Hand Gestures Classification and Image Processing using Convolution Neural Network Algorithm

Dr.Sk. Mahboob Basha[1],
H.C.Srivalli[2],
B.Jahnavi[3],
C.V.Basanth[4]
[1]Professor, NRI Institute of Technology, A.P, India-521212
[2]UG Scholar, Dept. of IT, NRI Institute of Technology, A.P-521212
[3]UG Scholar, Dept. of IT, NRI Institute of Technology, A.P-521212
[4]UG Scholar, Dept. of IT, NRI Institute of Technology, A.P-521212

Abstract:- The deaf community communicates primarily through the use of sign language. In general, sign language is much more figuratively formable for communication, which helps to advance and broaden the conversation. The ASL is regarded as the universal sign language, although there are numerous variations and other sign systems used in various parts of the world. There are fewer major ideas and concepts assigned. There are fewer principal ideas and assigned appearances in sign language. The main goal of this effort is to create a system of sign language that will benefit the deaf community and speed up the process of communication. The project's main objective is to build a classifier-based software model for sign language recognition. The strategy for this is to identify the gestures and use classifiers to assess the attributes. Principal component analysis is used for gesture recognition, and a classifier is used to assess the gesture features. The hand gesture has been used as a form of communication since the beginning of time. Recognition of hand gestures makes human-computer interaction (HCI) more versatile and convenient. Because of this, accurate character identification is crucial for a tranquil and error-free HCI. The majority of the hand gesture recognition (HGR) systems now in use have only taken a few straightforward discriminating motions into account for recognition performance, according to a literature review. This study uses robust modelling of static signs in the context of sign language recognition by using convolutional neural networks (CNNs) based on deep learning. In this study, CNN is used for HGR, which takes into account both the ASL alphabet and numbers simultaneously. The CNNs utilised for HGR are emphasized, along with their benefits and drawbacks. Modified Alex Net and modified VGG16 models for classification form the foundation of the CNN architecture. After feature extraction, a multiclass support vector machine (SVM) classifier is built, which is based on modified pre-trained VGG16 and Alex Net architectures. To achieve the highest recognition performance, the results are assessed using various layer features. Both the leave-one-subject-out and a random 70-30 method of cross-validation were used to test the accuracy of the HGR schemes. This work also emphasises how easily each character can be recognised and how similar their motions are to one another. To show how affordable this work is, the experiments are run on a basic CPU machine as opposed to cutting-edge GPU hardware. The proposed system outperformed several cutting-edge techniques with a recognition accuracy of 99.82%.

***Keywords*:-** *Sign language, ASL(American Sign Language) , Deaf Community , Gestures , Human Computer Interaction , Hand Gesture Recognition, CNN(Convolution Neuron Network), SVM(Support Vector Machine)*

I. INTRODUCTION

As the standard computer input devices that have been developed in the realm of technology have not altered much, this is due to the fact that these gadgets work well. As of today, as computers become more prevalent in our daily lives, it is becoming simpler and easier to introduce new hardware and software. We can only use keyboards, light pens, and these days, keypads, to interact with computers. Although they are very popular, these gadgets have speed limitations. Vision-based interfaces are becoming more common among users as technology advances, giving computers the ability to see. As a result, this evolution will prompt the creation of new device interfaces, which will enable these devices to process commands that cannot be entered into their present input mechanisms. Man-machine interaction, often known as human-computer interaction, is the term used to describe the relationship or interaction between humans and machines. This is a reference to how people and machines interact. When creating the HCI model, it is important to keep in mind the two key traits of functionality and usability.

Usability of the system is its ability to do the specific activities that the user accomplishes with accuracy, whereas functionality of the system is the services that are supplied to the user of the system by the system. The use of gestures for inter-human communication, including sign language, is also utilised for inter-human communication. Recognition of hand signals has recently attracted a lot of attention. Applications for hand gesture detection include operating machines and removing the mouse from video games. The sets of motions that make up sign language are its most important structures. Every gesture has a distinct meaning in sign language.

ISSN No:-2456-2165

In essence, sign language serves as a means of communication between hearing and deaf people. Using the sensors in the data gloves, sign language has become more prevalent as a result of technological innovation. Recognizing the representations of human hand motions is the major goal of sign language. Early on, recognition was challenging, but as technology has advanced, it has become simpler and more accurate. Communication is a necessary ability for community members to express themselves and live in harmony.

While members of the community can communicate via verbal and auditory languages, individuals who do not have these abilities can express themselves through sign language, which is a visual language. Sign language is a communication strategy in which persons with hearing challenges or issues communicate using body motions such as hands, arms, and gestures. Because most people in our culture are unfamiliar with the sign language used by people with hearing problems to communicate and express themselves, it is clear that these individuals struggle to express themselves in their everyday lives.

