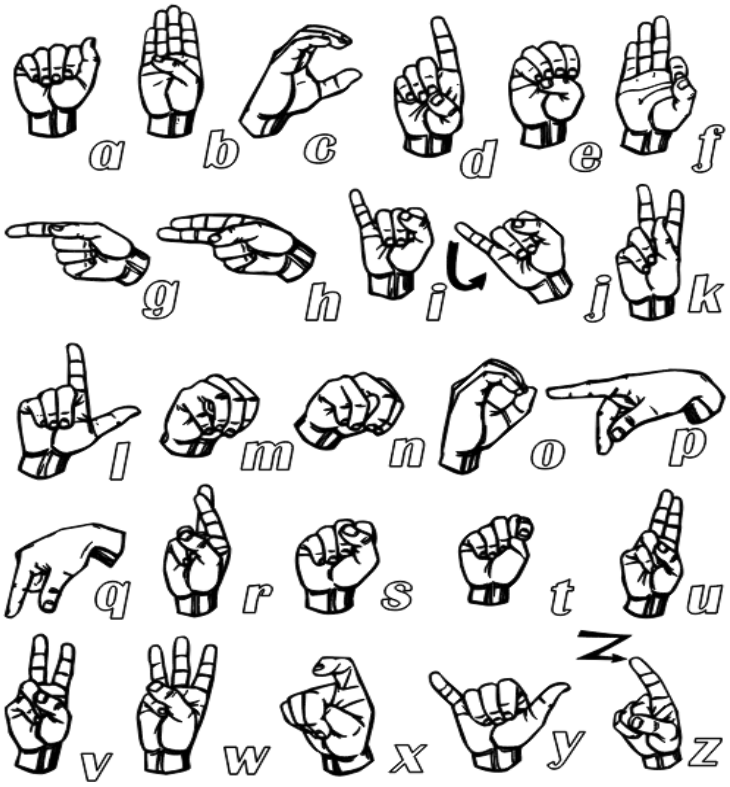


Figure 1.1: Signs for fingerspelling of English alphabets in the ASL [4]

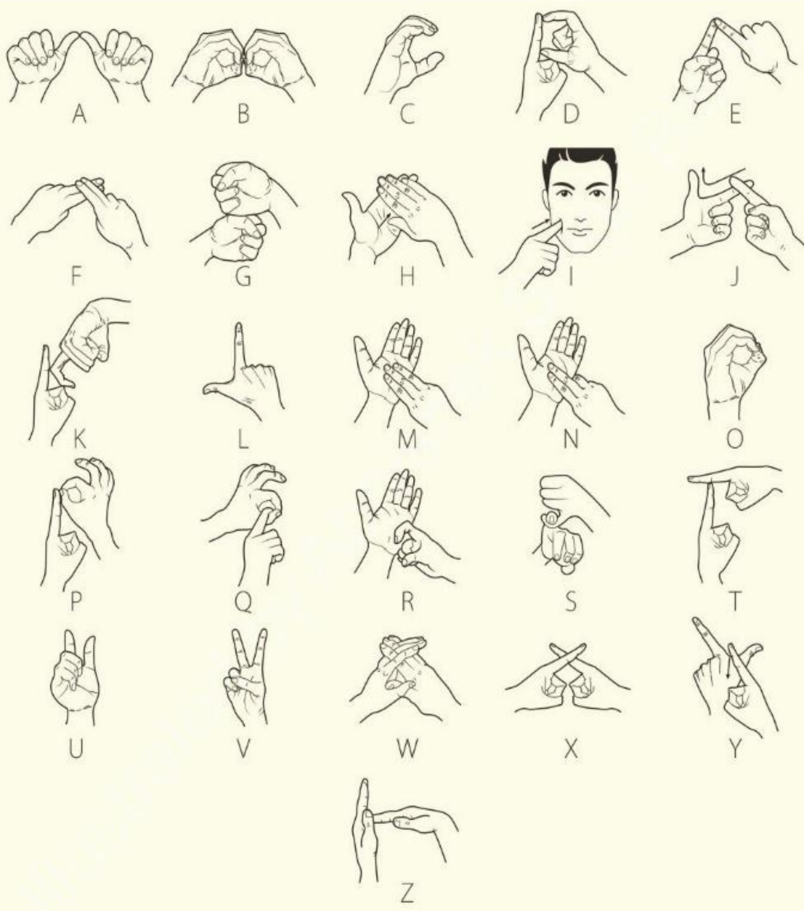


Figure 1.2: Signs for fingerspelling of English alphabets in the ISL [5].

Most people who can hear well are unaware of the gestures in SL, making interaction with others who have hearing and speech impairments difficult [8]. Researchers from all over the world have helped to develop numerous technologies that can help these people integrate into society while also overcoming communication gaps. In research and development, focus has been placed on the introduction of methodological tools for identifying the position, shape, and movements of a finger, wrist, arm, and forearm, as well as facial expressions.

RADAR waves [9], speakers [10], EEG sensors [11] - [14], Wi-Fi modules [15] - [18], RF waves [19], and other technologies have also been employed in SLR systems for the same.

The VBT, which employs cameras and image processing algorithms, is an AI-based solution to sign gesture identification problems. As a default technique of recording indicators, several researchers have employed a single camera. The signer's hand is segmented from gesture photos using techniques such as edge-based active contours [20], frame subtraction [21], wavelet transform [22], and others. RGB-D sensors, such as Microsoft Kinect [23] - [26], have been used by certain researchers to collect spatial information of the signer's hand and body. The jump motion sensor [27], [28] is another gadget discussed in the literature that records hand gestures on a 3D axis. The 3D information of the signer's hand and body is created from the precise photographs. This information may be used to determine a variety of important metrics, such as hand direction, wrist orientation, and joint angles.

Wearable technologies with sensors including Flex Sensors, Motion Processing Units, and pressure/contact sensors have been used for SLR in real-world conditions like strong illumination and complex backdrops in SBT. These sensors are a promising alternative since they are inexpensive, lightweight, and have the ability to process data on a low-power device such as an Arduino microcontroller. However, as the amount of expressions in a system grows, the possibilities of one of them being identical to another grow as well, potentially affecting the system's identification accuracy.

Certain aspects of SLR which must be considered are:

Processing Time: In practical application, the processing time taken for gesture recognition is essential for better user experience. In our system, we have developed a real-time application to capture gesture input using webcam and perform some pre-processing on the RAW image before sending it to pre-trained CNN model for classification.

Skin Tone Impact: In our application, the RGB picture obtained from the camera is transformed using the HSV filter in the pre-processing step to extract just human skin coloured regions from the image. However, humans have a broad range of skin tones, and different skin shades absorb light in a different manner, impacting gesture detection accuracy. As a consequence, for successful segmentation in varied lighting circumstances, different HSV values were necessary.

Impact of Intense Body Movements: The system's recognition rate is restricted during the real-time gesture detection process by the frame rate of the camera in question, which is typically 24-30 fps on a laptop webcam. If sudden motions are performed, acquired images may become blurred as a result of motion artefacts, resulting in a deformed image that is worthless for identification.

Background and lighting conditions: In picture classification tasks, both background colour and illumination circumstances are critical. It is simpler to remove the hand segment from a picture if the backdrop is uniformly coloured and does not match the signer's skin tone. The algorithm may fail to segment the hand region from the picture if any skin-color items are present in the ROI.

Photographs acquired in low-light environments have more noise than photos taken in well-lit settings. As a result, edge detection and segmentation are poor, resulting in lower identification accuracy.

The majority of gestures are classified as static or dynamic depending on whether the signer moves his or her arms. The topic of our paper is the identification of static gesture indicators. One of the most significant issues with SL gestures is that the appearance of one gesture might well be identical to that of another in terms of fingers flexion and wrist positioning, making the two signals appear identical when performed. Because of the ambiguity, poor classification and low accuracy are possible outcomes. Another issue is the dependency on signers. Differences in the capturing of values can be noticed by repeating the same sign numerous times since an individual cannot keep all of his or her hands/fingers in the same position.

CHAPTER 2

PRIOR WORK

Some of the existing work in literature in the domain of SLR using vision based techniques are shown in table 2.1.

Table 2.1. Some existing VBT for SLR

Ref.	Methodology	Conclusion	Data set
[29]	In this paper, authors have performed data augmentation by flipping, rotating (0º-30º), and brightness changes on the images. Then image features were extracted using pre-trained CNN models ResNet50, MobileNet-V2 architectures, and a combination of the two.	Trained for less epochs, leading to reduced performance. Accuracy increase observed by using combination of both models.	ArSL 2018 [38], [39]
[30]	To overcome issue with classes contain different sample sizes, they used Synthetic Minority Oversampling Technique, after which their model's accuracy increased. After normalisation and standard scaling, custom CNN model with 7 convolution layers was used for classification.	Synthetic oversampling works better than minority oversampling, and minority under-sampling, showcasing advantages of using data augmentation.	ArSL 2018
[31]	After image pre-processing and data normalisation, the data was passed to a CNN model with 4 convolution layers. Compared to system trained on reduced size database, using the complete database, and training for more epochs, the testing accuracy increased to 97.6%, however this significantly increased the training time by almost 9 times.	Increasing the number of layers in the CNN model increased the significantly, however the number of images used for the training process didn't had such a large impact on the model's performance.	ArSL 2018
[32]	After resizing images to 32*32, data augmentation using horizontal flipping of the images, data normalisation, and removing noisy images, 20,227, and 12,480 images were used for training and testing respectively of two custom CNN models with different layers.	In their implementation, they only used 20,227 for the training set, had they used about 40,000 images from the overall set, then their overall accuracy might increase from 96.6%.	ArSL 2018
[33]	To overcome the imbalanced number of images in classes, they used only 1000 randomly selected images from each class before performing data augmentation and normalisation. The data was then classified using EfficientNetB4 based CNN classifier trained for 30 epochs.	Their system took 10 hours to train the model which is considerably high, still they attained accuracy of 95%, which is low compared to other models in the study.	ArSL 2018
[34]	They used pre-trained VGG16 model for feature extraction from the images before passing it to fully connected dense layers of the CNN. GridSearchCV was used to find optimal values for	Considerably low accuracy of 94.33% was achieved on this dataset compared to other models for the same.	ArSL 2018

	hyperparameters tuning.		
[35]	They used Sobel operator to find ‘absolute gradient magnitude’ of the images for segmentation of the input images before passing to a 3 different transfer learning based models, out of which VGGNet achieves the highest classification accuracy.	Freezing weights of the pre-trained models led to model not being trained for the actual dataset, leading to 97% accuracy.	ArSL 2018
[36]	Initially they resized the images to 28*28, and then applied Gabor filter with 4 different sizes and 4 different orientations to obtain spectral representations, which were then passed to a CNN classifier for recognition.	Usage of Gabor filters increased the dataset by factor of 16, which led to a high recognition accuracy of 99.05 on this dataset.	ArSL 2018
[37]	After under-sampling to equalize the number of images per class, they performed data augmentation before passing it to pre-trained VGG16, and ResNet152 models, and trained for 100 epochs with the original weights frozen.	Even though attained validation accuracy of 99.6%, it increased by only 0.1 percent from epoch 60 to 100.	ArSL 2018
[40]	After image pre-processing, data augmentation, and standard scaling, the data was passed to a VGGNet16 model for training the model. The authors only trained their model for very few epochs due to fact the pre-trained model’s converged very rapidly.	Their model underperforms for ‘0’, ‘N’, ‘W’ gestures, if the model were trained for more epochs, it might have improved the classification accuracy.	Mass ey [50], [51]
[41]	In this paper, a pre-trained GoogLeNet architecture trained on the ILSVRC2012 dataset has been used for sign gesture classification.	Scalability remains concern as model has validation accuracy of 98% with 5 letters, 75% with 10 letters.	Mass ey
[42]	The input data was passed to 3 models: A custom CNN model, pre-trained VggNet16 model, and Inception-V3 model. It was found that the pre-trained models outperform the custom CNN model by about 4%. Further by using data augmentation accuracy increased by about 3.5%.	The selected features from the fused set resulted in high recognition accuracy of 99.63%.	Mass ey
[43]	Single Shot Multibox Detector to detect and crop hand region was used before background subtraction on cropped image. 10 different sets of random background images were added to make the classifier more robust. The pixel values were then rescaled to be in 0-1 range, and passed to a Wide Residual Network (WRN) classifier.	The performance increase of pre-trained WRN over a baseline CNN model further signifies the merits of pre-trained models trained over large datasets.	Mass ey
[44]	In this paper, multi-channel CNN based feature fusion technique is used as classifier. The input image is passed to one pipeline of the CNN as it is, and a Gabor filter is applied to the image before passing it to the second CNN pipeline for feature extraction.	The proposed ensemble model for feature extraction removes high frequency spatial dependencies and improves performance.	Mass ey
[45]	They used Media-pipe hands API to extract co-ordinates of 21 joint points on the hand of signer. Then distances between the sets of these joint points were calculated for a total of 190 distance vectors, similarly 210 angle vectors were calculated by using the relative distance between the joint points. Also, angle based vectors were also calculated to fetch the orientation of the hand. All these feature vectors were then passed to a SVM, and a light Gradient Boosting Machine (GBM) for training and classification. Testing accuracies of	Their system achieves per sample recognition time of only 14 ms, which is highly adequate for a real-time SLR system given that their system also achieves high recognition accuracy of 99.39% on this dataset, and respectively high accuracies on 2 more ASL datasets.	Mass ey

	99.39%, and 97.80% were obtained for SVM, and GBM respectively.		
[46]	They used a RCNN based model to detect hand region in image, from which 5 different regions were cropped out. After this, and Gaussian, and Salt-and-pepper noise were added to the images in 4 different varying proportions. The original image, cropped images, noise-added cropped images were passed to 3 different Restricted Boltzmann Machines, whose outputs were fused for final output generation.	Multi-channel data fusion and multiple transformations of each image results in high accuracy 99.31%. The drawback with their system is that it was run for 10,000 epochs, and the high design complexity.	Massey
[47]	YCbCr based hand region segmentation and Convex Hull Algorithm based hand shape detection for image cropping was used before passing to a CNN classifier.	Hand region segmentation reduces useless features, training time, as well as increases accuracy.	Massey
[28]	VGG16 for classification, GridSearchCV for optimal hyperparameters tuning.	Model has lower recognition accuracy.	Massey
[48]	After cropping and resizing images based on skin segmentation, they have used Local-Binary-Pattern (LBP), and Histogram-of-Oriented-Gradients (HOG) as feature extractor, and CNN, SVM as classifier. They ran multiple combination of these classifiers and feature extractors.	The HOG-LBP-SVM model provides better accuracy than the CNN-SVM and other models, but increases complexity.	Massey
[49]	The authors propose a SLR system based on CNN models built using transfer learning from image multiple classifier systems. The input images will be resized, background subtraction, and convex hull algorithm for image processing will be used.	Only a model is proposed for future implementation and does not include any implemented model for the same.	Massey
[52]	In this paper, sign images were acquired from video feed, and after pre-processing to extract hand region and converting the RGB image to HSV, CNN model was used for training and classification purpose.	Their model converges in only 5 epochs with a high learning rate of 0.1, with high accuracy of 98.55%.	ASL[54], [55]
[53]	In this paper, the authors have used same technique as in [52], however by using a different configuration of the CNN layers, they manage to achieve better test accuracy on the dataset. They further evaluated their model's performance on 2 more datasets, as shown.	Increasing number of epochs to 100 allowed them to get increased accuracy of 99.41% on dataset [54], 99.48% on dataset [56], and 99.38% on their own synthetic dataset [57].	ASL [54], [55] ASL [56] ASL [57]
[58]	Human skin colour based hand portion segmentation using HSV values was done before passing the images to a SVM classifier for training and classification. The system works on mobile phone which allows it to be available to masses for the purpose of real-time SLR.	However, accuracy of the system is only 80-90 percent for different alphabets, which makes it inadequate compared to other systems in our study.	ASL [58]
[59]	A Region Proposal Network was used for ROIs in the image then pooling was used to bring all those segments to same size, before passing to classifier.	Attention based model reduces processing load and increases accuracy by 5%.	ASL [65]
[60]	After converting the RGB images to grayscale, and using Canny edge detector for edge detection[61], the converted image was passed to multiple feature extractors including HOG, LBP, PCA, and ORB[62] feature detector which produces a 32-dimension feature vector. After converting these features to a bag-of-words [63] model, multiple	Combination of PCA with multi-layer-perceptron (MLP) classifier achieved highest classification accuracy.	ASL [65]

	machine learning based classifiers were used.		
[61]	They used multiple CNN based classifiers using transfer learning methods. It was found that the pre-trained models offer much better classification accuracy than custom CNN models, with ResNet50 model achieving highest accuracy.	A finding of this paper was that the performance of InceptionNet was lower than compared to baseline CNN model, and Resnet50 outperformed VGG16.	ASL [65]
[62]	After resizing, image was broken down into smaller patches. These patches were passed to Vision Transformer (ViT) encoder consisting of attention layer and MLP. The output of this layer was then passed to another MLP for final classification.	Due to low accuracy of 80.6%, this model cannot be considered to be an adequate.	ASL [65]
[63]	In this paper, image resizing to 50*50 pixels, data augmentation using Gaussian noise adding and image rotation was performed. The images were then passes to a CNN model with 4 convolution layer for training and classification.	Excellent accuracy of 99.89%, but per image recognition time of 0.236 seconds makes it unsuitable for a real-time system.	ASL [65]
[64]	In this paper, transfer learning model for gesture classification was tested using VGG16 model with and without Bimodal Distribution Removal (BDR). It was found that use of BDR leads to significant improvement in recognition accuracy.	Only 9% images from the dataset were used for training, and they got 68% accuracy.	ASL [65]
[42]	Combination of VggNet, Inception-V3 were used for this dataset too.	99.02% accuracy on dataset [65].	ASL [65]
[36]	Gabor filters with CNN	99.9 accuracy on MNIST dataset.	MNIST
[66]	Image data was passed to two pre-defined models with and without augmentation. It was observed that lowest accuracy was obtained by LeNet, accuracy increased by 6.74% in the CapsNet model fed with original dataset, while accuracy increased by another 6.62% using augmented data.	Data augmentation used during training process resulted in improved performance of the system.	MNIST [72]
[67]	A custom CNN model was proposed with 3 convolution layers, and ran for 10 epochs over the training set.	The system was evaluated using only 10 images, which is insignificant.	MNIST
[68]	In this paper, numeric data from xls files of the dataset were converted to PNG files, and then passed to two models: MobileNet, InceptionV3. These models were trained and then the trained model was used for SLR on a smartphone.	Average recognition time for a gesture was about 2.42 seconds, which is not suitable for real-time system.	MNIST
[69]	In this paper, Particle Swarm Optimisation (PSO) was used to optimise the parameters that affect classification accuracy of CNN based image classifier. Use of PSO results in accuracy of 99.53%.	Time required for the system to run for all the combinations of parameters is quite long, and not scalable for large datasets.	MNIST
[70]	This model uses a CNN classifier incorporating 7 convolution, and 2 dense layers for SLR.	The model's accuracy is low.	MNIST
[71]	In this paper, discrete wavelet transform was used to extract features from the images, to be classified by the CNN module.	Using wavelet transform to extract low level features from images results in high accuracy.	MNIST

The following are some of the key takeaways from the preceding table:

1) Transfer learning is an excellent strategy for solving the SL classification issue since it gives improved performance as well as reduced training time because these models have already been trained for greater performance on much bigger classification problems in the same category.
2) When pre-trained model weights are frozen, the models tend to be less accurate on the given task since they are more generalised.
3) Using RoI segmentation from an image improves classification accuracy and reduces training time by removing undesirable features from the data.
4) Oversampling strategies for adjusting for uneven numbers of entities increase performance in classes with unequal numbers of entities.

