ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY273

Sign Language Recognition Using Machine Learning

Dr. Bhuvaneshwari K V[1]
Associate Professor
Department of Information Science and Engineering,
BIET, Davanagere, Karnataka, India

Bindu A R[2]; Manvitha G K[3]
Nikitha N Chinchali[4]; Nisha K N[5]
U.G. Students
Department of Information Science and Engineering,
BIET, Davanagere, Karnataka, India

Abstract:- Communication is very important in human daily life and the most widely used type of communication is verbal communication. But there are people with hearing and speech impairment who cannot communicate verbally and the language which they use for communication is sign language. Many other languages, tools are being developed for inter-language translation from sign language to text. There has been a lot of research done in the field of American Sign Language but the work is limited in the case of Indian Sign Language. This is due to lack of standards and the variation in the language. The proposed system aims to recognize Indian Sign Language digit gestures and convert it into text. By using Machine Learning Techniques, sign language recognition leads to the development of a more accurate and robust system. As Deep learning techniques, ResNet100 and ensemble models continue to evolve, sign language recognition system plays a transformative role in bridging the communication gap between deaf and hearing individuals. It helps the user to recognize the sign language by using this proposed system.

Keywords:- *Sign Language, Convolutional Neural Networks, Residual Network , Random Forest Classifier, Ensemble Model.*

I. INTRODUCTION

The language known as sign language uses hand movements in place of spoken words to communicate. It is beneficial for the deaf to know sign language. It gives deaf people an easy method to communicate with others. It helps the deaf community feel accepted by society. Although most people who use signing are deaf, hearing people also use it, including those who are unable to talk normally, have problems with oral language because of a disability or disease, and have family members who are deaf. Sign language has a large and diversified community. There are hundreds of distinct sign languages spoken around the world, even though there isn't one global sign language. Here are a few sign languages that are used worldwide:

- *American Sign Language (ASL):*

With over 500,000 users in the US and Canada, ASL is one of the most popular sign languages in the world. ASL has its own unique grammar and hand patterns and was established in US schools for the deaf in the 18th century.

Deep learning is a type of machine learning that relies on artificial neural networks. It can recognize intricate links and patterns in data. Computer technology that is inspired by the functioning of the human brain is called deep learning. It learns and completes difficult tasks by using networks of synthetic neurons. During training, the layers of these networks, also known as neural networks, modify their connections. The term "deep" refers to the ability of these networks to comprehend and represent data at a higher level of detail through the use of several layers. To enhance the model's predictions, these connections are adjusted during the training phase. Deep learning is renowned for its ability to automatically identify significant elements in data without requiring human input. To provide precise insights and forecasts, deep learning models can identify intricate patterns in images, text, audio, and other types of data. Deep Learning has shown great promise in several domains, such as recommendation systems, speech recognition, natural language processing, and picture recognition. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transfer learning are a few of the well-liked Deep Learning designs.

Deep learning includes a variety of methods, each tailored to certain tasks and situations. The following are some basic methods used in the field of deep learning:

- *Convolutional Neural Networks (CNN):*

Dedicated to handling data that resembles a grid, like pictures. It makes use of convolutional layers to identify spatial hierarchies and patterns. Tasks in computer vision, such as object detection, image segmentation, and image recognition, make extensive use of it. Recurrent Neural Nets (RNN): Suitable for sequence data, enabling the persistence of information. To transfer data from one stage in the sequence to the next, it contains loops. It is employed in time-series analysis, speech recognition, and natural language processing activities. Transfer of Learning: It entails using previously trained models from one job and modifying them for a similar but distinct activity.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY273

An ensemble model builds upon several separate models to provide a prediction that is more reliable and powerful than any of the individual models working alone. An ensemble model called Random Forest is made up of several decision trees because of its scalability, resilience, and capacity to manage high-dimensional data, Random Forest is a robust and adaptable ensemble learning method that is frequently employed for both regression and classification applications. A random portion of the training data and a random subset of the characteristics are used to create each decision tree that makes up Random Forest. A random portion of the training data and a random subset of the characteristics are used to create each decision tree that makes up Random Forest.

