IT204 MINI PROJECT REPORT
ON

Visual Hand Gesture Recognition with Convolutional Neural Network (CNN)

Submitted By : Nithin S [221IT085]

Under the guidance of

Prof. Palla Parasuram Yadav
Mr. Magar Pankaj Kumar Dada

DEPARTMENT OF INFORMATION TECHNOLOGY
NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA
SURATHKAL, MANGALORE – 575025

NOVEMBER 2023

ACKNOWLEDGMENT

I would like to thank **Prof. Palla Parasuram Yadav** and my mentor **Mr. Magar Pankaj Kumar Dada** for giving me the opportunity to undertake the project. I thank them for their immense guidance, and appreciate their timely engagement.

I would like to extend special gratitude to the the assistants and lab coordinators of the Department for providing us the infrastructural facilities necessary to sustain the project.

ABSTRACT

Simply put, communication is the act of transferring information from one location, person, or group to another. Every communication has at least one sender, one message, and one recipient. This may appear straightforward, but communication is a very complex subject.The transmission of a message from sender to recipient can be influenced by a wide range of factors, including our emotions, the cultural situation, the communication medium used and even our location. Because of the complexities, good communication skills are highly valued by employers all over the world.American Sign Language (ASL) is the fourth most studied second language atAmerican universities and one of the most widely used languages in the United States. ASL is primarily used by people who are deaf or hard of hearing in North America. In the United States and Canada, there are between 250,000 and 500,000 ASL users, the majority of whom use ASL as their primary language. Knowing ASL allows us to communicate with a diverse range of hearing, hard of hearing, and deaf people, including students in mainstream and deaf school or university programs, as well as deaf or hard of hearing residents and business people in our community. To address this situation, we present a project that will benefit such people. This project proposes a solution based on Machine Learning that will recognize hand gestures and translate them into text. A webcam will be used to capture the region of interest, i.e., recognize the indicated hand motions. The translated text will thus be provided based on the recognized gestures.

TABLE OF CONTENTS

4 Experimental Design

5 Results and Analysis

6 Conclusion and Future Work

7 References

LIST OF FIGURES

1. INTRODUCTION

Communication is the process of exchanging information, views and expressions between two or more persons, in both verbal and non-verbal manner.Hand gestures are the non-verbal method of communication used along with verbal communication. A more organized form of hand gesture communication is known as sign language. In this language each alphabet of the English vocabulary is assigned a sign. A physically disabled person like the deaf and the dumb use this language to communicate with each other. The idea of this project is to design a system that can understand the sign language accurately so that the less fortunate people may communicate with the outside world without the need of an interpreter.

Sign language is a more organized and defined way of communication in which every word or alphabet is assigned some gesture. In American Sign Language (ASL) each alphabet of English vocabulary, A-Z, is assigned a unique gesture. Sign language is mostly used by the deaf, dumb or people with any other kind of disabilities.

Our aim is to design a Human Computer Interface (HCI) system that can understand sign language accurately so that the signing people may communicate with the non-signing people without the need of an interpreter.

1.1 Sign Language

Deaf people around the world communicate using sign language as distinct from spoken language in their every day a visual language that uses a system of manual, facial and body movements as the means of communication. Sign language is not an universal language, and different sign languages are used in different countries, like the many spoken languages all over the world. Some countries such as Belgium, the UK, the USA or India may have more than one sign language. Hundreds of sign languages are in used around the world, for instance, Japanese Sign Language, British Sign Language (BSL), Spanish Sign Language, Turkish Sign Language.

Sign language is a visual language and consists of 3 major components:

Fingerspelling	**Word level sign vocabulary**	**Non-manual features**
Used to spell words letter by letter .	Used for the majority of communication.	Facial expressions and tongue, mouth and body position.

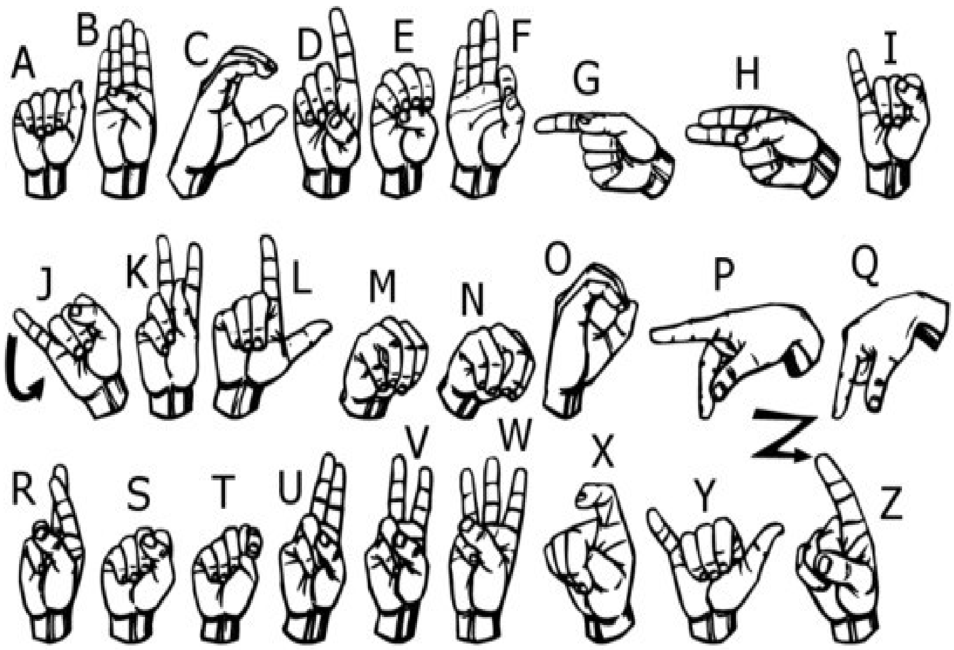

Fig 1: American Sign Language

2. LITERATURE SURVEY

In the recent years, there has been tremendous research on the hand sign language gesture recognition. The technology for gesture recognition is given below.

2.1 Vision-based

In visionbased methods computer camera is the input device for observing the information of hands or fingers. The Vision Based methods require only a camera, thus realizing a natural interaction between humans and computers without the use of any extra devices. These systems tend to complement biological vision by describing artificial vision systems that are implemented in software and/or hardware. This poses a challenging problem as these systems need to be background invariant, lighting insensitive, person and camera independent to achieve real time performance. Moreover, such systems must be optimized to meet the requirements, including accuracy and robustness.

The vision based hand gesture recognition system is shown in fig.

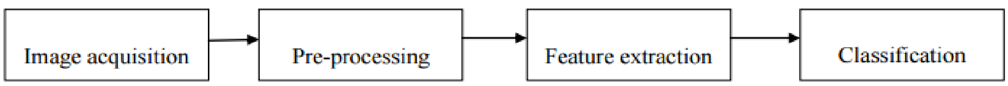

Fig 2: Block Diagram of vision based recognition system

Vision based analysis, is based on the way human beings perceive information about their surroundings, yet it is probably the most difficult to implement in a satisfactory way.

For Deaf signers who learned a different sign language before learning American Sign Language, qualities of their native language may show in their ASL production. Some examples of that varied production include fingerspelling towards the body, instead of away from it, and signing certain movement from bottom to top, instead of top to bottom.

Hearing people who learn American Sign Language also have noticeable differences in signing production. The most notable production difference of hearing people learning American Sign Language is their rhythm and arm posture.

2.1.1 M. Han, J. Chen, L. Li and Y. Chang, "Visual hand gesture recognition with convolution neural network," 2016

This paper reviews significant projects in the field beginning with finger-spelling hands such as “Ralph” (robotics), CyberGloves (virtual reality sensors to capture isolated and continuous signs), camera-based projects such as the CopyCat interactive American Sign Language game (computer vision), and sign recognition software (Hidden Markov Modeling and neural network systems). Avatars such as “Tessa” (Text and Sign Support Assistant; three-dimensional imaging) and spoken language to sign language translation systems such as Poland's project entitled “THETOS” (Text into Sign Language Automatic Translator, which operates in Polish; natural language processing) are addressed. Finally, the article considers synthesized sign, which is being added to educational material and has the potential to be developed by students themselves.The Problem based neural classifier, using the proposed descriptor, achieved an accuracy rate above 90%.

