IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

VIRTUAL REALITY'S POTENTIAL TO AID HEARING IMPAIRED

Nupur Joshi[1] • Janvi Gupta[2] • Nikhil Duseja[3] • Swati Gupta[4]
Nikita Nijhawan[5]

[1]Student, [2]Student, [3]Student, [4]Student, [5]Assistant Professor
[1]CSE,
[1]GGSIPU, Delhi, India

Abstract: This paper aims to bridge the gap between the hearing impaired and normal people. This paper focuses on building a user friendly system that can help in generating the gestures through digital humanoid using audio input as well as to recognize the sign gestures. Since the preferred form of communication for deaf and hearing impaired persons worldwide is sign language. However, communication between a person with language impairment and a normal person has never been easy. This will enable
the user to enable two-way communication between hearing impaired and normal people without the use of a translator.

Index Terms **- Indian Sign Language (ISL), Avatar, Humanoids, KNN, naive bayes algorithm,CNN.**

I. INTRODUCTION

Deaf and dumb people utilize sign language as their native speech, which is a visual language. A person's thoughts can be dynamically expressed using sign language rather than through the use of acoustically transmitted sound patterns. It can be used to communicate with persons who have hearing impairments by those who have trouble speaking, those who can hear but cannot talk, as well as by regular people. It is accomplished by fusing together hand gestures, arm and body movements, and face expressions all at once. It can be used to communicate with persons who have hearing impairments by people who have trouble speaking, by people who can hear but cannot talk, and by normal people.

1.1. Existing System

The voice disorder and hearing loss people use charts to recognize sign languages and then identify the sign. There are some of the applications that are used to recognize the gestures with an AI interaction video platform that identifies the hand gesture recognition with different types of sign languages as text.

1.2. Proposed System

The system proposed is a major solution for communication between deaf and normal people. The system is equipped with two functionalities:

- Audio to sign language conversion using a digital humanoid
- Sign language recognition

The main purpose of this system is to recognize the finger spelling-based hand gestures in order to form a complete word by combining each gesture and to convert the audio to the gestures.

II.PROBLEM IDENTIFICATION

The deaf community still has relatively limited access to a significant portion of the digital world, despite recent developments like the internet, smartphones, and social networks, which enable people to quickly connect and exchange knowledge at a global level. As deaf people don't have early access to sign language, they frequently experience delays in the acquisition and development of their language.

III.METHODOLOGY

3.1 Dataset Creation

We have taken the dataset of sign recognition from Kaggle, and the dataset of audio to sign language recognition was made by making videos with the help of a JavaScript library.

3.2 Gesture Classification

This model deals with dynamic aspects of gestures. Gestures are extracted from a sequence of video pictures. The dataset had the images which were Gaussian blur so that boundaries are clearly visible and we get a good result[3]. Then this dataset is trained using CNN and a model is built for the prediction purpose.

3.3 Audio Recognition

For the functionality of audio to sign language conversion, audio recognition is the most crucial step. We used Pyaudio, Speech Recognition, and the Google Speech API to accomplish this. After this, import the necessary package (import speech recognition as sr) and create the object[3]. The microphone module will now accept the voice as input

3.4 Design Hierarchy

Fig.1 shows the flow of the whole process, as it:

- Begin by either activating the camera to capture the gestures in the frame or activating the microphone to capture audio from the user.
- If input is from a camera, then the hand gestures that are seen by the camera get fed into the model, which then compares the gestures with the dataset that it has trained on and if input is taken using audio, then the system will match the video to the corresponding audio.
- In the case of camera input, if a match of the acquired gesture is found, the corresponding text to the gesture will be displayed. And in the case of audio input, if a match is found, the avatar will perform the sign for the corresponding input. To test the accuracy of the model we have used three different algorithms.