II. LITERATURE SURVEY

- [1] The basic concept of sign language recognition system and review of its existing techniques along with their comparison presented. The main objective of presenting this survey is to highlight the importance of vision based method with a specific focus on sign language. They covered most of the currently known methods for SLR tasks based on deep neural architectures that were developed over the past several years, and divided them into clusters based on their chief traits. The most common design deploys a CNN network to derive discriminative features from raw data, since this type of network offers the best properties for this task. In many cases, multiple types of networks were combined in order to improve final performance.
- [2] Work on American Sign Language (ASL) words share similar characteristics. These characteristics are usually during sign trajectory which yields similarity issues and hinders ubiquitous application. However, recognition of similar ASL words confused translation algorithms, which lead to misclassification. Based on fast fisher vector (FFV) and bi-directional Long-Short Term memory (Bi-LSTM) method, a large database of dynamic sign words recognition algorithm called bidirectional long-short term memory-fast fisher vector (FFV-Bi-LSTM) is designed. The performance or FFV-Bi-LSTM is further evaluated on ASL data set, leap motion dynamic hand gestures data set (LMDHG), and semaphoric hand gestures contained in the Shape Retrieval Contest (SHREC) dataset.
- [6] The research article investigates the impact of machine learning in the state of the sign language recognition and classification. It highlights the issues faced by the present recognition system for which the research frontier on sign language recognition intends the solutions. In this article, around 240 different approaches have been compared that explore sign language recognition for recognizing multilingual signs. The research done by various authors is also studied, and some of the important research articles are also discussed in this article. This article discussed how machine learning methods could benefit the field of automatic sign language recognition and the potential gaps that machine learning approaches need to address for the real-time sign language recognition.
- [8] Important application of hand gesture recognition that is translation of sign language. In sign language, the fingers' configuration, the hand's orientation, and the hand's relative position to the body are the primitives of structured expressions. The importance of hand gesture recognition has increased due to the rapid growth of the hearing-impaired population. In this paper, a system is proposed for dynamic hand gesture recognition using multiple deep learning.
- [9] The approach is to have a vision based system in which the sequence of images representing a word in ISL is translated to equivalent English word. The translation would be done by means of Deep learning algorithms namely convolutional neural nets and recurrent neural nets. The system will be analyzing sequence of images, hence CNNs will analyze each image and their sequence is analyzed by LSTM (which is an implementation of RNN). We divided dataset into training dataset and testing dataset, which obtained 73.60% accuracy. The image distributions are kept fairly different in training and testing datasets.
- [11] An alternative to written and direct communication languages used in India and the Indian subcontinent is Indian Sign Language (ISL). People who are deaf or mute and are unable to hear or talk frequently use it. Compared to other sign languages used in developed nations, the ISL is a novel sign language. Given its current application characteristics, automatic recognition of any sign language, including ISL, is necessary. ISL automation will benefit both communities—those who can exclusively communicate in ISL and those who do not know the language at all—because unblessed individuals frequently have trouble interacting in public settings like airports, train stations, banks, and hospitals.
- [12] The basis of every human interaction, whether it be personal or professional, is communication. It is among the necessities for surviving in a community. Without a clear, mutually understood language, verbal communication is impossible. In India, sign language is used for communication by about 26% of the disabled population. Therefore, it is imperative to close the communication gap that exists between the general public and those who are speech challenged. The objective is to create a pair of sensor gloves that can translate motions used in Indian Sign Language (ISL) into audible speech.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY273

III. SYSTEM DESIGN

This Chapter discusses the system architecture of sign language recognition using deep learning techniques and ensemble model.

A. Architecture Design

As shown in the Figure 1 The Proposed system mainly consist of two modules namely Training phase and Testing phase.