The scientific community has long recognised this need and has been working to create sign language devices to assist hearing-impaired persons in communicating. Despite the fact that the development of such technologies might be challenging owing to the existence of various sign languages and a lack of funding, large annotated datasets, as well as recent breakthroughs in AI and machine learning, have all significantly contributed to the automation and improvement of such technologies.

The objective of sign language recognition (SLR) is to create sophisticated machine learning algorithms that reliably categorise human articulations into individual signs or continuous phrases. Currently, the absence of large annotated datasets restricts the accuracy and generalisation capabilities of SLR algorithms, as well as the difficulty in recognising sign boundaries in continuous SLR scenarios.

II. TECHNOLOGIES USED

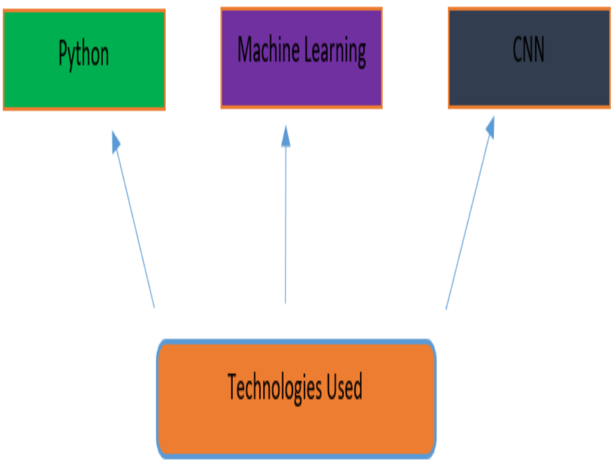

Fig 1 Technologies used

➢ *Python*

Python is an object-oriented, dynamically semantic, high-level, interpreted programming language. Rapid Application Development, as well as usage as a scripting or glue language to bring existing components together, find its high-level built-in data structures, coupled with dynamic type and dynamic binding, to be particularly appealing. Python's straightforward syntax promotes readability, which lowers the expense of software maintenance. Python's support for modules and packages promotes the modularity of programmes and the reuse of code. Both the comprehensive standard library and the Python interpreter are freely distributable and are accessible in source or binary form for all popular systems from fig [1] .

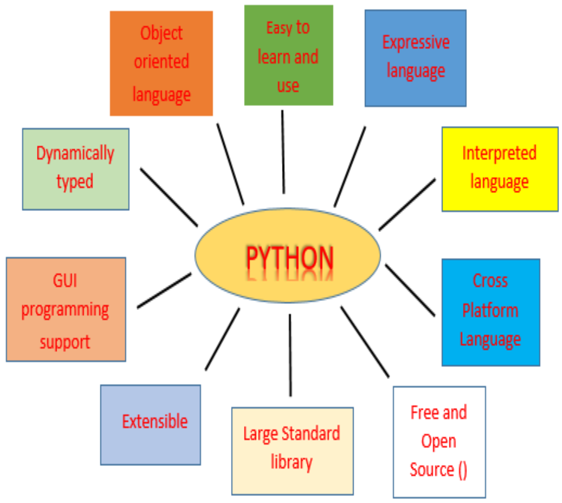

Fig 2 Features of Python Language

➢ *Machine Learning*

Basic and advanced machine learning principles are covered in the machine learning lesson. Both students and professionals in the workforce can benefit from our machine learning training from fig[3].

Machine learning, a developing technique, allows computers to automatically learn from historical data. Machine learning is the process of creating mathematical models using historical data or information and generating predictions using a variety of techniques. It is being utilised for many different tasks, including image identification, speech recognition, email filtering, Facebook auto-tagging, recommender systems, and more.

This machine learning lesson provides an overview of the field as well as a variety of machine learning approaches, including supervised, unsupervised, and reinforcement learning. Regression and classification models, clustering techniques, hidden Markov models, and other sequential models will all be covered.

