Indian Journal of Computer Science and Technology
Volume 2, Issue 3 (September-December 2023), PP 01-07.
https://www.doi.org/10.59256/indjcst.20230203001
www.indjcst.com
ISSN No: 2583-5300

Ml Based Hand Gesture Recognition

V V Nivedha[1], KS Suganya[2]
[1]Department of Computer Science and Engineering, KCG College of Technology, Chennai, India.
[2]Assistant professor, Department of Computer Science and Engineering, KCG College of Technology, Chennai, India.

Abstract: The projects focus on Gesture recognition using the media pipe hands Lite framework where the custom data set of 10 gesture Images is trained with an in built media pipe hands model which contains 2CNN models-A palm detection model running on single Shot detection architecture and a hand landmark generator running on regression model architecture. The data set is successfully trained and tested with the proposed method and an accuracy of 98 percent is obtained.

Keywords: Gesture Recognition, Media pipe, Hands De-tection, CNN, Regression Model, Hand Landmark Detection, Machine Learning, Image Processing.

I.INTRODUCTION

In today's world, being part of something that constantly has an upgrade in life, something to always Update yourself on, and most importantly the impact technology has on our everyday life is just absolutely ecstatic. I have always been amused of how useful technology can be in Our day today lives. Machine learning is one such Boone in the field of technology. In this project, we would be implementing gesture recognition-a highly interesting and important topic. There have been various research methods and algorithms proposed to recognize gestures using machine learning. We will be looking at one such method in this paper, gesture recognition using the frame work media pipe. We would be preparing a custom hand gesture data set and then train the media pipe hands model to give us recognized gestures. This project helps the hearing and speech impaired people communicate effectively with others by showing the gestures in front of the webcam when the code is being run. This inturnenables thenon-speech/hearing impaired person to communicate and understand with more clarity and precision. Although a lot of methods have been proposed to enable the sign language recognition, there are not many in terms of emotions-gesture recognition.

II.RELATEDWORK

In our pursuit of gaining comprehensive in sights in to Sign Language Recognition and the various techniques employed in this domain, we conducted a thorough review of several research papers. These papers collectively contribute tothe over arching theme of our project:

a) Kumudtripathi, Nehabaranwal, And g.C.NANDI(2015)[1]proposed a continuous Indian Sign Language (ISL) gesture recognition system, addressing the chal- lenge of recognizing sign language gestures with in continuous sequences. They employed agradient-based key frame extraction method and achieved improved accuracy by using features obtained through Orientation Histogram (OH)and Principal Component Analysis (PCA).:

b)Archanas.Ghotkarandgajanank. KHARATE (2015) [2] delved into dynamic hand gesture recognition for Indian Sign Language and developed algorithms for word recognition, including rule-based and Dynamic Time Warping-based methods. Their innovative approach to sentence interpretation using inverted indexing tackled the challenges of continuous sentence recognition.

c)M.M.GHARASUIEANDH.SEYEDARABI(2013)[3] introduced a real-time system for recognizing dynamic hand gestures using Hidden Markov Models (HMMs). They successfully recognized key gestures related to English numbers in real-time, distinguishing between key and link gestures.

d)KAIRONGWANG,BINGJIAXIAO,JINYAOXIA,ANDDANLI(2015)[4]presented a dynamic hand gesture recog- nitional gorithm that utilized an improved Code book(CB) modeling method and spatial moments for feature extraction, achieving satisfactory recognition accuracy.

e)H.FRANCKE,J.RUIZ-DEL-SOLAR,ANDR.VERSCHAE(2014)[5]proposed arobustreal-time hand gesture detection and recognition system that employed boosted classifiers, skin segmentation, and hand tracking. Innovative training techniques, such as active learning and bootstrap, significantly improved system performance.

f)HARIPRABHATGUPTA,HARESHSCHUDGAR,ANDSIDDHARTHA(2016)[6]addressed continuous hand gesture recognition for Human-Machine Interaction (HMI)using accelerometerandgyroscopesensorsin smart devices. Their approach included gesture coding and spotting algorithms, facilitating the recognition of predefined gestures.