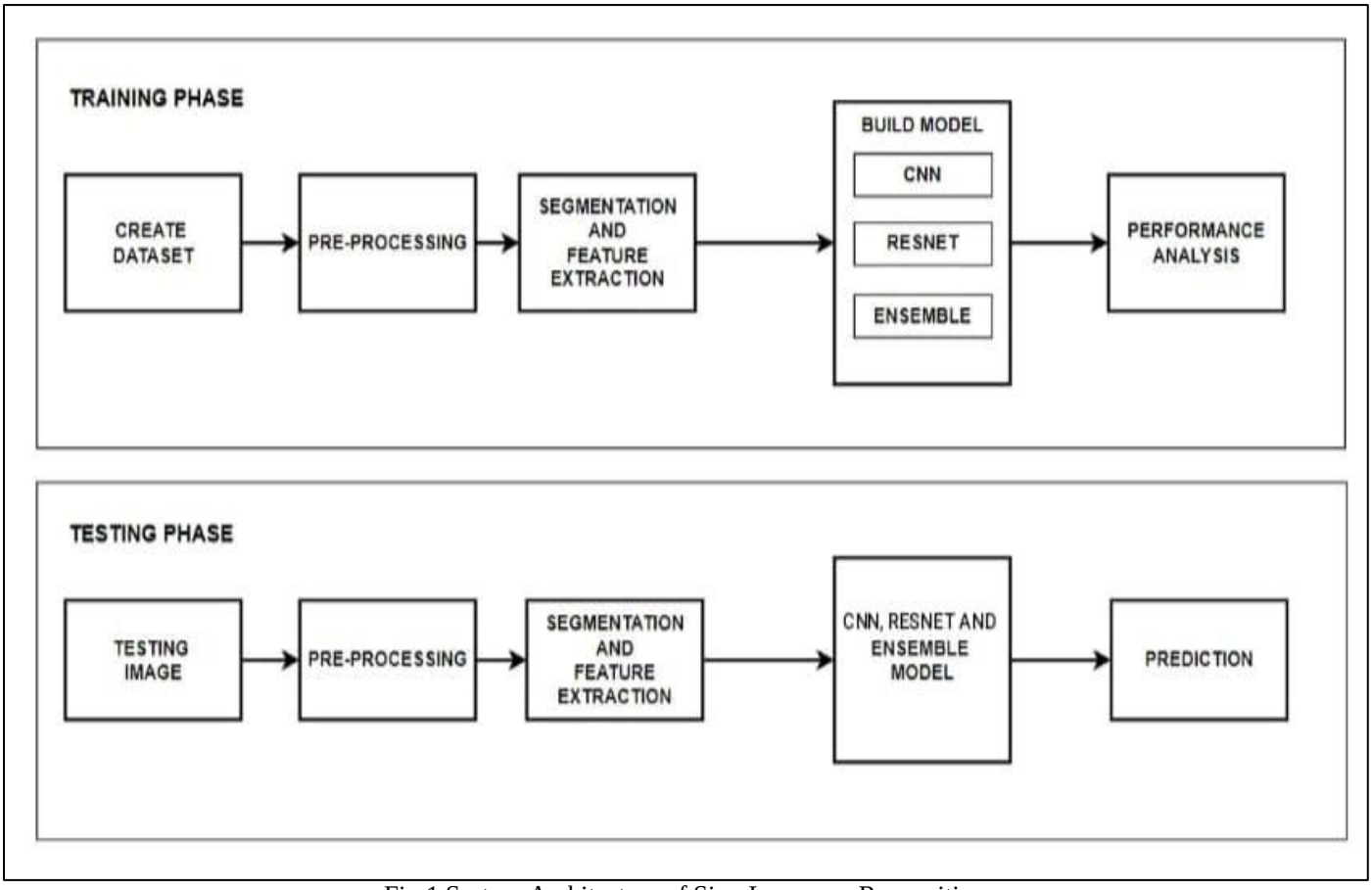

Fig 1 System Architecture of Sign Language Recognition

➢ *Training Phase*

Training phase includes preparation of data set, pre-processing, segmentation, feature extraction and building a model.

- *Preparing the Data Set*

The first stage is to collect a wide range of sign language gesture. This dataset should contain a variety of signs. Next, annotate the dataset with relevant labels for each sign.

✓ *Data Collection:*

Collect a broad variety of hand gesture images or videos. To improve the model's robustness, ensure that the lighting conditions, backdrops, and hand locations vary.

✓ *Dataset Splitting:*

In this step, the dataset is split into training and testing sets. This division avoids over fitting and aids in assessing the model's performance on untested data.

- *Data Preprocessing*

A vital stage in the machine learning process is data preprocessing, which includes preparing raw data for model training by cleaning and formatting it. The objectives of this procedure are to improve the quality of the dataset, and remove superfluous data. By addressing variances such variations in lighting, background, or hand locations, normalizing the data aids in achieving uniformity. The careful preparation of the dataset guarantees that the machine learning algorithm can successfully identify patterns and produce accurate predictions, which enhances model performance. Data preprocessing includes steps like resizing, gray scale conversion and converting images from BGR to RGB to enhance the foreground images and reduce the computation complexities.

- *Segmentation and Feature Extraction*

Hand gestures and background can be distinguished from one another by using temporal segmentation techniques to identify distinct sign boundaries. In order to help with the precise separation of signs within a sequence,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY273

openCV library is used. Hand gesture segmentation applies the segmentation mask to the original image, and saves the resulting masked image. The process of feature extraction, which comes after segmentation, involves taking pertinent information out of the segmented data. Capturing hand and finger postures and movements is a typical element in the context of sign language recognition. These elements are essential for conveying the distinct qualities of every sign. In this proposed system, feature extraction is done using mediapipe library. A sign language recognition system's overall accuracy and resilience are greatly enhanced by efficient feature extraction.