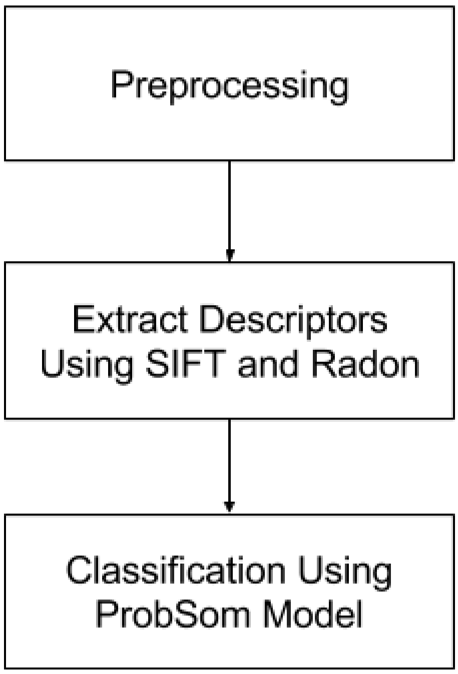

Fig 3: Block Diagram of Hand Gesture Recognition System for THETOS

2.1.2 S. Meshram, R. Singh, P. Pal and S. K. Singh, "Convolution Neural Network based Hand Gesture Recognition System," 2023

This paper introduced software which presents a system prototype that is able to automatically recognize sign language to help deaf and dumb people to communicate more effectively with each other or normal people. Pattern recognition and Gesture recognition are the developing fields of research.24 different alphabets were considered in this paper where 96% recognition rate was obtained.

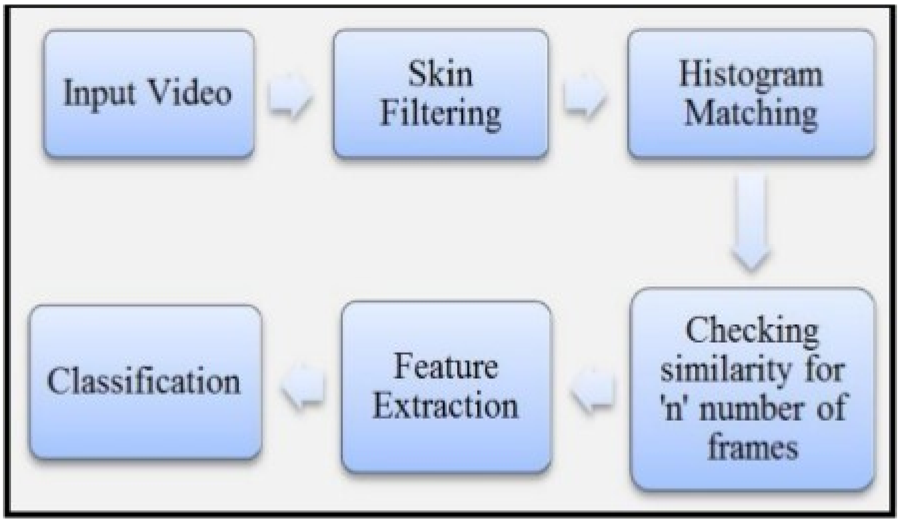

Fig 4: System Overview

2.1.3 Ashfaq T, Khurshid K (2016) Classification of hand gestures using gabor filter with bayesian and na ve bayes classifier.

This paper reviews different steps in an automated sign language recognition (SLR) system. Developing a system that can read and interpret a sign must be trained using a large dataset and the best algorithm.
As a basic SLR system, an isolated recognition model is developed. The model is based on vision-based isolated hand gesture detection and recognition. Assessment of ML-based SLR model was conducted with the help of 4 candidates under a controlled environment.
The model made use of a convex hull for feature extraction and KNN for classification. The model yielded 65% accuracy.

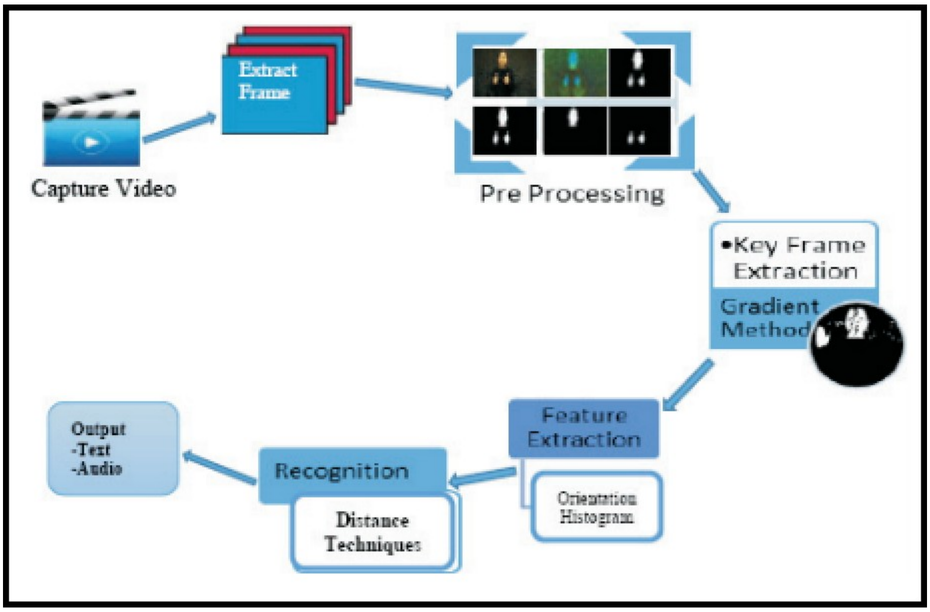

Fig 5: General Diagram of the Work

2.1.4 Chaikhumpha T, Chomphuwiset P (2018) Real—time two hand gesture recognition with condensation and hidden Markov models.

This paper deals with robust modeling of static signs in the context of sign language recognition using deep learning-based convolutional neural networks (CNN). In this research, total 35,000 sign images of 100 static signs are collected from different users. The efficiency of the proposed system is evaluated on approximately 50 CNN models.

The results are also evaluated on the basis of different optimizers, and it has been observed that the proposed approach has achieved the highest training accuracy of 99.72% and 99.90% on colored and grayscale images, respectively.

The performance of the proposed system has also been evaluated on the basis of precision, recall and F-score. The system also demonstrates its effectiveness over the earlier works in which only a few hand signs are considered for recognition.

2.1.5 . Deshpande, Aditi, Ansh Shriwas, Vaishnavi Deshmukh, and Shubhangi Kale. "Sign Language Recognition System using CNN." In 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), pp. 906-911. IEEE, 2023.

This paper demonstrates the statistical techniques for recognition of ISL gestures in real time which comprises both the hands. A video database was created by the authors and utilized which contained several videos for large number of signs. Direction histogram is the feature used for classification due to its appeal for illumination and orientation invariance. Two different approaches utilized for recognition were Euclidean distance and Knearest neighbor metrics.

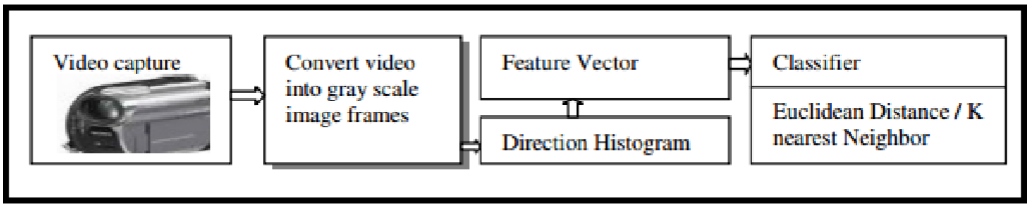

Fig 6: Methodology for real time ISL classification

2.1.6 Kalsh, Er Aditi, and N. S. Garewal. "Sign language recognition system." International Journal of ComputationalEngineering Research (IJCER) 3, no. 6 (2013).

This is the first identifiable academic literature review of sign language recognition systems. It provides an academic database of literature between the duration of 2007–2017 and proposes a classification scheme to classify the research articles.