g)NOORTUBAIZ(2015):[7]introduced a glove-based Arabic sign language recognition system using sensor-based data gloves. This approach achieved high sentence recognition rates while eliminating the limitations of vision-based systems.
h)M.K.BHUYAN(2012)[8]proposed a gesture recognition approach based on finite state machines(FSMs) and video object planes.This method reduced gesture video sequences into representative key frames for efficient recognition.

i)JOYEETASINGHAANDKARENDAS(2015)[9]presented an automatic system for recognizing Indian Sign Language alphabets in continuous video sequences, achieving a high recognition rate through a combination of preprocessing, feature extraction, and classification techniques.

j)SUHARJITO, GUNAWAN,ANDH.THIRACITTA (2018)[10]explored the use of a modified Convolutional Neural Network (CNN) model for Sign Language Recognition, in corporating transfer learning from Action Recognition. While achieving high training accuracy, the paper acknowledged challenges in validation accuracy. These referenced articles collectively contribute to the extensive body of knowledge in Sign Language Recognition, forming the foundation for our project's research and development.

III.PROPOSEDSYSTEM

A. Techniques for recognition and training

1) Machinelearning: Machine learning is an art of computer science that comes under Artificial intelligence. As illustrated in the figure1, This is used in the development of analytical models that use different form so fin put data increating meaningful and understand able in sights. Machine learning models constantly learn from the data to find the optimal point of accuracy in its outputs.

Atitscore, machine learning involves training a model on a data set, allowing it to learn patterns and relationships in the data. The model can the nusethislearning to make predictions or decisions on new data that it has not seen before. There are several different types of machine learning, including supervised learning, unsupervised learning, and reinforcement learning.

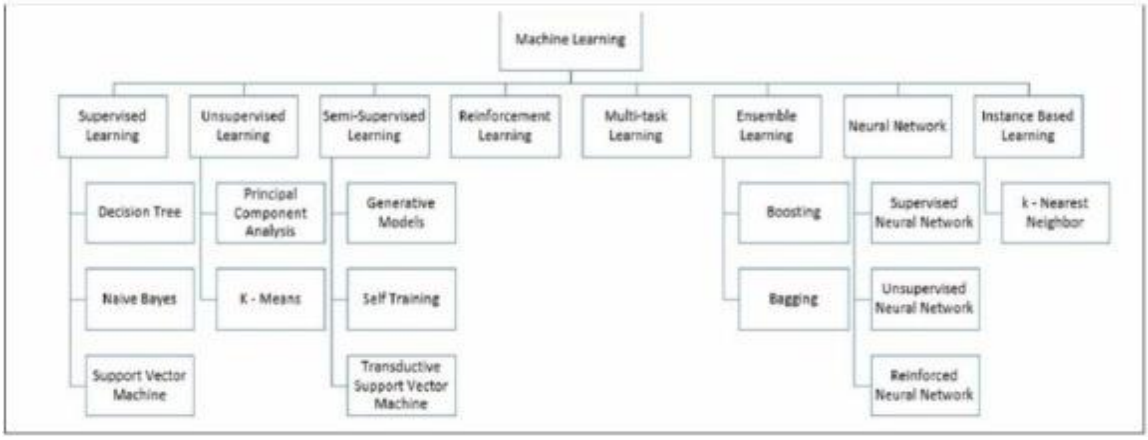

Fig.1. MachineLearning.

2)Convolution neural network: The convolutional neural network (CNN) is a basic ML model which uses in the images, the training input data sets and the nassigns weights to each and every parameter based on the situation considered and thus, based up on the assigned scores it differentiates one image from the other. These weightsdenote the importance of an object in an image.

This Image recognition feature performed by the CNN can be brokend own in to fewer steps. To begin with, the Convolutional layer captures all the low-level features such as the edge detection, sharpness, brightness, color exhibited by the image and other observable features. Then, when there are further more layers added on to pof this, the architecture can read the high-level features as well, which helps the model understand every aspect of the picture considered along with its parameters of consideration add formula for CNN. There are multiple layers in the CNN: Convolutional layer, Pooling
Layer, Fully Connected layer.