- *Building a Model*

Building a data model in deep learning involves selecting an appropriate architecture for the specific task. Convolutional Neural Networks (CNNs) and ResNet100 are types of Deep learning models, each suited for different types of data and tasks. Ensemble model has also been built in this proposed system where further comparison is done between these models to determine which model possesses high performance.

✓ *Training the Model:*

The pre-processed dataset, obtained after dividing the data into training and testing sets, is used to train the model. After that, it is checked to make sure all pertinent characteristics have been retrieved and cleaned, subdivided, and handled correctly. Depending on the type of job (classification, regression, etc.), an appropriate loss function is selected, and an optimization technique is used to modify the model's parameters during training in order to minimize this loss. To avoid over fitting, the model's performance is evaluated during training on a validation set.

A reliable and efficient model for sign language recognition can be created by choosing a suitable deep learning model architecture (such as CNNs, Residual network and Random Forest classifier) and training it on a prepared dataset. The selection between convolutional neural networks (CNNs), Residual Network and Random forest classifier is based upon the type of input, task-specific requirements, and accuracy attained from each model.

➢ *Testing Phase*

Testing Phase includes steps like preprocessing, Segmentation, feature extraction, testing a model, and finally prediction.

To evaluate the performance of a deep learning model, use a separate test dataset. The method typically involves the following steps:

✓ *Metrics Evaluation:*

After training and optimizing the model, it is tested on the testing set using a variety of metrics.

✓ *Accuracy:*

The proportion of instances properly predicted to all instances.

✓ *Precision:*

The ratio of correctly predicted positives to all predicted positives, which shows how well the model avoids false positives.

✓ *Recall (Sensitivity):*

A measure of how well the model captures all pertinent cases, calculated as the ratio of true positive predictions to all actual positives.

✓ *F1 Score:*

A balanced metric produced by taking the harmonic mean of recall and precision.

B. System Architecture of Trained Models

The Proposed system uses three models are Convolutional neural network, Residual network (ResNet), Random forest classifier (Ensemble model).

➢ *Convolutional Neural Network*

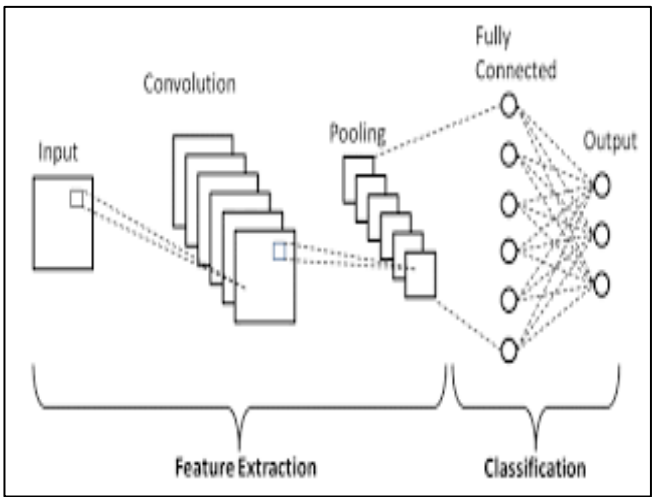

Fig 2 System Architecture of CNN

- *Input Layer:*

The raw image dataset is sent to this layer. The input could be a series of frames for dynamic signals or a grayscale or RGB image depicting a single hand position, depending on the approach used.

- *Convolutional Layer:*

These layers take the input image's features and extract them. A small filter, known as a kernel, is used in convolution to compute dot products between the input data and the filter at each location on the image as it slides across it. This makes it easier to spot patterns and small details in the picture.These characteristics may be used to identify hand posture, finger configuration, and hand placement in relation to the body in sign language. varied levels of abstraction in the features can be learned by utilizing multiple convolutional layers with varied filter sizes.

- *Pooling Layer:*

Pooling layers can be utilized to introduce some degree of invariance to little variations in hand pose and lower the dimensionality of the feature maps.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY273

- *Max Pooling:*

In max pooling, we choose a window size [for example, a window of size 2*2] and only accept the maximum of four values. Well, close this window and repeat the process until you have an activation matrix that is half the size of its original.

- *Average Pooling:*

In average pooling, we make use of all values in a window.