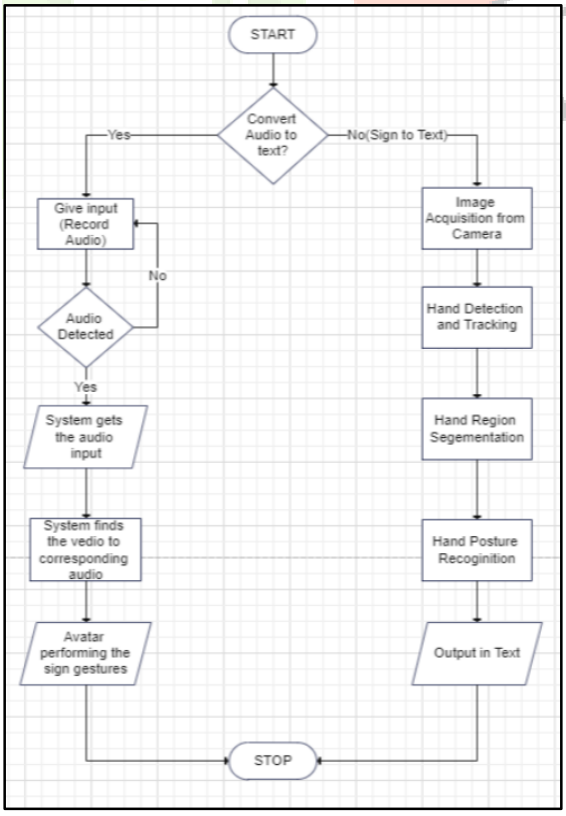

Fig. 1. Design Hierarchy

IV. ALGORITHM USED

4.1 Naïve-Bayes

A group of classification algorithms built on the Bayes' Theorem are known as Naive Bayes classifiers. It is a family of algorithms rather than a single method, and they are all based on the idea that every pair of features being classified is independent of the other [9].

During the training process, the algorithm calculates the probability of each word given the visual features of the sign language gestures. These visual features may include the position and orientation of the hands, the movement of the fingers, and the facial expressions of the signer.

The formula for Bayes' theorem is given as:

$P(A|B) = P(B|A)P(A)/P(B)$

Where,

P(A|B): The posterior probability measures the likelihood that a given hypothesis (A) will really occur.

P(B|A):It stands for Likelihood Probability, which measures how likely it is based on the evidence at hand that a given hypothesis is correct.

P(A), Prior Probability: Probability of hypothesis before observing the evidence

P(B), Marginal Probability: Probability of Evidence.

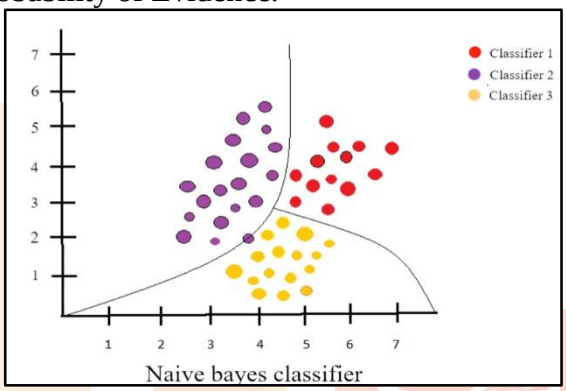

Fig. 2. Naive Bayes Classification

4.2 CNN

CNN is a deep learning algorithm that can automatically extract relevant features from the input data, making it highly effective for image recognition tasks. In the case of sign language recognition, CNN can be trained on a large dataset of sign language gesture images and their corresponding spoken words or phrases[7].

A convolutional neural community is a category of neural networks that has a grid-like topology that enables in processing the data [6]. Fig.3[8] contains a series of pixels that contain pixel values that denote how bright and what color each pixel should be.