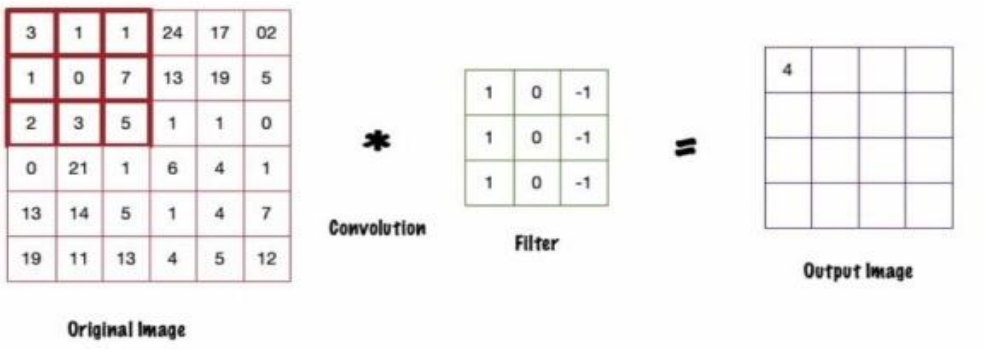

Fig.2.ConvolutionLayer

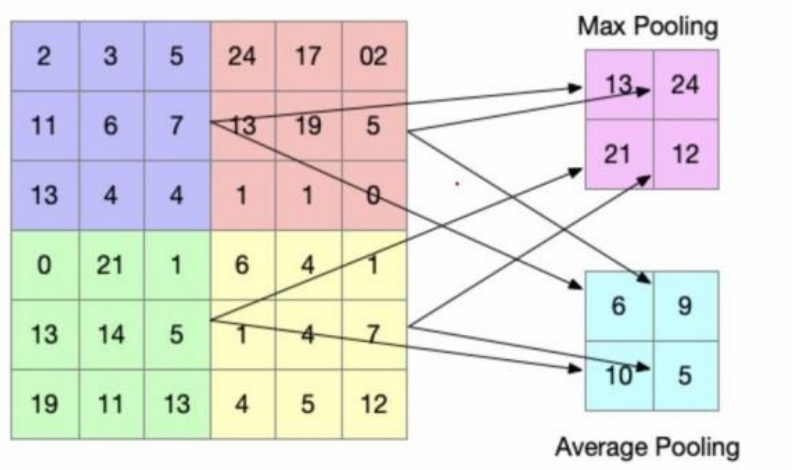

Fig.3.PoolingLayer

Convolutional layer:

The convolutionallay erasillustrated in Figure2, is a part of CNN in which the majority of the task liesin recording and analyzing the characteristics in a given input data set. It will have an output data set, which will be high if among the two matrices considered the one having a high value is in the same place. This calculation is because the layer functions by considering a dot product of the filter with the random size of the input image. As a result, dot product of the output can inform us if the pixel pattern in the under lying image corresponds to the pixel pattern described by our filter or just a dot product of the weights.

Pooling layer:

The Pooling Layer, as illustrated in Figure 3,Similar to the way dictionary mapping works, image processing also works by having multiple feature maps where the entire image will be replicated and have a mapping structure of the original input data image. Whenever there is an operation of the convolution allayer on these data sets, these feature maps will associate the input data with the filter. They are majorly used in performing two features in image recognition: Max pooling and Average pooling.

Fully connected layers:

These layers are usually not very significant but also at the same time is used in improving the overall class scores. They are contained at the end of the network layer and stays hidden. These layers kick in as soon as the Pooling layer data maps the input data with the output and the final image recognition model is ready to go. If the layer finds some anomalies in the mapping, it will optimize the scores of various parameters and either re-run the algorithm or adapt the model to the new scores as illustrated in the Figure 4.