✓ *Fully Connected Layer:*

Based on the features extracted in the preceeding layers, picture categorization in the CNN takes place int the FC layer. These layers are similar to those used in classic neural networks. They connect the outputs of the convolutional and pooling layers to all neurons in the next layer. Fully-connected layers are commonly used for classification tasks, with the final output layer representing the probabilities of the image falling into distinct classes.

✓ *Final Output Layer:*

The final layer gets the output of the last fully-connected layer, which serves as a high-dimensional representation of the input data retrieved by preceding layers (convolution and pooling). The output layer converts the internal representation into the required output format.

➢ *Residual Network (ResNet)*

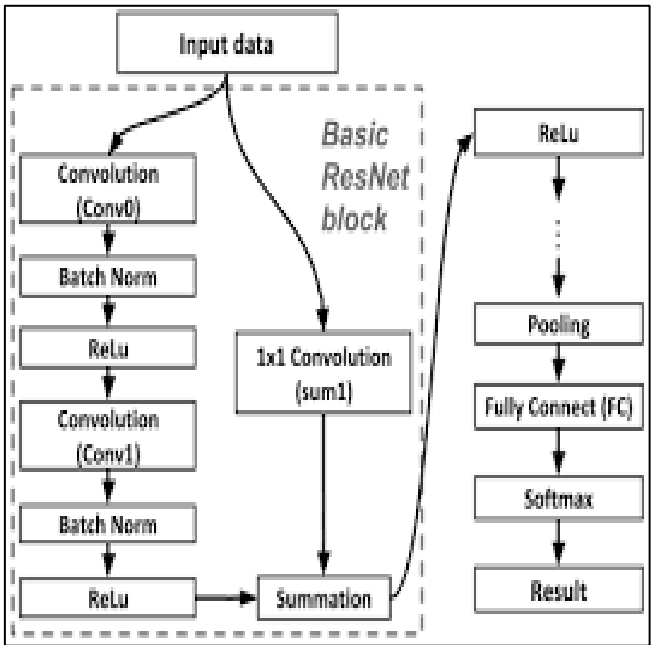

Fig 3 System Architecture of Residual Network

- *Preprocessing:*

The first stage is to prepare the input data for the network. Common methods include: Hand Segmentation is separating the hand from the image. Normalization is scaling the pixel intensity values to a specific range for better network training and Background Subtraction is Removing the background clutter to focus on the signer's hand region.

- *Initial Convolution:*

The preprocessed image/frame is processed by a single convolutional layer that extracts low-level information related to hand posture, such as edges and shape.

- *ResNet Stage:*

This is the core of the architecture and involves stacking multiple ResNet stages. A set of residual blocks makes up each stage.

➢ *Residual Block (in Every Stage):*

- *Input Transformation:*

The input from the preceding block (or the first stage's convolution) is processed by two or three convolutional layers. The goal of these tiers is to extract more intricate traits.

- *Skip Connection:*

The residual block's unaltered input is added to the convolutional layers' output via this direct connection. This guarantees that in addition to the necessary transformations learned by the convolutions, the network can learn the identity mapping.

- *Non-Linearity:*

The summed output from the skip connection is subjected to an activation function (such as ReLU) and convolutions. This introduces non-linearity and improves the network's ability to learn complex patterns.

✓ *Pooling Layer (Between Stages):*

To lessen the dimensionality of the feature maps, a pooling layer can be applied after each ResNet stage—possibly the final one. This improves processing efficiency and can introduce some level of invariance to small variations in hand pose. A global average pooling layer compiles the data from the complete feature map after the last ResNet stage. Before the data is fed into the fully-connected layers, it is further reduced in this way.

✓ *Fully Connected Layers:*

The pooled features from the previous phase are applied to these layers. The network learns to classify the features into several sign language categories using a sequence of fully-connected layers with decreasing dimensionality. The odds that the input belongs to particular signs in the vocabulary are represented in the final output layer.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY273

