Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 19 No. 9 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v19i09.39927

SHORT PAPER

Static Hand Gesture Recognition Using Novel Convolutional Neural Network and Support Vector Machine

Parsipogu Glory Veronica(✉), Ravi Kumar Mokkapati, Lakshmi Prasanna Jagupilla, Chella Santhosh

[1]Department of Electronics and Communication Engineering in Koneru Lakshmaiah Education Foundation (Deemed to be University), Guntur, Andhra Pradesh, India

gloryveronica99@gmail.com

ABSTRACT

Hand tracking and identification through visual means pose a challenging problem. To simplify the identification of hand gestures, some systems have incorporated position markers or colored bands, which are not ideal for controlling robots due to their inconvenience. The motion recognition problem can be solved by combining object identification, recognition, and tracking using image processing techniques. A wide variety of target detection and recognition image processing methods are available. This paper proposes novel CNN-based methods to create a user-free hand gesture detection system. The use of synthetic techniques is recommended to improve recognition accuracy. The proposed method offers several advantages over existing methods, including higher accuracy and real-time hand gesture recognition suitable for sign language recognition and human-computer interaction. The CNN automatically extracts high-level characteristics from the source picture, and the SVM is used to classify these features. This study employed a CNN to automatically extract traits from raw EMG images, which is different from conventional feature extractors. The SVM classifier then determines which hand gestures are being made. Our tests demonstrate that the proposed strategy achieves superior accuracy compared to using only CNN.

KEYWORDS

novel convolutional neural network, hand gesture recognition, data augmentation, support vector machine, human-computer interaction, sign language.

1 INTRODUCTION

Static hand gesture recognition has gained considerable interest in recent years because of its potential applications in many areas, including virtual reality, sign language recognition, and human-computer interaction. This research, proposes a novel approach combining a convolutional neural network (CNN) and support vector machine (SVM) for accurate and efficient static hand gesture recognition.

Veronica, P.G., Mokkapati, R.K., Jagupilla, L.P., Santhosh, C. (2023). Static Hand Gesture Recognition Using Novel Convolutional Neural Network and Support Vector Machine. *International Journal of Online and Biomedical Engineering (iJOE)*, 19(9), pp. 131–141. https://doi.org/10.3991/ijoe.v19i09.39927

Article submitted 2023-03-27. Resubmitted 2023-04-25. Final acceptance 2023-04-25. Final version published as submitted by the authors.

The proposed system comprises the feature extraction stage and the classification stage. At the step of feature extraction, we first pre-process the hand gesture images and extract the region of interest (ROI) using skin color segmentation [1]. Then, we apply several image processing techniques, such as smoothing, thresholding, and contour detection, to enhance the quality of the ROIs. Finally, we extract relevant features from the pre-processed ROIs using CNN [2]. In the classification stage, we employ SVM to classify the extracted features into different hand gestures. SVM is a powerful machine learning algorithm that can handle high-dimensional data and is well-suited for binary classification tasks. We compare our approach using a variety of cutting-edge techniques, such as deep learning and conventional machine learning algorithms models. The results show that our approach achieves higher accuracy and lower computational complexity than the existing methods [3][4].

The use of a novel CNN architecture for feature extraction is one of our work's major advances. Unlike the traditional CNNs that use completely connected layers after convolutional layers, our CNN architecture uses a combination of a global average pooling layer followed by convolutional and pooling layers [5]. The typical pooling layer worldwide averages the feature maps across the spatial dimensions, which decreases the number of parameters and boosts the effectiveness of the model. Another contribution of our work utilizes SVM to classify data. SVM is currently widely used in static hand gesture recognition, but our approach extends the use of SVM by combining it with a novel CNN architecture for feature extraction [6][7]. This combination of CNN and SVM achieves higher accuracy and lower computational complexity than the traditional SVM-based methods.