CHAPTER 3

PRELIMINARY

Transfer learning[73] utilizing customized pre-trained EfficientNetB2[74] based CNN[75], [76] architectures was utilised to detect Sign language gestures[77]. In the sections that follow, the design of these components is explained. Also mentioned below is thorough information regarding the techniques and Python libraries that we employed in our suggested implementation, such as OpenCV[78], Keras[79], TensorFlow[80], and others.

3.1 Sign Language: Hand signals, body posture, and facial expression are all used in this method of communication. It is the Deaf and Hard-of-Hearing community's primary mode of communication. Static gestures, which are invariant to motion throughout the duration of the gesture, and dynamic gestures, which are concentrated on movement, are two types of SL gestures. As illustrated in figure 3.1, a type 0 sign is a two-handed dynamic gesture in which both hands are active. Signs can be executed with one both hands, they can be static or dynamic, and additional categorizations of signs are available. Only the primary hand performs motion in the Type 1 sign, which is a two-handed dynamic sign.

3.2 Convolutional neural networks

They're amongst the top popular deep learning algorithms for analysing visual images. CNN needs less pre-processing than other image categorization approaches. CNN is a feed-forward artificial neural network composed of several neural network layers, each with numerous neurons. The network learns feature extraction filters that are typically hand-modelled in other paradigms. The four types of processes in a CNN are convolution, pooling, flattening, and completely connected layers. The convolutional layer commonly captures low-level properties such as colour, edges, and gradient direction. To minimise the feature space of the convolved features, a pooling layer is utilised. This procedure reduces the amount of processing time required to work with the data by reducing dimensionality.

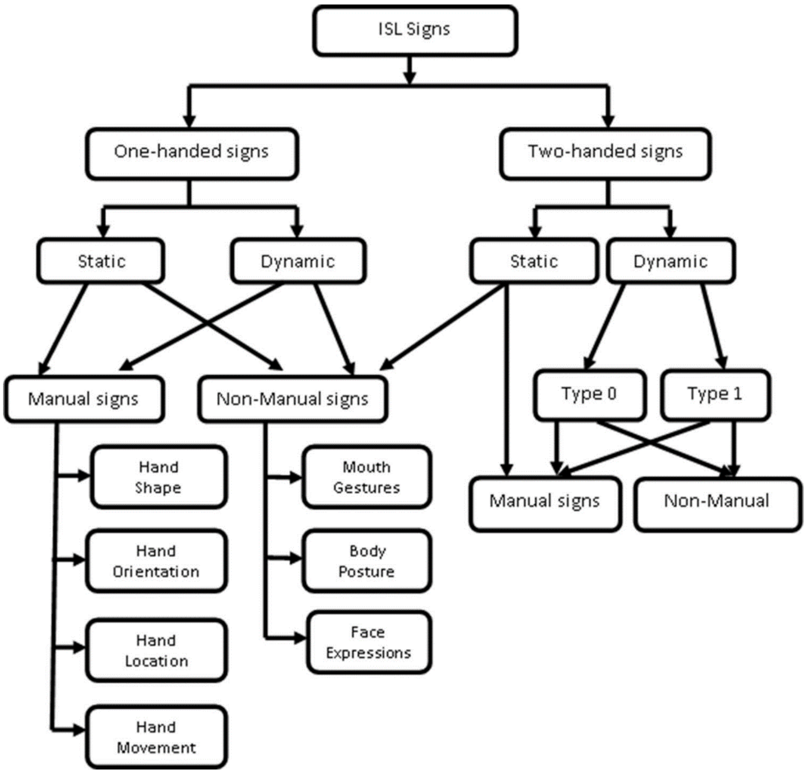


Figure 3.1: Systematic representation of Hierarchy of SL gestures[81]

It also has the advantage of maintaining dominant traits, which are immutable in terms of positioning and rotation, throughout the model training process. Following the processing of the input image, higher-level attributes can be used for classification. The image gets flattened into a one-dimensional vector as a result of this. The flattened output is sent into a CNN fully connected layer, which assigns labels to the features obtained by the convolutional/pooling layer. After training with SoftMax[82] classification, the model may give a likelihood of identification of objects in the image.

3.3 Transfer Learning

Transfer learning for imagery categorization is based on the idea that if a system is developed on a big and varied dataset, it may be used as a general model of the visual world. This is important for a wide range of reasons, including:

- Useful Learnt Traits: The models have learned to recognise generic elements from images after being trained on millions of photos in thousands of categories.
- State-of-Art Performance: The models achieved cutting accuracy and tend to excel in the visual identification task for which they were designed.
- Simple to Use: The model weights may be obtained for free, and many libraries support easy APIs for directly obtaining and utilising the frameworks..
- This technique also has the advantage of requiring substantially less data and time.

3.4 EfficientNet

CNNs are usually designed with a fixed resource cost and then scaled up when additional resources become available to achieve better accuracy. Traditional model scaling procedures, on the other hand, are notoriously unreliable. Some models have a depth scale, while others have a width scale. To obtain better results, some models simply employ higher-resolution pictures. This approach of arbitrarily scaling models needs manual adjustments and a large number of man-hours, with limited or no efficiency gain.

EfficientNet uses a simple yet effective method called compound coefficient to scale up models. Rather than arbitrarily expanding width, depth, or resolution, compound scaling raises each parameter using a specified group of scaling factors. The researchers behind efficient developed 7 models[B0 to B7] with varying sizes that outperformed and outperformed most CNNs[83].

The architecture of EfficientNetB2 is shown in figure 3.2, figure 3.3.

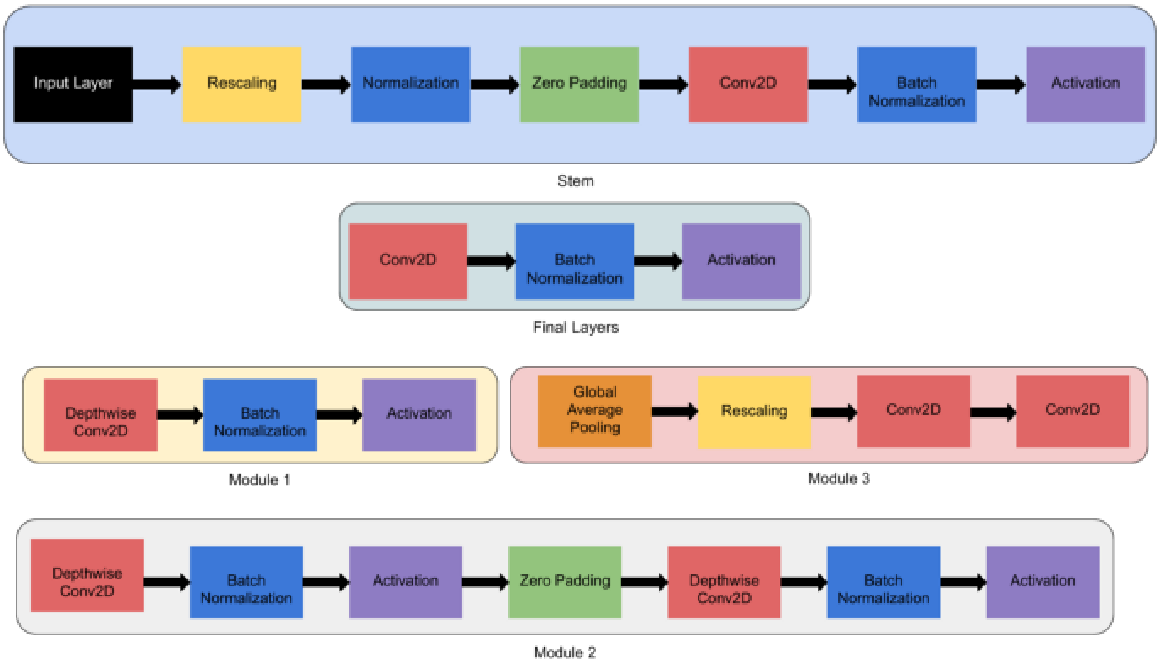


Figure 3.2: Contents of modules in EfficientNetB2 [83]

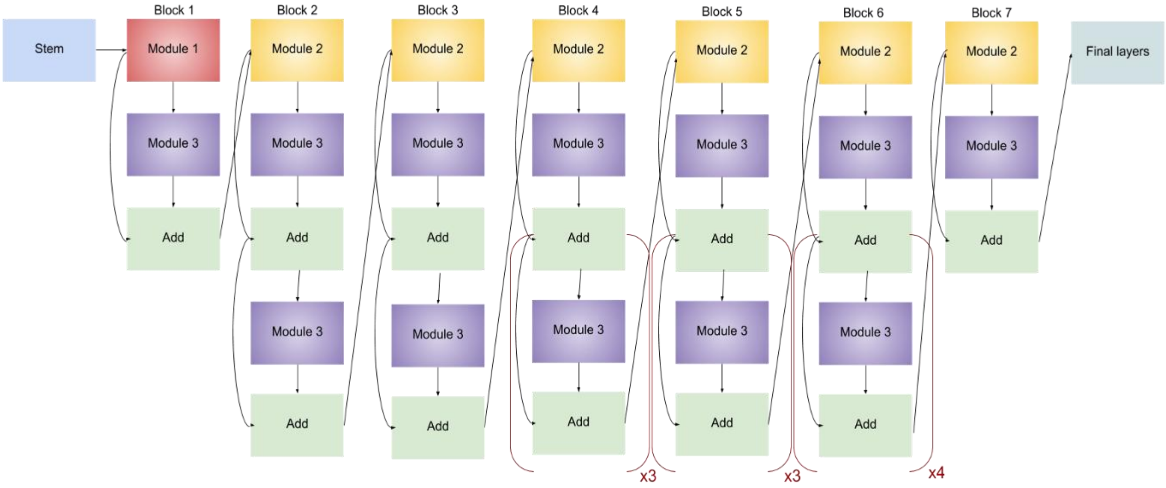


Figure 3.3: Architecture of EfficeintNetB2[83]

CHAPTER 4

PROPOSED WORK

4.1 PROBLEM STATEMENT

We typically communicate our ideas, emotions, and facts through vocalisation. However, not everyone is endowed with the ability to vocally share our views with others. The deaf and mute community has a difficult time conveying their feelings and perspectives to others; sign language is their most expressive mode of communication, but the general public is unaware of sign language, therefore the mute and deaf have difficulty connecting with the bulk of the people. A technology that can correctly translate sign language motions to speech and conversely in real-time is required to resolve these communication challenges. While there are ways for consistently and accurately transcribing written and verbal language to SL instantaneously, the same cannot be said for translating SL to textual and/or vocal formats. Existing systems either don't provide bidirectional communication, aren't real-time, have limited identification accuracy, or need stable environmental circumstances. Some systems need the purchase of extra hardware, such as pricey sensors, which raises the price. As a result, we created a vision-based SLR system that recognises objects in real time and with great accuracy.

4.2 PROPOSED SOLUTION

Our system is made up of the following components: EfficientNetB2 is the pre-trained model. To get the final categorization results, we added some custom layers after the original model. To get the best performance out of the system, certain settings were tweaked. To get at the final architecture, several design considerations are made. More detailed decisions are made on the pre-trained model, optimization strategies, and hyper-parameters to test and analyse. The resulting model has about 340 layers, and more than 8 million trainable parameters.

We also created a real-time SLR application as part of this research. The OpenCV library was used to retrieve images from webcam video feed. When the application is in prediction mode, three windows emerge during runtime, as seen in figure 4.1, 4.2. The first window is the photo capture window, which contains a fixed 192*192-pixel rectangle in which the user must perform the signs. The cropped picture is converted to HSV and presented in the preview window using user given parameters during program execution by utilising the trackbars in third window after taking the picture from the designated region, shown as stacked window in figure 4.1. The program then stores the modified image and feeds it to the CNN model that has already been trained to classify it. The network is trained with augmented data from many ASL datasets used in the study. Finally, the anticipated character appears in the image capture window's preset region. A third control panel with track bars, as illustrated in figure 4.2, is used to change the HSV threshold for skin colour based feature extraction during runtime. This is necessary since skin colour seems varied and appears to be of different hues in different lighting circumstances. Figure 8 shows a data flow diagram demonstrating the real-time system's working. The data flow diagram depicting the training process of the system is shown in figure 4.4, while the data flow of real-time SLR system is shown in figure 4.3.

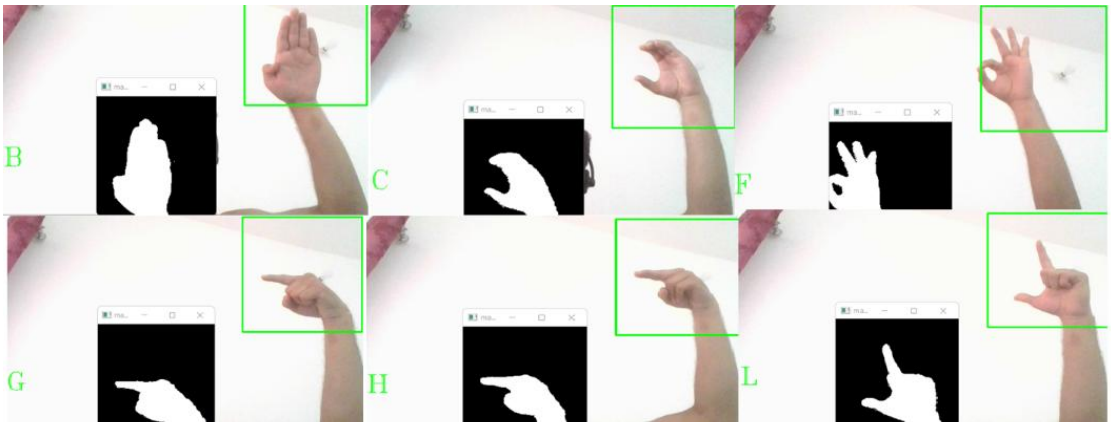
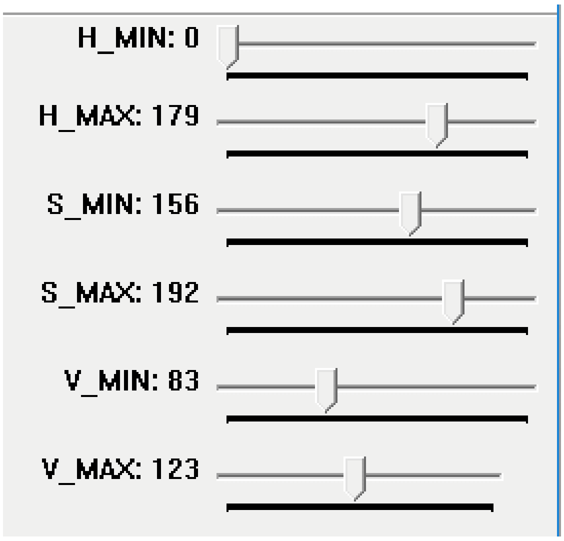


Figure 4.1. Some output images of first two windows of application from the runtime.