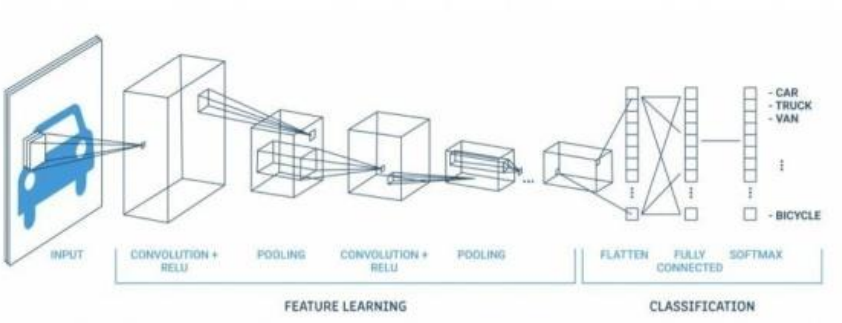

Fig.4.FullyConnectedLayers

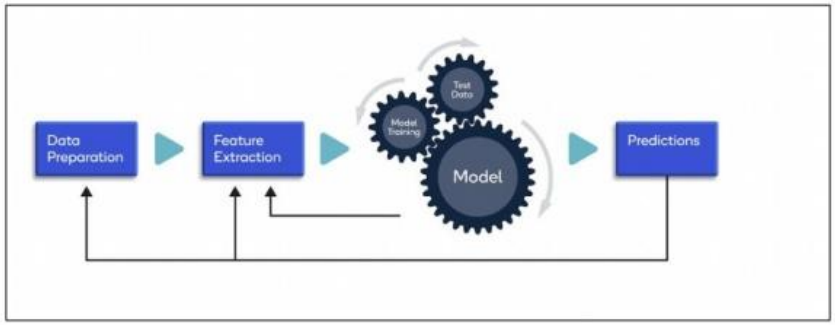

Fig.5. Image Processing Structure

B. Image processing

Image processing is a crucial technique used to analyze and extract valuable information from images, playing a pivotal role in various industries undergoing automation during the fourth industrial revolution. Many companies are turning to machine learning for its precision and cost-cutting potential. This technology finds applications in tasks like image information extraction, object classification, categorization, visualization, and pattern recognition.

Today's advanced cameras and image-capturing devices rival human perception. Smart phones employ facial recognition for real-time 3D object over lays, while cars use 360 degree cameras for constant environmental monitoring and autonomous driving. Digital image processing is the preferred method due to its speed, allowing computers to process raw image data swiftly, identify features, characteristics, masks, and points within images. In contrast, analog methods are slower and less efficient.

The focus here is on digital image processing, particularly using Convolutional Neural Networks (CNNs) due to their effectiveness. CNNs address a common challenge in image processing, where input data must match the model's resolution standards to prevent excessive hidden layers and parameter complexity. CNNs employ three key layers: Convolution, Pooling, and fully connected layers.

The specific projects cope revolves around gesture recognition, utilizing CNNs to achieve the desired results. As human gestures are the input, this topic is explored further. The image processing work flow involves these layers, ensuring efficient and accurate recognition of gestures in a wide range of applications.

An illustration of the Image Processing Structure is shown in the Figure 5.

C. Gesture recognition

Gesture recognition involves identifying human gestures and triggering actions accordingly, from symbol depiction to movement or even dancing. This process captures and analyzes complex gestures, interprets them, and translates them into action able commands. Instead of traditional input methods like typing, gestures can be used, with motion sensors in cameras translating the actions into machine language for ML models. The success of gesture recognition relies on the chosen pipe line or architecture. The project employs Google's efficient and hassle-free Media pipe frame work, utilizing hand tracking neural network pipe lines for accurate gesture recognition.

1) STEPS INVOLVED IN RECOGNITION: The process of gesture recognition involves several key steps to ensure the development of an accurate machine learning model for image recognition.

Creating a Dataset: The first crucial step is to prepare and curate the right datasets. These datasets should have images with resolutions compatible with the machine learning model's requirements. Two essential types of datasets are used:

1) Training Data: This labelled dataset is used to train the machine learning model. The model uses this data to learn and adjust its parameter site ratively. Typically, this dataset accounts for up to 80 percent of the total data.