1.1 Background

Static hand gesture recognition in the field of computer vision, recognition is a key area for study and human-computer interaction. It involves the automatic recognition of hand gestures made by a user in a still image or video frame [8]. A type of deep neural network called convolutional neural networks (CNNs) has been extensively utilized in image recognition tasks. They are composed of multiple layers of interconnected processing nodes that can extract and identify visual features from input images. In hand motion identification, a CNN can be trained to identify visual features characteristic of different hand gestures. The output of the CNN can then be fed into an SVM to classify the hand gesture. This approach is commonly known as a CNN-SVM pipeline [9][10]. Researchers have proposed various novel CNN and SVM approaches to address these challenges. These approaches aim to improve accuracy and reduce computational complexity by incorporating new techniques for feature extraction and classification [11][12].

1.2 Problem statement

The purpose of the statement is to design an accurate and efficient static hand gesture recognition system that can recognize and categorize hand movements from images or videos using a new convolutional neural network (CNN) and support vector machine (SVM). The system seeks to be useful in a variety of domains, including human-computer interaction, gaming, robotics, and sign language recognition. The system should be able to handle a variety of hand shapes, sizes, orientations, and lighting situations, and it should be able to work in real time with minimal latency. The ultimate objective is to develop a dependable and user-friendly gesture detection system that may improve user experience and accessibility by enabling users to engage with technology through natural and intuitive hand movements.

2 METHODOLOGY

2.1 Novel CNN architecture for feature extraction

To comprehend the Custom model's design, we need to know that it comprises two sets of convolution layers along with a combination of average pooling. Following these layers, there are three fully connected layers [13]. Finally, the Softmax classifier is utilized to classify the images into their respective classes. To understand the architecture of the Custom model, it is important to note that it is composed of two sets of convolution layers combined with average pooling [14]. After these layers, there are three fully connected layers, and ultimately, the Softmax classifier is employed to classify the images into their respective classes. Following that, there is a convolution layer with sixteen 5×5 filters, resulting in a modification of the feature map to 29×29×16 [15][16]. The output size is determined using a similar approach. Subsequently, another average pooling or subsampling layer is implemented, reducing the feature map's size by half to 14×14×16. However, we have made changes to the architecture to suit our model. Thus, our Custom model comprises 3 fully connected layers, 2 pooling layers, and 2 convolution layers. This is the ultimate architecture of the Lenet-5 model as shown in Figure 1.

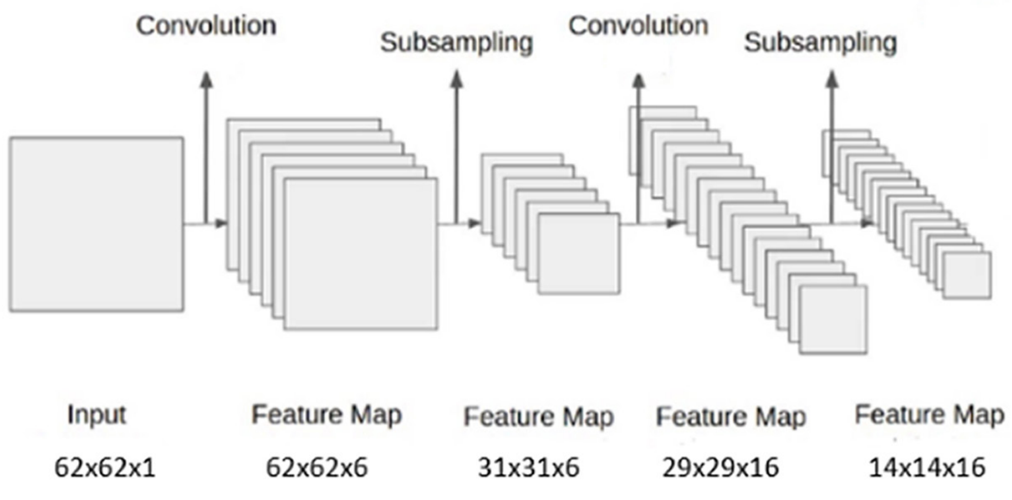

Fig. 1. Architecture for custom model

2.2 Proposed work

To achieve this goal, the following sub-problems need to be addressed as shown in Figure 2:

1. **Dataset collection:** Dataset collection entails capturing and accumulating hand gesture images or videos through sensors or cameras to train the models [17].