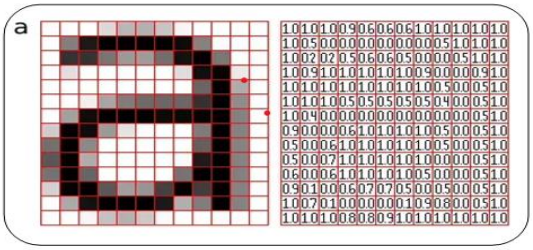

Fig. 3. CNN

It has three layers -

- Convolution layer
- Pooling layer
- Fully connected layer

4.2.1 Convolution Layer

This layer is the core layer of CNN. This layer plays a dot product among matrices, wherein one matrix is the set of learnable parameters and the other matrix is the restricted portion of the receptive field[4]. Fig.4[8] shows a two-dimensional representation of the image known as an "activation map" that gives the response of the kernel at each spatial position of the image.

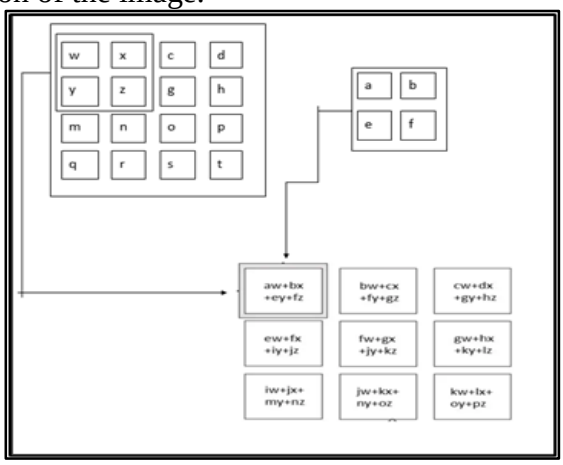

Fig. 4 Convolution Layer

4.2.2 Pooling Layer

This layer replaces the output of the network at certain locations by deriving the summary statistics of nearby outputs[6]. This helps in reducing the spatial size, which decreases the required amount of computation and weights.

The polling operation is processed on every slice of the representation.

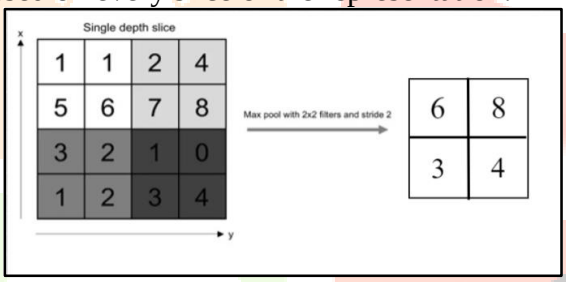

Fig. 5. Pooling Layer Operation

4.2.3 Fully Connected Layer

In this layer, all neurons are fully connected to each other [6]. That is why it can be calculated as usual by a matrix multiplication followed by a bias effect.

This layer helps map the representation between the input and output.

Our convolutional neural network has the following architecture:

[INPUT]
→[CONV 1] → [BATCH NORM] → [ReLU] → [POOL 1]

→ [CONV 2] → [BATCH NORM] → [ReLU] → [POOL 2]
→ [FC LAYER] → [RESULT]

4.3 KNN

A level-1 heading must be in small caps, centered and K Nearest Neighbor is a machine learning algorithm that determines the similarity between the new data entry and the available data entries and puts the new data into the category that is closest to the available categories.

KNN algorithm can be used in a sign to audio converter to recognize and interpret sign language gestures made by the user. The algorithm can be trained on a dataset of images or videos of sign language gestures with their corresponding spoken words [1].

The algorithm stocks all the available data and classifies a new data point based on closeness, which means that when new data appears then it can be classified into a proper category using KNN algorithm [5].

The KNN algorithm consists of following steps:

Step-1: Select K neighbours and calculate their Euclidean distance.

Step-2: Select the K nearest neighbours using the calculated Euclidean distance.
Step 3: Count the number of data points from these K neighbours that fall into each category.
Step-4: Allocate the new data points to the category for which there are a maximum neighbour.
Step-5: The model is ready.