H_MIN: 0

H_MAX: 179

S_MIN: 156

S_MAX: 192

V_MIN: 83

V_MAX: 123

Figure 4.2. HSV parameters control trackbars window.

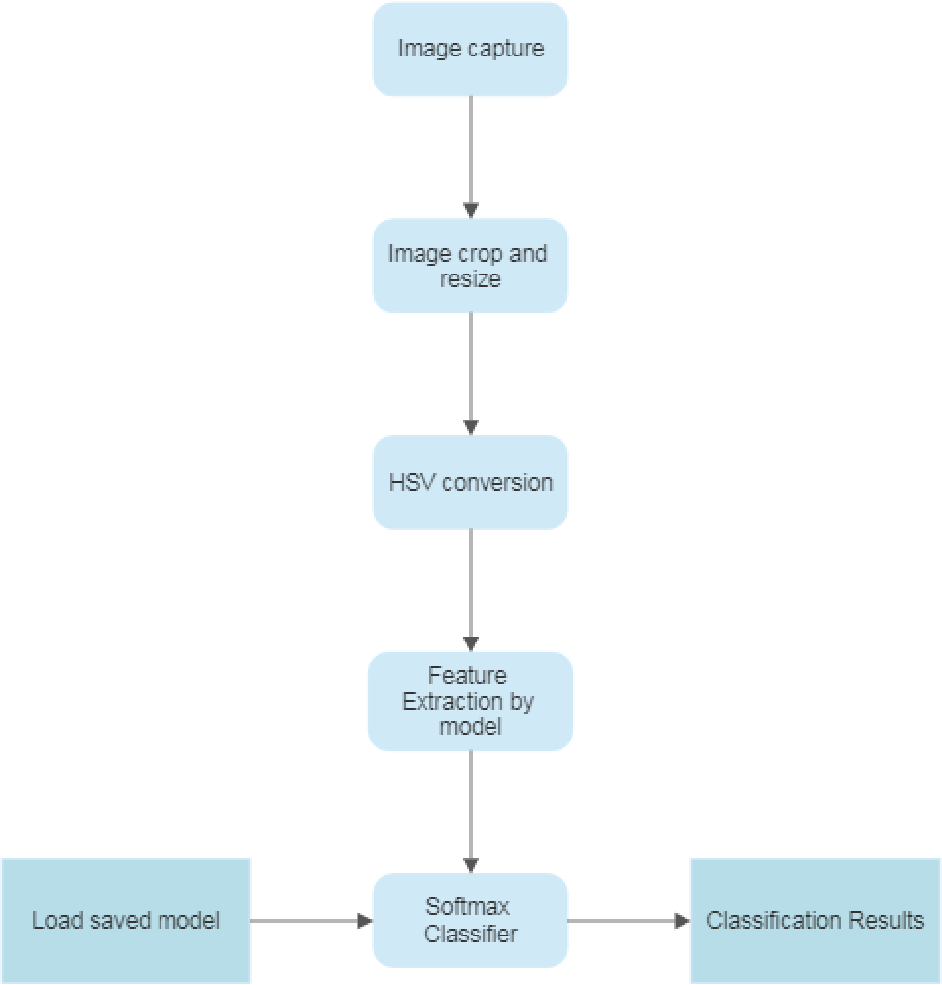


Figure 4.3. Data flow of real-time SLR system.

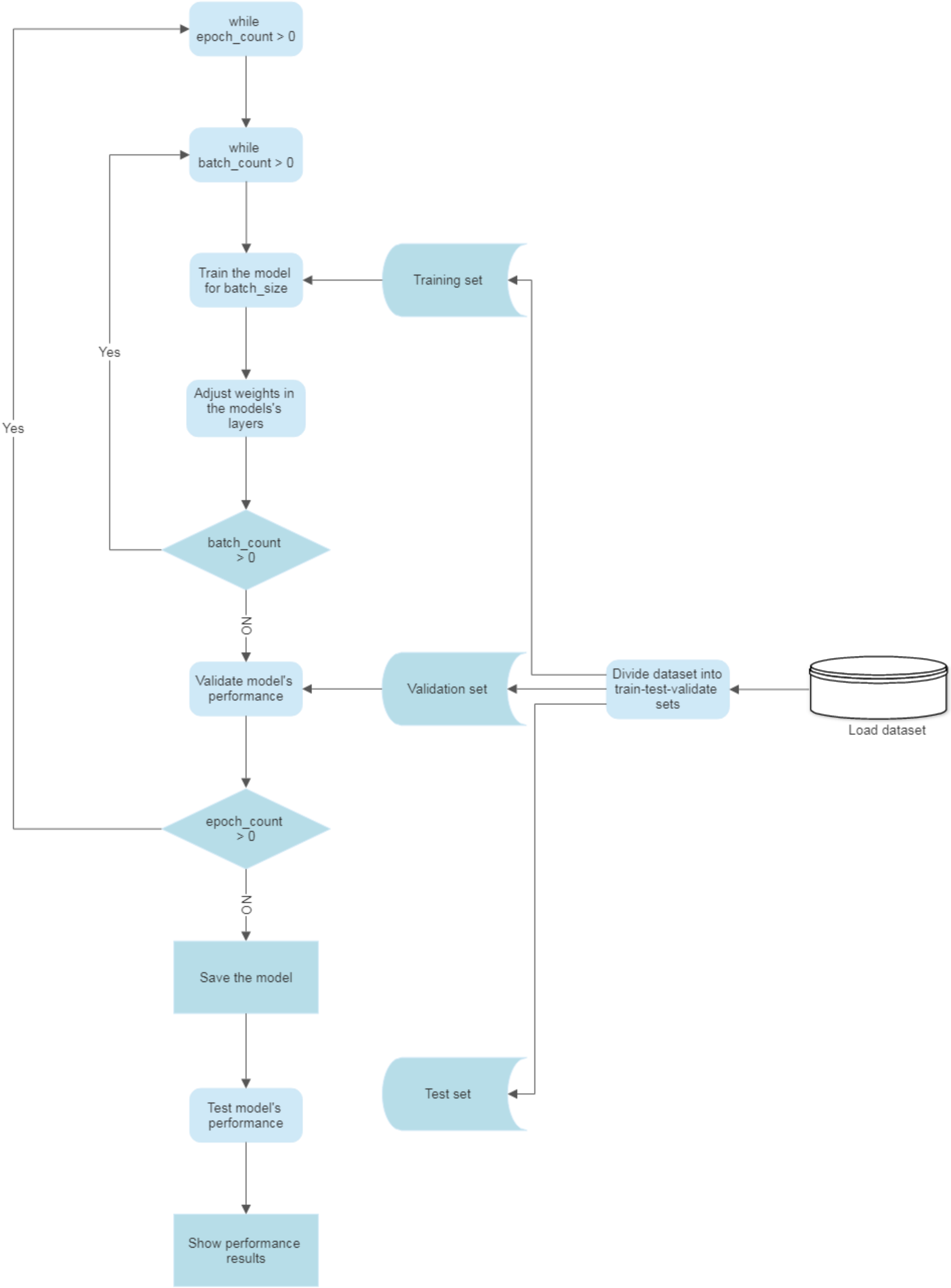


Figure 4.4. Data flow for training process of proposed model.

CHAPTER 5

EXPERIMENTS AND RESULTS

In this section, we describe the SLR datasets that we have included in our implementation, and implementation details.

5.1. Datasets

Our experiments are carried out on eight different Sign languages datasets, including 6 datasets of ASL and one dataset each of ArSL, PSL. Details about these datasets is shown in table 1. For all the datasets, if the train and test sets were not already in segregated form, then the data was split in 80-20 ratio for training-test sets respectively. After this, for all the datasets, the training set was split into 95-5 ratio for training and validation data respectively. Overview of details of all the dataset used is present in table 5.1.

a) The MNIST[72] dataset is divided into train-test files containing grayscale images, CSV files of ASL hand gestures represent a label (0-25) as a one-to-one map for each alphabetic letter A-Z (excluding 9 = J and 25 =Z which require motion).Each image is 28 pixels in height and 28 pixels in width, for a total of 784 pixels in total. Each pixel has a single pixel-value associated with it, indicating the lightness or darkness of that pixel, with higher numbers meaning darker. This pixel-value is an integer between 0 and 255, inclusive.The training data set, (train.csv), has 786 columns. The first column, called "Id", is the image identification number. The second column, called "label", is the hand gesture that was shown by the user. The rest of the columns contain the pixel-values of the associated image.The number of repetitions for each sign is not the same, but varies a lot, as shown below:

   MNIST test_set contents: A : 331 ,B : 432 ,C : 310 ,D : 245 ,E : 498 ,F : 247 ,G : 348 ,H : 436 ,I : 288 ,K : 331 ,L : 209 ,M : 394 ,N : 291 ,O : 246 ,P : 347 ,Q : 164 ,R : 144 ,S : 246 ,T : 248 ,U : 266 ,V : 346 ,W : 206 ,X : 267 ,Y : 332

MNIST train_set contents: A : 1126, B : 1010, C : 1144, D : 1196, E : 957, F : 1204, G : 1090, H : 1013, I : 1162, K : 1114, L : 1241, M : 1055, N : 1151, O : 1196, P : 1088, Q : 1279, R : 1294, S : 1199, T : 1186, U : 1161, V : 1082, W : 1225, X : 1164, Y : 1118

b) PSL[58] dataset contains 640*480 resolution images of 37 classes of Urdu words, with each class containing between 39-50 images each. After splitting into training, test, and validation sets, each class ended up having between 30-40 images for the training set, and between 8-10 images for each class of the test set. The images in the dataset are of cropped hand regions containing hand gestures only. These gestures were performed in front of uniform coloured background, and in slightly variable lighting conditions, with varying distance from the camera. Some images also contain shoulder and arm regions of the signer in addition to the hand. The number of signers involved in creation of the dataset have not been explicitly mentioned in the article, but from observation it can be assumed that only 1 person was involved in creation of the entire dataset. Some sample images from this dataset after cropping are shown in figure 5.1.

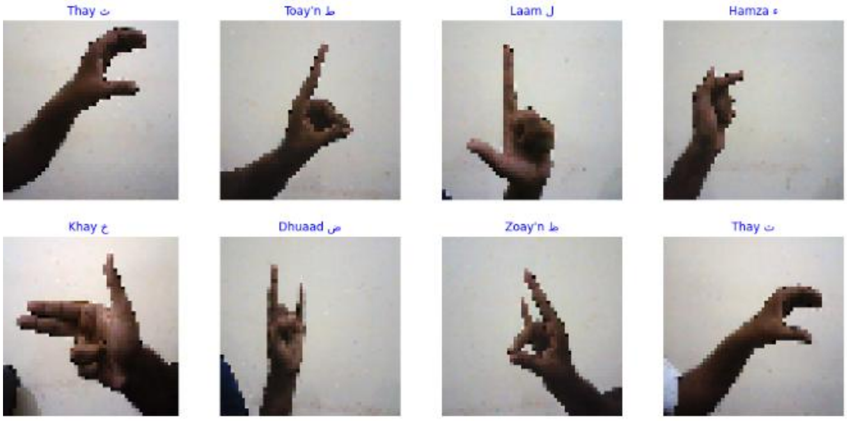


Figure 5.1: Sample images from the PSL dataset [58].

c) ArSL2018[38] dataset contains 64*64 resolution images of 32 classes of Arabic words, with each class containing between 1293-2114 images each. The images in the dataset are of cropped hand regions containing hand gestures only. These gestures were performed in front of uniform coloured background, and in slightly variable lighting conditions, with varying distance from the camera. Major changes in hand orientation w.r.t. the camera are present in the dataset. Some sample images from this dataset after cropping are shown in figure 5.2.

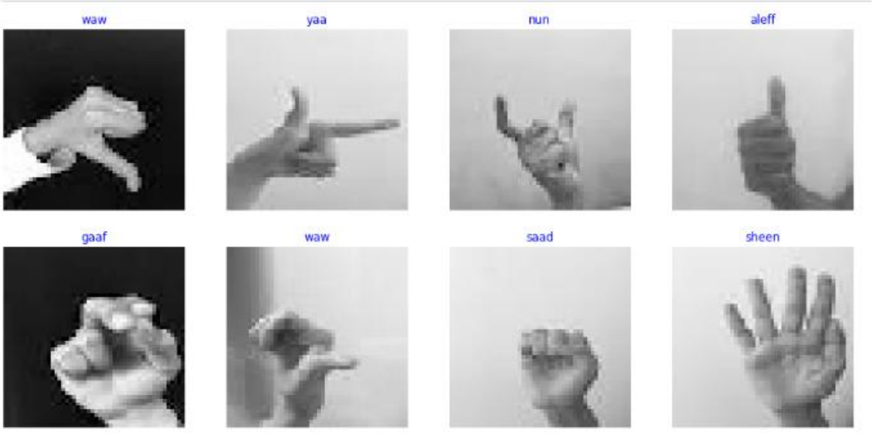


Figure 5.2: Sample images from the ArSL2018 dataset.

d) Massey[50] University gesture dataset contains 64*64 resolution images of 26 classes of English alphabets, and 10 classes for numbers 1-10, for a total of 36 classes. The images in the dataset are of cropped hand regions containing hand gestures only. These gestures were performed in front of uniform coloured background, and in well-lit conditions, with slightly varying distance from the camera. The images have varying resolution with image height varying approximately between 300-600 pixels, and image width varying approximately between 190-300 pixels. Some sample images from this dataset after cropping are shown in figure 5.3.

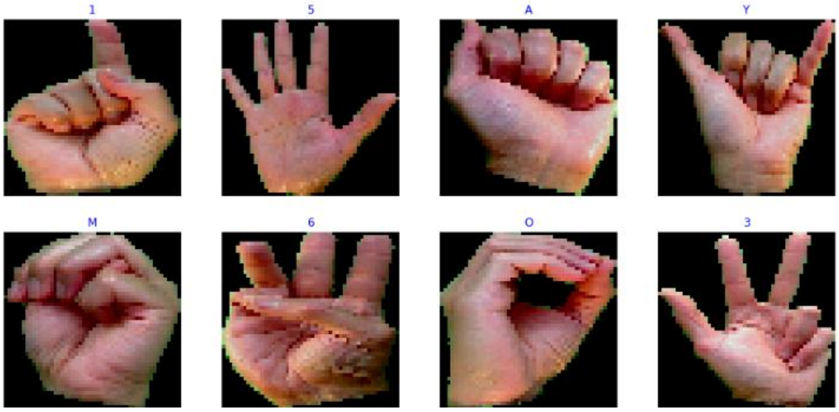


Figure 5.3: Sample images from the Massey university dataset.

e) In dataset [65], the training data set contains 87,000 images of ASL, which are 200*200 pixels. There are 29 classes, of which 26 are for the letters A-Z and 3 classes for SPACE, DELETE and NOTHING. The test data set contains a mere 29 images with 1 image for each class, due to this, the train data was split between train and test sets, after which each class contains 2400 images for training set, and 600 images each for testing set. The images in the dataset are of cropped hand regions containing hand gestures only. These gestures were performed in front of varying background conditions, and in variable lighting conditions, with varying distance from the camera. Major changes in hand orientation w.r.t. the camera are present in the dataset. Some sample images from this dataset after cropping are shown in figure 5.4.

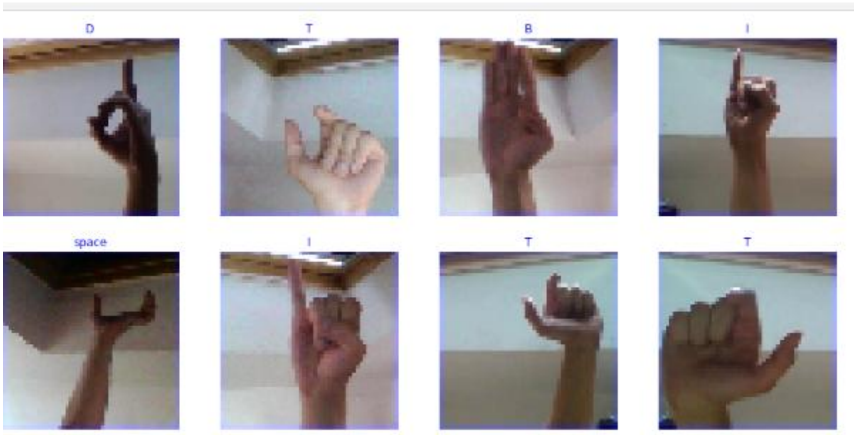


Figure 5.4: Sample images from the ASL dataset [65].

f) In dataset [54], the training data set contains 52,000 images of ASL, which are 64*64 pixels. There are 26 classes for the alphabets A-Z. Each class contains 1750 images for training set, and 250 images each for testing set. The images in the dataset are of cropped hand regions containing hand gestures only. These gestures were performed with slightly varying distance from the camera. Minor changes in hand orientation w.r.t. the camera are present in the dataset. In this dataset, the images were captured using camera and converted to black and white images using HSV skin colour based segregation. Some sample images from this dataset after cropping are shown in figure 5.5.