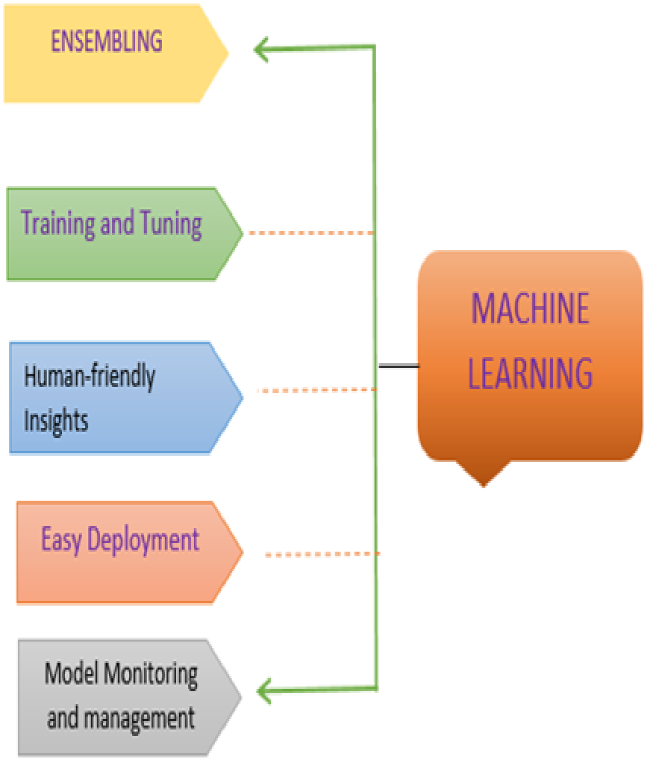

Fig 3 Features of Machine Learning

➢ *CNN*

Convolutional Neural Networks (CNNs) are deep neural networks that are used to analyze data with a topology resembling a grid, such as photographs that may be represented as a 2-D array of pixels. Convolution, Non-Linearity (Relu), Pooling, and Classification are the four fundamental processes that make up a CNN model (Fully-connected layer).

Convolution: Convolution is a technique used to take features out of a picture. By learning picture attributes from tiny squares of input data, it retains the spatial connection between pixels. Relu generally comes after it from fig[4].

Relu: It is an operation that zeroes out all negative pixel values in the feature map on an element-by-element basis. Its objective is to make a convolution network non-linearfrom fig[4].

Pooling: Downsampling, commonly known as pooling, lowers the dimensionality of each feature map while preserving crucial information from fig[4].

A multi-layer perceptron with a fully linked layer that employs the softmax function in the output layer. Its goal is to classify the input image into distinct classes using training data and characteristics from previous layers.In order to build a CNN model, these layers are combined. A fully linked layer makes up the final layer.

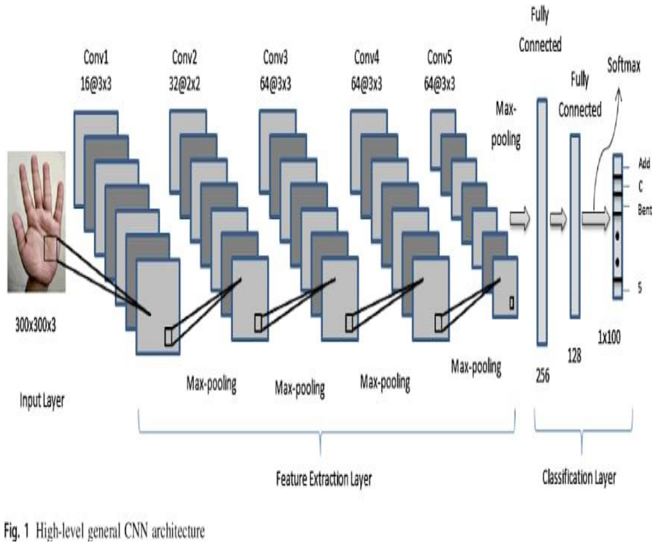

Fig 4 Working of CNN model

III. SOFTWARE REQUIREMENT SPECIFICATION

➢ *Functional Specifications:*

- This feature will translate the recognized gesture into the textual meaning of the gesture and display the translated text to the user from fig[3].

- System will recognize the appropriate movement of the hands and will search its database to match the movement with the pre-defined gestures. After matching system, will add the meaning of the sign to the opened file.

- *Normal Flow of Events:*

✓ User selects the communication mode from the main menu
✓ User opens a file □ User performs the movement
✓ Gesture is recognized and has a match
✓ The text is added to the file and displayed to the user and also speech out the words or sentence.

- *Sentence Level Translation:*

Sign languages are distinct languages with their own linguistic frameworks, just like spoken languages. Only when SEE (Signed Exact English) is available can a system that detects individual signs one at a time translate. A sign that appears later in the statement, such as when the signed sentence turns out to be a question, may radically alter the translation and the words employed. Additionally, making the signer stop and wait for the outcome after each sign could lead to a bad user experience.

- *Tracking the Face:*

All non-manual signals are important for tracking the face when signing. The face is particularly significant because it offers a variety of sign sites and because facial expressions have grammatical connotations that must be taken into account. A true sign language translation cannot ever come from a project that ignores the face.

