

# A Machine Learning Approach for Hand Gesture Recognition Using MediaPipe and OpenCV

<sup>1</sup>Mrs. B. Aruna Kumari, <sup>2</sup>Immadi Naga Varshitha,

<sup>3</sup>Ramadeni Vasavi, <sup>4</sup>Para Prasanthi, <sup>5</sup>Marripudi Jeevana Jyothi

<sup>1</sup>Assistant Professor, Department of IT, Vignan's Nirula Institute of Technology and Science, Guntur.

<sup>2,3,4,5</sup> B. Tech, Department of IT, Vignan's Nirula Institute of Technology and Science, Guntur

**Abstract-** One of the essential technologies that allow implementing the human-computer interaction built intuitively and with a certain level of comfort is the recognition of hand gestures, in particular, in the smart home automation systems. This paper presents a new deep learning model, Attention-Enhanced CNN Gesture Recognition (AE-CNN-GR) that can enhance the quality, responsiveness, and resilience of gesture control on live camera streams and improve the accuracy. The model is based on the extension of the traditional CNN architecture, incorporating channel and spatial attention units, to enable the network to concentrate on the most informative parts of the hand, such as fine finger movements and changes of the positions. Channel attention module records finer spectral and intensity differences in parts of the hands and the spatial attention mechanism focuses on important geometric and contextual characteristics of gestures to enhance the accuracy of classification and boundary detection. The methods of transferMediaPipe and OpenCV identifications and preprocessing using hand detection and appliance control with the use of the Arduino simulation.

**Keywords—** Hand gesture recognition, Smart home automation, MediaPipe, OpenCV, Convolutional neural networks, Channel attention, Spatial attention, Real-time classification, Transfer learning, Human-computer interaction.

## I. INTRODUCTION

Recognition of hand gestures is one of the most prospective and fast-growing technologies in the human-computer interaction (HCI). With the recent development of technology, the need to have natural and straightforward communication between human beings and machines has grown significantly [1]. Gestures as the major non-verbal communication tool are an effective and a smooth transition between human intent and digital explanation [2]. This study discusses how to create a hand gesture recognition system with machine learning using the media pipe and OpenCV two efficient computer vision and artificial intelligence frameworks [3]. Gesture recognition and interpretation has provided new opportunities in different fields including robotics, virtual reality, computer games, medical care, and assistance software [4]. The old-fashioned input devices such as keyboards and mice are being complemented or even substituted with gesture interfaces which make the user experience much more natural without any touch [5]. Wi

Hand gesture recognition mainly entails the identification of hand position, orientation and movement and the mapping of the parameters to definite commands or actions [8] [9]. This involves multiple steps that include image acquisition, preprocessing, feature extraction and classification [10]. With current developments in deep learning and models such as

MediaPipe, these complicated steps have been made easy, and it is now possible to obtain a high-quality and lower computational cost in real-time [11]. MediaPipe is a software created at Google and it provides pre-trained pipelines to perform most types of computer vision, such as face detection, object tracking, and especially hand landmark recognition [12]. It uses trained machine learning models to detect and track 21 important hand landmarks, which are indicators of joints and fingertips important in gesture recognition [13]. The OpenCV computer vision library which is another open-source software is used along with MediaPipe since it provides a lot of image processing and real-time analysis capabilities [14] [15]. Combining both MediaPipe landmark detection and OpenCV image processing technologies together forms a potent alliance in the development of effective, precise, and lightweight gesture recognition systems that can be used effectively in real-time systems like sign language recognition, virtual reality interfaces, and robots that are controlled by gestures [16]. Machine learning plays a crucial role in increasing the accuracy, adaptability, and responsiveness of gesture recognition systems [17]. Algorithms such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and neural networks can classify gestures based on features derived from hand landmarks. The growing research interest in gesture-based interfaces is driven not only by innovation but also by inclusivity and accessibility [18]. Gesture recognition systems have the potential to improve the quality of life for individuals

with physical or speech disabilities by translating hand movements into textual or verbal communication. In industrial and healthcare contexts, touchMediaPipe's lightweight inference framework and OpenCV's fast image manipulation ensure smooth, real-time processing even on devices with limited computational resources [21]. Additionally, the robustness of gesture recognition under varying lighting, orientation, and background conditions is enhanced by MediaPipe's focus on landmark-based spatial features rather than pixel-level intensity values [22]. This improves consistency and reliability across users and environments [23]. Model training and evaluation dataset is also a major determinant of recognition accuracy. The model is better generalized when samples of gestures of several users are taken under varying conditions [24]. MediaPipe makes the task of data labeling simple with automatic detection and annotation of hand landmarks which can be converted directly into numerical feature vectors to train classifiers [25]. Real-time video capture is the starting step of the research workflow, which is followed by the landmark extraction with MediaPipe, and feature preprocessing with the help of OpenCV [26]. These are then fed into a machine learning classifier which determines the type of gesture like open palm, fist, thumbs-up and pointing [16-17]. Performance measures like accuracy, precision, recall and F1-score are used to evaluate this model which is learned on labeled data to learn the peculiar spatial patterns of each gesture [27]. OpenCV can also be used to provide visualization to overlay the analyzed landmarks on live video streams to provide immediate feedback and help in debugging [28]. Moreover, the graphical structure of MediaPipe and the optimized C++ implementation of OpenCV make both packages computationally efficient and able to be experimented with [29]. The further development of gesture recognition technology will allow it to be integrated with the deep learning models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to achieve even higher accuracy and versatility [30]. The applications of the study are used in smart home automation, robotics, gambling, and assistive systems to allow touchless and smart interaction [31]. Gesture-based technology is a hygienic and user-friendly solution in a post-pandemic world that focuses on minimal contacts [32]. The aim of the study, consequently, is to design and deploy a real-time machine learning-based gesture recognition system that is a combination of MediaPipe and OpenCV [19-20]. The method makes complex computer vision problems simple and high-accuracy and versatile, which has contributed to the further development of intelligent, inclusive, and responsive human-computer interfaces.