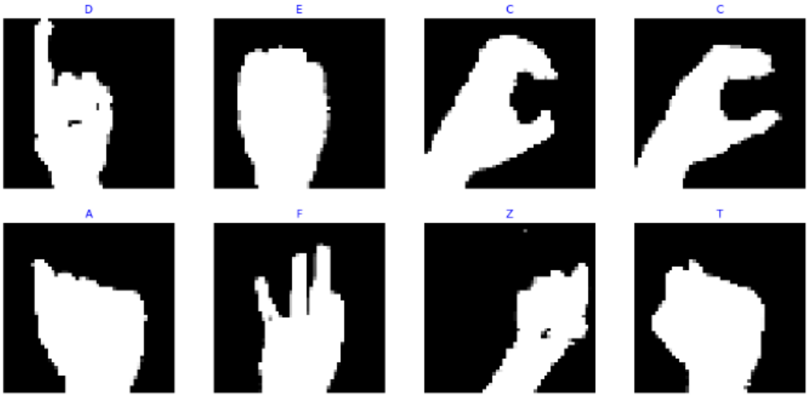


Figure 5.5: Sample images from dataset [54].

g) In dataset [56], the training data set contains 105,600 images of ASL, which are 50*50 pixels. There are 44 classes for the alphabets A-Z. Each class contains 2400 images in total, which contain 1920 images each for training set, and 480 images each for testing set. The images in the dataset are of cropped hand regions containing hand gestures only. These gestures were performed with slightly varying distance from the camera. Minor changes in hand orientation w.r.t. the camera are present in the dataset. In this dataset, the images were captured using camera and converted to black and white images using HSV skin colour based segregation. Some sample images from this dataset after cropping are shown in figure 5.6.

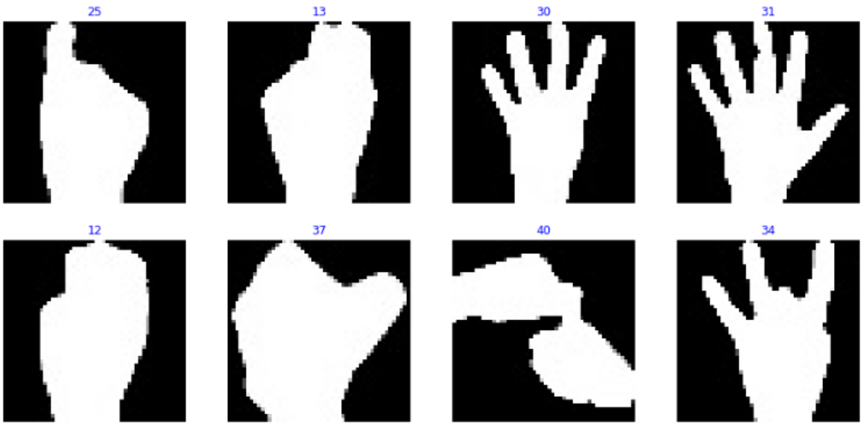


Figure 5.6: Sample images from dataset [56].

h) The dataset [57] is identical in structure to dataset [54], except for presence of 3500 images per class in training set, and 500 imager per class for testing set. Some sample images from this dataset after cropping are shown in figure 5.7.

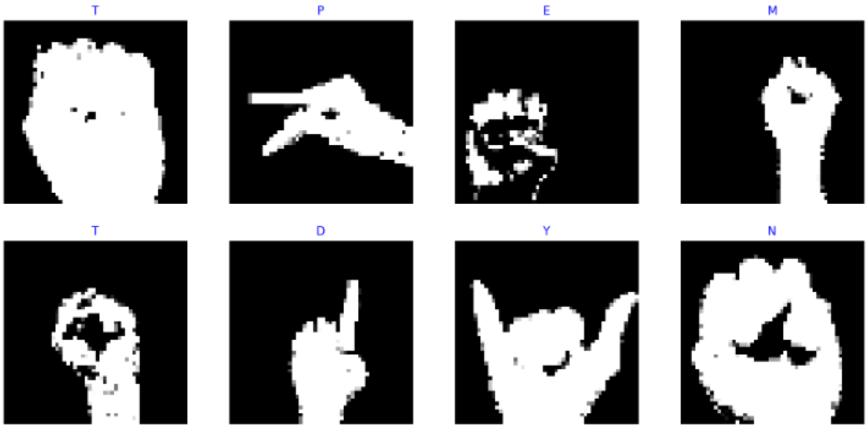


Figure 5.7: Sample images from dataset [57].

Table 5.1: Details of the datasets used in our study

Dataset	Sign Language	Type of dataset	Vocabulary size	Number of Signers	Total number of images
ArSL[38]	Arabic	Grayscale Images	32	40	54,049
Massey [50]	American	RGB Images	36	5	2,425
[54], [55]	American	HSV-BW Images	26	1	52,000
[56]	American	HSV-BW Images	44	1	105,600
[57]	American	HSV-BW Images	26	1	104,000
PSL[58]	Pakistani	RGB Images	37	NA	1,509
[65]	American	RGB Images	29	NA	87,000
MNIST[72]	American	CSV file	24	NA	27,455

Note: NA -> data not provided by authors.

5.2 Datasets Limitations

1) In most of the cases, the whole datasets were recorded with the same background conditions at the same place.
2) In the dataset's classes, there are insufficient orientation, posture, noise, lightning, and zooming variations.
3) The number of signers involved in dataset creation were usually very low.
4) In some datasets, the number of images differs from one class to another, e.g. in [38], [72].

5.3 Classification results and Performance Comparison

Eight separate SL datasets were used in the evaluation process of our proposed model. The details of these datasets are present in methodology section. In the cases if train data set is very large, we have chosen to trim the train dataset to max 1000 samples per class and the validation dataset to max 50 samples per class in order to reduce training time. All the images were resized to 50*50 pixels before feeding to the model. According to the results of the experiments, our proposed Effi-CNN model out-performed almost other existing models in test accuracy. Performance of proposed model on the aforementioned datasets is as shown below, and comparison of Effi-CNN with the best approach in literature, and training time taken is presented in table 5.2.

A) Performance on PSL dataset

On this dataset, we ran our model for 40 epochs, and got test accuracy between 99.34% and 99.6% on different runs of the model. Training and validation metrics for the course are as shown in figure 5.8. The classification report of the model on this dataset is shown in figure 5.9. In [58], the authors have used SVM for classification, and obtained testing accuracy of 90% on this dataset.

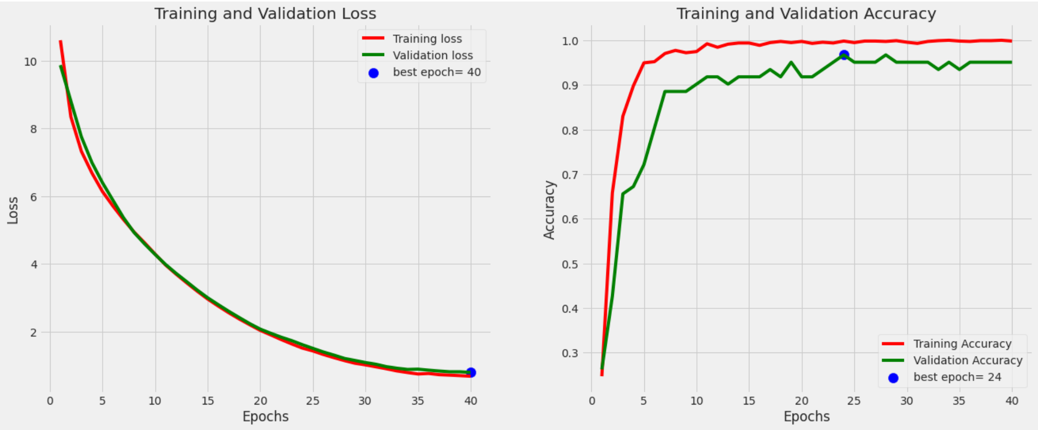
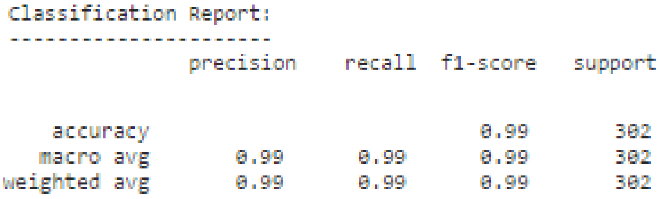


Figure 5.8: Training and validation performance of proposed model on PSL dataset.



```
Classification Report:
----------------------
                precision    recall  f1-score   support


    accuracy                           0.99       302
   macro avg       0.99      0.99      0.99       302
weighted avg       0.99      0.99      0.99       302
```

Figure 5.9: Classification report on the test set for PSL dataset.

B) Performance on ArSL2018 dataset

On this dataset, we ran our model for 40 epochs, and got test accuracy between 98.66% and 98.92% on different runs of the model. Training and validation metrics for the course are as shown in figure 5.10. The classification report of the model on this dataset is shown in figure 5.11. Figure 5.12 shows the number of errors per class on the test set. Performance comparison of our proposed model with others existing systems on ArSL dataset is shown in figure 5.13.

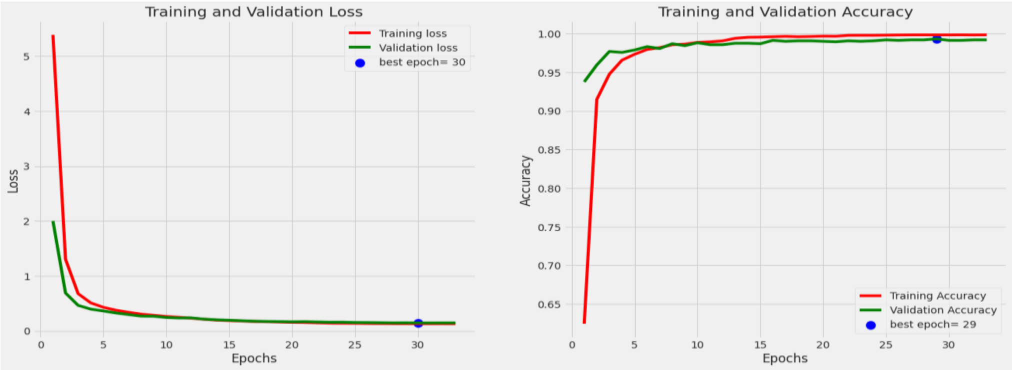
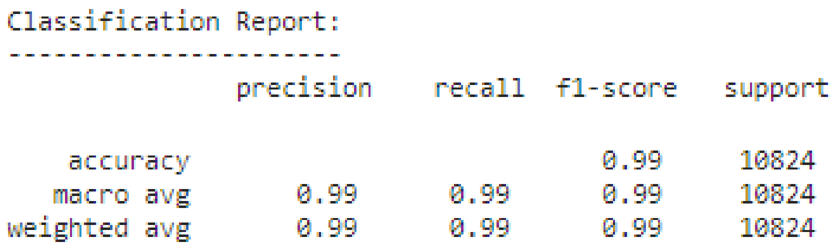


Figure 5.10: Training and validation performance of proposed model on ArSL dataset.



```
Classification Report:
----------------------
               precision    recall  f1-score   support

    accuracy                           0.99     10824
   macro avg       0.99      0.99      0.99     10824
weighted avg       0.99      0.99      0.99     10824
```

Figure 5.11: Classification report on the test set for ArSL dataset.

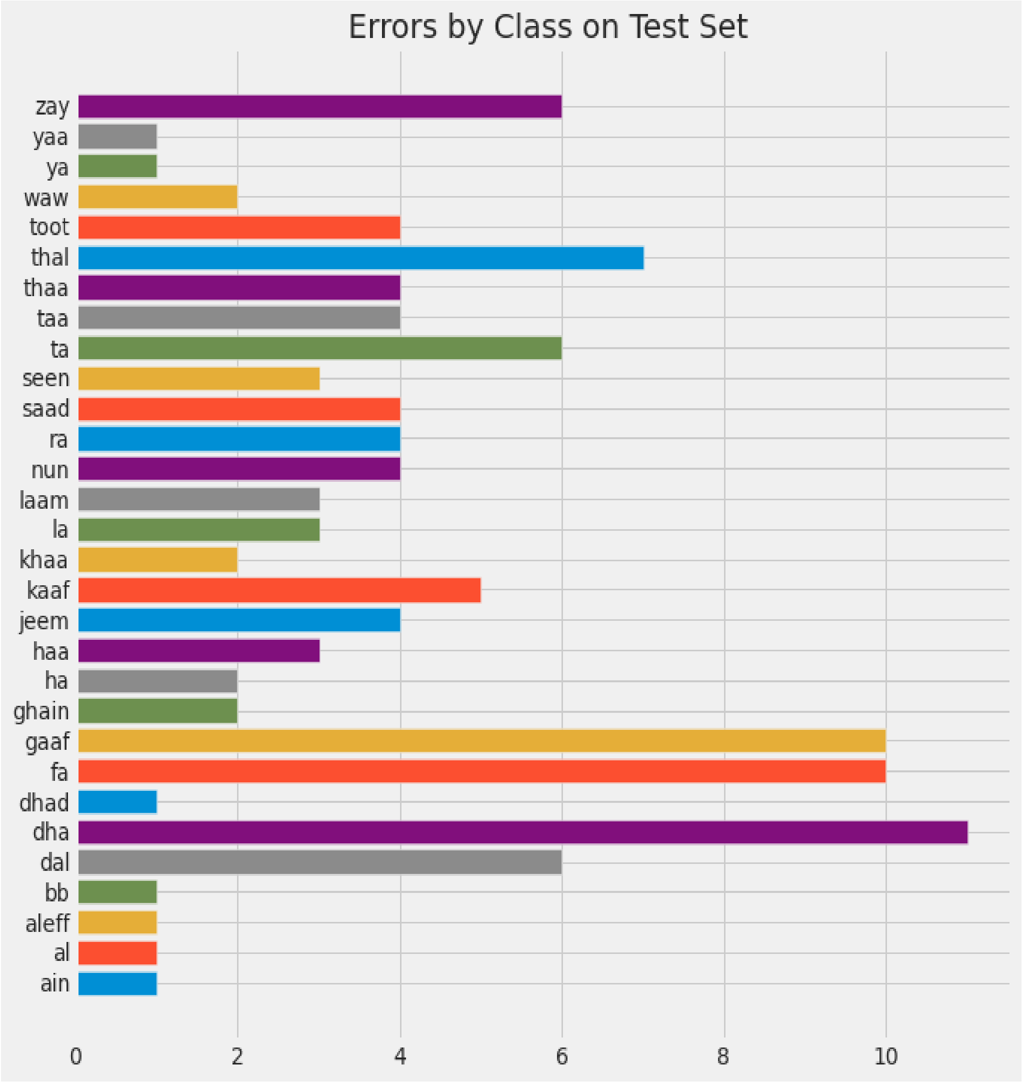


Figure 5.12: Number of errors per class on the test set for ArSL dataset.

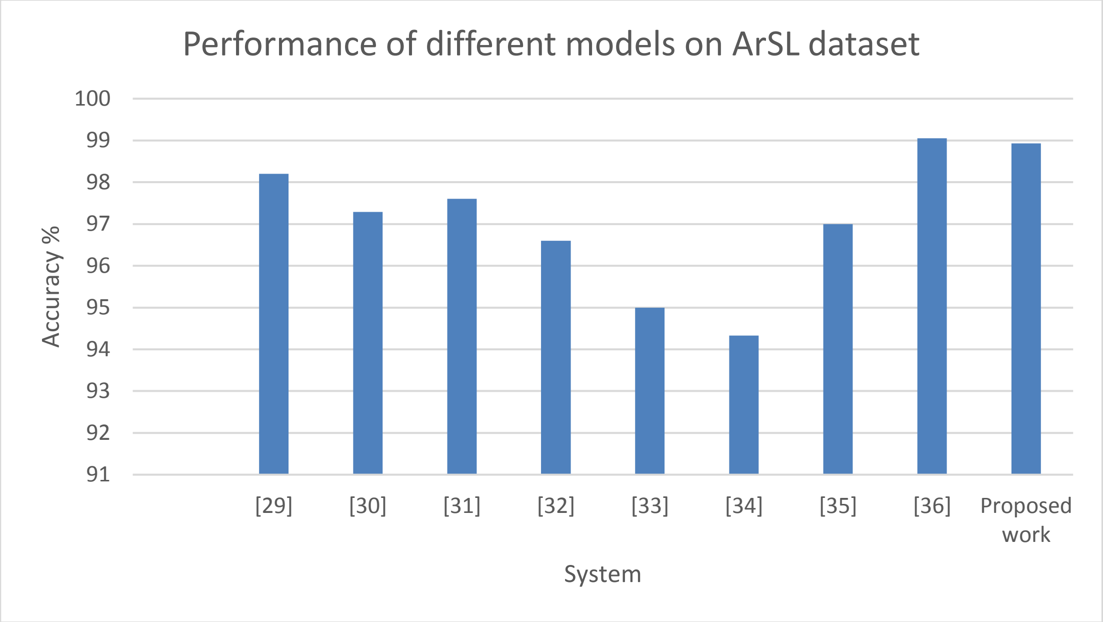


Figure 5.13: Performance of different models on ArSL dataset.