2. **Pre-processing:** Pre-processing the data to ensure consistency in lighting conditions, hand orientation, and background.
3. **Feature extraction:** Developing a novel CNN architecture to extract relevant features from the pre-processed data that are characteristic of different hand gestures [18].
4. **Training and optimization:** Training the CNN using the pre-processed data to learn discriminative features and optimize the classification performance. This involves fine-tuning the CNN parameters, selecting appropriate hyperparameters, and optimizing the loss function [19].

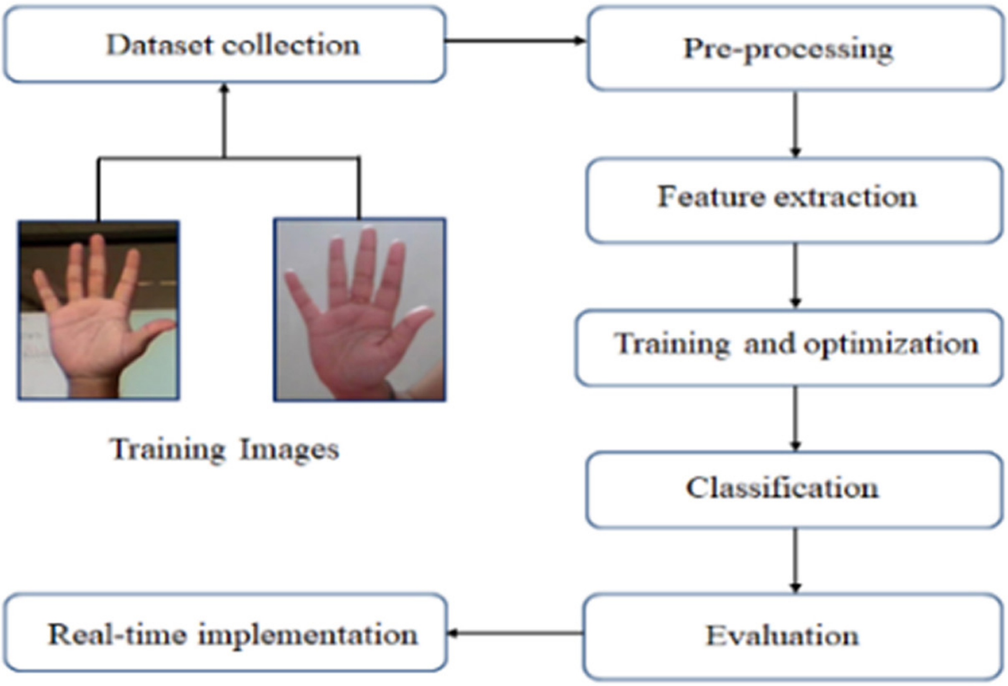

Fig. 2. Proposed hand gesture method

5. **Classification:** Developing an SVM classifier that can accurately classify based on the characteristics taken from the CNN, and hand gestures.
6. **Evaluation:** Evaluate the performance of the system using standard measures like F1 score, recall, precision, and accuracy, and compare it to existing state-of-the-art methods [20].
7. **Real-time implementation:** Optimizing the system to increase and decrease computational complexity speed and efficiency, enabling it intended for use in real-time apps.

2.3 Implementation details

i. **Capture the image from the web camera:** These systems use the web camera to take pictures of a hand or hands, which are subsequently analyzed to identify various hand motions.
ii. **Detect hand-in images using data augmentation:** By creating new instances from the current ones, a technique known as data augmentation can be utilized to increase the size of a dataset [21].

iii. **Crop the image and remove the part under the wrist region and background:** The following step is frequently to crop the image to get rid of any extraneous portions after identifying the hand region in a picture.
iv. **Dataset Collection:** Assemble a collection of hand gesture photographs with associated labels for each gesture.
v. **Data Pre-processing:** To ensure that all of the photographs are the same size, pre-process the data by normalizing the pixel values to [0,1] and resizing the images.
vi. **Data Augmentation:** Use data augmentation methods like random rotation, translation, and scaling to create more images to expand the dataset [22].