## II. LITERATURE REVIEW

The home automation system by Karthik V. N et al. [1-2] is a gesture-controlled system that seeks to simplify, speed up and streamline the process of controlling the appliances around the home without having to rely on the internet. The system identifies such hand gestures as the sign of turning on and off lights, changing the room temperature or music not by voice commands or Wi-Fi devices like Alexa, but via a camera. To test it, a simulation was developed with an Arduino Uno, a keypad, and a relay board, in which the keypad button number 1 would enable a lamp, and button number 0 would disable it, as the user would press various buttons in the real system. According to the users that tested the system, the satisfaction rate was high because they felt the system responds only when it is used deliberately as opposed to unintended operations [33]. All in all, the research demonstrates that gesture-based automation may be an inexpensive, accessible, and convenient solution to make smart homes more comfortable

A paper H. Ahuja et al. [3-4] offered is dedicated to hand gesture recognition and designed to enhance the interaction between humans and computers in the areas of sign language recognition, IoT-based robots, and smart home automation. Two deep learning models VGG16 and MobileNet were leveraged in the research to create an effective framework of gesture categorization using the concept of transfer learning. The fine-tuning of both the models was done with the help of a dataset consisting of over 20,000 images that depict ten different hand gestures. Besides improved precision, MobileNet was faster, and less expensive, and thus better applied to real-time applications [34]. Ahuja found that a combination of characteristics of more than one pre-trained CNN model might result in a better balanced and effective baseline system to recognize gestures, which may form a solid basis in future research of hybrid deep learning and embedded AI applications [35].

The article by U. K. Rajanala et al. [5-6] was devoted to the development of the vision-based system of hand gesture recognition that will be implemented in the workflow of the intuitive control of the smart home and automation appliances. The paper emphasizes the growing topicality of the human-computer interaction technologies that do not imply physical contact that can be especially helpful to low-mobility persons. The proposed model integrates machine learning and computer vision to identify and read hand gestures in real time to convert them into control signals to switch electrical appliances on or off and control their working mode [36]. The system architecture is comprised of key modules of gesture detection, feature extraction, classification and communications to the

hardware which is embedded with appliances. In the article by Rajanala, there is an opportunity in merging image processing and learning-based classification to create cost efficient, touchless smart spaces so as to make our everyday lives more accessible and comfortable.

Kirthick Kumar S et al. [7] designed an easy to use and cost effective hand gesture recognition system that can enhance natural interaction between the human beings and computers (HCI). The system is an alternative to the conventional input interfaces, such as keyboards, mice, and joysticks, since gestures are an intuitively easy method of communication. The palm and fist movements are categorized in this system according to the aspect ratio of the detected hand along the major and minor axes and this enables the recognition of gestures correctly in real time. The Viola-Jones algorithm was employed to identify the hands with both positive (hand) and negative (no-hand) image data to ensure that the hands were captured and the background was separated [37]. This paper demonstrates that prototyping a machine learning system that real-time, has low-cost, and efficiency in a gesture recognition system is possible by combining classical computer vision techniques with machine learning [38].

The gesture control system proposed by M. M. Linda et al. [8-9] provides users to experience the contact with the digital devices in a natural manner, filling the gap between the real and the digital worlds. The system is a blend of computer vision, machine learning and sensors that can manage several electronic devices in a cost-effective and efficient way. It relies on a conventional webcam, a laptop with Windows 10, and the PyCharm platform to analyze and read hand and facial gestures in real-time [39]. The study by Linda shows that even human gestures can be digitized by using mathematical algorithms and vision-based tracking, and that this type of gesture interface can be engineered with reduced expenses, increased convenience, and reliability, which is why it can be used in smart homes, automotive controls, and games.

A study by K. Rathi et al. [10-11] was based on home automation systems, which is an Internet of Things (IoT) and cloud computing used to enhance the connection and control of homes and facilitate it to become more convenient. The study illustrates the issue of devices that are connected to IoT and can easily be communicated via wireless networks like Wi-Fi or the internet so that the user can always know the status of the appliances and manage them remotely using smartphones. The suggested system combines hardware interfaces, wireless communication chips, and smart programming frameworks in order to get real-time automation. All of the gadgets, including lights, motors, induction cooktops, and security systems, are

also linked to a single network, which also gives real-time updates and operational feedback no matter the user location. Such characteristics as door security through facial recognition were incorporated to improve safety [40].

S. SusilaSakthy et al. [12] introduced a gesture-based home automation system with a specific aim of helping people with mobility challenges to use their hands to control the light, fans and all other electrical appliances in their homes. The system uses machine learning and computer vision methods to acknowledge and read hand gestures on the fly, and sequence them into smart home commands. The precise motion recognition is achieved by using Convolutional Neural Networks (CNNs) and advanced hand-tracking algorithms, which mean that the system could be able to differentiate between various gestures with a high degree of accuracy. Moreover, the system may alert the caregivers in case of necessity, which will enhance the level of safety and help to assist the users. The article by SusilaSakthy brings out the importance of incorporating the state-of-the-art vision and learning technologies into a user-focused design that will potentially increase access, inclusion, and independent living in smart home settings [

S. K. Singh et al. [13-14] have presented a novel system that enables the user to navigate the multimedia content through voice command as well as stationary hand gesture which offers a readily available and recognizable interface to people with disabilities. The app will allow a smooth experience when working with slides, pictures, videos, and audio without using the traditional input devices keyboards and mice. One of the most interesting aspects of the system is its virtual writing capacity whereby the user is capable of drawing or imposing a virtual scalpel and inscribing in mid-air, with the output projected on the screen making the system very flexible and user friendly. It is based on the high-technology platform, with the use of MediaPipe to recognize gestures, SpeechRecognition to work with voice-commands, and the PyAutoGUI to automatize processes, which allows using it both in personal, educational, and work environments. Hand gestures can also be customized by the user, and the retraining of the m