➢ *Random Forest Classifier (Ensemble Model)*

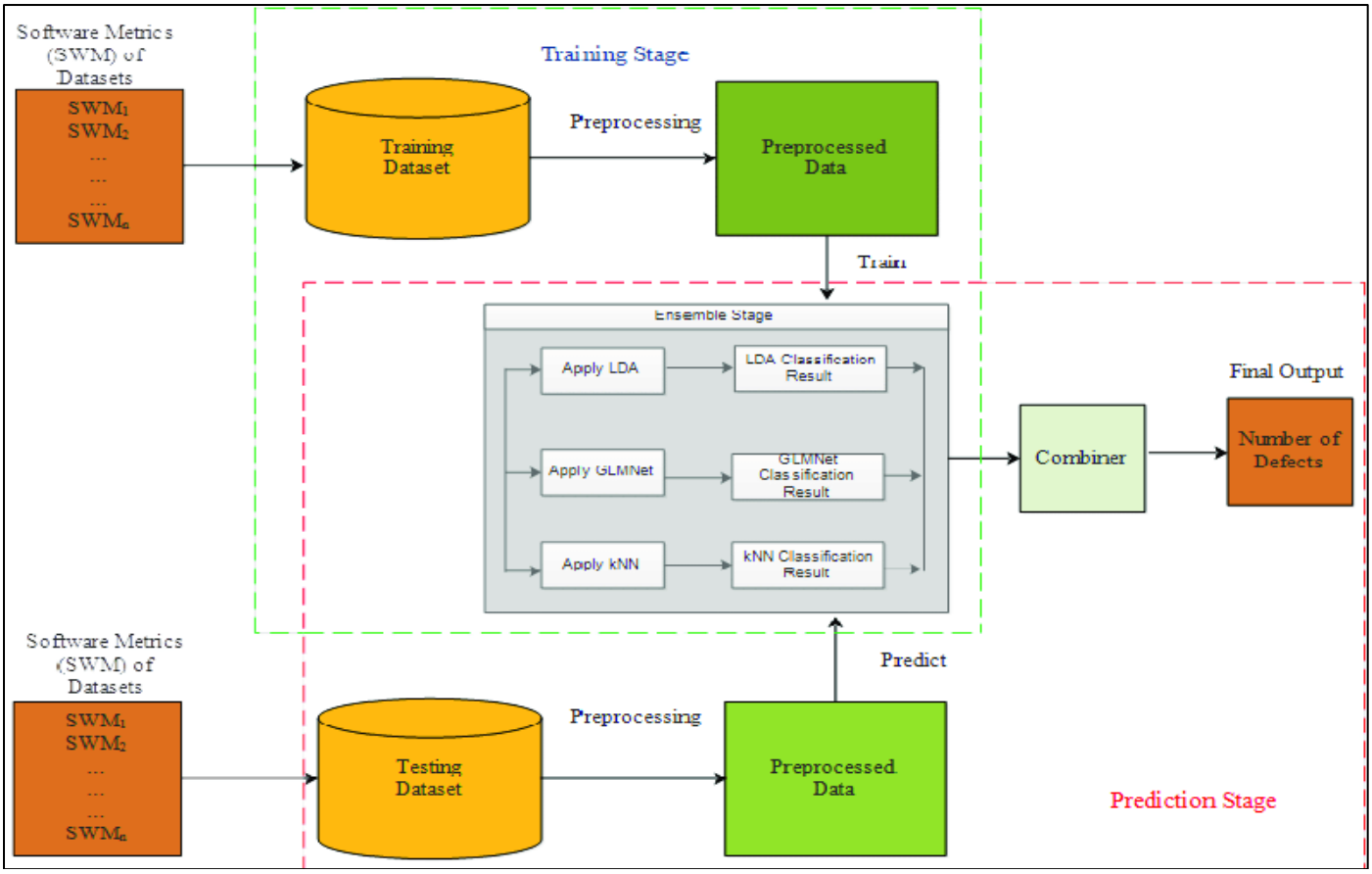

Fig 4 System Architecture of Ensemble Model

- *Data Preparation:*

The first stage involves dividing the dataset into sets for testing and training. The random forest model is constructed using the training set, and its performance is assessed using the testing set.

- *Random Sampling:*

In order to build each decision tree, the random forest algorithm randomly chooses a subset of the training data. This method is known as bagging or bootstrap aggregating. A distinct subset of the data is used to train each tree in the forest.

- *Feature Sampling:*

The method chooses a subset of features at random for every decision tree in addition to randomly sampling data. This lessens overfitting and adds variation to the decision trees.

- *Constructing Decision Trees :*

A subset of the characteristics and data are used to construct each decision tree in the random forest. A tree-like structure is produced by the algorithm's recursive division of the data into smaller subsets according to the values of the chosen features.

- *Voting for Prediction :*

After constructing the decision trees, the method employs them to forecast new data points. The random forest classifier aggregates the predictions of all the decision trees by obtaining the majority vote on the class predictions.

- *Model Evaluation:*

Finally, the random forest classifier's performance is assessed on the testing set. Accuracy, precision, recall, and F1 score are examples of commonly used evaluation metrics.