2) Testing Data: This dataset is used to evaluate the trained model's performance on unseen data. Adjustments and fine tuning are done based on the model's performance with this dataset.

Choosing a Model: The choice of machine learning model depends on the specific problem, type of dataset and desired outputs. Various machine learning algorithms are available, such as K-means for unlabeled data, regression for prediction, and categorical classification. In real-world scenarios with large datasets and complex problems, there is no one-size- fits-all model. Parameters, variable types, and data nature determine the suitable algorithm.

Supervised Learning: This paradigm involves the model learning from input-output pairs through historical data. It encompasses regression for continuous output and classification for discrete datasets. Common algorithms include Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN), line arregression, and decision trees. Supervised learning excels when there isample input data, as the model refines its parameter stomatch historical out puts.

In the context of this project, CNN models have proven highly effective for image processing and classification, specifically for gesture recognition. These models learn to recognize gestures by comparing input images to labelled historical data, making them a suitable choice for this application.

D. Architectures

Single Shot Detector Model (SSD): The SSD model is a powerful and efficient solution for object detection across multiple categories. It comprises two main components: the back bone model, which extracts features from the input data using a pre-trained image classification network, and the SSD head, consisting of stacked convolutional layers that generate bounding boxes around objects. The SSD model detects different object classes and a spectratiosa teach grid in its feature map. It assigns cores to these detections and applies non-maximal suppression to obtain the final detections. The training of SSD involves optimizing loss function that in corporate localization and classification components, making it arobust choice for object detection tasks.

Regression Model: Regression models, including neural network regression, are statistical models that predict relationships between dependent and independent variables. They use lines or curves to estimate future values based on input data. Neural network regression introduces non-linearity by mapping node outputs to multiple values, making it suitable for complex prediction tasks. Each neuron in the model is connected too thersthrough weighted connections, allowing them to collectively determine outputs efficiently. This approach is useful for applications like gesture recognition, where predicting precise out comes based on input data is essential.

E. Train the algorithm to construct the Model

Training the machine learning algorithm is a crucial step in the model development process. It involves splitting the data set into training and testing portions, and then feeding the training data in to the chosen algorithm. The model iteratively adjusts its parameters untilit achieves high accuracy. In supervised learning, the model learns from labeled input- output pairs, minimizing differences between predicted and documented results. The process aims for minimal human intervention, focusing on selecting the right model parameters and refining the input data set to optimize results. The model's performance is evaluated by comparing its outputs to the expected results, allowing for continuous improvement.

F. Testing and improvisations

Once the model has been trained successfully, the next important step is to use the input dataset and test if the model produces the required outputs. It is also important to check if the model that we have developed using the test and train can handle the adversarial data, as what seems normal to a human eye might cause the biggest misclassification to a machine learning model. Once we check the testing data with the outputs, we can then check for the output accuracy and model performance using various factors.

For testing our model, we use the following formulae and calculate the accuracy. One can say that a model can perform well, by checking the accuracy value. We would be checking the accuracy of our model for the custom data set based on the formula given below:

There are a few parameters to be taken in to consideration while calculating the accuracy:

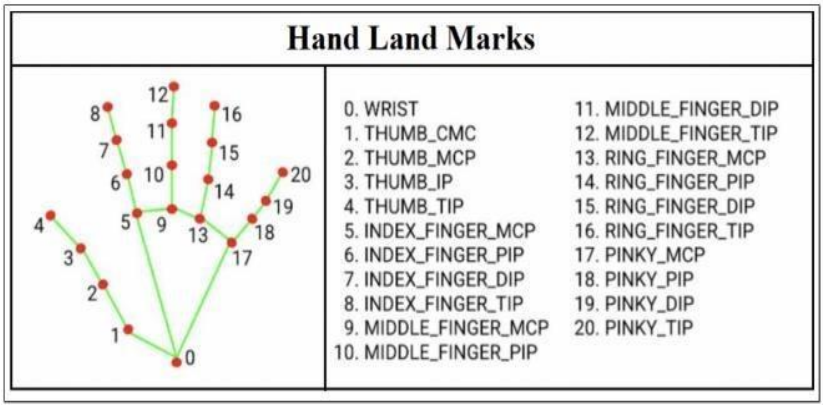

Fig.6. Hand Lank Marks

Precision-Precision gives the percentage of correct positive results out of the total positive results predicted by the model precision=true positives/ true positives +false positives