C) Performance on Massey University dataset

On this dataset, we ran our model for 50 epochs, and got test accuracy of 99.60%. Training and validation metrics for the course are as shown in figure 5.14. The classification report of the model on this dataset is shown in figure 5.15. Performance comparison of our proposed model with others existing systems on Massey university dataset is shown in figure 5.16.

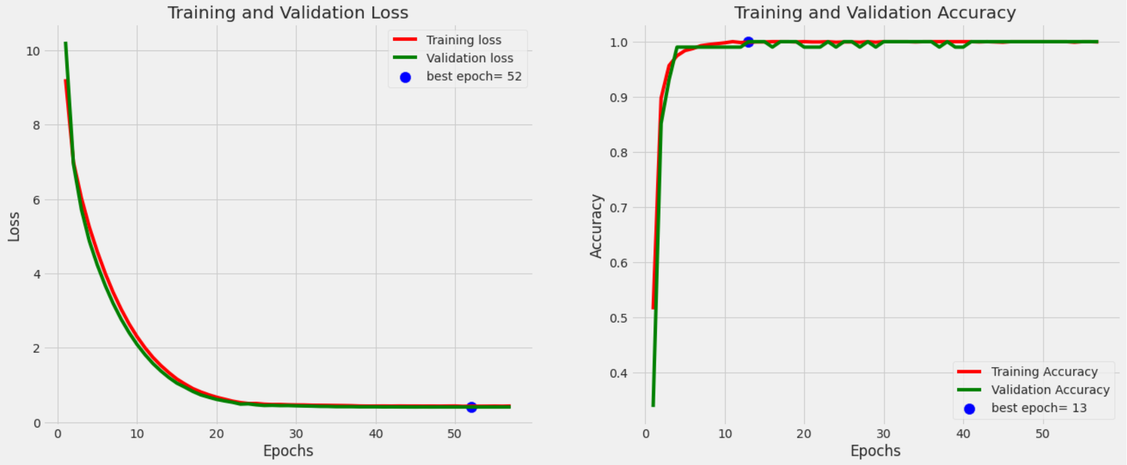
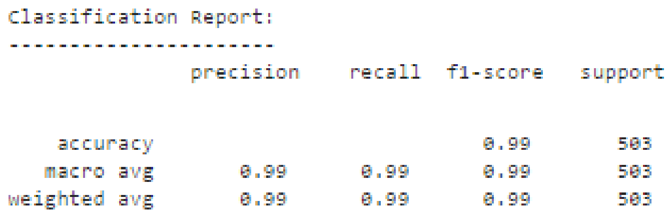


Figure 5.14: Training and validation performance of proposed model on Massey University dataset.



```
Classification Report:
----------------------
              precision    recall  f1-score   support

    accuracy                           0.99       503
   macro avg       0.99      0.99      0.99       503
weighted avg       0.99      0.99      0.99       503
```

Figure 5.15: Classification report on the test set for Massey university dataset.

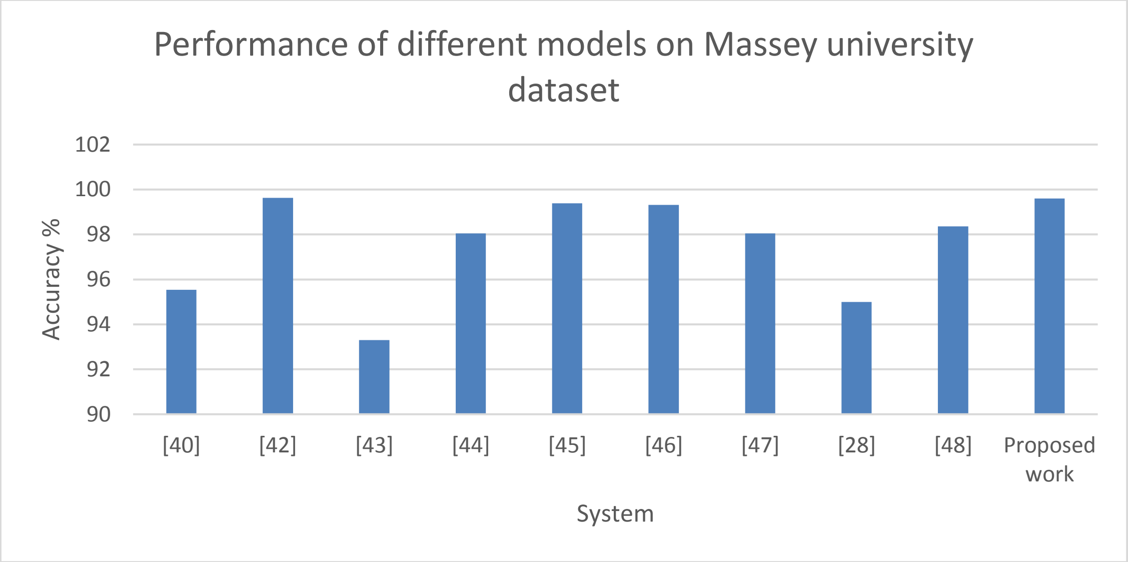


Figure 5.16: Performance of different models on Massey University dataset.

D) Performance on ASL dataset[65] with 87,000 images

On this dataset, we ran our model for 50 epochs, and got test accuracy of 99.99%. Training and validation metrics for the course are as shown in figure 5.17. The classification report of the model on this dataset is shown in figure 5.18. Confusion matrix for the test set is shown in figure 5.19. Performance comparison of our proposed model with others existing systems on dataset [65] is shown in figure 5.20.

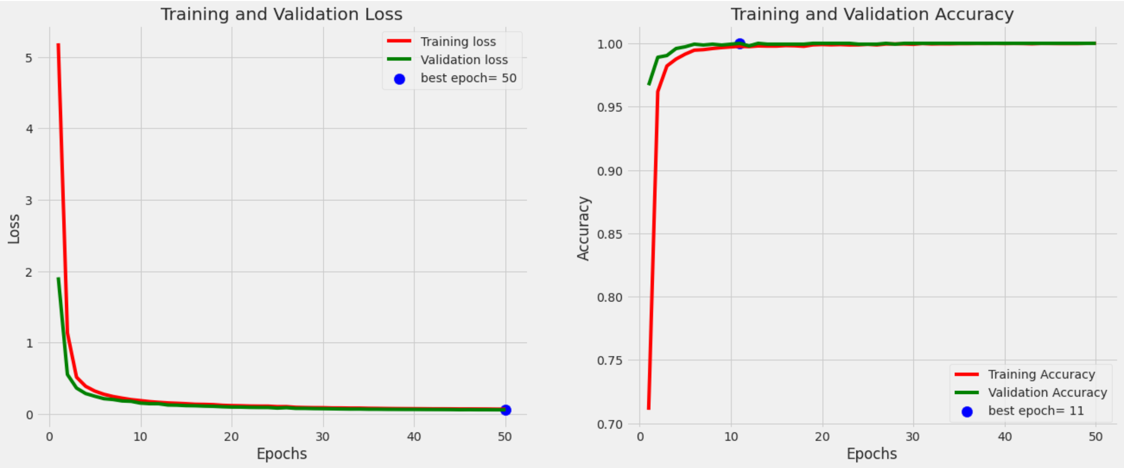
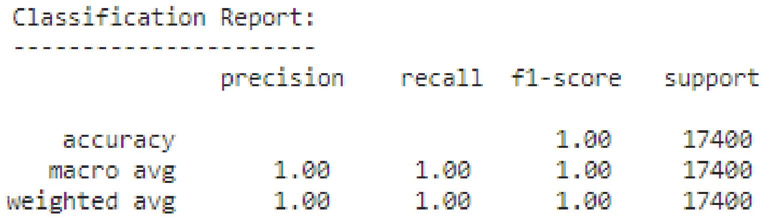


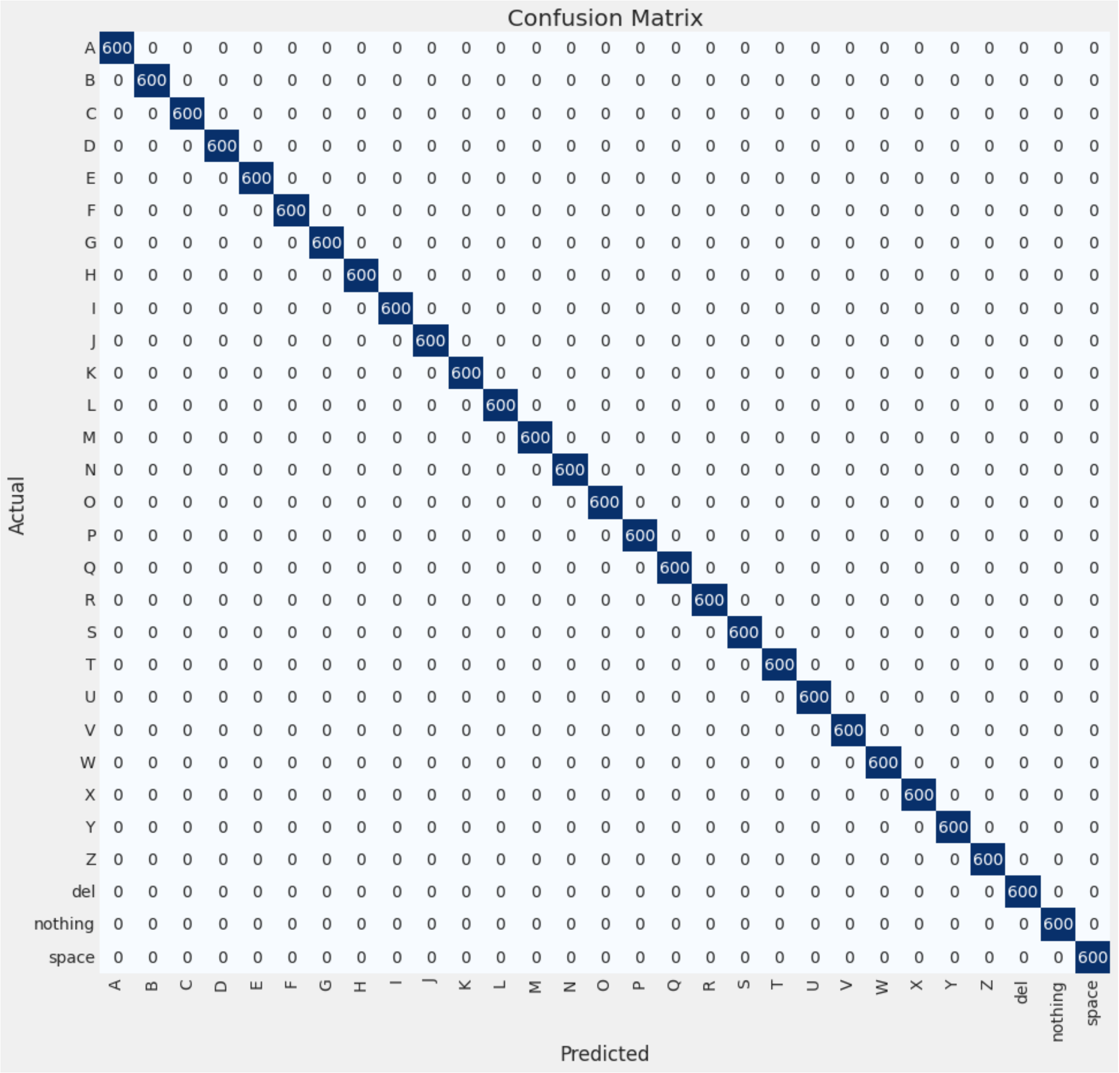
Figure 5.17: Training and validation performance of proposed model on dataset [65].



```
 Classification Report:
 ----------------------
                precision    recall  f1-score   support

    accuracy                           1.00     17400
   macro avg       1.00      1.00      1.00     17400
weighted avg       1.00      1.00      1.00     17400
```

Figure 5.18: Classification report on the test set for dataset [65].



Confusion Matrix

Actual \ Predicted	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	del	nothing	space
A	600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0
Z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0
del	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0
nothing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0
space	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600

Figure 5.19: Confusion matrix for the test set on dataset [65].

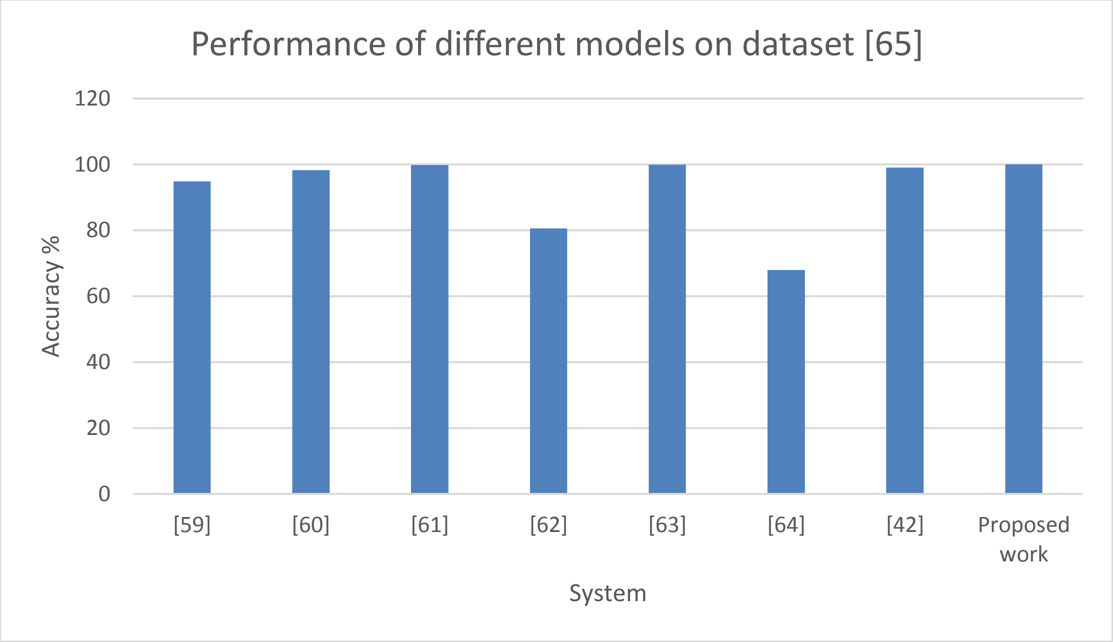


Figure 5.20: Performance of different models on dataset [65].

E) Performance on ASL dataset [54] with 52,000 images

On this dataset, we ran our model for 40 epochs, and got test accuracy between 99.89% and 99.93% on different runs of the model. Training and validation metrics for the course are as shown in figure 5.21. The classification report of the model on this dataset is shown in figure 5.22. Performance comparison of our proposed model with others existing systems on dataset [54] is shown in figure 5.23.

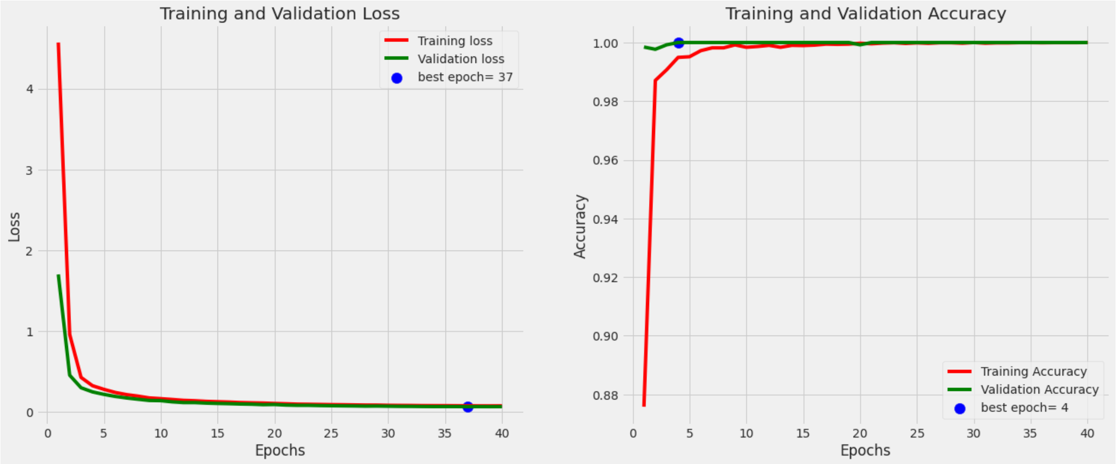
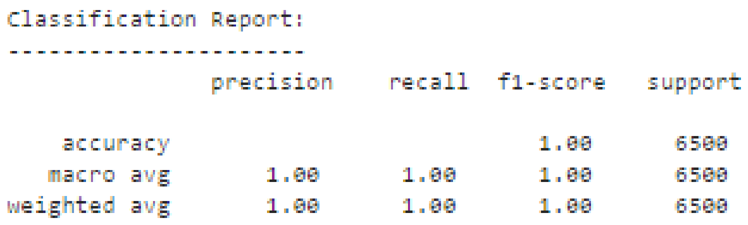


Figure 5.21: Training and validation performance of proposed model on dataset [54].



```
Classification Report:
----------------------
              precision    recall  f1-score   support

    accuracy                           1.00      6500
   macro avg       1.00      1.00      1.00      6500
weighted avg       1.00      1.00      1.00      6500
```

Figure 5.22: Classification report on the test set for dataset [54].