Three hundred and ninety six research articles were identified and reviewed for their direct relevance to sign language recognition systems. One hundred and seventeen research articles were subsequently selected, reviewed and classified. Each of 117 selected papers was categorized on the basis of twenty five sign languages and were further compared on the basis of six dimensions (data acquisition techniques, static/dynamic signs, signing mode, single/double handed signs, classification technique and recognition rate). .

Sl No	Research Paper	Merits	Demerits
1.	Visual hand gesture recognition with convolution neural network	Earth mover's distance classification algorithm	Low Accuracy in Low Lighting Conditions
2.	Convolution Neural Network based Hand Gesture Recognition System	CNN classification approach based on CC algorithm with enhancement technique	Low Accuracy in high noisy conditions.
3.	Classification of hand gestures using gabor filter with bayesian and naïve bayes classifier.	Convexity algorithm	Static hand gesture algorithm.
4.	Two hand gesture recognition with condensation and hidden Markov models.	Hidden Markov models	Low Accuracy for an untrained hand.
5.	Sign Language Recognition System using CNN.	Support vector machine	Unstable in few systems.
6.	Sign language recognition system.	SVM classification approach based on CC algorithm with enhancement technique	Old Approach, Low accuracy.

3. ALGORITHMS

3.1 Convolutional Neural Network (CNN)

Neural networks, as its name suggests, is a machine learning technique which is modeled after the brain structure. It comprises of a network of learning units called neurons. These neurons learn how to convert **input signals** (e.g. picture of a cat) into corresponding **output signals** (e.g. the label "cat"), forming the basis of automated recognition.

A convolutional neural network (CNN, or ConvNet) is a type of feed-forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex.

CNNs have repetitive blocks of neurons that are applied across space (for images) or time (for audio signals etc). For images, these blocks of neurons can be interpreted as 2D convolutional kernels, repeatedly applied over each patch of the image. For speech, they can be seen as the 1D convolutional kernels applied across time-windows. At training time, the weights for these repeated blocks are 'shared', i.e. the weight gradients learned over various image patches are averaged.

3.1.1 CNN Summarized in 4 Steps

There are four main steps in CNN: convolution, subsampling, activation and full connectedness.

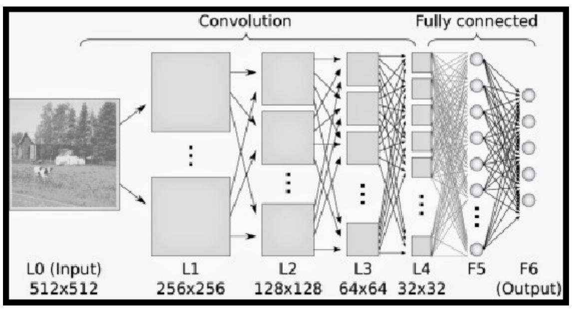

Fig 7: Convolutional neural network

3.1.1.1 Convolution

The first layers that receive an input signal are called convolution filters. Convolution is a process where the network tries to label the input signal by referring to what it has learned in the past. If the input signal looks like previous cat images it has seen before, the "cat" reference signal will be mixed into, or convolved with, the input signal. The resulting output signal is then passed on to the next layer.

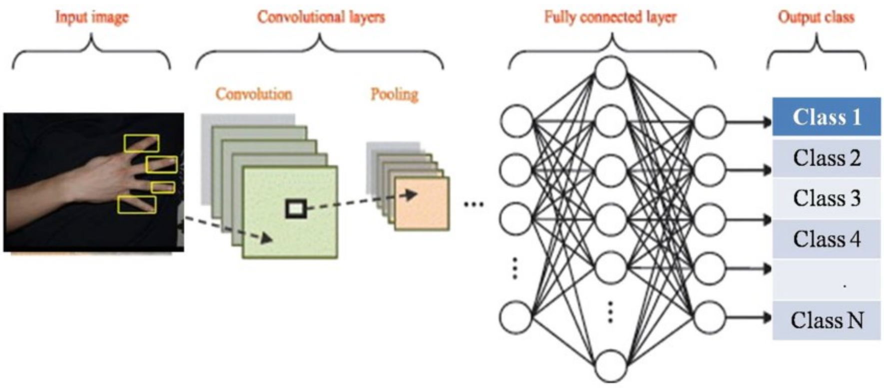

Fig 8: Internal architecture of Proposed CNN classifier for hand gesture recognition

Convolution has the nice property of being **translational invariant**. Intuitively, this means that each convolution filter represents a feature of interest (e.g whiskers, fur), and the CNN algorithm learns which features comprise the resulting reference (i.e. cat). The output signal strength is not dependent on where the features are located, but simply whether the features are present. Hence, a cat could be sitting in different positions, and the CNN algorithm would still be able to recognize it.

For e.g suppose we convolve a 32x32x3 (32x32 image with 3 channels R,G and B respectively) with a 5x5x3 filter. We take the 5*5*3 filter and slide it over the complete image and along the way take the dot product between the filter and chunks of the input image.

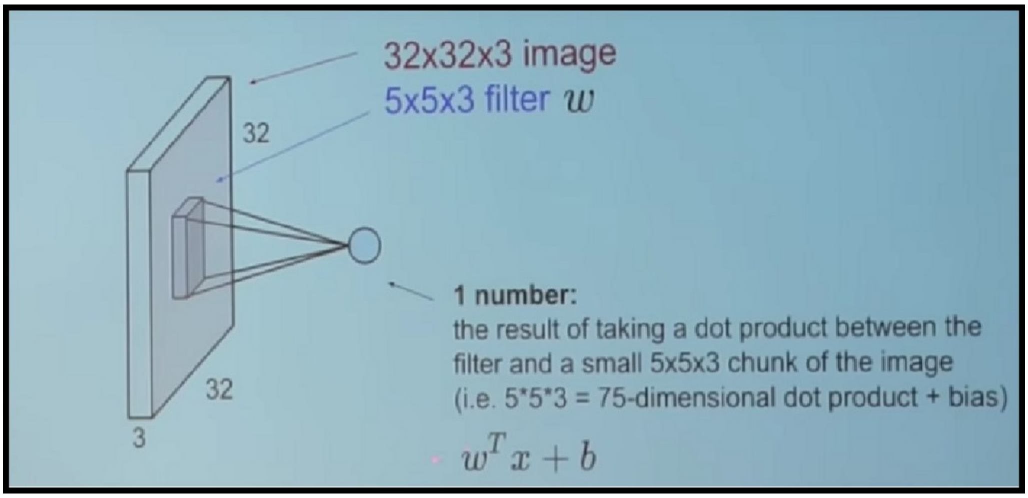

Fig 9: Dot Product of Filter with single chunk of Input Image

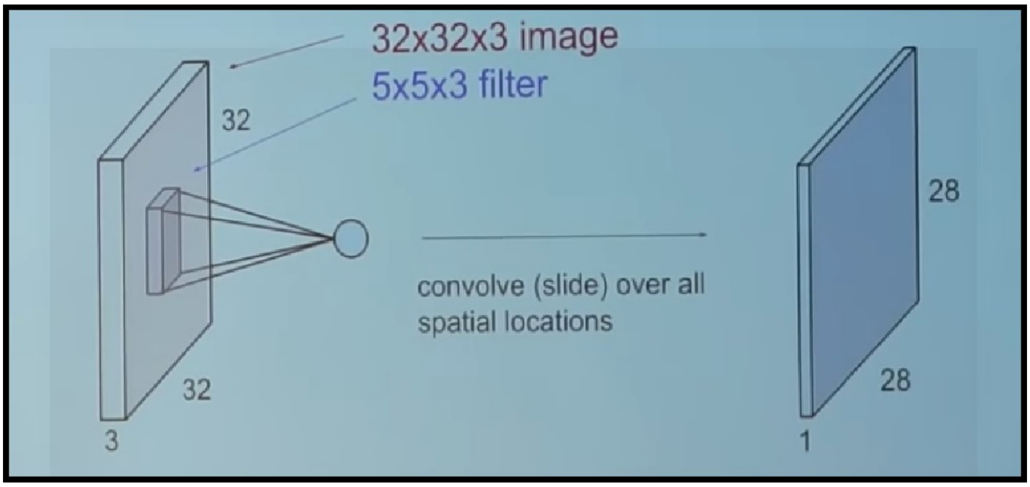

Fig 10: Dot Product or Convolve over all possible 5x5 spatial location in Input Image

The convolution layer is the main building block of a convolutional neural network. The convolution layer comprises of a set of independent filters (6 in the example shown). Each filter is independently convolved with the image and we end up with 6 feature maps of shape 28*28*1.