Recall-it Recall gives the percentage of correct positive results out of all positive results

Recall=true positives/ true positive + False Positives

G. Library function

Media pipe

The ability to perceive the coordinates and movement of hands is a very important aspect in improving the success rate and efficiency of data modeling. In our case, it is highly crucial for signal nguage understanding and hand gesture recognition. Media pipe is a frame work by Google that enables open-source cross-platform for various machine learning methods and models in Image processing. It has detection methodologies built for various action tracking. However, we will be using the frame work pipe line in hand gesture recognition from Media pipe.

Hands–Media pipe

Media Pipe Hands utilizes an ML pipe line consisting of two models working together. We are using the hand tracking neural network pipe line-lite on our custom data set to predict

2D and 3D handl and marks on our custom images.

The proposed model contains two in built CNN models- A palm detection model which uses the singles hot detection model architecture and A hand land mark generator which uses a regression model architecture.

Methodology of media pipe-hands

First, when the palm detection model detects an image, it creates a box around the image. then, the palm detector removes the background and returns a cropped image. The palm detection model processes the image in the initial frame or if the hand is absent from subsequent frames in an effort to simplify the model and boost its real-time performance. The hand land mark model then generates 3D hand key points from the cropped image. The landmark model works through

21 key point coordinates of the hand by using the regression algorithm. The 21 coordinate points can be seen below in Figure 6. The landmark consists of the coordinates (x, y) and the relative depth (z), where z represents the camera's distance from the wrist and larger values indicate closer proximity to the camera. [17] This in turn predicts the final coordinates of the image. The final returned coordinates from the model are then matched with the stored coordinates of our custom images and the model returns the gesture name.

IV. EXPERIMENTAL RESULTS

A. Output

The Recognitions using the technique of interest -"Media pipe Hands" were documented and are given below in Figures 7 and 8

Fig.7. Recognition of gestures

Fig.8. Recognition of gestures

We can see from the above images that the model has given us accurate recognition outputs making this a successful attempt. The accuracy of the model will be discussed in the next section to document the efficiency of the model.

B. Accuracy

The model was trained and tested for accuracy and the results are as follows:
From Figure 9, we can see the training and testing accuracy to be precise and unwavering. Further, to calculate the overall

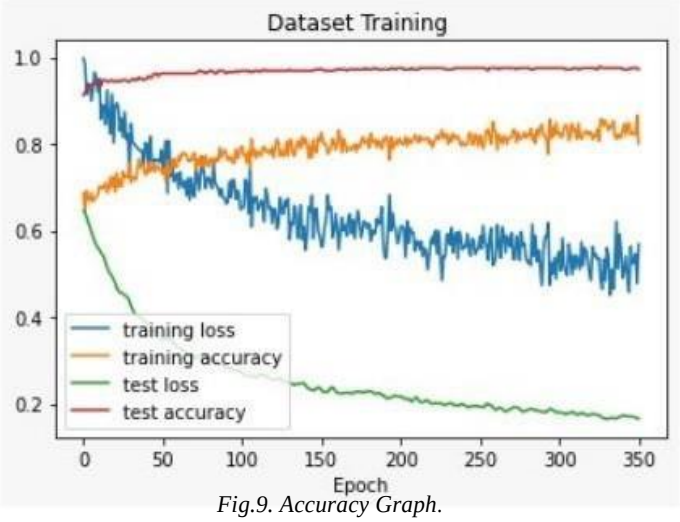

Fig.9. Accuracy Graph.

Classification Report			
Gesture Number	Precision	Recall	F1-score
0	0.97	1.00	0.98
1	1.00	1.00	1.00
2	1.00	1.00	1.00
3	1.00	1.00	1.00
4	1.00	0.95	0.97
5	1.00	1.00	1.00
6	0.94	0.97	0.95
7	0.95	0.98	0.96
8	1.00	0.96	0.98
9	1.00	1.00	1.00
Accuracy	0.98		

Fig. 10. Accuracy Table.