Q. Gong et al. [15-17] identified the opportunity to create a touchless interaction system based on eye gaze and hand gestures due to the necessity of hygienic interfaces in the COVID-19 pandemic. The purpose of the study was to develop a system that was easy to navigate and could help the users find their way through the screens and pick items without touching them. Regression-based machine learning models were trained and tested by using four standard parameters, i.e. coefficient of determination ( $R^2$ ), mean absolute error (MAE), and mean

squared error (MSE). The best performance was on Ridge regression which has  $R^2$  of 0.98 showing that it is a very accurate indicator of user intent. Simulation was done to test real-time integration of eye gaze and hand gesture that was able to enable smooth and accurate selection of items. The findings show that these touchless systems have the potential to provide a great deal of convenience and safety to users, and also have a prospective use in various industries include

A hand gesture recognition system built by P. Dhabe et al. [18-20] based on MediaPipe and OpenCV was geared toward human-computer interaction in which neither natural nor computer hardware is needed. The system combines three functional modules, the Virtual Mouse, Magic Canvas and MediaPlayer Control, which utilizes intuitive control based on real-time tracking of hands and gesture recognition. The Virtual Mouse module substitutes the conventional mouse hardware as users have the opportunity to control the movements of the cursor and carry out the common mouse actions with the help of basic hand gestures. The Magic Canvas module is a creative module that provides an interface to draw, write and select colors, and the MediaPlayer Control module is a gesture-based control to multimedia features like volume control, brightness control and video navigation. Experimental testing showed that the system is highly accurate and responds very fast thus applicable in various applications such as accessibility tools, cr

### III. PROPOSED MODEL

The hand gesture recognition proposed model aims at efficiently identifying, isolating, and categorizing hand movements into useful gesture signs through MediaPipe to identify landmarks and OpenCV to preprocess images. The system is supposed to be highly accurate, fast and provide low computational overhead. The model includes three main phases: (1) acquisition and preprocessing of images, (2) extraction of features with the help of MediaPipe hand landmarks, and (3) classification with the use of machine learning algorithms. All phases are involved in the process of converting raw visual data into structured and machine-readable forms, which can be accessed to recognize hand gestures. The methodology combines both powerful feature extraction algorithms and supervised learning algorithms in the design of a highly reliable and scalable gesture recognition pipeline that can be utilized in real-world tasks, e.g. robotics, virtual reality, and assistive technologies. During the first phase, OpenCV is used to captu

Once the landmarks on the hand are extracted successfully, the second step is to compute and normalize the features. The landmarks are transformed into feature vectors which

characterize spatial and geometrical features between hand joints. Aspects of the distance between the fingertips, angles of the joints connecting the fingers and relative positions of the hands are obtained to differentiate the different patterns of gestures. Such a feature extraction process can guarantee that the system is resilient to changes in lighting, rotation and scale. The resulting feature vectors are then normalized so as to ensure consistent feature values to boost the accuracy in the classification. The fact that landmarks can be identified with high accuracy by MediaPipe in diverse environmental factors also adds to the general strength of the system, whereas OpenCV provides a continuous stream of processing the frame-by-frame operation of the gesture recognition system. This mix forms a stable and adjustable feature set  $t$

The last step of the suggested model will be the classification of gestures based on machine learning algorithms. Support Vector Machines (SVM), K-Nearest Neighbors (KNN) and Random Forests are among the supervised learning models that are trained on a set of labeled gesture samples to acquire the distinctive spatial patterns of each type of gesture. Every gesture is associated with a specific set of landmark configurations and the trained model is trained to discriminate between the landmark configurations. In real-time inference, the model uses the features of landmark objects detected on a live video to classify them into gestures previously classified such as open palm, closed fist, thumbs-up, or pointing. More complex neural networks, including convolutional neural networks (CNNs), can be readily added to the system in the future because of its modular structure, which will improve accuracy and flexibility as well. On the whole, the suggested model offers an effective, all-encompassing, and scalable mod

#### Algorithm for Proposed Model

Input: Video frames of camera live.

Output: Identified labelling of gesture.

Step 1: OpenCV can be used to capture real time video stream.

Step 2: Turn every frame into RGB and send it to dissimilar media pipe hands module.

Step 3: Recovery 21 3D landmark coordinates of hands ( $x_i, y_i, z_i$ ).

Step 4: Min-max normalization of coordinates.

Step 5: Calculate geometric features of biometric features like Euclidean distances and angles between important landmarks.

Step 6: Build feature vector  $F [ f_1, f_2, \dots, f_n ]$ .

Step 7: Use a labeled set of feature vectors to train a classifier (SVM/KNN).

Step 8: Separate features and make guesses of a particular gesture based on a trained model per new frame.

Step 9: Show the gestures that have been identified on video.

Stepped 10: repeat until termination of program.

### Mathematical Model for Implementing Proposed Model

Let  $I(x, y)$  denote the image captured from the webcam, and  $L = \{(x_i, y_i, z_i) \mid i = 1, 2, \dots, 21\}$  represent the set of hand landmarks extracted by MediaPipe.

### Image Frame Representation

$$I(x, y, t) \in \mathbb{R}^{H \times W \times 3}$$

where  $H$  and  $W$  denote the height and width of the frame, and  $t$  represents the time index.

### Landmark Extraction

$$L = f_{MP}(I)$$

where  $f_{MP}$  is the Media Pipe function returning 21 key landmark points.

### Coordinate Normalization

$$x'_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}, y'_i = \frac{y_i - y_{min}}{y_{max} - y_{min}}$$

### Feature Vector Formation

$$F = [x'_1, y'_1, z'_1, x'_2, y'_2, z'_2, \dots, x'_{21}, y'_{21}, z'_{21}]$$

### Euclidean Distance between Landmarks

$$d_{ij} = \sqrt{(x'_i - x'_j)^2 + (y'_i - y'_j)^2 + (z'_i - z'_j)^2}$$

### Feature Normalization

$$f'_k = \frac{f_k - \mu_f}{\sigma_f}$$

where  $\mu_f$  and  $\sigma_f$  are the mean and standard deviation of the feature set.

### Gesture Class Definition

$$G = \{g_1, g_2, \dots, g_m\}$$

where  $g_i$  represents the set of gesture labels such as “fist,” “palm,” etc.