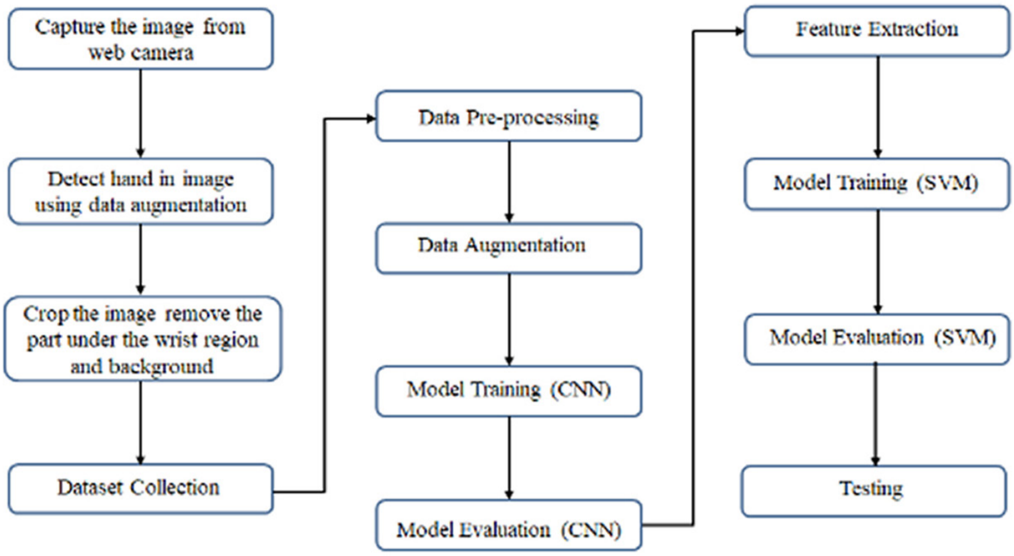

Fig. 3. Algorithm of hand gesture recognition using a novel CNN and SVM

vii. **Model Training (CNN):** Train a unique architecture-based CNN using the dataset to identify static hand motions. Convolutional layers should be Pooling and completely interconnected layers in the CNN come next.
viii. **Model Evaluation (CNN):** Evaluation of the learned CNN model's results using a validation set. For the highest accuracy, adjust the CNN's hyperparameters, including the learning rate, quantity, size, and thickness of filters, as well as layers [23].
ix. **Feature Extraction:** For each image in the dataset, extract features from the trained CNN model's output. A Support Vector Machine These features serve as the classifier's data [24].
x. **Model Training (SVM):** On the retrieved features and related labels, train an SVM classifier. To get the best accuracy, adjust the SVM's hyperparameters, including the kernel function, regularisation parameter, and gamma.
xi. **Model Evaluation(SVM):** Assess the effectiveness of the trained SVM model using the validation set. To get the best accuracy, tweak the SVM's hyperparameters [25].
xii. **Testing:** To assess the effectiveness of the completed model, run it on a test set.

Figure 3 demonstrates the suggested hand gesture detection algorithm using a novel CNN and SVM.

3 TRAINING & TESTING

For training and testing, we use Kernel regulizers a machine-learning approach that aids in the prevention of overfitting in models that employ kernel techniques, such as support vector machines (SVMs). Also, this technique will identify a function that transfers the input data to a high-dimensional feature space where the data may be separated into distinct classes more easily. Such as support vector machines (SVMs). Also, this technique will identify a function that transfers the input data to a high-dimensional feature space where the data may be separated into distinct classes more easily. However, this might result in overfitting, in which the model gets too complicated and too closely fits the training data, resulting in poor generalization of new data. A mathematical function like SoftMax is used to turn a vector of real values into probability distribution at the output layer of the neural network over various classes. The image data generator used in this case enables real-time data augmentation during training, which can increase the model's performance and generalization capabilities and also expand the training data set's size. This whole process is done for the Lenet-5 architecture and Custom model architecture. The lenet-5 contains 5 layers with learnable parameters considering 32x32 pixels as input with 3 convolution layers, two average pooling layers, and two fully connected layers with a SoftMax classifier. Also, for Custom model architecture, we consider 64x64 pixels which have 2 convolution layers and 2 pooling layers, 3 fully connected layers. We have created our own dataset with these 5 hand gestures (Fist, Five, Okay, Rad & Thumbs up) which are used as training datasets and testing images given as an array of images in Figure 4.