- *Engagement of the Deaf Community:*

No good product can be made without the input of the ultimate end customers, and this is especially true in this particular sector. This is one true evidence that differentiates dreamers from serious competitors. The team will quickly understand the two reasons mentioned above among many others if there are deaf participants in the project. The inclusion of deaf team members, in our opinion, is the key to overall success.

➢ *Non Functional Specifications:*

- Only one person should use the Leap motion controller to do ISL. A dual-core 2.66 GHz or faster processor, either 32 bit (x86) or 64 bit (x64), should be able to operate our system. It shouldn't have more RAM than 2 GB.
- For the time being, we'll control the input stream using a Leap motion controller and write C code in Visual Studio. Real-time continuous gesture recognition techniques will be the foundation of the software architecture.

- *Performance and scalability*. How fast does the system return results? How much will this performance change with higher workloads?
- *Portability and compatibility*. Which hardware operating systems, and browsers, along with their versions does the software run on? Does it conflict with other applications and processes within these environments?
- *Reliability, maintainability, availability*. How often does the system experience critical failures? How much time does it take to fix the issue when it arises? And how is user availability time compared to downtime
- *Security*. How well are the system and its data protected against attacks?
- *Localization*. Is the system compatible with local specifics?
- *Usability*. How easy is it for a customer to use the system?

IV. EXISTING SYSTEM

A module for the existing system's dumb person feature was created utilizing flex sensors, which the user's hand is attached to. The flex sensor on this module responds to each finger bend separately. That value is used by the controller to start speaking. The APR Kit has recorded a separate voice for each flex sensor, and it will play that voice for each indication. Additionally, in another system now in use, precision is only achieved by working on a select few alphabets rather than words or complete phrases.

➢ *Limitations of Existing System:*

- In existing system its restricted to only 10 voice announcements it may reduce product capacity
- One of the major problem of the existing system is Dumb person should always carry the hardware with him
- User cant do any other work with flex sensor on fingers and also sensors placed straight
- The controller may think that the user is giving command and finally it may result in unwanted results and less hardware lifetime

V. PROPOSED SYSTEM

A gesture or sign picture should be sent to the system in the proposed system by a dumb or inept person. The system evaluates the sign input using a MATLAB image processing technique before classifying the input to the recognized identity. The machine then begins the speech media when the input picture matches the supplied datasets. A written version of the output will also be shown. The concept of converting sign language into speech and text has a functional prototype. This work develops an application that will aid society in enhancing communication between deaf and mute people using an image processing technique. We made advantage of the free, publicly available American Sign Language (ASL) data set from MNIST, which is available on Kaggle.This dataset contains 7172 test photos and 27455 training images, each of which is a square of 28 pixels by 28 pixels. The 25 classes of the English alphabet, from A to Y, are represented by these images. (Z has no class designations because of gestures.) The Kaggle dataset's training data is given in CSV format and has 27455 rows and 785 columns. The first column of the dataset contains the class name for the image, while the other 784 columns comprise the 28 × 28 pixels. The test set of data follows the same paradigm.

➢ *Advantages of Proposed System:*

- When comparing with existing system user can give more signs
- The module provides two way communications which helps in easy interaction between normal people and disables
- Easy to interface
- Flexible to work.