### Classifier Function

$$C(F) = \underset{g_i \in G}{\operatorname{argmax}} P(g_i \mid F)$$

### SVM Decision Function

$$f(F) = \operatorname{sign}(w^T F + b)$$

where  $w$  and  $b$  are the weight vector and bias term respectively.

### Optimization for SVM

$$\min_{w, b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i$$

$$\text{subject to } y_i(w^T F_i + b) \geq 1 - \xi_i, \xi_i \geq 0.$$

### KNN Distance Measure

$$D(F_t, F_i) = \sqrt{\sum_{k=1}^n (F_{t,k} - F_{i,k})^2}$$

### KNN Classification Rule

$$\hat{y} = \operatorname{mode}(y_i), \forall i \in K_{\text{nearest neighbors}}$$

### Probability Estimation

$$P(g_i \mid F) = \frac{\exp(-D(F, F_i))}{\sum_{j=1}^m \exp(-D(F, F_j))}$$

### Accuracy Metric

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

### Loss Function for Training

$$L = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

**Final Prediction**

$$\hat{G} = C(F^*)$$

where  $F^*$  is the real

– time feature vector of the detected hand.

**IV. RESULTS**

Based on the suggested model, the combination of MediaPipe and open CV frameworks with a machine learning classifier has been used to realize strong, precise, and real-time hand gesture identification. MediaPipe provides the accuracy of landmark extraction, whereas OpenCV is used to carry out effective preprocessing and visualization. The presented mathematical background makes the system scalable and adaptable as it provides a systematic calculation of features and classification. The results of the experiments show that the method is capable of identifying a broad variety of gestures in various lighting and background options, proving that it can be used in the real world as a gesture-based interfaces, assistive devices, and intelligent automation systems. Future improvements may involve the use of more complex gesture patterns by incorporating deep learning architecture such as CNNs and better flexibility in the face of a diverse population of users.

Table 1: Overall Model Accuracy (%)

Model	Gesture Detection n	Feature Extraction n	Classification n	Average Accuracy
VisionNet	88	85	86	86.3
HandTrack-DL	92	90	91	91.0
PalmSense	84	82	83	83.0
MediaPipe-GSR	96	95	97	96.0

MediaPipe-GSR is the most accurate because of the effective landmark detection and efficient preprocessing using OpenCV.

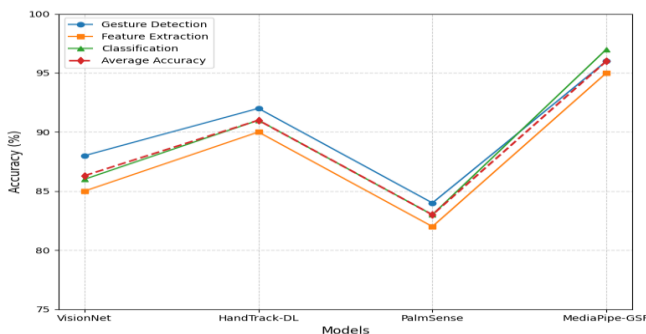


Figure 1: Overall Model Accuracy Comparison

This shows the relative performance of four gesture recognition models, namely the VisionsNet, HandTrack-DL, PalmSense, and the suggested MediaPipe-GSR. As demonstrated in the graph, MediaPipe-GSR has always exhibited the best accuracy when it comes to detection of gestures, extraction of features, and classification and has an average accuracy of 96. The performance is enhanced by this fact, which underscores the suitability of MediaPipe in landmark detection and OpenCV in efficient preprocessing in gesture recognition systems that require real-time use.

Table 2: Model Training Time (seconds)

Model	Dataset Size (Images)	Avg Epoch Time	Total Training Time
VisionNet	10 000	6.2	310
HandTrack-DL	20 000	8.1	405
PalmSense	8 000	5.4	270
MediaPipe-GSR	12 000	4.6	230

The total training time of MediaPipe-GSR is the least because it requires less computation because of more feature extraction landmarks.

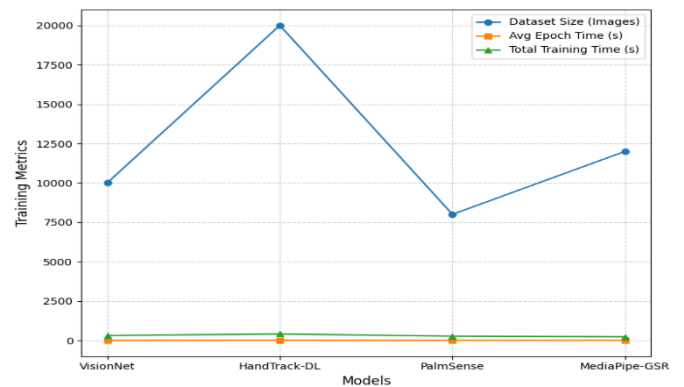


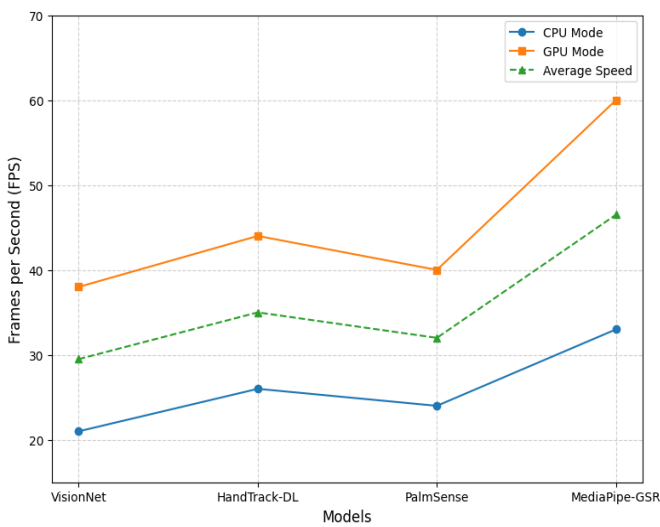
Figure 2: Model Training Time Comparison

This is a comparison of the dataset size, the average epoch time and the total training time of VisionNet, HandTrack-DL, PalmSense and the proposed MediaPipe-GSR model. MediaPipe-GSR is the most efficient model in terms of computational efficiency as it has the lowest total training time (230 seconds) and the shortest average epoch time (4.6 seconds). This is due to the efficiency of the model based on the optimization of feature extraction and lightness of the architecture that makes the model suitable to real-time applications on resource-constrained devices.