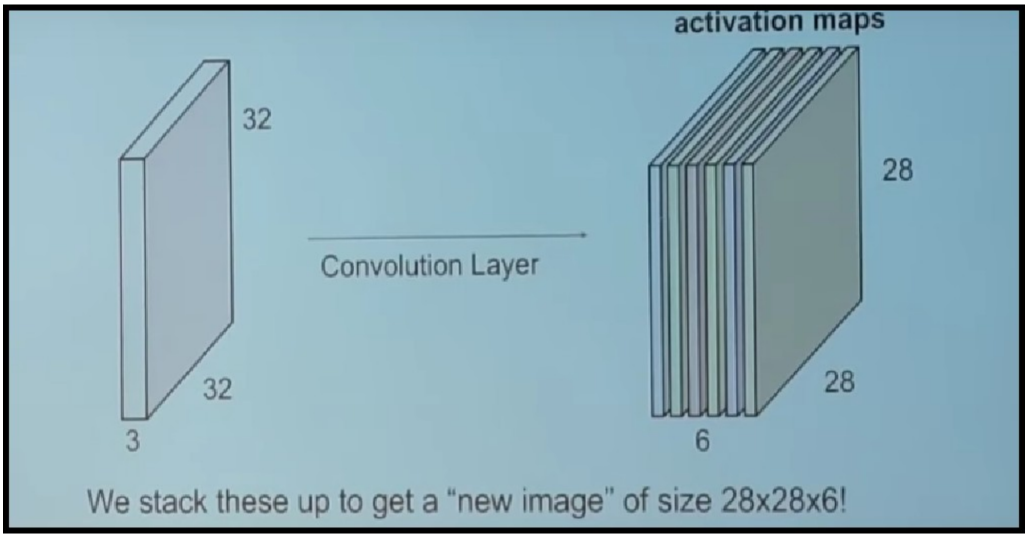

Fig 11: Input Image Convolving with a Convolutional layer of 6 independent filters

The CNN may consists of several Convolutional layers each of which can have similar or different number of independent filters. For example the following diagram shows the effect of two Convolutional layers having 6 and 10 filters respectively.

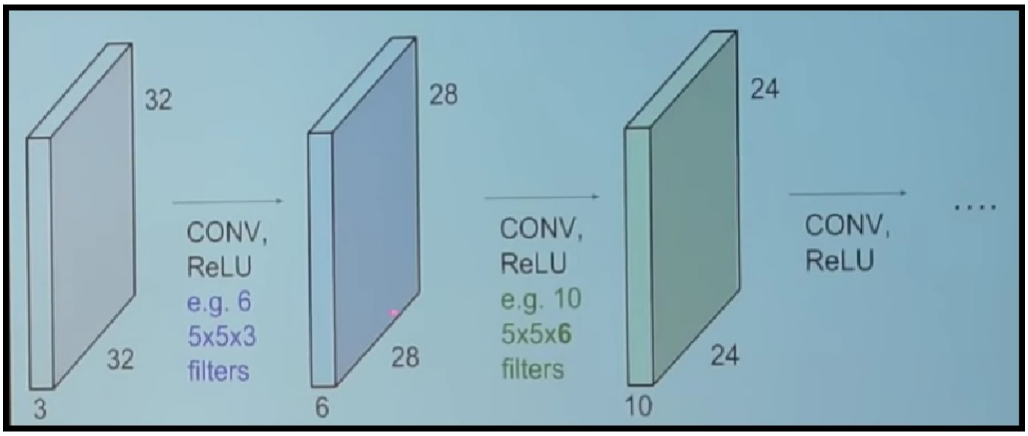

Fig 12: Input Image Convolving with two Convolutional layers having 6 and 10 filters respectively

All these filters are initialized randomly and become our parameters which will be learned by the network subsequently.

3.1.1.2 Subsampling

Inputs from the convolution layer can be "smoothened" to reduce the sensitivity of the filters to noise and variations. This smoothing process is called **subsampling**, and can be achieved by taking averages or taking the maximum over a sample of the signal. Examples of subsampling methods (for image signals) include reducing the size of the image, or reducing the color contrast across red, green, blue (RGB) channels.

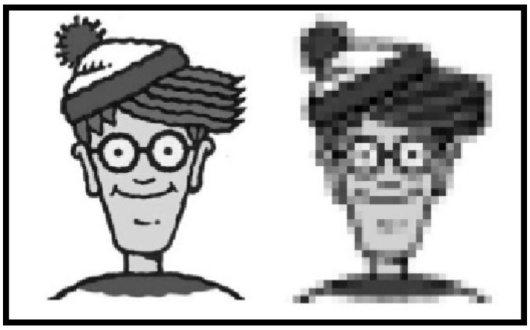

Fig 13: Sub sampling Wally by 10 times. This creates a lower resolution image.

3.1.1.3 Pooling

A pooling layer is another building block of a CNN.

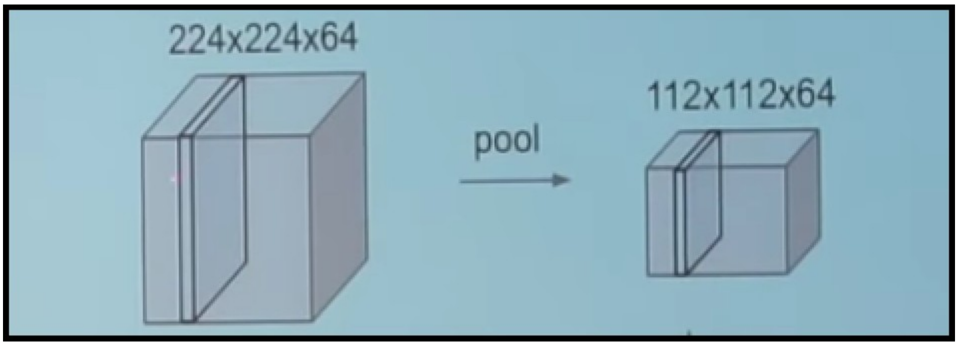

Fig 14: Pooling to reduce size from 224x224 to 112x112

Its function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network. Pooling layer operates on each feature map independently.

The most common approach used in pooling is max pooling in which maximum of a region taken as its representative. For example in the following diagram a 2x2 region is replaced by the maximum value in it.

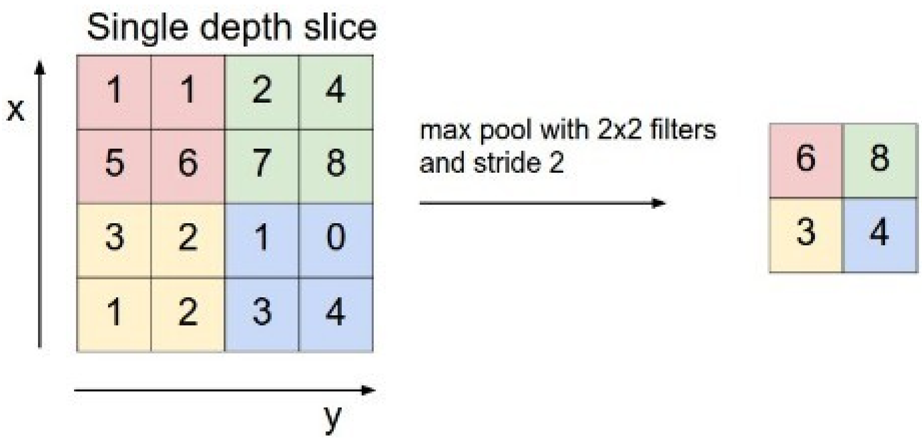

Fig 15: Max Pooling

3.1.1.4 Activation

The activation layer controls how the signal flows from one layer to the next, emulating how neurons are fired in our brain. Output signals which are strongly associated with past references would activate more neurons, enabling signals to be propagated more efficiently for identification.