VI. SYSTEM ARCHITECTURE

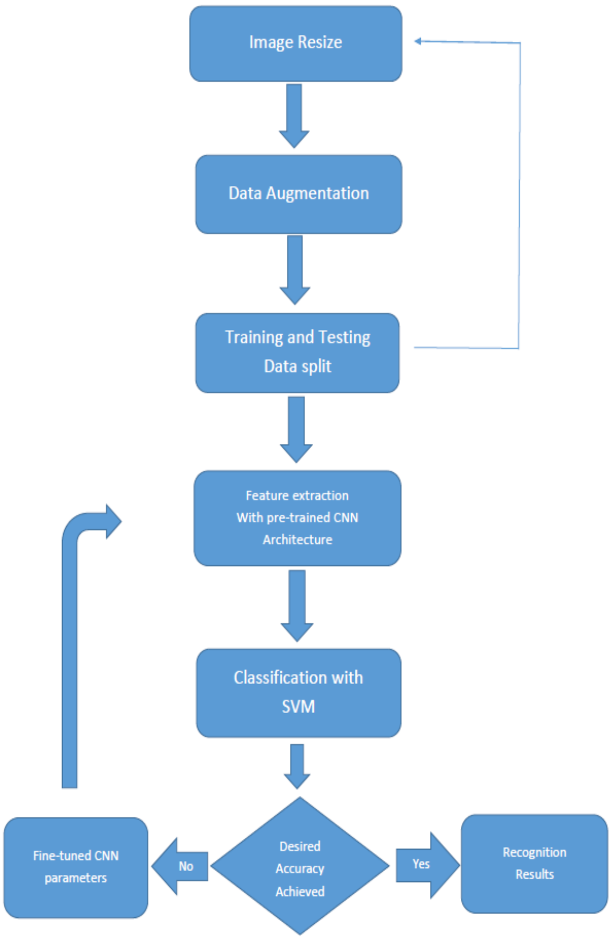

Fig 5 System Architecture

➢ *Future Scope:*

In order to expand the model's recognition of alphabetical features while maintaining high accuracy, we intend to use more alphabets in our datasets. In order to help blind people, we would also like to improve the system by incorporating speech recognition. In order to communicate, more than 70 million deaf people worldwide use sign language. They can learn jobs, access resources, and participate in their communities by using sign language.

VII. CONCLUSION

In the fields of artificial intelligence, machine learning, and computer vision,many advances have been made. They have significantly improved how we use their techniques in our daily lives and how we see the world around us.Numerous studies have been conducted on the recognition of sign gestures using different algorithms such as ANN, LSTM, and 3D CNN. However, the majority of them demand more processing capability. On the contrary, our study has a surprising accuracy of over 90% while requiring little computing power. In order to obtain features (binary pixels) and enhance the system, we proposed in our research to normalize and dynamically resize our photos to 64 pixels. CNN is used to categorize the ten alphabetical American signs.

REFERENCES

[1]. https://peda.net/id/08f8c4a8511
[2]. K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4896-4899, doi: 10.1109/ BigData.2018.8622141.
[3]. CABRERA, MARIA & BOGADO, JUAN & FermÃ-n, Leonardo & AcuÃ±a, Raul & RALEV, DIMITAR. (2012). GLOVE-BASED GESTURE RECOGNITION SYSTEM. 10.1142/9789814415958_0095.
[4]. He, Siming. (2019). Research of a Sign Language Translation System Based on Deep Learning. 392-396. 10.1109/AIAM48774.2019.00083.
[5]. International Conference on Trendz in Information Sciences and Computing (TISC). : 30-35, 2012.
[6]. Herath, H.C.M. & W.A.L.V.Kumari, & Senevirathne, W.A.P.B & Dissanayake, Maheshi. (2013). IMAGE BASED SIGN LANGUAGE RECOGNITION SYSTEM FOR SINHALA SIGN LANGUAGE
[7]. M. Geetha and U. C. Manjusha, , A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation, International Journal on Computer Science and Engineering (IJCSE), vol. 4, no. 3, pp. 406-415. 2012.
[8]. Pigou L., Dieleman S., Kindermans PJ., Schrauwen B. (2015) Sign Language Recognition Using Convolutional Neural Networks. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision – ECCV 9.2014 Workshops. ECCV 2014. Lecture Notes in Computer Science, vol 8925. Springer, Cham. https://doi.org/10.1007/978-3-319-16178-5_40
[9]. Escalera, S., BarÃ³, X., GonzÃ lez, J., Bautista, M., Madadi, M., Reyes, M., . . . Guyon, I. (2014). ChaLearn Looking at People Challenge 2014: Dataset and Results. Workshop at the European Conference on Computer Vision (pp. 459-473). Springer, . Cham.
[10]. Huang, J., Zhou, W., & Li, H. (2015). Sign Language Recognition using 3D convolutional neural networks. IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). Turin: IEEE.

BIOGRAPHIES

Dr.Sk.Mahaboob Basha is presently working as Professor in the department of Information Technolgy at NRI Institute of Technology, Vijayawada. He received his M.Tech degree from Jawaharlal Nehru Technological University, Kakinada(JNTUK) and Ph.D in Computer Science and Engineering from Acharya Nagarjuna University(ANU). He has published over 10 research papers in International Conferences and Journals. He has more than 20 years of experience in teaching..

H.C.Srivalli is currently studying B.Tech with specification of Information Technology in NRI Institute of Technology. She done a mini project on Sign language Classification and she has done 2 NPTEL certifications She has finished an internship at black bucks.

B.Tech with specification of Information Technology in NRI Institute of Technology. She done a mini project Sign language Classification and She has finished an internship at Verzeo and black bucks and one NPTEL certification

C.V.Basanth is currently studying B.Tech with specification of Information Technology in NRI Institute of Technology. he did a mini project Sign language Classification. He earned two NPTEL certificates in total. After serving as an intern at Black Bucks