IV. IMPLEMENTATION

A. Create Dataset :

The dataset is created manually by collecting images from the webcam for all digits from 0 to 9 that require only single hand for the gesture in the Indian sign Language. Images are stored in the png format. For each digit in Indian sign language separate directories are created to store the respective digit. Each directory contains 100 images of the respective digit in Indian sign language. Since we have taken 0 to 9 digit computes to form totally 1000 images. Out of 1000 images, 800% of the data i.e, 800 images considered for training phase and 200 images for testing phase. Collected dataset of hand gestures representing ISL..

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY273

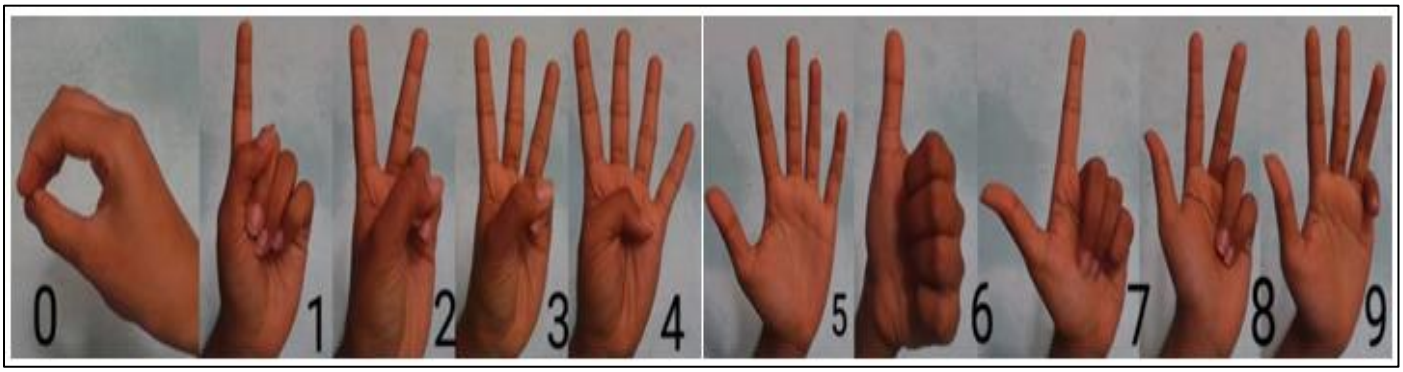

Fig 5 Hand Gestures Dataset of ISL

B. Preprocessing :

Preprocessing is used to enhance the quality and the foreground image. To implement this following are the steps used in preprocessing:

➢ *Resizing:*

Resizing an image refers to the process of changing its overall dimensions, either making it larger or smaller.

➢ *Grayscale:*

A grayscale image is a digital image that contains only shades of gray, ranging from pure black (absence of light) to pure white (full intensity).

➢ *Converting image to RGB:*

Changing the order of the colour channels that represent the image data. Both BGR and RGB are colour models.

- *Segmentation :*

Segmentation mainly involves separating the hand region from the background image. The isolated hand region is then evaluated to determine the exact gesture being sent. Segmentation applies the segmentation mask to the original image, and saves the resulting masked image.

- *Feature Extraction :*

Feature Extraction is the process of identifying and extracting informative characteristics from the hand region in an image. The technique to detect and track the locations of particular points on an individual's hand is known as hand landmarking. These locations, which are sometimes referred to as landmarks or key points, can be the wrist, the tips and bases of the fingers, or other hand points. One can use landmarks to recognize the various indications that the person is making. By using the MediaPipe library can identify hand landmarks.

V. RESULT AND CONCLUSION

CNN, Residual network (ResNet), and Ensemble model performance in the suggested system were assessed and compared. When compared to the other two models, the ensemble model provides the most accurate performance measurement.

Table 1 Represents the Accuracy, Precision, F1 Score and Recall of all the Three Models CNN, Residual Network (ResNet) and Random Forest Classifier (Ensemble Model)

Model	Accuracy	Precision	F1 score	Recall
CNN	0.94	0.95	0.94	0.94
ResNet	0.96	0.97	0.96	0.96
Ensemble	0.99	0.98	0.98	0.99

Image in the figure 6 represents the prediction of hand gesture in Indian sign language with better accuracy.