CNN is compatible with a wide variety of complex activation functions to model signal propagation, the most common function being the Rectified Linear Unit (ReLU), which is favored for its faster training speed.

3.1.1.5 Fully Connected

The last layers in the network are fully connected, meaning that neurons of preceding layers are connected to every neuron in subsequent layers. This mimics high level reasoning where all possible pathways from the input to output are considered.

3.1.1.6 (During Training) Loss

When training the neural network, there is additional layer called the loss layer. This layer provides feedback to the neural network on whether it identified inputs correctly, and if not, how far off its guesses were. This helps to guide the neural network to reinforce the right concepts as it trains. This is always the last layer during training.

3.2 Implementation

Algorithms used in training CNN are analogous to studying for exams using flash cards. First, you draw several flashcards and check if you have mastered the concepts on each card. For cards with concepts that you already knowdiscard them. For those cards with concepts that you are unsure of, put them back into the pile. Repeat this process until you are fairly certain that you know enough concepts to do well in the exam. This method allows you to focus on less familiar concepts by revisiting them often. Formally, these algorithms are called gradient descent algorithms for forward pass learning. Modern deep learning algorithm uses a variation called stochastic gradient descent, where instead of drawing the flashcards sequentially, you draw them at random. If similar topics are drawn in sequence, the learners might overestimate how well they know the topic. The random approach helps to minimize any form of bias in the learning of topics.

Learning algorithms require feedback. This is done using a **validation set** where the CNN would make predictions and compare them with the true labels or ground truth. The predictions which errors are made are then fed backwards to the CNN to refine the weights learned, in a so called backwards pass. Formally, this algorithm is called **backpropagation of errors**, and it requires functions in the CNN to be differentiable (almost).

CNNs are too complex to implement from scratch. Today, machine learning practitioners often utilize toolboxes developed such as Caffe, Torch, MatConvNet and Tensor flow for their work.

4 EXPERIMENTAL DESIGN

4.1 Tech Stack

- Linux Environment
- Python Programming Language
- CNN Architecture
- Open CV
- Tensorflow
- Keras
- Numpy

4.2 Dataset Used

We have used a robust dataset of images of hand gestures for all alphabets having 3000 images for each letter.

Fig 16: Dataset of letter F

4.3 Methodology

- All the images in the dataset will be convoluted using CNN architecture which will eventually train a keras .h5 model. There is seperate python code to do this.
- This trained ML model will recognize hand gestures of the American Sign Langauge in real time through the webcam.