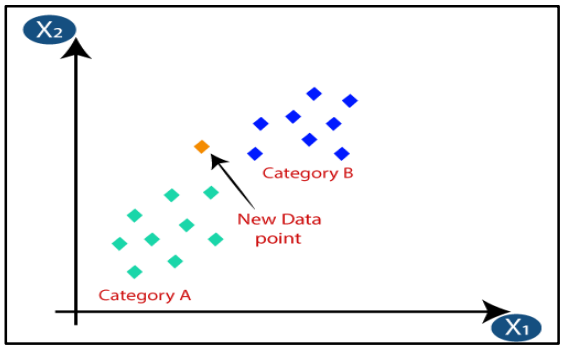

Fig. 6. KNN Classification Adapted from [5]

V. EXPERIMENT USED

The dataset is divided into two sets: a training dataset and a testing dataset. The system is trained to recognize 26 signs. Three classifiers, CNN, KNN, and Naive Bayes, have been used to check the accuracy of the images. The results showed that CNN performed better with fewer features.

5.1 CNN Performance

CNN performance accuracy curves are an important tool for evaluating and improving the performance of CNN models by analyzing these curves, researchers and practitioners can identify potential issues with their models and make changes to improve their accuracy and generalizing ability.

One layer with 32 filters and window size 3*3

The figures from 5 to 10 show the results of CNN, in which we got an overall accuracy of greater than 95% on the training set at the last epoch and a testing accuracy of more than 97%. The total epochs are 10. There was a training loss of 0.0143 in the last epoch of 0.0143 in the last epoch.

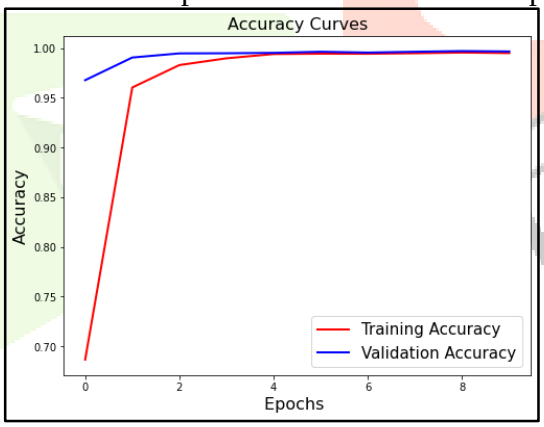

Fig. 7. Accuracy Curve(one layer)

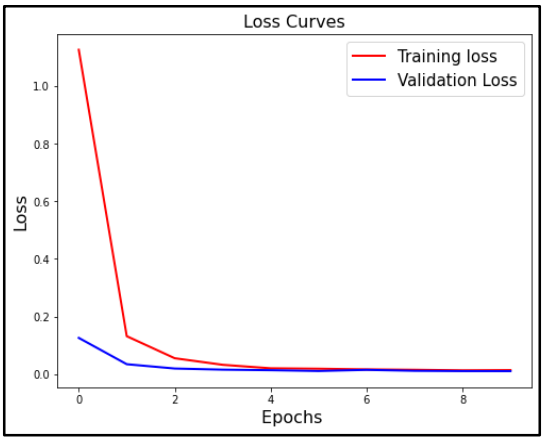

Fig. 8. Loss Curve (One Layer)

Fig.9 and Fig.10 shows the two layer with 32 filters and window size 3*3.

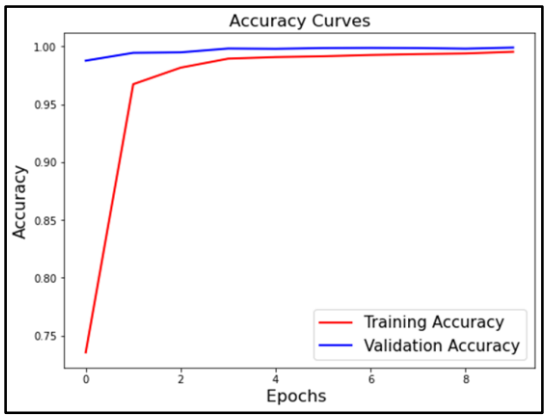

Fig. 9. Accuracy Curve(Two Layer)