(a) (b) (c) (d) (e)

Fig. 4. Hand gestures a) Fist, b) Five, c) Okay, d) Rad, e) Thumbs

4 EXPERIMENTAL RESULTS

The experiment's findings are presented in this part using Novel CNN & SVM, where the lenet-5 and Custom Model were used to obtain the output for five different classes. The experimental results indicate that the model achieved a high level of accuracy. From Figures 5 and 6. We plotted the accuracy and validation accuracy from historical epochs using the Lenet-5 Model and displayed them on a graph with the x-axis labeled as "epoch" and the y-axis labeled as "accuracy". From Figure 5, the graph shows that the model's accuracy reached 95% & from Figure 6, the graph displays the model's loss.

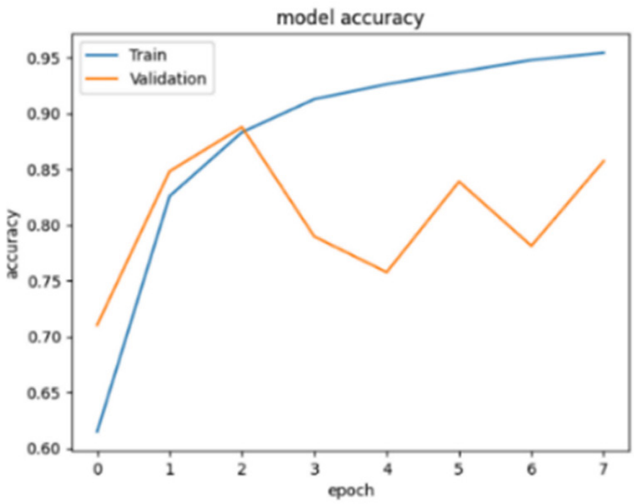

Fig. 5. Reliability of the lenet-5 models

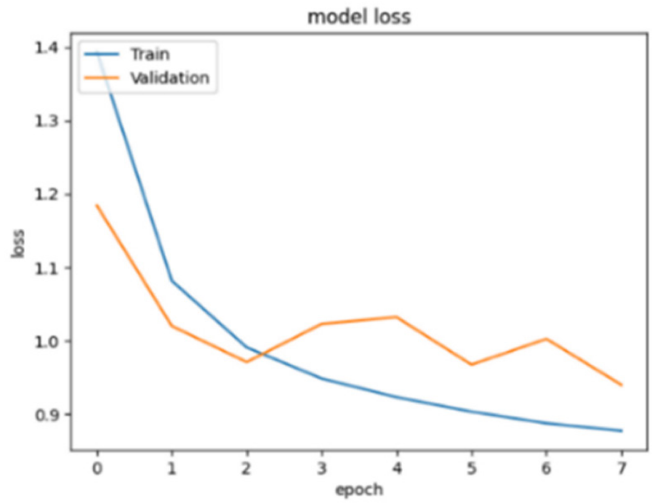

Fig. 6. Loss of the lenet-5 models

This section discusses the results of an experiment utilizing a CNN configuration. The experimental outcomes demonstrate that the suggested method of using Novel CNN & SVM for Hand Gesture Recognition can be tailored to the dataset, as evidenced by the high accuracy achieved by the model. Using the Custom Model, we plotted the accuracy and validation accuracy from historical epochs on a graph, with the x-axis labeled as "epoch" and the y-axis labeled as "accuracy". Figure 7 The graph demonstrates the model's 99% accuracy and Figure 8 The graph shows that the model loss is 83%. The above graphs show comparative graphs of accuracy and loss. The graphs demonstrate the accuracy of Train and Validation from historical epochs, with Train's accuracy being greater than validation accuracy at each epoch. Furthermore, the trained accuracy model advanced more quickly than the validated accuracy model. The loss of Train was similarly lower than the accuracy model's validation loss.