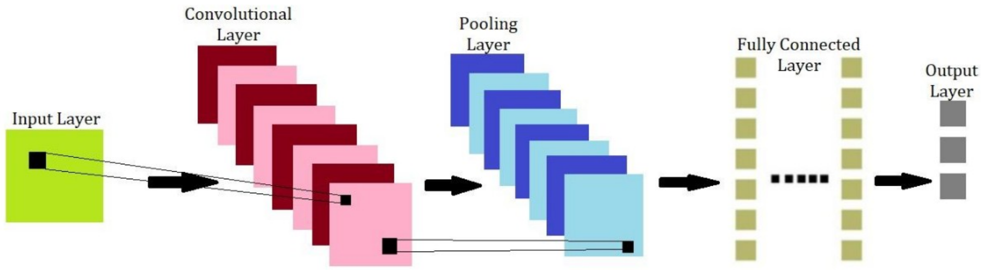

Fig 17: Our CNN model

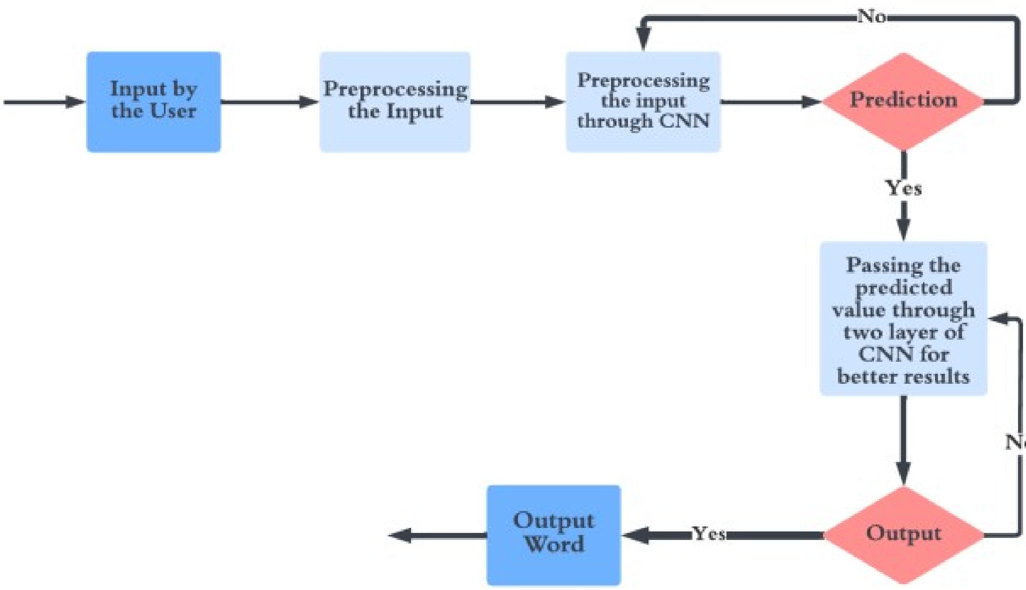

Fig 18: Flowchart of Methodology

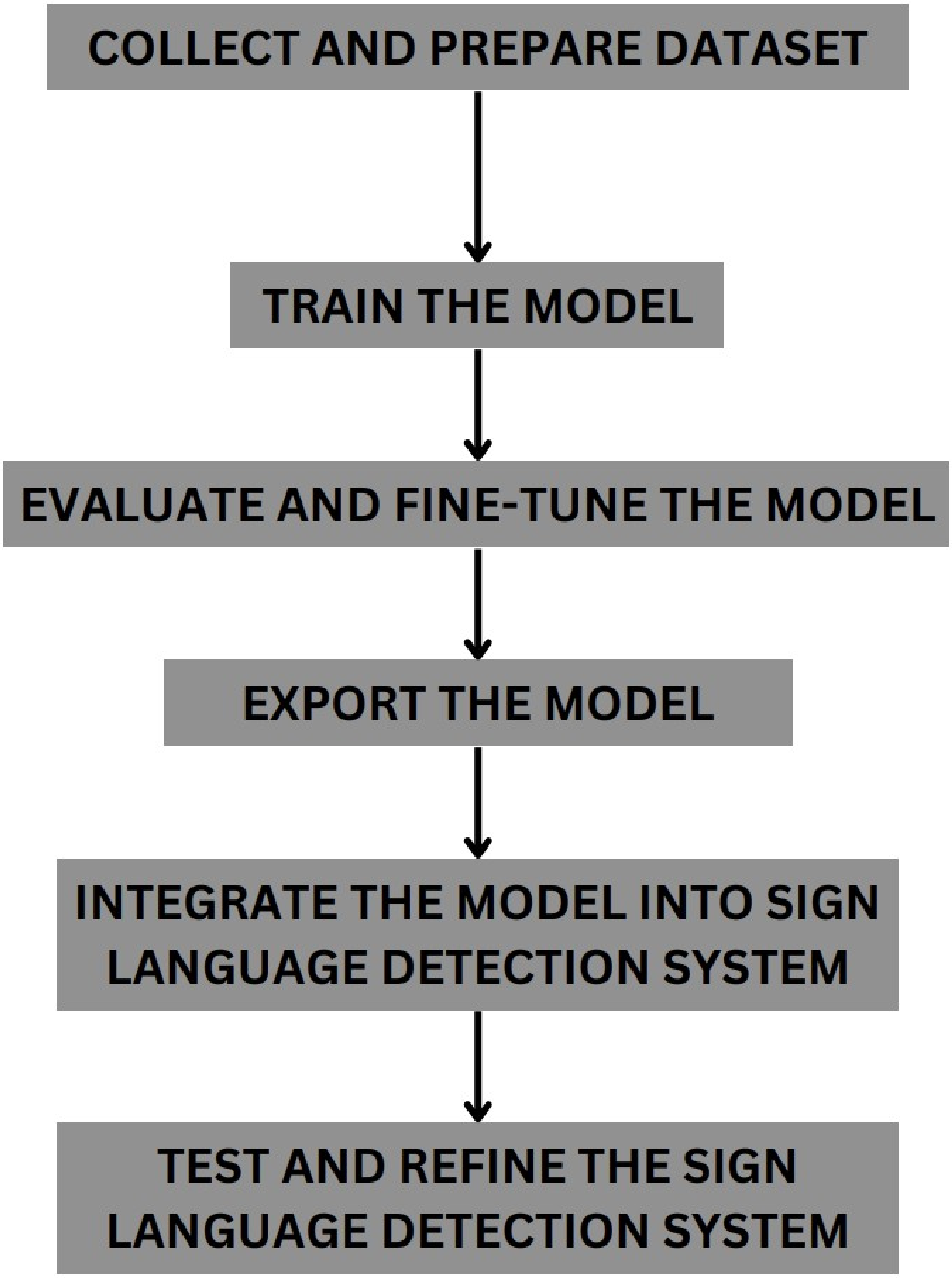

Fig 19 : Block Diagram of Methodology

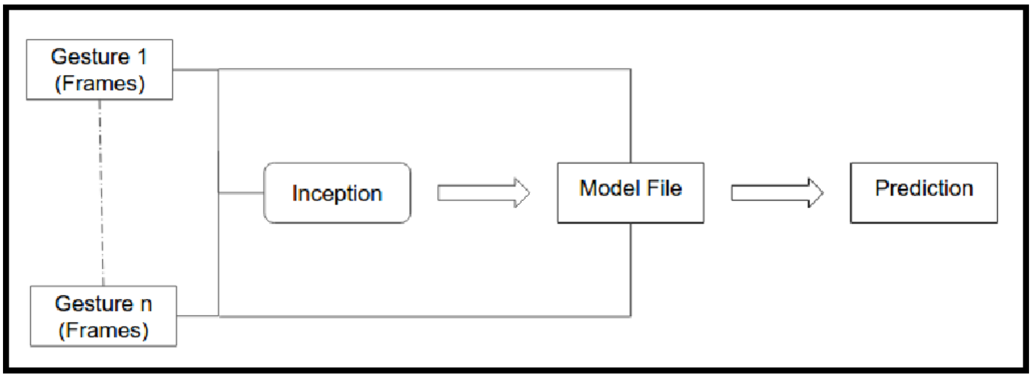

Fig 20 : Block Diagram of Code execution process

Use the train.py to train your sign language detection model using the uploaded dataset. Experiment with different model architectures and hyperparameters to optimize model performance. Monitor the training process and iterate as needed to achieve satisfactory model accuracy.

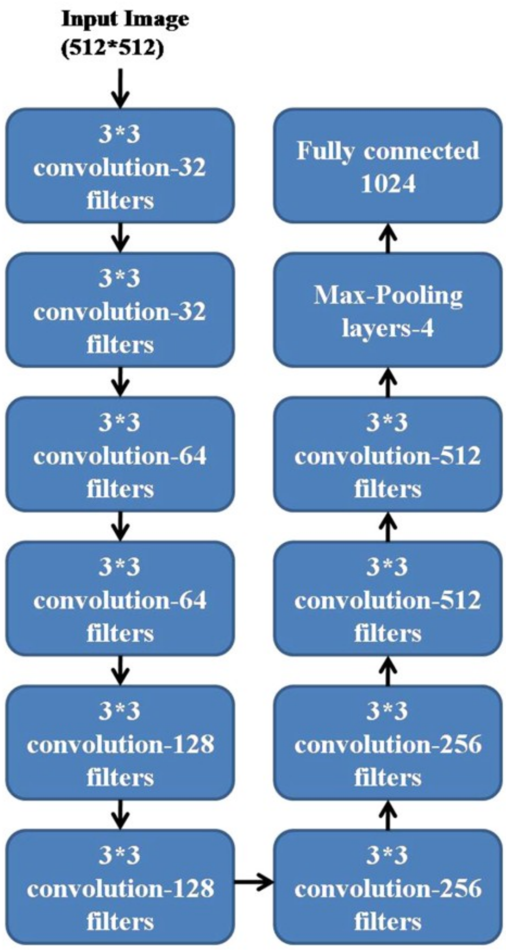

Fig 21: Developed CNN architecture used in hand gesture reconition

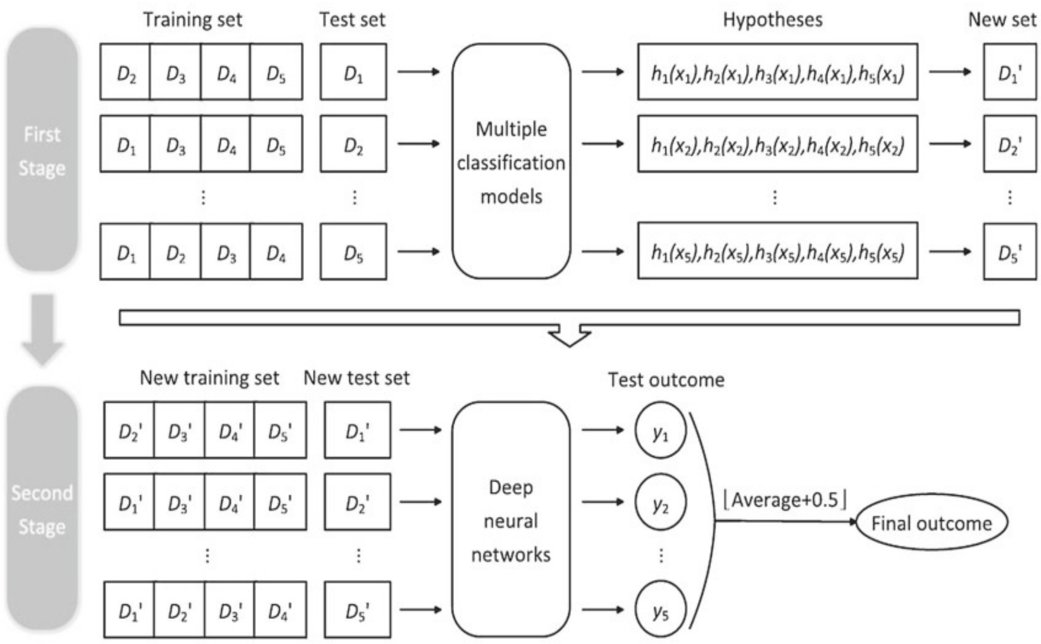

Fig 22: Deep learning classification process

Evaluate the trained model using the validation and testing sets to assess its accuracy and generalization performance. Fine-tune the model by adjusting model parameters or collecting more data, if necessary, to improve model accuracy.

4.4 Novelty

Dataset of above 3000 images.
Convoluted the dataset to create a .h5 keras ML model
No Handtracking module is being used.
Increased the efficiency above 95%.
Recognizes gestures which seem very similar to each other
Matplotlib used to prepare accuracy parameters.

4.5 Limitations

The Trained ML model recognized all letters in the ASL sign language with exceptional accuracy except the letters ‘J’ and ‘Z’ as these two letters are not gesture bases but movement based, recognising which is out of the scope of the trained ML Model.