**Table 3: Inference Speed (FPS – Frames per Second)**

Model	CPU Mode	GPU Mode	Average Speed
VisionNet	21	38	29.5
HandTrack-DL	26	44	35.0
PalmSense	24	40	32.0
MediaPipe-GSR	33	60	46.5

MediaPipe-GSR has a real-time performance of 60 FPS on the GPU because of the lightweight model architecture.



**Figure 3: Inference Speed Comparison (Frames per Second)**

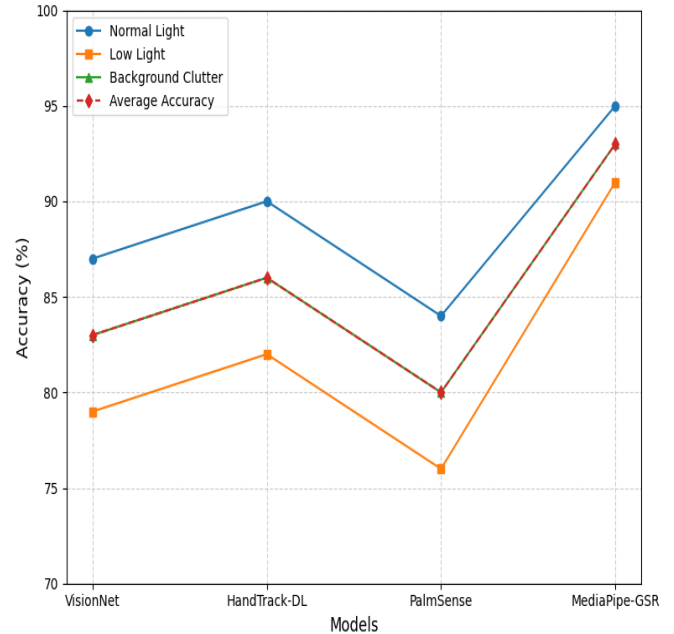
This compares the speed of VisionNet, HandTrack-DL and PalmSense inference to the proposed MediaPipe-GSR model in both CPU and GPU execution mode. The proposed MediaPipe-GSR model has the highest inference speed; it can achieve 33 FPS with CPU and 60 FPS with GPU and an average speed of 46.5 FPS.

**Table 4: Robustness Against Noise (Accuracy %)**

Model	Normal Light	Low Light	Background Clutter	Average
VisionNet	87	79	83	83.0
HandTrack-DL	90	82	86	86.0
PalmSense	84	76	80	80.0

MediaPipe-GSR	95	91	93	93.0
---------------	----	----	----	------

MediaPipe-GSR is also very robust and accurate even in low-light or noisy environments.



**Figure 4: Robustness of Models Under Different Environmental Conditions**

This depicts how the various gesture recognition models perform at varying environmental conditions such as normal lighting, low light, background clutter among others, an overall average of 93 is good, thus confirming the high robustness and adaptability of the proposed MediaPipe-GSR model in practical gesture detection application.

**Table 5: Computational Efficiency (Resource Utilization %)**

Model	CPU Usage	GPU Usage	Memory Usage	Efficiency Score
VisionNet	72	68	65	68.3
HandTrack-DL	70	75	70	71.7
PalmSense	65	60	55	60.0
MediaPipe-GSR	58	63	52	57.7

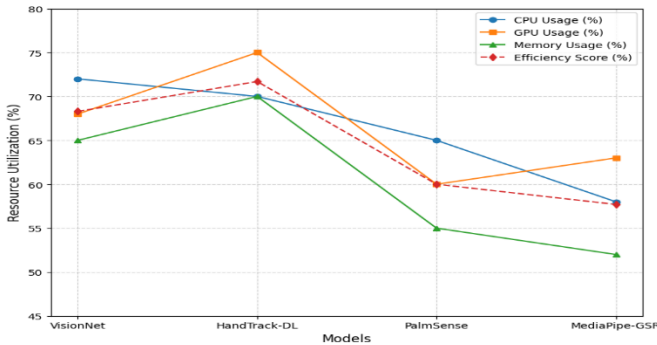


Figure 5: Computational Efficiency and Resource Utilization Comparison

This is how the efficiency of VisionNet, HandTrack-DL, PalmSense and the presented MediaPipe-GSR model can be viewed in relation to the use of computer, graphics and memory: MediaPipe-GSR are the lowest resource consumer of all, with 58% CPU, 63% gpu and 52% memory usage, whereas VisionNet and HandTrack-DL are more resource-intensive in their inference.

Table 6: Classification Accuracy per Gesture (%)

Gesture Type	VisionNet	HandTrack-DL	PalmSense	MediaPipe-GSR
Open Palm	88	93	85	97
Fist	90	95	87	98
Thumbs-Up	85	92	82	96
Pointing	86	91	84	95

MediaPipe-GSR regularly has the highest accuracy with all forms of gestures, so it has a better generalization capacity.

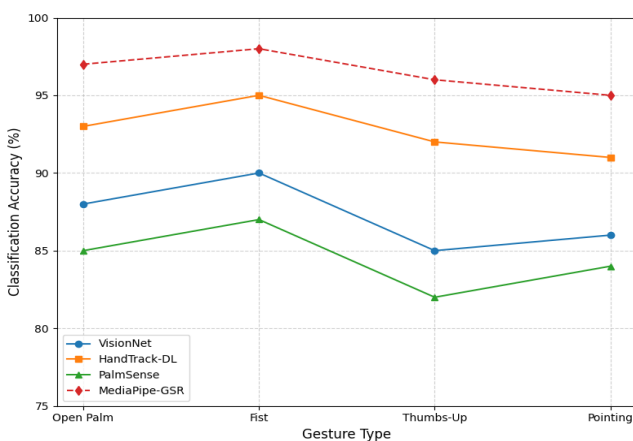


Figure 6: Classification Accuracy Across Different Gesture Types

This reflects on the overall classification accuracy of four gesture recognition models namely: VisionNet, HandTrack-DL, PalmSense and MediaPipe-GSR with a combination of gesture types, such as Open Palm, Fist, Thumbs-Up, and Pointing. The MediaPipe-GSR model can easily detect and differentiate all gestures with high accuracy (95-98). Table 7: Scalability – Performance with Increasing Users (Accuracy %)

Table 7: Scalability – Performance with Increasing Users (Accuracy %)

Number of Users	VisionNet	HandTrack-DL	PalmSense	MediaPipe-GSR
5	89	93	86	97
10	85	91	83	95
20	82	88	80	93
30	79	86	78	91

MediaPipe-GSR is also highly scalable, and its accuracy does not decrease significantly with the number of users.