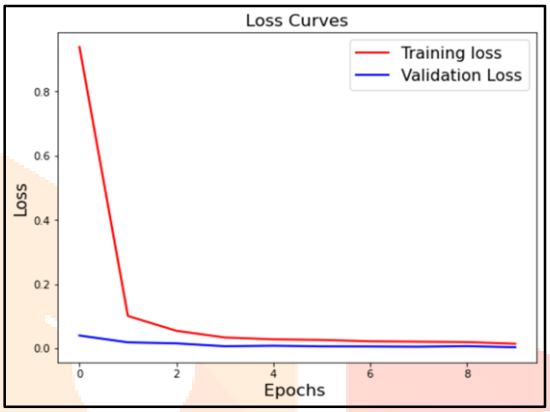

Fig. 10. Loss Curve(Two Layer)

Fig.11 and Fig.12 shows the three layer with 32 filters and window size 3*3.

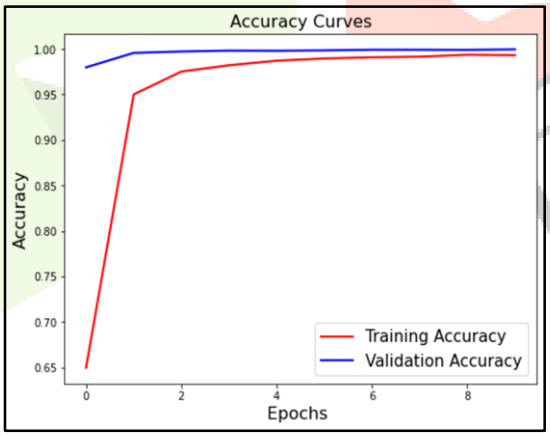

Fig. 11. Accuracy Curve(Three Layer)

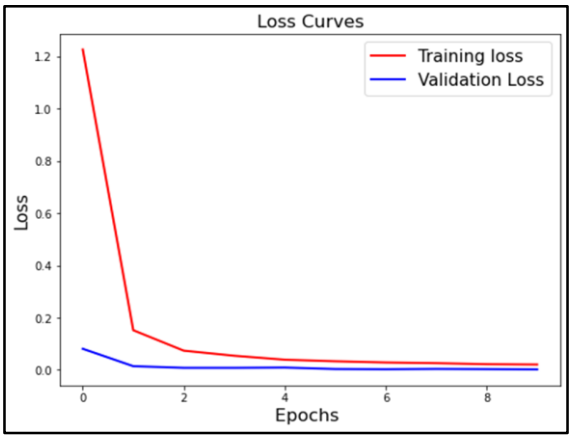

Fig. 12. Loss Curve(Three Layer)

5.2 KNN Performance

Figure 11 shows the KNN results, which have an average accuracy of 92%. It is observed that an increasing K value results in a decreasing accuracy score.

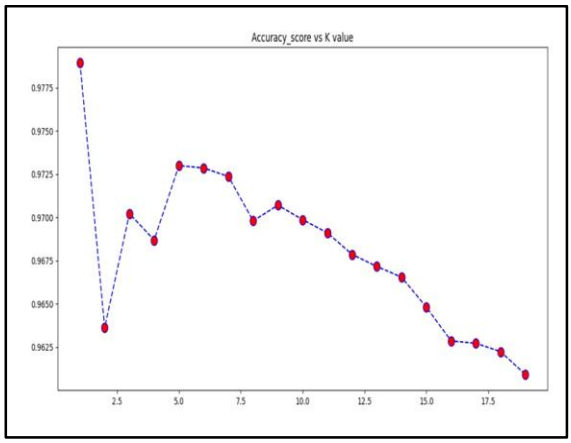

Fig. 13. KNN Accuracy Curve

5.3 Naïve Bayes Performance

The confusion matrix in Fig.12 shows the true positive on the x-axis and the false positive on the y-axis. It provides an average accuracy of 77% and precision of 92% when tested using the data.