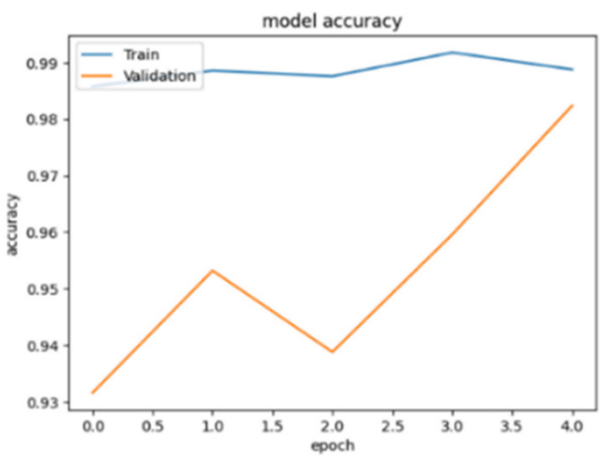

Fig. 7. Reliability of the custom models

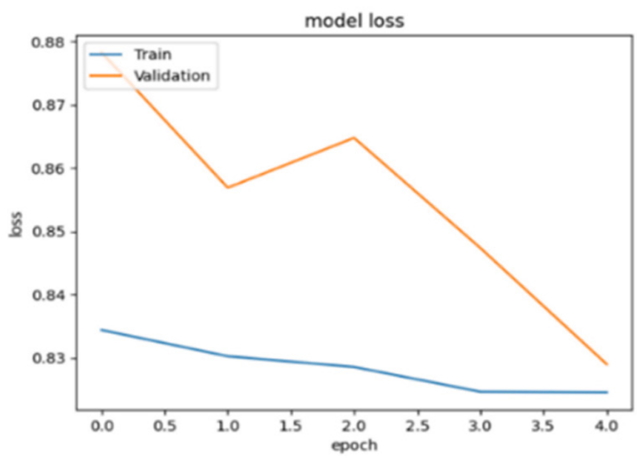

Fig. 8. Loss of the custom models

4.1 Confusion matrix

The performance of a classification method is determined using a confusion matrix for multiclass classification. It takes the form of a matrix. The confusion matrix compares actual and expected values. As the input has 5 values, we get a 5x5 matrix.