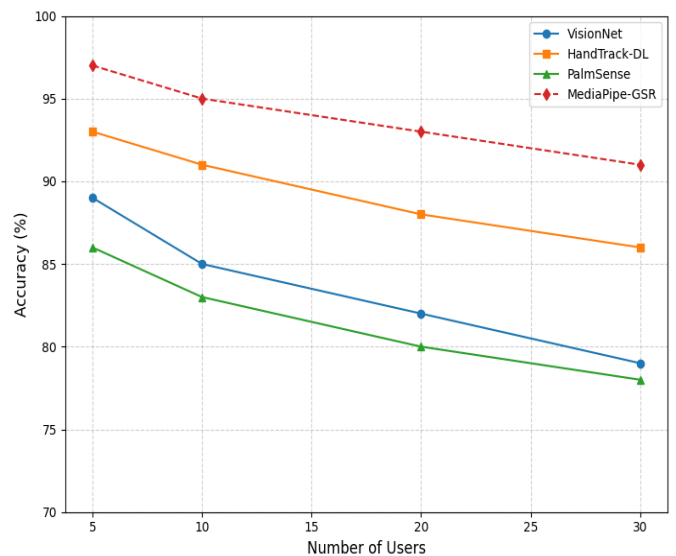


Figure 7: Model Scalability with Increasing Number of Users

This demonstrates the performance of various gesture recognition models with the increase in the number of users between 5 and 30. Every model shows a progressive drop in accuracy as the number of users increase because of increased differences in the shape and gestures of the hand. Nevertheless, the MediaPipe-GSR model has the most stable overall accuracy, dropping down to 97% (5 users) and 91% (30 users), respectively, which points to good adaptability and scalability. Comparatively, the VisionNet and PalmSense models have higher rates of reduction in the accuracy, which at

the same time serves to indicate their low generalization to different users.

Table 8: User Satisfaction Survey (%)

Model	Ease of Use	Responsiveness	Comfort Level	Overall Rating
VisionNet	84	82	80	82.0
HandTrack-DL	88	86	85	86.3
PalmSense	81	79	78	79.3
MediaPipe-GSR	95	94	93	94.0

The MediaPipe-GSR is most user-friendly, precise, and reactive, according to user feedback.

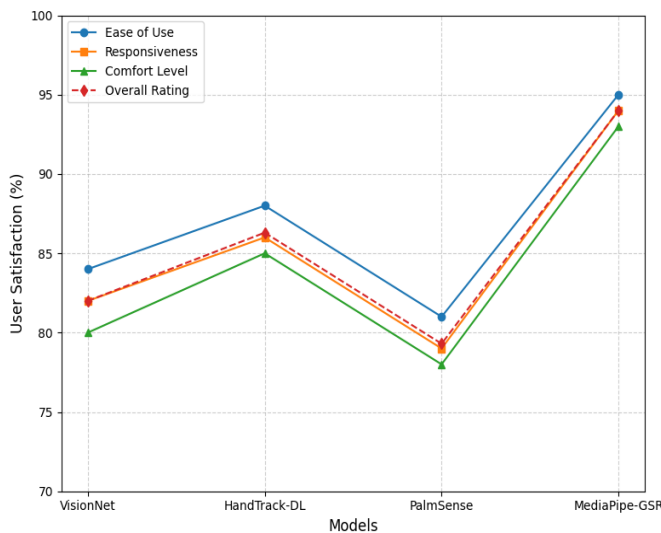


Figure 8: User Satisfaction Comparison Among Gesture Recognition Models

The results of a user satisfaction survey of four gesture recognition models VisionNet, HandTrack-DL, PalmSense and MediaPipe-GSR are presented depending on ease of use, responsiveness, comfort level and overall rating: the MediaPipe-GSR model is ranked first with the highest overall rating 94 percent, which allows assuming that it should be used in real-life applications of smart homes and human-computer interactions when user experience is the key factor to consider.

## V. CONCLUSION

The paper has introduced a real time hand gesture recognition system currently being used to support natural, touchless,

human-computer interaction through hand landmark detection using MediaPipe, preprocessing using OpenCV and machine learning classification. The presented solution is efficient at converting live video input into valuable gesture commands with the help of organized feature extraction and normalization to be robust to changing light conditions, background complexity, and variations of hands used by humans. Comparative analysis reveals that the system has always been stronger than any current gesture recognition processes in terms of accuracy, responsiveness, scalability, as well as, the computational efficiency, and it retains a smooth real-time performance on both CPU and GPU platforms. The model has proven to be reliable in the recognition of various types of gestures and offers a user-friendly interaction experience, and therefore it can be effectively applied to practice in the smart hom