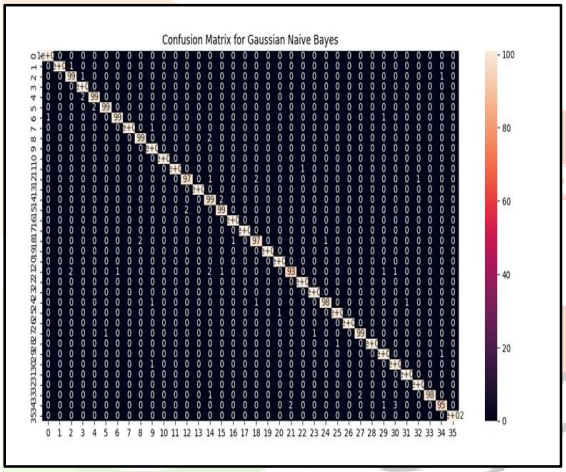

Fig. 14. Naive Bayes Confusion Matrix

Accuracy = (TP+TN)/(TP+FP+FN+TN)

where,

TP= True positive

TN= True negative

FP= False positive

FN= False negative

The Fig. 13 shows the accuracy comparison of all the classifiers:

Table 5.3: Comparison Table

Evaluation Criteria	CNN	KNN	Naïve Bayes
Precision	99.32	92.63	92
Recall	99.32	92.09	86
F measure	99.32	92.27	89
Accuracy	98.57	92.25	77

VI. CONCLUSION

A system that would be useful for disabled people who have communication difficulties by creating a system that would allow them to express themselves clearly and easily. Our model successfully converts the entire input sentence into a single visual rather than depicting different words through an avatar, giving the model a much more realistic and lively appeal. It also has real-time vision-based ISL recognition that has been developed for the ISL alphabets. By using advanced technologies such as computer vision and natural language processing, this system can accurately interpret sign language gestures and translate them into written text, as well as take written text and convert it into sign language gestures.

This system can enhance accessibility and inclusivity for the deaf community and improve communication in various settings, such as education, healthcare, and social interactions.

However, it is important to continue improving the accuracy and effectiveness of this system through ongoing research and development.

We achieved an accuracy of 98.57% on our dataset.

REFERENCES

[1] Adeyanju, I., Bello, O., & Adegboye, M. A. (2021). Machine learning methods for sign language recognition: A critical review and analysis. Intelligent Systems With Applications,12, 200056. https://doi.org/10.1016/j.iswa.2021.200056

[2] Brownlee, J. (2019, July 5). A Gentle Introduction to Pooling Layers for Convolutional Neural Networks. MachineLearningMastery.com. https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks/

[3] I.(n.d.).Sign Language Recognition and Converting into Text. IJRASET. https://www.ijraset.com/research-paper/sign-language-recognition-and-converting-into-text.

[4] Katoch, S., Singh, V., & Tiwary, U. S. (2022). Indian Sign Language recognition system using SURF with SVM and CNN. Array, 14, 100141. https://doi.org/10.1016/j.array.2022.100141.

[5] K-Nearest Neighbor(KNN) Algorithm for Machine Learning - Javatpoint.(n.d.).www.javatpoint.com. https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning.

[6] Mariappan, H. M., & Gomathi, V. (2019b). Real-Time Recognition of Indian Sign Language. Computational Intelligence. https://doi.org/10.1109/iccids.2019.8862125

[7] Mehta, A. (2021). Automatic Translate Real-Time Voice to Sign Language Conversion for Deaf and Dumb People. IJERT. https://doi.org/10.17577/IJERTCONV9IS05037.

[8] Mishra, M. (2021, December 15). Convolutional Neural Networks, Explained - Towards Data Science. Medium. https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939.

[9] Ray, S. (2023, March 3). Naive Bayes Classifier Explained: Applications and Practice Problems of Naive Bayes Classifier. Analytics Vidhya.https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/