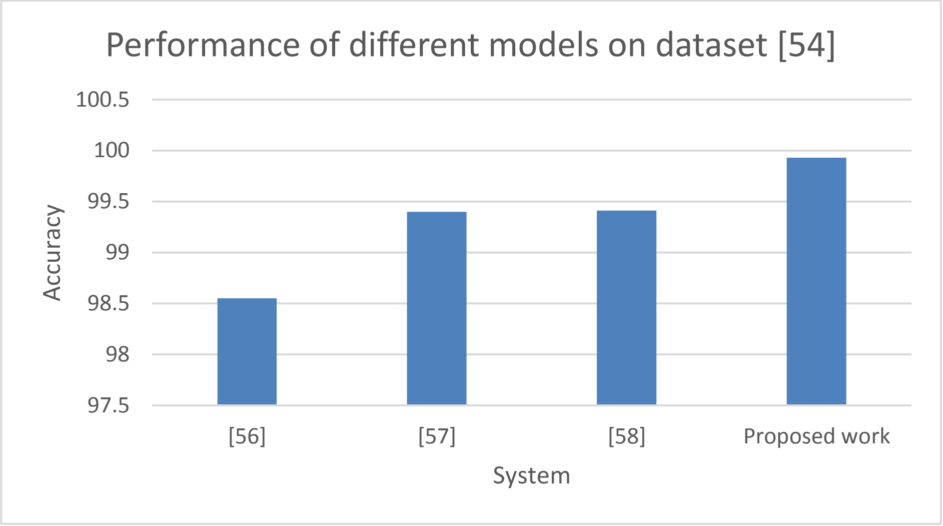


Figure 5.23: Performance of different models on dataset [54].

F) A Performance on ASL dataset [57] with 104,000 images

On this dataset, we ran our model for 40 epochs, and got test accuracy between 99.29% and 99.31% on different runs of the model. Training and validation metrics for the course are as shown in figure 5.24. The classification report of the model on this dataset is shown in figure 5.25, confusion matrix in figure 5.26. In [53], the authors have used CNN for classification, and obtained testing accuracy of 99.38% on this dataset.

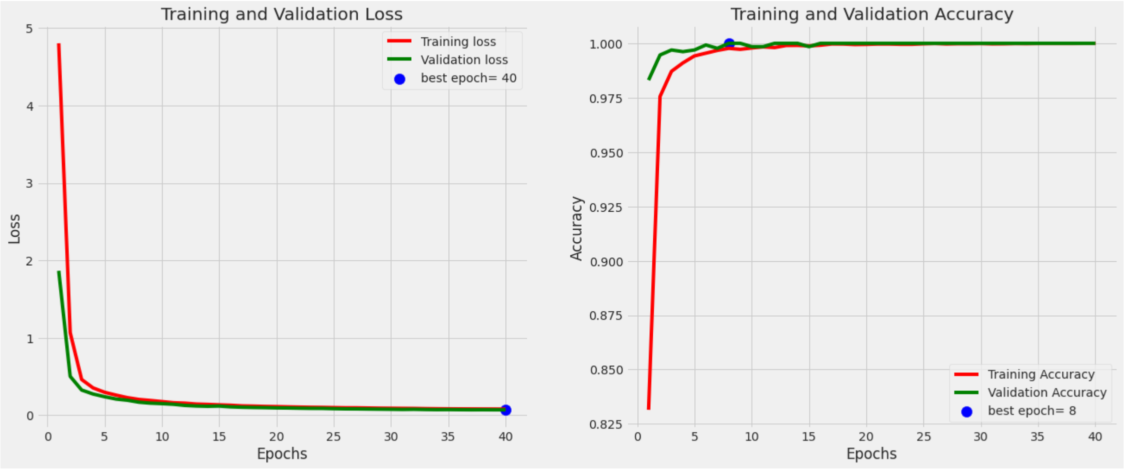
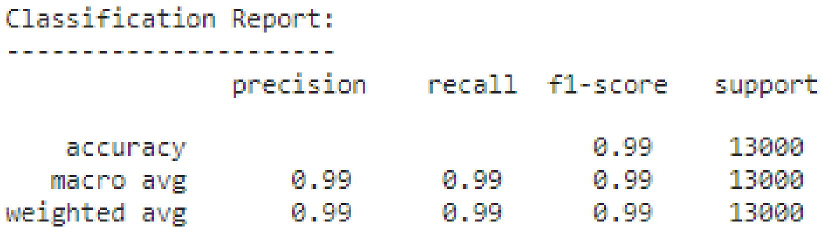


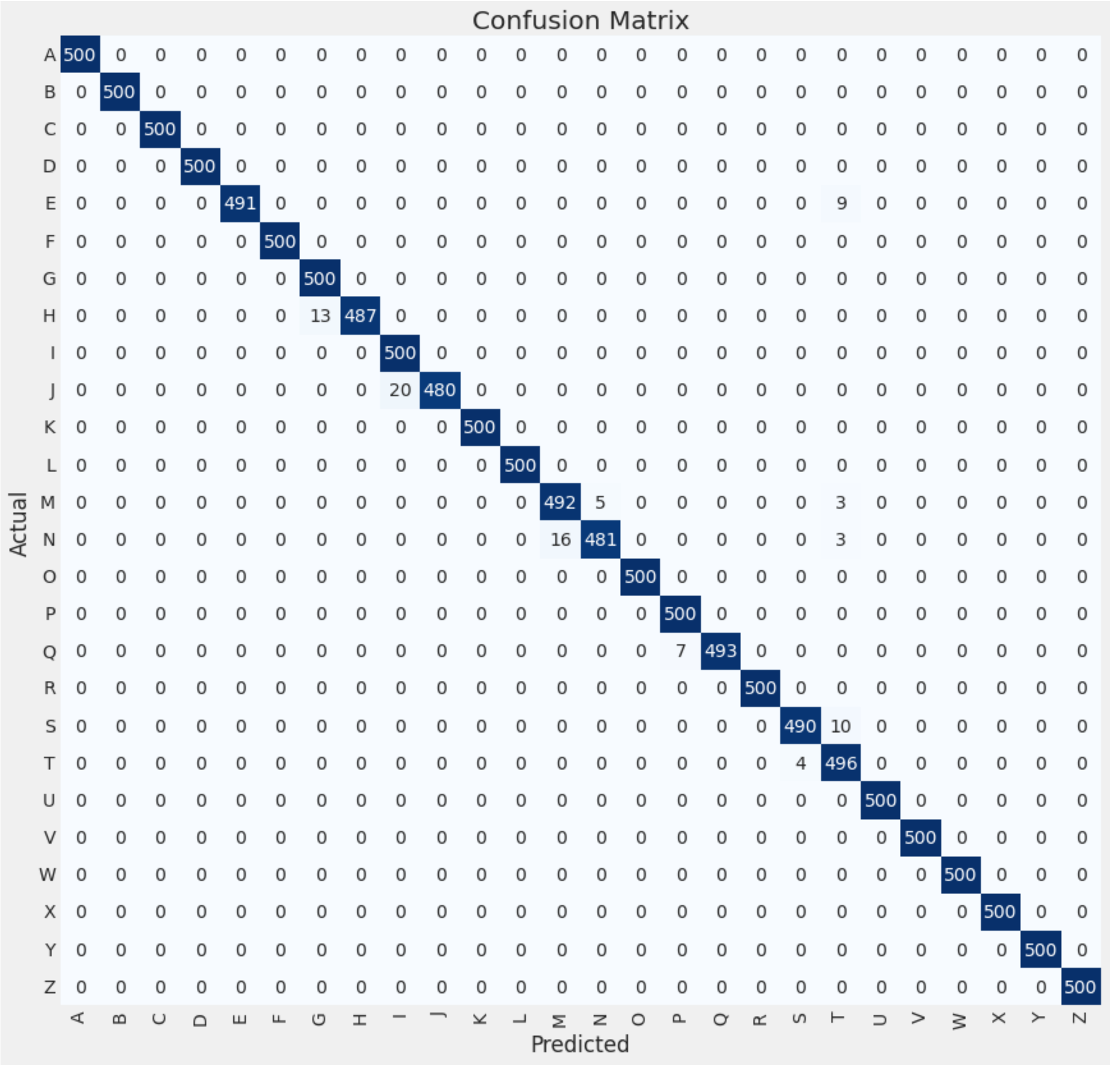
Figure 5.24: Training and validation performance of proposed model on dataset [57].



```
Classification Report:
----------------------
              precision    recall  f1-score   support

    accuracy                           0.99     13000
   macro avg       0.99      0.99      0.99     13000
weighted avg       0.99      0.99      0.99     13000
```

Figure 5.25: Classification report on the test set for dataset [57].



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	491	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0
F	0	0	0	0	0	500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	13	487	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	20	480	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	500	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	492	5	0	0	0	0	0	3	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	16	481	0	0	0	0	0	3	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	500	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	500	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	493	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	500	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	490	10	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	496	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	500	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	500	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	500	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	500	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	500	0
Z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	500

Figure 5.26: Confusion matrix for the test set on dataset [57].

G) Performance on ASL dataset [56] with 105,600 images

On this dataset, we ran our model for 40 epochs, and got test accuracy of 99.99%. Training and validation metrics for the course are as shown in fig. 5.27. The classification report of the model on this dataset is shown in figure 5.28. In [53], the authors have used CNN for classification, and obtained testing accuracy of 99.48% on this dataset.

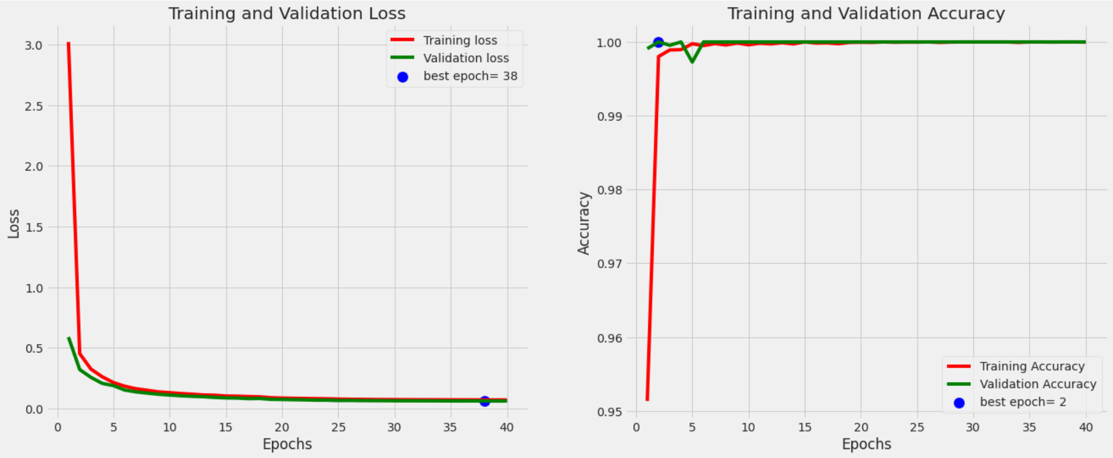
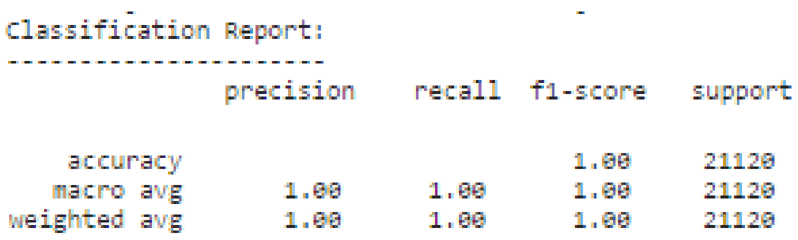


Figure 5.27: Training and validation performance of proposed model on dataset [56].



```
Classification Report:
----------------------
              precision    recall  f1-score   support

    accuracy                           1.00     21120
   macro avg       1.00      1.00      1.00     21120
weighted avg       1.00      1.00      1.00     21120
```

Figure 5.28: Classification report on the test set for dataset [56].

H) Performance on MNIST dataset.

On this dataset, we ran our model for 50 epochs, and got test accuracy between 99.94% and 99.98% on different runs of the model. Training and validation metrics for the course are as shown in figure 5.29. Performance comparison of our proposed model with others existing systems on the MNIST dataset is shown in figure 5.30.

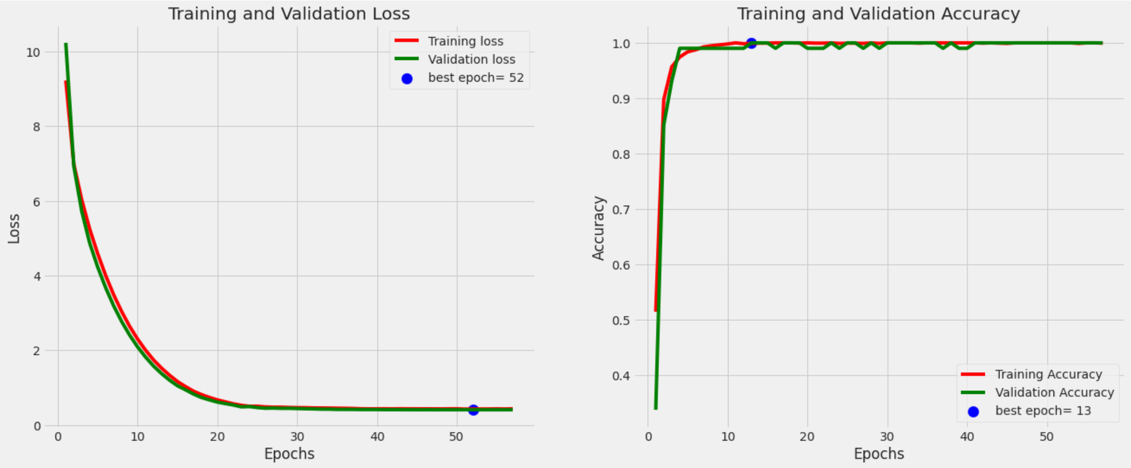


Figure 5.29: Training and validation performance of proposed model on MNIST dataset.

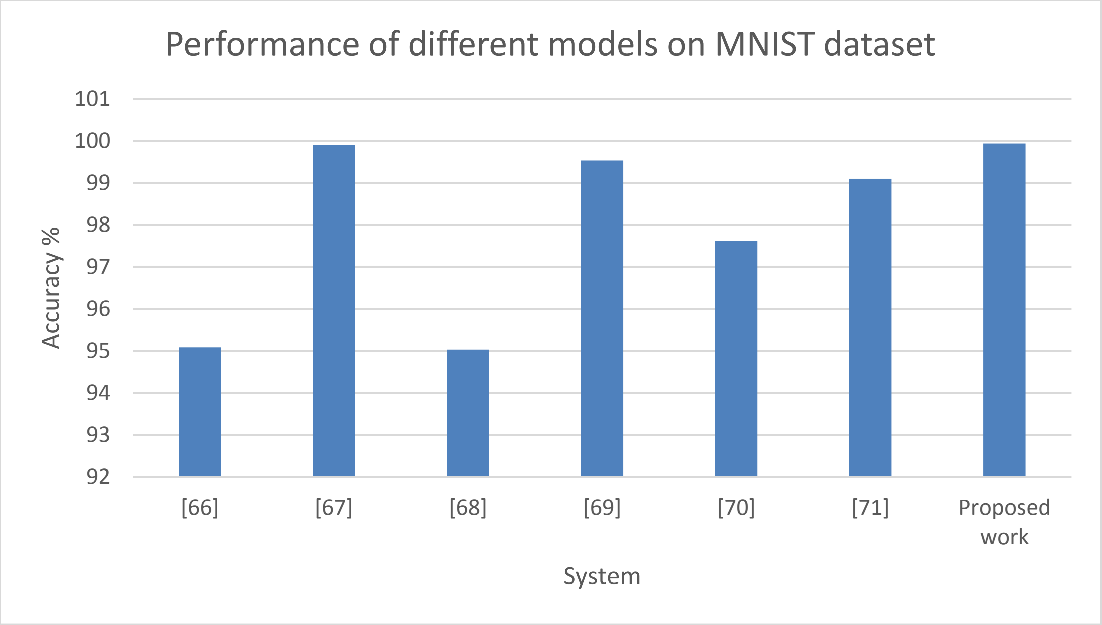


Figure 5.30: Performance of different models on the MNIST dataset.

Table 5.2: Performance analysis and comparison of our system on all the datasets

Dataset	Accuracy	Best Accuracy in literature	Training epochs	Training time
MNIST[72]	99.94 – 99.98	99.90	40	12 min, 24 sec
PSL[58]	99.34 – 99.60	90.00	40	26 min, 48 sec
ArSL[38]	98.66 – 98.93	99.05	40	44 min, 33 sec
Massey[50]	99.60 – 99.63	99.63	50	24 min, 8 sec
[65]	99.99	99.89	50	59 min, 11 sec
[54]	99.89 – 99.93	99.41	40	33 min, 12 sec
[56]	99.99	99.48	40	59 min, 54 sec
[57]	99.29 – 99.31	99.38	40	36 min, 31 sec

For the real-time system, on an 'AMD A8-7410 APU with Radeon Graphics Quad Core 2.2 GHz system with 4 GB DDR-3 RAM' the processing delays were found to be within an acceptable limit of 0.01 seconds, which is sufficient for a real-time system. It's worth mentioning that only one CPU core is active during the runtime, with about 30 percent load.