5. RESULTS AND ANALYSIS

5.1 Final Results

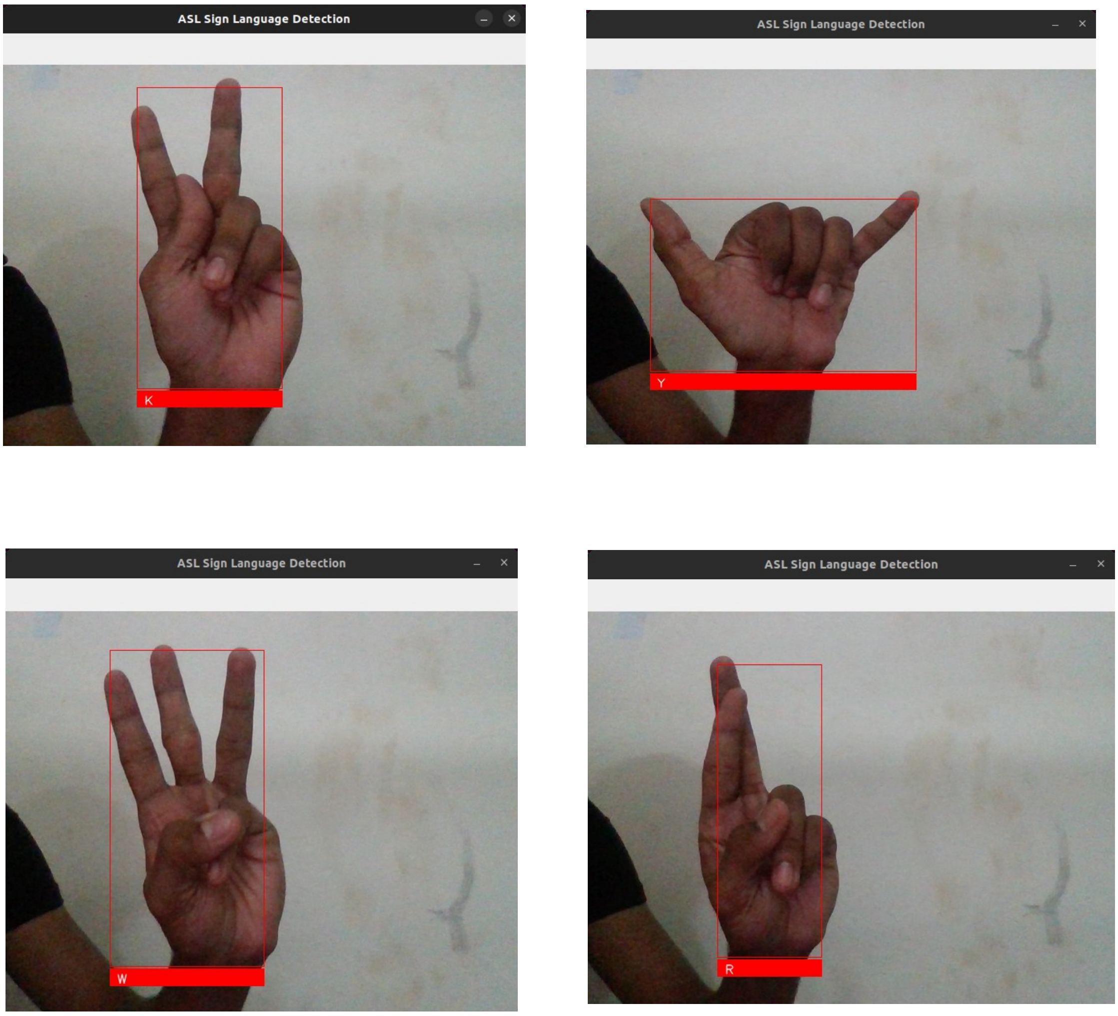

Fig 23: Recognition of ASL gestures

The methodology presented above boasts an impressive accuracy rate of 95.70%. In a rigorous trial involving 100 signs, the ML Model demonstrated its prowess by accurately recognizing 96 of them. This exceptional performance can be attributed to the robust dataset comprising over 78,000 images, with a substantial 3,000 images dedicated to each alphabet

5.2 Analysis of Results

Accuracy of recognising letters after 1000 iterations is tabulated:

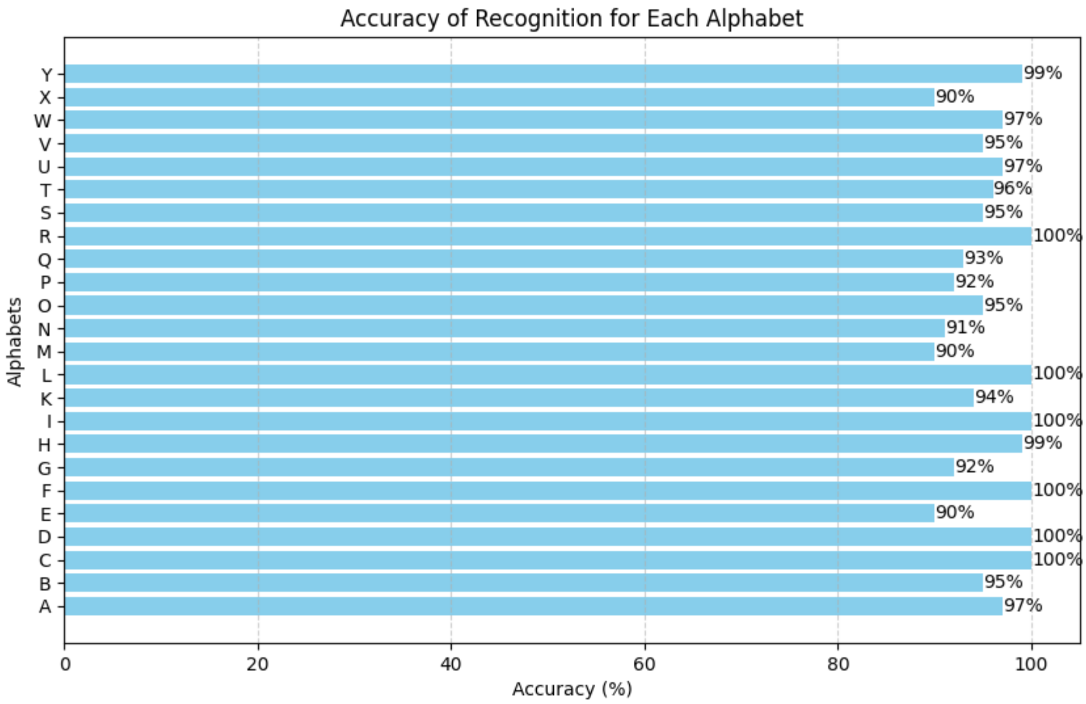

Fig 24: Accuracy Plot of recogition rates of alphabets

Average Accuracy is 95.7%.

The CNN provides a robust and systematic methodology to classify the type of hand gestures. Fig.2 shows the architecture of the CNN. Dataset consists of 12000 pictures, in which 10000 pictures are used for training the neural network and 2000 image for testing recognition accuracy of the CNN. The total iteration of updating parameters in convolutional neural network is 12000 with 10 mini-batches.

The proposed system also had the satisfactory results on the transitive gestures in a continuous motion using the proposed rules. In the future, a high-level semantic analysis will be applied to the current system to enhance the recognition capability for complex human tasks.