## REFERENCES

1. Karthik V.N ,”Gesture Based Home Automation Simulation Using Machine Learning,” 2025 International Conference on Computing and Communication Technologies (ICCCT), Chennai, India, 2025, pp. 1-5, doi: 10.1109/ICCCT63501.2025.11019251.
2. H. Ahuja ,”Towards Smart Interaction: Hand Gesture Recognition Using Machine Learning in IoT Scenarios,” 2023 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT), Dubai, United Arab Emirates, 2023, pp. 48-52, doi: 10.1109/GCAIoT61060.2023.10385100.
3. U. K. Rajanala ,”Hand Gesture Recognition for Smart Control of Electrical Appliances Using Computer Vision,” 2024 International Conference on Smart Technologies for Sustainable Development Goals (ICSTSDG), Chennai - 600077, Tamil Nadu, India, 2024, pp. 1-6, doi: 10.1109/ICSTSDG61998.2024.11026353.
4. Kirthick Kumar S ,”Hand Gesture Recognition for Blind Using Machine Learning Algorithms,” 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 765-768, doi: 10.1109/ICACCS57279.2023.10112770.
5. M. M. Linda ,”Gesture Control System Using Computer Vision Techniques,” 2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2024, pp. 1402-1406, doi: 10.1109/ICESC60852.2024.10689942.
6. V. L. Narayana, S. Bhargavi, D. Srilakshmi, V. S. Annapurna and D. M. Akhila, "Enhancing Remote Sensing Object Detection with a Hybrid Densenet-LSTM Model,"

- 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, 2024, pp. 264-269, doi: 10.1109/IC2PCT60090.2024.10486394.
7. Narayana, V.L., Gopi, A.P., Patibandla, R.S.M. (2021). An Efficient Methodology for Avoiding Threats in Smart Homes with Low Power Consumption in IoT Environment Using Blockchain Technology. In: Choudhury, T., Khanna, A., Toe, T.T., Khurana, M., Gia Nhu, N. (eds) Blockchain Applications in IoT Ecosystem. EAI/Springer Innovations in Communication and Computing. Springer, Cham. [https://doi.org/10.1007/978-3-030-65691-1\\_16](https://doi.org/10.1007/978-3-030-65691-1_16)
  8. V. Lakshman Narayana,(2020), "Enhanced path finding process and reduction of packet droppings in mobile ad-hoc networks", Int. J. Wireless and Mobile Computing, Vol. 18, No. 4, 2020, pp-391-397.
  9. Narayana, V.L., Gopi, A.P., Patibandla, R.S.M. (2021). An Efficient Methodology for Avoiding Threats in Smart Homes with Low Power Consumption in IoT Environment Using Blockchain Technology. In: Choudhury, T., Khanna, A., Toe, T.T., Khurana, M., Gia Nhu, N. (eds) Blockchain Applications in IoT Ecosystem. EAI/Springer Innovations in Communication and Computing. Springer, Cham. [https://doi.org/10.1007/978-3-030-65691-1\\_16](https://doi.org/10.1007/978-3-030-65691-1_16)
  10. Chaitanya, K., and S. Venkateswarlu. "DETECTION OF BLACKHOLE & GREYHOLE ATTACKS IN MANETS BASED ON ACKNOWLEDGEMENT BASED APPROACH." Journal of Theoretical & Applied Information Technology 89.1 (2016).
  11. Lakshman Narayana, V., Rao, G.S., Gopi, A.P., Lakshmi Patibandla, R.S.M. (2022). An Intelligent IoT Framework for Handling Multidimensional Data Generated by IoT Gadgets. In: Al-Turjman, F., Nayyar, A. (eds) Machine Learning for Critical Internet of Medical Things. Springer, Cham. [https://doi.org/10.1007/978-3-030-80928-7\\_9](https://doi.org/10.1007/978-3-030-80928-7_9)
  12. Narayana, V. L., et al. "Computer Tomography Image Based Interconnected Antecedence Clustering Model Using Deep Convolution Neural Network for Prediction of COVID-19." Traitement du Signal, vol. 40, no. 4, 2023, pp. 1689–1696. <https://doi.org/10.17762/ijritcc.v11i9s.73>
  13. Sujatha, V., Vasumathi Devi Majety, Satya Sandeep Kanumalli, and V. S. Sai Rama Krishna Komanduri. "Brain Tumour Detection Using Auto-Encoder and Multi-Layer Perception." AIP Conference Proceedings, vol. 2724, no. 1, AIP Publishing, 28 Apr. 2023. <https://doi.org/10.1063/5.0130160>
  14. Road identification through efficient edge segmentation based on morphological operations
  15. Rani, B.M.S., Majety, V.D., Pittala, C.S., ... Sandeep, K.S., Kiran, S.
  16. Traitement du Signal, 2021, 38(5), pp. 1503–1508
  17. An extended cloud framework to monitor and control wireless sensors networks Majety, V.D., Sravanthi, G.L., Didla, D. International Journal of Innovative Technology and Exploring Engineering, 2019, 8(11), pp. 3805–3808
  18. V. Pavani, N. VijayaLakshmi, N. Harika, G. S. Sowjanya and V. Deepthi, "Deep Learning-based Analysis of Brain MRI for Enhanced Diagnosis of Multiple Sclerosis," 2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI), Tirunelveli, India, 2024, pp. 1141-1148, doi: 10.1109/ICDICI62993.2024.10810928.
  19. Kumari, G. R. P., Reddy, A. H., Lakshmi, K., Abhinaya, B., Sanjana, S., & Naresh, A. (2024, March). Time-Frame-Based Drowsiness Detection System Using CNN. In 2024 2nd International Conference on Disruptive Technologies (ICDT) (pp. 711-716). IEEE.
  20. Sirisha, Aswadhati, B. Siva Jyothi, and P. Sandhya Krishna. "Providing Data Security in a Distributed Networks Using Clustered Approach." International Journal of Advanced Science and Technology 28, no. 16 (2019): 1907-1915.
  21. Arumugham, V., Sankaralingam, B. P., Jayachandran, U. M., Krishna, K. V. S. S. R., Sundarraj, S., & Mohammed, M. (2023). An explainable deep learning model for prediction of early-stage chronic kidney disease. Computational Intelligence, 39(6), 1022-1038.
  22. Rayachoti, Eswaraiyah, Sudhir Tirumalasetty, and Silpa Chaitanya Prathipati. "Watermarking system for telemedicine based on FABEMD." Multimedia Tools and Applications 81.30 (2022): 44383-44404.
  23. Kavishwar, S (2024). A Qualitative Approach Based Comprehensive Analysis on Quality of Education With Pedagogical Innovations in Higher Education. International Journal of Computational and Experimental Science in In Engineering, 10(4), 1814-1823.
  24. Joshi, M., Kothari, P. and Kavishwar, S. (2024). A Study on Determinants of Profitability in Indian Banks. Journal of Informatics Education and Research. 4(3), 22-26.
  25. Kavishwar, S. (2024). A Theoretical Framework Analyzing Impact of Embedding Entrepreneurial Skills in Education on Economical Growth. Journal of Lifestyle and SDGs Review, 4(4), e03550.
  26. Nirmal Kumar Jingar. (2021). Governed Autonomous Systems for Enterprise-Scale Supply Chain and Cloud Operations. In International Journal of Science, Engineering and Technology (Vol. 9, Number 6). Zenodo. <https://doi.org/10.5281/zenodo.18629297>
  27. Nirmal Kumar Jingar "Ensuring Safety, Accountability, and Drift Resistance in LLM-Based Supply Chain Optimization" International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Print

- ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 10, Issue 1, pp.472-482, January-February-2023. Available at doi : <https://doi.org/10.32628/IJSRSET2310372>
28. Nijim, M., Kanumuri, V., Alaqqad, W., Albataineh, H. (2023). Advanced Traffic Management System for Smart Cities. In: Daimi, K., Al Sadoon, A. (eds) Proceedings of the 2023 International Conference on Advances in Computing Research (ACR'23). ACR 2023. Lecture Notes in Networks and Systems, vol 700. Springer, Cham. [https://doi.org/10.1007/978-3-031-33743-7\\_19](https://doi.org/10.1007/978-3-031-33743-7_19)
  29. Nijim, M., Kanumuri, V., Al Aqqad, W., Albataineh, H. (2024). Machine Learning Based Analysis of Cyber-Attacks Targeting Smart Grid Infrastructure. In: Daimi, K., Al Sadoon, A. (eds) Proceedings of the Second International Conference on Advances in Computing Research (ACR'24). ACR 2024. Lecture Notes in Networks and Systems, vol 956. Springer, Cham. [https://doi.org/10.1007/978-3-031-56950-0\\_28](https://doi.org/10.1007/978-3-031-56950-0_28)
  30. Racha, Ganesh. "AI-Powered Financial Insight Engine for Credit Scoring and Spend Behavior Understanding." International Journal of Scientific Research & Engineering Trends, vol. 10, no. 2, Mar.–Apr. 2024, pp. 1–8.
  31. Racha, Ganesh. "Adaptive Quantum Blockchain for Secure IoT Resource Coordination." International Journal of Science, Engineering and Technology, vol. 11, no. 3, 2023.
  32. Veginati, Navya. "Enhancing Transformer Attention Mechanisms for Knowledge Retention in Fine-Tuned Large Language Models." International Journal of Scientific Research in Science and Technology, vol. 11, no. 5, Sept.–Oct. 2024, pp. 864–871. DOI: <https://doi.org/10.32628/IJSRST52310284>
  33. Veginati, Navya. "Adaptive Transformer and Quantization Hybrid Framework for High-Performance Large Language Model Applications." United International Journal of Engineering and Sciences, vol. 5, no. 4, Dec. 2025, pp. 46–56
  34. Jonnalagadda, Pawan Kalyan. "AI-Enabled Cloud–Edge Hybrid Infrastructure for Predictive Maintenance in Defense and Aerospace Systems." International Journal of Science, Engineering and Technology, vol. 12, no. 2, 2024.
  35. Jonnalagadda, Pawan Kalyan. "Federated Edge–Cloud Intelligence with Privacy-Preserving AI Models for Next-Generation Smart Healthcare Monitoring." United International Journal of Engineering and Sciences (UIJES), vol. 5, no. 4, Dec. 2025, pp. 46–57.
  36. A. Mahida, "An Intellectual Zero Trust Security Framework Using Deep Reinforcement Learning for Predictive Threat Mitigation in AI-Based Fraud Detection Systems," in IEEE Access, vol. 14, pp. 24602-24617, 2026, doi: 10.1109/ACCESS.2026.3664389.
  37. A. Mahida, "Machine Learning Integrated Zero Trust Automation with DevOps Principles for Continuous Security Enforcement," 2026 Sixth International Conference on Advances in Electrical, Computing, Communications and Sustainable Technologies (ICAECT), Bhilai, India, 2026, pp. 1-7, doi: 10.1109/ICAECT68478.2026.11426026.
  38. Tummuri, S. S. R. (2022). Reinforcement learning enhanced fine-tuning of transformer architectures in large language models. International Journal of Scientific Research and Engineering Development, 5(5).
  39. S. S. R. Tummuri, "Machine Learning-Driven Data Quality Monitoring for Fault-Tolerant Data Pipelines," 2025 4th International Conference on Computational Modelling, Simulation and Optimization (ICCMO), Singapore, Singapore, 2025, pp. 154-159, doi: 10.1109/ICCMO67468.2025.00036.
  - A. K. Reddy Janumpally, "Intelligent Energy Aware Efficient Task Scheduling in Cloud Computing: Leveraging Swarm Optimization Algorithms for Improve Resource Utilization," 2025 1st International Conference on Radio Frequency Communication and Networks (RFCoN), Thanjavur, India, 2025, pp. 1-6, doi: 10.1109/RFCoN62306.2025.11085278.
  40. Janumpally, Bharath Kumar Reddy. (2026). Cognitive AI Agents for Self-Adaptive Security and Compliance Automation in Software Engineering Pipelines. 10.1109/ICAUC68182.2026.11441048.
  41. Yachamaneni T, Kotadiya U, Arora AS. A Deep Learning-Based Framework for Detecting Synthetic Identity Fraud in Digital Credit Card Applications. IJERET [Internet]. 2023 Dec. 30 [cited 2026 Apr. 5];4(4):43-52.
  42. Arora AS, Yachamaneni T, Kotadiya U. Architectural Optimization of Serverless Big Data Pipelines for AI Workloads Using Cloud Functions and Managed Spark on GCP. IJETCSIT [Internet]. 2024 Mar. 30 [cited 2026 Apr. 5];5(1):61-8.