CHAPTER 6

CONCLUSION AND FUTURESCOPE

This paper presents a SLR system for vision-based .A deep learning-based Effi-CNN model using transfer learning with EfficientNetB2 as the base model is used for vision-based static gesture classification. The suggested vision-based system does not require any external equipment, such as Kinect sensors, making it more practical to use. This study's ability to recognise unique SL signs with good recognition performance over state-of-the-art methodologies is a remarkable addition. This work's performance is evaluated using eight publicly available SL datasets, compared to other state-of-the-art methodologies, the Effi-CNN model either outperforms or achieves competitive results.

By evaluating the probability of a class as mostly likely, the model attempts to categorise the input image based on patterns learned during training. We must continue to build future work that incorporates all punctuation marks and all sorts of joint letter representation in order to appropriately translate hand-sign-spelled words and sentences into written language. This research, on the other hand, will serve as a starting point for SLR researchers. The system can be used to translate sign language and non-sign language communication, as well as for human–computer–machine interface and robot control. Our current study focuses on static gesture detection; however, we hope to recognise dynamic sign language movements using video frames, which will be a difficult effort.

Furthermore, our current strategy is based on a clearly defined zone of interest. If object detection was applied to locate the hand region, which is required for practical applications, there would be no requirement for a predetermined ROI. For a better experience, if the system's text output can be translated to voice utilising a text-to-speech tool like Google's Cloud Text-To-Speech.

REFERENCES

[1]. National Institute on Aging. n.d. Hearing Loss: A Common Problem for Older Adults. [online] Available at: <https://www.nia.nih.gov/health/hearing-loss-common-problem-older-adults> [Accessed 24 May 2022].

[2]. Disabled World. (2022, April 7). Deaf Communication: Sign Language and Assistive Hearing Devices. Disabled World. Retrieved May 24, 2022 from www.disabled-world.com/disability/types/hearing/communication/

[3]. Ai-Media creating accessibility, one word at a time. 2021. Sign Language Alphabets From Around The World - ASL - Ai-Media. [online] Available at: <https://www.ai-media.tv/ai-media-blog/sign-language-alphabets-from-around-the-world/> [Accessed 24 May 2022].

[4]. En.wikipedia.org. Fingerspelling - Wikipedia. [online] Available at: <https://en.wikipedia.org/wiki/Fingerspelling> [Accessed 28 April 2022].

[5]. Islrtc.nic.in. Poster of the Manual Alphabet in ISL | Indian Sign Language Research and Training Center (ISLRTC), Government of India. [online] Available at: <http://www.islrtc.nic.in/poster-manual-alphabet-isl> [Accessed 26 April 2022].

[6]. Lifeprint.com. n.d. American Sign Language (ASL). [online] Available at: <https://www.lifeprint.com/asl101/pages-layout/grammar.htm> [Accessed 24 May 2022].

[7]. Who.int. 2021. Deafness and hearing loss. [online] Available at: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss#:~:text=Over%205%25%20of%20the%20world's.will%20have%20disabling%20hearing%20loss> [Accessed 24 May 2022]..

[8]. Lackner, A., 2017. Comprehending the Complexities of Sign Language. [online] The Wire. Available at: <https://thewire.in/culture/amazing-complexity-sign-language> [Accessed 24 May 2022].

[9]. B. Li, J. Yang, Y. Yang, C. Li and Y. Zhang, "Sign Language/Gesture Recognition Based on Cumulative Distribution Density Features Using UWB Radar," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-13, 2021, Art no. 2511113, doi: 10.1109/TIM.2021.3092072.

[10]. Yang, Y., Li, J., Li, B. et al. MDHandNet: a lightweight deep neural network for hand gesture/sign language recognition based on micro-doppler images. World Wide Web (2022). https://doi.org/10.1007/s11280-021-00985-1

[11]. P. Wang, Y. Zhou, Z. Li, S. Huang and D. Zhang, "Neural Decoding of Chinese Sign Language With Machine Learning for Brain–Computer Interfaces," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 29, pp. 2721-2732, 2021, doi: 10.1109/TNSRE.2021.3137340.

[12]. E. A. Malaia, S. C. Borneman, J. Krebs and R. B. Wilbur, "Low-Frequency Entrainment to Visual Motion Underlies Sign Language Comprehension," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 29, pp. 2456-2463, 2021, doi: 10.1109/TNSRE.2021.3127724.

[13]. D. AlQattan and F. Sepulveda, "Towards sign language recognition using EEG-based motor imagery brain computer interface," 2017 5th International Winter Conference on Brain-Computer Interface (BCI), 2017, pp. 5-8, doi: 10.1109/IWW-BCI.2017.7858143.

[14]. Emily Kubicek, Lorna C. Quandt, Sensorimotor system engagement during ASL sign perception: An EEG study in deaf signers and hearing non-signers, Cortex, Volume 119, 2019, Pages 457-469, ISSN 0010-9452, https://doi.org/10.1016/j.cortex.2019.07.016.

[15]. Lei Zhang, Yixiang Zhang, and Xiaolong Zheng. 2020. WiSign: Ubiquitous American Sign Language Recognition Using Commercial Wi-Fi Devices. ACM Trans. Intell. Syst. Technol. 11, 3, Article 31 (June 2020), 24 pages. https://doi.org/10.1145/3377553

[16]. Jiacheng Shang and Jie Wu. 2017. A Robust Sign Language Recognition System with Multiple Wi-Fi Devices. In Proceedings of the Workshop on Mobility in the Evolving Internet Architecture (MobiArch '17). Association for Computing Machinery, New York, NY, USA, 19–24. https://doi.org/10.1145/3097620.3097624

[17]. J. Shang and J. Wu, "A Robust Sign Language Recognition System with Sparsely Labeled Instances Using Wi-Fi Signals," 2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), 2017, pp. 99-107, doi: 10.1109/MASS.2017.41.

[18]. Hasmath Farhana Thariq Ahmed, Hafisoh Ahmad, Kulasekharan Narasingamurthi, Houda Harkat, Swee King Phang, DF-WiSLR: Device-Free Wi-Fi-based Sign Language Recognition, Pervasive and Mobile Computing, Volume 69, 2020, 101289, ISSN 1574-1192, https://doi.org/10.1016/j.pmcj.2020.101289.

[19]. S. Z. Gurbuz et al., "American Sign Language Recognition Using RF Sensing," in IEEE Sensors Journal, vol. 21, no. 3, pp. 3763-3775, 1 Feb.1, 2021, doi: 10.1109/JSEN.2020.3022376.

[20]. K. Amrutha and P. Prabu, "ML Based Sign Language Recognition System," 2021 International Conference on Innovative Trends in Information Technology (ICITIIT), 2021, pp. 1-6, doi: 10.1109/ICITIIT51526.2021.9399594.

[21]. Liang-Guo Zhang, Yiqiang Chen, Gaolin Fang, Xilin Chen, and Wen Gao. 2004. A vision-based sign language recognition system using tied-mixture density HMM. In Proceedings of the 6th international conference on Multimodal interfaces (ICMI '04). Association for Computing Machinery, New York, NY, USA, 198–204. https://doi.org/10.1145/1027933.1027967

[22]. Ali Karami, Bahman Zanj, Azadeh Kiani Sarkaleh, Persian sign language (PSL) recognition using wavelet transform and neural networks, Expert Systems with Applications, Volume 38, Issue 3, 2011, Pages 2661-2667, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2010.08.056.

[23]. C. -C. Wang et al., "Real-Time Block-Based Embedded CNN for Gesture Classification on an FPGA," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 68, no. 10, pp. 4182-4193, Oct. 2021, doi: 10.1109/TCSI.2021.3100109.

[24]. O. M. Sincan and H. Y. Keles, "AUTSL: A Large Scale Multi-Modal Turkish Sign Language Dataset and Baseline Methods," in IEEE Access, vol. 8, pp. 181340-181355, 2020, doi: 10.1109/ACCESS.2020.3028072.

[25]. G. Plouffe and A. Cretu, "Static and Dynamic Hand Gesture Recognition in Depth Data Using Dynamic Time Warping," in IEEE Transactions on Instrumentation and Measurement, vol. 65, no. 2, pp. 305-316, Feb. 2016, doi: 10.1109/TIM.2015.2498560.

[26]. C. Sun, T. Zhang, B. Bao, C. Xu and T. Mei, "Discriminative Exemplar Coding for Sign Language Recognition With Kinect," in IEEE Transactions on Cybernetics, vol. 43, no. 5, pp. 1418-1428, Oct. 2013, doi: 10.1109/TCYB.2013.2265337.

[27]. D. Avola, M. Bernardi, L. Cinque, G. L. Foresti and C. Massaroni, "Exploiting Recurrent Neural Networks and Leap Motion Controller for the Recognition of Sign Language and Semaphoric Hand Gestures," in IEEE Transactions on Multimedia, vol. 21, no. 1, pp. 234-245, Jan. 2019, doi: 10.1109/TMM.2018.2856094.

[28]. E. Rho, K. Chan, E. J. Varoy and N. Giacaman, "An Experiential Learning Approach to Learning Manual Communication Through a Virtual Reality Environment," in IEEE Transactions on Learning Technologies, vol. 13, no. 3, pp. 477-490, 1 July-Sept. 2020, doi: 10.1109/TLT.2020.2988523.

[29]. Abeer Alnuaim, Mohammed Zakariah, Wesam Atef Hatamleh, Hussam Tarazi, Vikas Tripathi, Enoch Tetteh Amoatey, "Human-Computer Interaction with Hand Gesture Recognition Using ResNet and MobileNet", Computational Intelligence and Neuroscience, vol. 2022, Article ID 8777355, 16 pages, 2022. https://doi.org/10.1155/2022/8777355

[30]. Alani, Ali A.; Cosma, Georgina (2021): ArSL-CNN: a convolutional neural network for Arabic sign language gesture recognition. Loughborough University. Journal contribution. https://hdl.handle.net/2134/16878787.v1 https://doi.org/10.11591/ijeecs.v22.i2.pp1096-1107

[31]. Latif G, Mohammad N, AlKhalaf R, AlKhalaf R, Alghazo J, Khan M. An automatic Arabic sign language recognition system based on deep CNN: an assistive system for the deaf and hard of hearing. International Journal of Computing and Digital Systems. 2020 Jul 1;9(4):715-http://dx.doi.org/10.12785/ijcds/09041824.

[32]. Alshomrani, Shroog, i in. „Arabic and American Sign Languages Alphabet Recognition by Convolutional Neural Network”. Advances in Science and Technology. Research Journal, t. 15, nr 4, 4, Stowarzyszenie Inżynierów i Techników Mechaników Polskich, 2021, s. 136–48.

[33]. Mohammed Zakariah, Yousef Ajmi Alotaibi, Deepika Koundal, Yanhui Guo, Mohammad Mamun Elahi, "Sign Language Recognition for Arabic Alphabets Using Transfer Learning Technique", Computational Intelligence and Neuroscience, vol. 2022, Article ID 4567989, 15 pages, 2022. https://doi.org/10.1155/2022/4567989

[34]. M. F. Nurnoby, E. -S. M. El-Alfy and H. Luqman, "Evaluation of CNN Models with Transfer Learning for Recognition of Sign Language Alphabets with Complex Background," 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT), 2020, pp. 1-6, doi: 10.1109/3ICT51146.2020.9311989.

[35]. Duwairi RM, Halloush ZA. Automatic recognition of Arabic alphabets sign language using deep learning. International Journal of Electrical & Computer Engineering (2088-8708). 2022 Jun 1;12(3).

[36]. Luqman, H., El-Alfy, ES.M. & BinMakhashen, G.M. Joint space representation and recognition of sign language fingerspelling using Gabor filter and convolutional neural network. Multimed Tools Appl 80, 10213–10234 (2021). https://doi.org/10.1007/s11042-020-09994-0

[37]. Saleh, Y., & Issa, G. F. (2020). Arabic Sign Language Recognition through Deep Neural Networks Fine-Tuning. International Journal of Online and Biomedical Engineering (iJOE), 16(05), pp. 71–83. https://doi.org/10.3991/ijoe.v16i05.13087

[38]. Latif, G., Alghazo, J., Mohammad, N., AlKhalaf, R. and AlKhalaf, R., 2018. Arabic Alphabets Sign Language Dataset (ArASL). [online] Mendeley Data. Available at: <https://data.mendeley.com/datasets/y7pckrw6z2/1> [Accessed 24 May 2022].

[39]. Ghazanfar Latif, Nazeeruddin Mohammad, Jaafar Alghazo, Roaa AlKhalaf, Rawan AlKhalaf, ArASL: Arabic Alphabets Sign Language Dataset, Data in Brief, Volume 23, 2019, 103777, ISSN 2352-3409, https://doi.org/10.1016/j.dib.2019.103777.

[40]. Masood, Sarfaraz & Thuwal, Harish & Srivastava, Adhyan. (2018). American Sign Language Character Recognition Using Convolution Neural Network. 10.1007/978-981-10-5547-8_42.

[41]. Garcia, B. and Viesca, S.A., 2016. Real-time American sign language recognition with convolutional neural networks. Convolutional Neural Networks for Visual Recognition, 2, pp.225-232.

[42]. Rajan RG. Transfer-learning analysis for sign language classification models. Turkish Journal of Computer and Mathematics Education (TURCOMAT). 2021 Apr 24;12(9):1423-33.

[43]. Kania, Kacper & Markowska-Kaczmar, Urszula. (2018). American Sign Language Fingerspelling Recognition Using Wide Residual Networks. 10.1007/978-3-319-91253-0_10.

[44]. Sérgio F. Chevtchenko, Rafaella F. Vale, Valmir Macario, Filipe R. Cordeiro, A convolutional neural network with feature fusion for real-time hand posture recognition, Applied Soft Computing, Volume 73, 2018, Pages 748-766, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2018.09.010.

[45]. Shin J, Matsuoka A, Hasan MAM, Srizon AY. American Sign Language Alphabet Recognition by Extracting Feature from Hand Pose Estimation. Sensors. 2021; 21(17):5856. https://doi.org/10.3390/s21175856

[46]. Rastgoo, R.; Kiani, K.; Escalera, S. Multi-Modal Deep Hand Sign Language Recognition in Still Images Using Restricted Boltzmann Machine. Entropy 2018, 20, 809. https://doi.org/10.3390/e20110809

[47]. M. Taskiran, M. Killioglu and N. Kahraman, "A Real-Time System for Recognition of American Sign Language by using Deep Learning," 2018 41st International Conference on Telecommunications and Signal Processing (TSP), 2018, pp. 1-5, doi: 10.1109/TSP.2018.8441304.

[48]. H. B. D. Nguyen and H. N. Do, "Deep Learning for American Sign Language Fingerspelling Recognition System," 2019 26th International Conference on Telecommunications (ICT), 2019, pp. 314-318, doi: 10.1109/ICT.2019.8798856.

[49]. Salian, Shashank & Dokare, Indu & Serai, Dhiren & Suresh, Aditya & Ganorkar, Pranav. (2017). Proposed system for sign language recognition. 058-062. 10.1109/ICCPEIC.2017.8290339

[50]. Massey.ac.nz. 2012. Massey University. [online] Available at: <https://www.massey.ac.nz/~albarcza/gesture_dataset2012.html> [Accessed 24 May 2022].

[51]. Barczak, A.L.C., Reyes, N.H., Abastillas, M., Piccio, A. and Susnjak, T., 2011. A new 2D static hand gesture colour image dataset for ASL gestures. Research Letters in the Information and Mathematical Sciences, 2011, Vol. 15, pp. 12–20. Available online at http://iims.massey.ac.nz/research/letters/

[52]. J. R. V. Jeny, A. Anjana, K. Monica, T. Sumanth and A. Mamatha, "Hand Gesture Recognition for Sign Language Using Convolutional Neural Network," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), 2021, pp. 1713-1721, doi: 10.1109/ICOEI51242.2021.9453072.1713-1721). IEEE.

[53]. Ahmed KASAPBAŞI, Ahmed Eltayeb AHMED ELBUSHRA, Omar AL-HARDANEE, Arif YILMAZ, DeepASLR: A CNN based human computer interface for American Sign Language recognition for hearing-impaired individuals, Computer Methods and Programs in Biomedicine Update, Volume 2, 2022, 100048, ISSN 2666-9900, https://doi.org/10.1016/j.cmpbup.2021.100048.2022 Jan 10:100048.

[54]. J. R. V. Jeny, A. Anjana, K. Monica, T. Sumanth and A. Mamatha, "Hand Gesture Recognition for Sign Language Using Convolutional Neural Network," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), 2021, pp. 1713-1721, doi: 10.1109/ICOEI51242.2021.9453072.1713-1721). IEEE.

[55]. GitHub. 2018. GitHub - rrupeshh/Simple-Sign-Language-Detector: Simple Sign Language Detector. [online] Available at: <https://github.com/rrupeshh/Simple-Sign-Language-Detector> [Accessed 24 May 2022].

[56]. GitHub. 2018. GitHub - EvilPort2/Sign-Language: A very simple CNN project.. [online] Available at: <https://github.com/EvilPort2/Sign-Language> [Accessed 24 May 2022].