Confusion Matrix

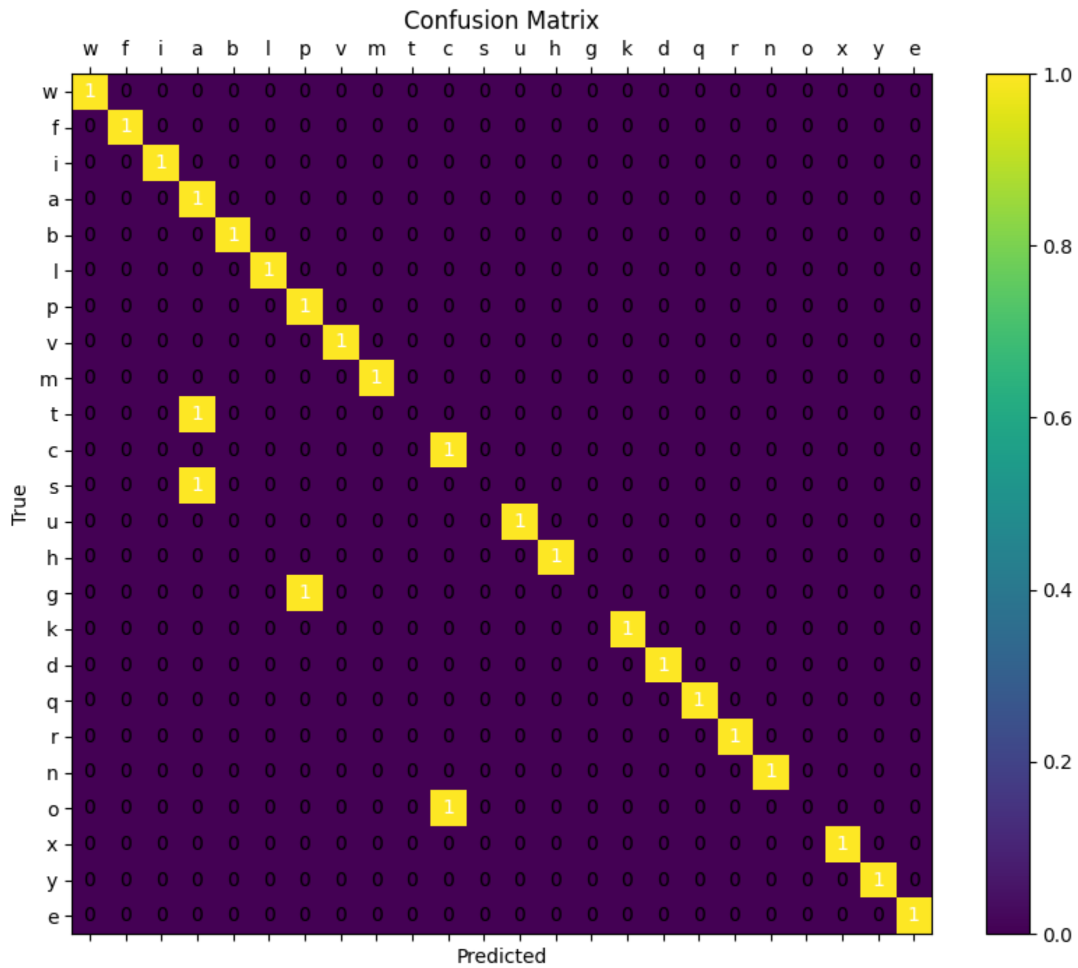

Fig 25: Confusion Matrix

Fig. 9 show the accuracy confusion matrices uses ten-fold cross validation based on our two stage CNN. Each cell in the matrix is computed as the average of the corresponding values of all confusion matrices. The low standard deviations are invisible here, as most values are either zero or quite low.

In the best case we achieved an accuracy of 99.6% for gesture 8 due to no similar gestures. While in the worst case we only achieved an accuracy of 88.3% for gesture 5. There may be finger blockings resulting in misjudging of finger numbers which leads to a lower recognition rate for gestures with more fingers.

Recognition Rates (%)

Testers	a	b	c	d	e	f	g	Ave.
Trained Tester	100	100	100	100	98	100	95	99
Untrained Tester 1	100	99	99	98	97	97	82	96
Untrained Tester 2	98	98	99	99	96	95	93	96.85
Untrained Tester 3	97	94	100	97	91	93	83	93.57

The Recognition Rates are impressive even with an Untrained testers hand-gestures.

Performance Analysis Parameters	Experimental Results (%)
Sensitivity	98.1
Specificity	93.4
Accuracy	95.7
Recognition Rate	98.7

The analysis work is performed using recognition time which can be computed by the total execution time for recognizing the single hand gesture image for an automated process. It is essential for real-time applications for pro cessing the hand gestures under different environmental conditions

The genereated model needs good lighting to recognising with the above mentioned accuracy . Dim Light conditions will reduce the accuray.The model also requires the webcam to capture hand gestures at a minimum resolution of 720p for desired results.

We obtained an accuracy of 95.7 %. This shows that CNN can be successfully used to learn spatial and temporal features and classify Sign Language Gestures. The accuracy of recognition is mainly due to the robustness of the dataset used and the accurate implementation of CNN.

6. CONCLUSION AND FUTURE WORK

Hand gestures are a powerful way for human communication, with lots of potential applications in the area of human computer interaction. Visionbased hand gesture recognition techniques have many proven advantages compared with traditional devices. However, hand gesture recognition is a difficult problem and the current work is only a small contribution towards achieving the results needed in the field of sign language gesture recognition. This report presented a visionbased system able to interpret isolated hand gestures from the American Sign Language(ASL).

We wish to extend our work further in recognising continuous sign language gestures with better accuracy. This method for individual gestures can also be **extended for sentence level sign language.**

Collect more diverse and extensive datasets, including a broader range of hand shapes, sizes, and backgrounds. This can help improve the model's robustness and generalization to different user scenarios.

Incorporate user feedback mechanisms to continually improve the model. This could involve allowing users to provide feedback on recognition errors, which can be used to update and fine-tune the model over time.

Enhance the model's adaptability to varying environmental conditions, such as changes in lighting or background. This may involve developing algorithms that can automatically adjust to different settings.

Create a mobile application that allows users to interact with the ASL hang gesture recognition system on their smartphones or other portable devices.

Investigate the **adaptability of the model to different cultural variations** in ASL. This may involve collaborating with users from diverse linguistic backgrounds to ensure the model's effectiveness for a wide range of users.

7. REFERENCES

[1] Ronchetti, Franco, Facundo Quiroga, César Armando Estrebou, and Laura Cristina Lanzarini. "Handshape recognition for argentinian sign language using probsom." Journal of Computer Science & Technology 16 (2016).

[2] Singha, Joyeeta, and Karen Das. "Automatic Indian Sign Language Recognition for Continuous Video Sequence." ADBU Journal of Engineering Technology 2, no. 1 (2015).

[3] Tripathi, Kumud, and Neha Baranwal GC Nandi. "Continuous Indian Sign Language Gesture Recognition and Sentence Formation." Procedia Computer Science 54 (2015): 523531.

[4] Nandy, Anup, Jay Shankar Prasad, Soumik Mondal, Pavan Chakraborty, and Gora Chand Nandi. "Recognition of isolated indian sign language gesture in real time." Information Processing and Management (2010): 102107.

[5] Ashfaq T, Khurshid K (2016) Classification of hand gestures using gabor filter with bayesian and naïve bayes classifier. Int J Adv Comput Sci Appl 7(3):276–279

[6] Chaikhumpha T, Chomphuwiset P (2018) Real—time two hand gesture recognition with condensation and hidden Markovmodels. In: 2018 international workshop on advanced imagetechnology (IWAIT), Chiang Mai, 2018, pp 1–4

[7] Ronchetti, Franco, Facundo Quiroga, César Armando Estrebou, Laura Cristina Lanzarini, and Alejandro Rosete. "LSA64: An Argentinian Sign Language Dataset." In XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016). 2016.

[8] Kapil Y, Bhattacharya J (2016) Real-time hand gesture detection andrecognition for human computer interaction. In: Berretti S, Thampi S, Srivastava P (eds) Intelligent systems technologiesand applications. Springer, Cham, pp 559–567

[9] Liu K, Chen C, Jafari R, Kehtarnavaz N (2014) Fusion of inertial and depth sensor data for robust hand gesture recognition. Sens J IEEE 14(6):1898–1903

[10] Cooper, Helen, Brian Holt, and Richard Bowden. "Sign language recognition." In Visual Analysis of Humans, pp. 539562. Springer London, 2011.

[11] Zhang, Chenyang, Xiaodong Yang, and YingLi Tian. "Histogram of 3D facets: A characteristic descriptor for hand gesture recognition." In Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, pp. 18. IEEE, 2013.

[12] Cooper, Helen, EngJon Ong, Nicolas Pugeault, and Richard Bowden. "Sign language recognition using subunits." Journal of Machine Learning Research 13, no. Jul (2012): 22052231.

[13] Rahman MH, Afrin J (2013) Article: hand gesture recognition using multiclass support vector machine. Int J Comput Appl 74(1):39–43

[14] Park S, Yu S, Kim J, Kim S, Lee S (2012) 3D hand tracking using Kalman filter in depth space. EURASIP J Adv Signal Process 2012(1), 36

[15] Marium A, Rao D, Crasta DR, Acharya K, D'Souza R (2017) Hand gesture recognition using webcam. Am J Intell Syst 7(3):90–94 Mitra S, Acharya T (2007) Gesture recognition: a survey. IEEE Trans Syst Man Cybern Part C Appl Rev 37(3):311–324