Accuracy of the model, The result of the calculation is as follows:

From the table depicted in 10, we can see that the model is 98 percent accurate and has been found to be very efficient.

V.CONCLUSION

As an image processing technique, Gesture recognition using media pipe has been successfully implemented providing us with promising accuracy. The coordinate-based recognition system has proved to be an accurate way to distinguish and recognize images as each angle and position is saved as a new coordinate. This in turn has provided us with maximum number of correct recognitions. The importance of the topic was to enable good us ability and simple understanding of the gestures and it has shown significant results and has paved way for massive enhancements in the future for the speech/hearing impaired. With advancing technologies and improving environments, this proposed project does have scope for additional implementations and future enhancements. The methodology can further be carried out in an app exclusively for the speech / hearing impaired empowering them to flourish in their fields of interest.

REFERENCES

1. *Mahesh,B.Machinelearningalgorithms-areview.InternationalJournal ofScienceandResearch(IJSR)9.,pp.381-386,(2020). ViharKurama-BlogML-basedImageProcessing,(2021)*
2. *GeethuGNathandArunCS,"RealTimeSignLanguageInterpreter," 2017InternationalConferenceonElectrical,Instrumentation,andCommunicationEngineering(ICEICE2017).*
3. *KumudTripathi,NehaBaranwalandG.C.Nandi,"ContinuousIn-dianSignLanguageGestureRecognitionandSentenceFormation", EleventhInternationalMulti-ConferenceonInformationProcessing-2015(IMCIP-2015),ProcediaComputerScience54(2015)523–531.*
4. *ManasaSrinivasaHSandSureshaHS,"ImplementationofRealTime HandGestureRecognition,"InternationalJournalofInnovativeResearch*
5. *inComputerandCommunicationEngineering,Vol.3,Issue5,May2015.*
6. *JoyeetaSinghaandKarenDas,"AutomaticIndianSignLanguageRecognitionforContinuousVideoSequence,"ADBUJournalofEngineeringTechnology2015Volume2Issue1.*
7. *ArchanaS.GhotkarandGajananK.Kharate,"DynamicHandGestureRecognitionandNovelSentenceInterpretationAlgorithmforIndian SignLanguageUsingMicrosoftKinectSensor,"JournalofPattern RecognitionResearch1(2015)24-38.*
8. *M.K.Bhuyan,"FSM-basedrecognitionofdynamichandgesturesvia gesturesummarizationusingkeyvideoobjectplanes,"WorldAcademy ofScience,EngineeringandTechnologyVol:62012-08-23.42*
9. *M.M.GharasuieandH.Seyedarabi,"Real-timeDynamicHandGesture RecognitionusingHiddenMarkovModels,"20138thIranianConferenceonMachineVisionandImageProcessing(MVIP).*
10. *KairongWang,BingjiaXiao,JinyaoXia,andDanLi,"ADynamicHand GestureRecognitionAlgorithmUsingCodebookModelandSpatial Moments,"20157thInternationalConferenceonIntelligentHuman- MachineSystemsandCybernetics.*
11. *FranckeH.,Ruiz-del-SolarJ.andVerschaeR.,"Real-TimeHandGes- tureDetectionandRecognitionUsingBoostedClassifiersandActive Learning,"AdvancesinImageandVideoTechnology.PSIVT2007. LectureNotesinComputerScience,vol4872.Springer,Berlin,Hei- delberg. JagreetKaur,BlogMachineLearningModelTestingTrainingandTool, February2022*
12. *Unext-blog10PopularRegressionAlgorithmsInMachineLearningof2022,2022*
13. *Harris,Moh,andAliSuryaperdanaAgoes.In2ndInternationalSeminar ofScienceandAppliedTechnology(ISSAT2021),AtlantisPress. ApplyingHandGestureRecognitionforUserGuideApplicationUsing MediaPipe.,pp.101-108,2021*