[57]. GitHub. 2020. GitHub - Ahmed-KASAPBASI/Success_Team_ASL: Project of Deep Learning (ASL). [online] Available at: <https://github.com/Ahmed-KASAPBASI/Success_Team_ASL> [Accessed 24 May 2022].

[58]. Ali Imran, Abdul Razzaq, Irfan Ahmad Baig, Aamir Hussain, Sharaiz Shahid, Tausif-ur Rehman, Dataset of Pakistan Sign Language and Automatic Recognition of Hand Configuration of Urdu Alphabet through Machine Learning, Data in Brief, Volume 36, 2021, 107021, ISSN 2352-3409, https://doi.org/10.1016/j.dib.2021.107021.

[59]. Upadhyay, S., Sharma, R.K. and Rana, P.S., 2020. Sign Language Recognition with Visual Attention. EasyChair Preprint, (2312).

[60]. Ashish Sharma, Anmol Mittal, Savitoj Singh, Vasudev Awatramani, Hand Gesture Recognition using Image Processing and Feature Extraction Techniques, Procedia Computer Science, Volume 173, 2020, Pages 181-190, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.06.022.

[61]. Grandhi, C., Liu, S. and Rahoria, D., 2021. American Sign Language Recognition using Deep Learning. In International Conference 2021.

[62]. S. N. Reddy Karna, J. S. Kode, S. Nadipalli and S. Yadav, "American Sign Language Static Gesture Recognition using Deep Learning and Computer Vision," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), 2021, pp. 1432-1437, doi: 10.1109/ICOSEC51865.2021.9591845.

[63]. Elsayed N, ElSayed Z, Maida AS. Vision-Based American Sign Language Classification Approach via Deep Learning. arXiv preprint arXiv:2204.04235. 2022 Apr 8.

[64]. [48] Rakshit A. Smart learners: Choosing What to Learn Using Bimodal Distribution Removal. Availabale at: http://users.cecs.anu.edu.au/~Tom.Gedeon/conf/ABCs2018/paper/ABCs2018_paper_10.pdf

[65]. Kaggle.com. 2018. ASL Alphabet. [online] Available at: <https://www.kaggle.com/grassknoted/asl-alphabet> [Accessed 24 May 2022].

[66]. M. Bilgin and K. Mutludoğan, "American Sign Language Character Recognition with Capsule Networks," 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2019, pp. 1-6, doi: 10.1109/ISMSIT.2019.8932829.

[67]. G. M. B. Makhashen, H. A. Luqman and E. M. El-Alfy, "Using Gabor filter bank with downsampling and SVM for visual sign language alphabet recognition," 2nd Smart Cities Symposium (SCS 2019), 2019, pp. 1-6, doi: 10.1049/cp.2019.0188.

[68]. Rathi, D., 2018. Optimization of transfer learning for sign language recognition targeting mobile platform. arXiv preprint arXiv:1805.06618.

[69]. Fregoso, Jonathan, Claudia I. Gonzalez, and Gabriela E. Martinez. 2021. "Optimization of Convolutional Neural Networks Architectures Using PSO for Sign Language Recognition" Axioms 10, no. 3: 139. https://doi.org/10.3390/axioms10030139

[70]. M. M. Hasan, A. Y. Srizon, A. Sayeed and M. A. M. Hasan, "Classification of Sign Language Characters by Applying a Deep Convolutional Neural Network," 2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT), 2020, pp. 434-438, doi: 10.1109/ICAICT51780.2020.9333456.

[71]. Beniwal, R., Nag, B., Saraswat, A., Gulati, P. (2022). Static Hand Sign Recognition Using Wavelet Transform and Convolutional Neural Network. In: Luhach, A.K., Poonia, R.C., Gao, XZ., Singh Jat, D. (eds) Second International Conference on Sustainable Technologies for Computational Intelligence. Advances in Intelligent Systems and Computing, vol 1235. Springer, Singapore. https://doi.org/10.1007/978-981-16-4641-6_13

[72]. Kaggle.com. 2020. Sign Language MNIST | Kaggle. [online] Available at: <https://www.kaggle.com/c/sign-language-mnist> [Accessed 24 May 2022].

[73]. TensorFlow. n.d. Transfer learning and fine-tuning | TensorFlow Core. [online] Available at: <https://www.tensorflow.org/tutorials/images/transfer_learning> [Accessed 24 May 2022].

[74]. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning 2019 May 24 (pp. 6105-6114). PMLR.

[75]. Scince Direct. n.d. Deep convolutional neural network–based image classification. [online] Available at :<https://www.sciencedirect.com/topics/engineering/convolutional-neural-network> [Accessed 24 May 2022].

[76]. En.wikipedia.org. 2022. Sign language - Wikipedia. [online] Available at: <https://en.wikipedia.org/wiki/Sign_language> [Accessed 24 May 2022].

[77]. LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. "Gradient- based learning applied to document recognition." Proceedings of the IEEE 86, no. 11 (1998): 2278-2324

[78]. OpenCV. n.d. About - OpenCV. [online] Available at: <https://opencv.org/about/> [Accessed 24 May 2022].

[79]. Chollet, F. & others, 2015. Keras. Available at: https://github.com/fchollet/keras. Software available from https://keras.io

[80]. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M. and Ghemawat, S., 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467. Software available from https://tensorflow.org.

[81]. Ankita Wadhawan, Parteek Kumar, “Sign language recognition systems: a decade systematic literature review,” Archives of computational methods in Engineering, 2019.

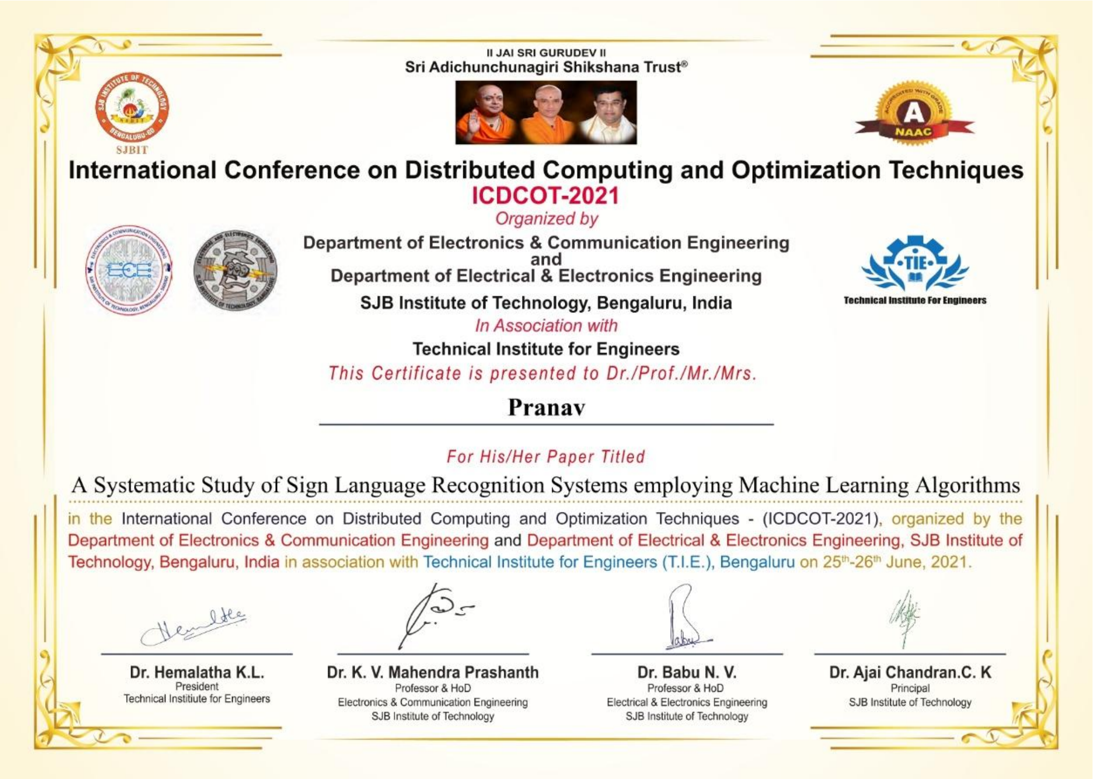
[82]. Google Developers. 2020. Multi-Class Neural Networks: Softmax | Machine Learning Crash Course | Google Developers. [online] Available at: <https://developers.google.com/machine-learning/crash-course/multi-class-neural-networks/softmax> [Accessed 24 May 2022].

[83]. Agarwal, V., 2020. Complete Architectural Details of all EfficientNet Models. [online] Medium. Available at: <https://towardsdatascience.com/complete-architectural-details-of-all-efficientnet-models-5fd5b736142> [Accessed 24 May 2022].

LIST OF PUBLICATIONS

1 Pranav, Rahul Katarya, "A Systematic Study of Sign Language Recognition Systems employing Machine Learning Algorithms". Accepted and presented at the **International Conference on Distributed Computing and Optimization Techniques (ICDCOT 2021)**.

Indexed by Scopus.
Paper Id: ICDCOT100



II JAI SRI GURUDEV II
Sri Adichunchunagiri Shikshana Trust®

SJBIT

International Conference on Distributed Computing and Optimization Techniques
ICDCOT-2021

Organized by

Department of Electronics & Communication Engineering
and
Department of Electrical & Electronics Engineering
SJB Institute of Technology, Bengaluru, India

In Association with

Technical Institute for Engineers

Technical Institute For Engineers

This Certificate is presented to Dr./Prof./Mr./Mrs.

Pranav

For His/Her Paper Titled

A Systematic Study of Sign Language Recognition Systems employing Machine Learning Algorithms

in the International Conference on Distributed Computing and Optimization Techniques - (ICDCOT-2021), organized by the Department of Electronics & Communication Engineering and Department of Electrical & Electronics Engineering, SJB Institute of Technology, Bengaluru, India in association with Technical Institute for Engineers (T.I.E.), Bengaluru on 25th-26th June, 2021.

Dr. Hemalatha K.L.
President
Technical Instiitute for Engineers

Dr. K. V. Mahendra Prashanth
Professor & HoD
Electronics & Communication Engineering
SJB Institute of Technology

Dr. Babu N. V.
Professor & HoD
Electrical & Electronics Engineering
SJB Institute of Technology

Dr. Ajai Chandran.C. K
Principal
SJB Institute of Technology

2 Pranav, Rahul Katarya, “Optimal Sign language recogniton employing multi-layer CNN”. Accepted for presentation and publication at the **4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) 2022.**

Indexed by Scopus.
Paper Id: 815

Abstract- We generally convey our ideas, thoughts, and facts through vocal communication. However, not everyone is gifted with the ability to express oneself. verbally. The deaf and mute populations have a difficult time expressing their thoughts and ideas to others; Sign Language (SL) is their most expressive mode of communication, but the majority of the general population is illiterate in SL; as a result, the mute and deaf have difficulty communicating with the rest of the world. A device that can reliably translate SL gestures to voice and vice versa in real-time is required to overcome this communication barrier. The current solutions are not real-time, have poor recognition accuracy, and need static environmental conditions. Some systems might need additional hardware components, such as expensive sensors, which boost the cost. The current methods allowing silent people to communicate their thoughts to others may be divided into two categories: systems based on computer vision and systems based on electrical sensors, each with its own set of benefits and drawbacks. In comparison to previous state-of-the-art approaches, we offer a Convolution neural network (CNN) based SLR system with many layers that gives great accuracy of 99.89%. Overall, we believe the study will serve as road map for future research in the domain of SLR.

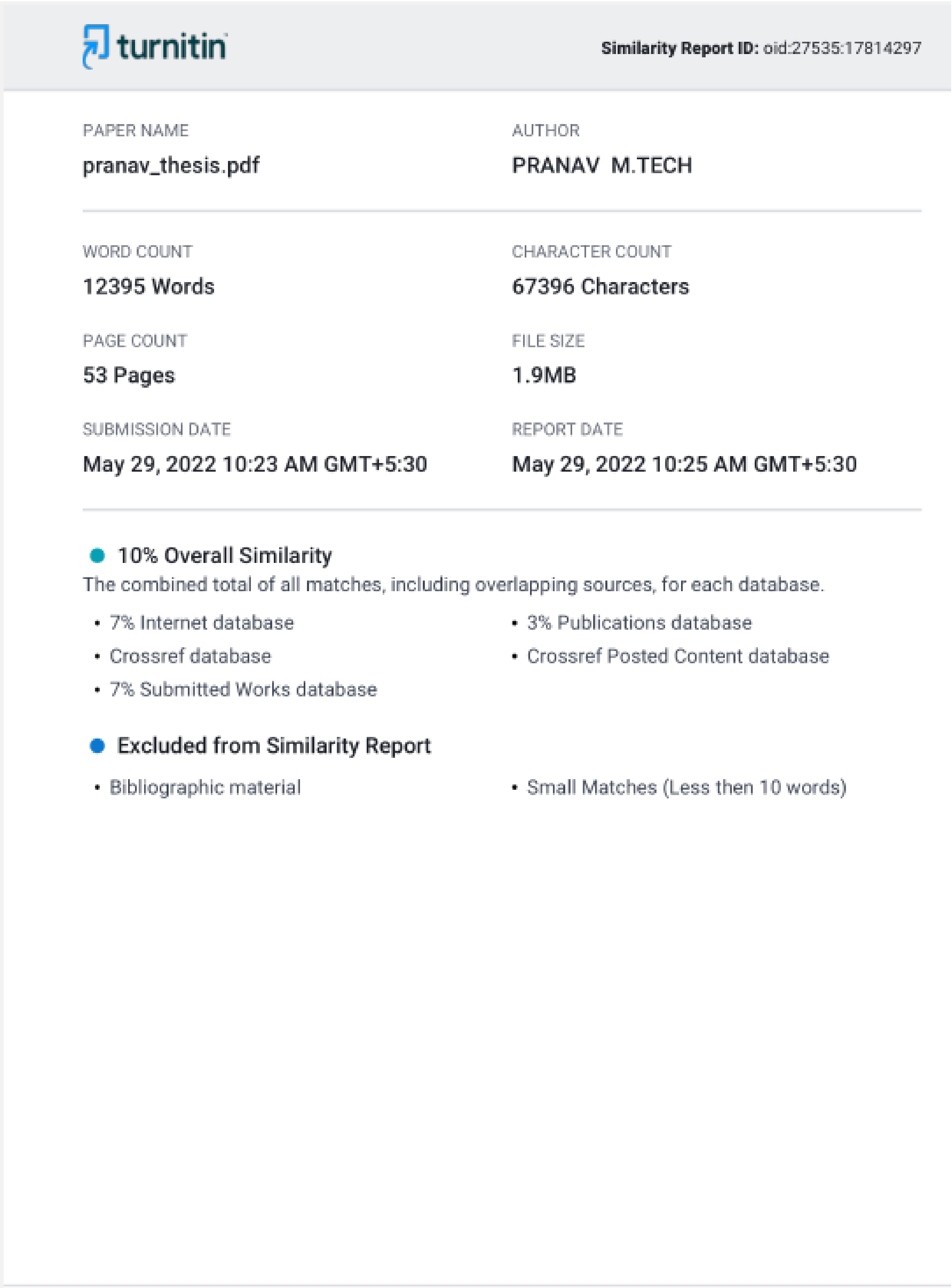
3 Pranav, Rahul Katarya, “Effi-CNN: real-time vision-based system for interpretation of sign language using CNN and Transfer Learning”. Submitted to ‘**Image and Vision Computing’ journal ISSN 0262-8856.**

Indexed by: Web of Science, Abstracts in Artificial Intelligence,
ACM Computing Reviews, Current Contents, Current Contents,
Ergonomics Abstracts, INSPEC, Research Alert, Science Citation
Index, CompuScience, Scopus.

Abstract- The deaf and mute population has a difficult time conveying their thoughts and ideas to others; sign language is their most expressive mode of communication, but the general public is callow of sign language, therefore the mute and deaf have difficulty communicating with others. A system that can correctly translate sign language motions to speech and vice versa in real-time is required to overcome this communication barrier.

Effi-CNN, a vision-based Sign Language Recognition (SLR) system, is proposed in this paper. To convert sign gesture photos into words, our system employs transfer learning with EfficientNetB2 as base model. We have also developed a system which coverts sign gestures to text in real time.

Our approach was evaluated on eight publically available datasets, including the Massey University gesture dataset, ArSL2018 dataset, MNIST-ASL dataset, and others. Comparing our results to state-of-the art algorithms, the experimental findings show that our technique is more successful. The results show that our Effi-CNN surpasses most of current existing solutions, and it has the ability to categorise a large number of gestures with a low rate of error.



turnitin

Similarity Report ID: oid:27535:17814297

PAPER NAME

pranav_thesis.pdf

AUTHOR

PRANAV M.TECH

WORD COUNT 12395 Words	CHARACTER COUNT 67396 Characters
PAGE COUNT 53 Pages	FILE SIZE 1.9MB
SUBMISSION DATE May 29, 2022 10:23 AM GMT+5:30	REPORT DATE May 29, 2022 10:25 AM GMT+5:30

● 10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 7% Internet database
- Crossref database
- 7% Submitted Works database
- 3% Publications database
- Crossref Posted Content database

● Excluded from Similarity Report

- Bibliographic material
- Small Matches (Less then 10 words)