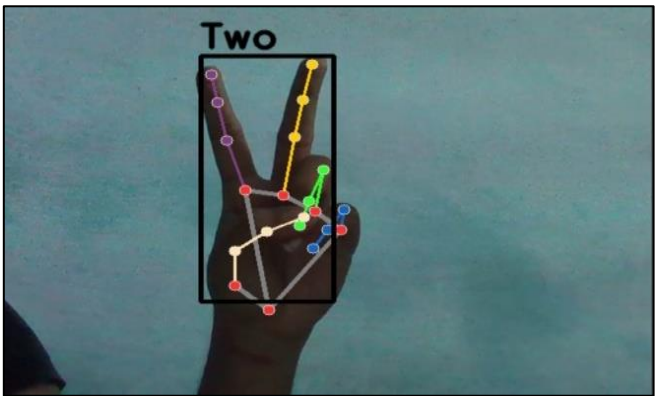

Fig 6 Image of Hand Gesture Representing Indian Sign Language

In conclusion, the development of a sign language recognition system is a significant step towards fostering inclusive communication for the deaf and hard-of-hearing communities. It can provide real-time translation of sign language into spoken language or text. It Enabling deaf individuals to interact and convey their messages more easily in various situations like education, employment, and social settings.By leveraging advancements in deep learning and human-computer interaction, such a system has the potential to bridge the communication gap between individuals who use sign language and those who do not use sign language.

In term of future improving the preprocessing to predict gestures even in low light conditions with a higher accuracy. The proposed system can an be further built as a web/mobile application for the users and it's only works for sign language gestures of a single hand. So, it can be enhanced to take gestures using both hands.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY273

REFERENCES

[1]. Sakshi Sharma,Sukhwinder Singh, "Vision-based sign language recognition system: A Comprehensive Review" Published in 2020 International Conference on Inventive Computation Technologies (ICICT) published on June 2020.

[2]. D Sathyanarayanan, T. Srinivasa Reddy, A. Sathish, P. Geetha, J.R. Arunkumar, S. Prem Kumar Deepak, "American Sign Language Recognition System for Numerical and Alphabets", 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), pp.1-6, 2023.

[3]. M. Alfonse, A. Ali, A. S. Elons, N. L. Badr and M. Aboul-Ela, "Arabic sign language benchmark database for different heterogeneous sensors", Proc. 5th Int. Conf. Inf. Commun. Technol. Accessibility (ICTA), pp. 1-9, Dec. 2016.

[4]. M. E. R. Grif and A. B. M. R. E. Prikhodko, ‘‘Recognition of Russian and Indian sign languages based on machine learning,’’ Anal. Data Process. Syst., vol. 3, no. 83, pp. 53–74, 2021.

[5]. R. Elakkiya, "Retraction note to: Machine learning based sign language recognition: A review and its research frontier", J. Ambient Intell. Humanized Comput., vol. 12, no. 7, pp. 7205-7224, Jul. 2022.

[6]. Muneer Al-Hammadi, Ghulam Muhammad, Wadood Abdul, Mansour Alsulaiman, Mohammed A. Bencherif, Tareq S. Alrayes, Hassan Mathkour, Aand Mohamed Amine Mekhtiche- "Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation" - - Grant No. 5-18-03-001-0003.

[7]. Riad Souissi, Thariq Khalid,Muhammad Al-Qurishi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," Published in IEEE Access Vol.9, September 2021.

[8]. J. C. Núñez, R. Cabido, J. J. Pantrigo, A. S. Montemayor, and J. F. Vélez, ‘‘Convolutional neural networks and long short-term memory for skeletonbased human activity and hand gesture recognition,’’ Pattern Recognit., vol. 76, pp. 80–94, Apr. 2018.

[9]. R. Cui, H. Liu, and C. Zhang,‘‘A deep neural framework for continuous sign language recognition by iterative training,’’ IEEE Trans. Multimedia, vol. 21, no. 7, pp. 1880–1891, Jul. 2019.

[10]. E. Rajalakshmi, R. Elakkiya, A. L. Prikhodko, M. G. Grif, M. A. Bakaev, J. R. Saini, K. Kotecha, and V. Subramaniyaswamy, ‘‘Static and dynamic isolated Indian and Russian sign language recognition with spatial and temporal feature detection using hybrid neural network,’’ ACM Trans. Asian Low-Resource Lang. Inf. Process., vol. 22, no. 1, pp. 1–23, Jan. 2023

[11]. A. K. Sahoo, G. S. Mishra, K. K. Ravulakollu, *ARPN J. Enq. Appl. Sci.* 2014, **9**, 116. "Indian sign language recognition using ensemble based classifier combination", Feb. 2022

[12]. Ajay S, Ajith Potluri, Sara Mohan George, Gaurav R, Anusri S, "Indian Sign Language Recognition Using Random Forest Classifier", IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), July 2021