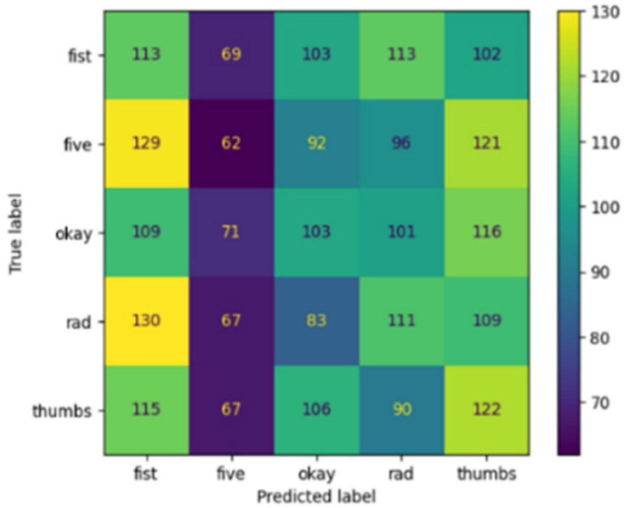

Fig. 9. Confusion matrix of lenet-5 model

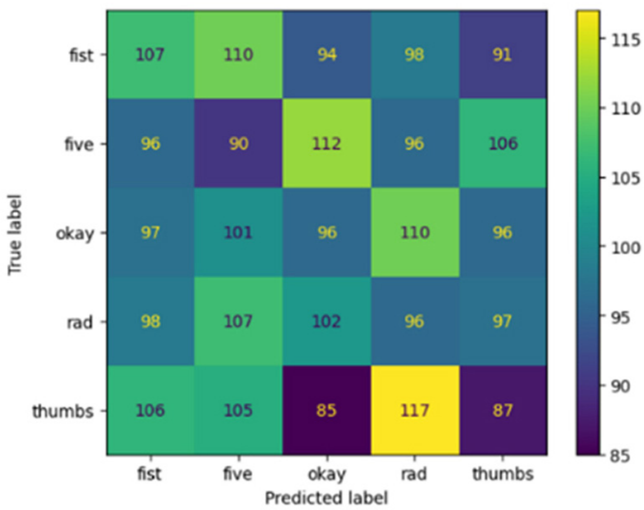

Fig. 10. Confusion matrix of custom model

Predicting the test set using the model, changing the dimension of the array. Getting the actual labels. Plotting the confusion matrix for actual and predicted values along with the display labels. Finally plotting and displaying the confusion matrix in Figures 9 and 10. The use of the Lenet-5 Model and Custom Model for Static Hand Gesture Recognition Using a Novel Convolutional Neural Network and Support Vector Machine demonstrates the adaptability of the suggested strategy for the dataset, with the achieved accuracy of 95% and 99%.

5 CONCLUSION

In this study, we developed a brand-new CNN-based SVM gesture detection technique that can be used in robotic systems or any other system. In this study, we can infer that CNN employs a data-driven methodology and that adding new data substantially influences deep learning. Even if the system effectively recognizes motions, some expansion is still feasible. As a result, it is possible to recognize gestures more precisely. The list of gestures that can be recognized can be expanded. The backdrop was expected to be less complicated. Consequently, another enhancement may be the ability to recognize motions against complicated backgrounds. The outcome demonstrates a 99% improvement in the suggested algorithm's accuracy. Due to the low performance of the suggested strategy, the gesture dataset is not large enough. As a result, we may increase the precision of the detection and identification processes at the beginning and conclusion of the gesture.

6 FUTURE WORK

The following actions will be taken in the future to quicken the second frame rate, enhance accuracy by boosting input picture resolution, or use techniques from our prior research, and this system is unable to recognize motions done with both hands. The identification of movements done with both hands might thus be another area of future exploration.

7 REFERENCES

[1] J. P. Sahoo, A. J. Prakash, P. Pławiak, and S. Samantray, "Real-Time Hand Gesture Recognition using Fine-Tuned Convolutional Neural Network," Sensors, vol. 22, no. 3, 2022. https://doi.org/10.3390/s22030706

[2] M. Z. Islam, M. S. Hossain, R. Ul Islam, and K. Andersson, "Static hand gesture recognition using a convolutional neural network with data augmentation," 2019 Jt. 8th Int. Conf. Informatics, Electron. Vision, ICIEV 2019 3rd Int. Conf. Imaging, Vis. Pattern Recognition, icIVPR 2019 with Int. Conf. Act. Behav. Comput. ABC 2019, pp. 324–329, 2019. https://doi.org/10.1109/ICIEV.2019.8858563

[3] R. F. Pinto, C. D. B. Borges, A. M. A. Almeida, and I. C. Paula, "Static Hand Gesture Recognition Based on Convolutional Neural Networks," J. Electron. Comput. Eng., vol. 2019, 2019. https://doi.org/10.1155/2019/4167890

[4] S. R, R. N, and A. R, "A Smart Vision Based Single Handed Gesture Recognition System using Deep Neural Networks," 2021. https://doi.org/10.4108/eai.7-6-2021.2308783

[5] A. Mujahid et al., "Real-Time Hand Gesture Recognition Based on Deep Learning YOLOv3 Model," Appl. Sci., vol. 11, no. 9, 2021. https://doi.org/10.3390/app11094164

[6] H. Chen, R. Tong, M. Chen, Y. Fang, and H. Liu, "A Hybrid Cnn-SVM Classifier for Hand Gesture Recognition with Surface Emg Signals," Proc. - Int. Conf. Mach. Learn. Cybern., vol. 2, pp. 619–624, 2018. https://doi.org/10.1109/ICMLC.2018.8526976

[7] R. K. Singh, A. K. Mishra, and R. Mishra, "Hand Gesture Identification for Improving Accuracy using Convolutional Neural Network (CNN)," Sci. Temper, vol. 13, no. 2, pp. 336–344, 2022. https://doi.org/10.58414/SCIENTIFICTEMPER.13.2.2022.336-344

[8] A. Kika and A. Koni, "Hand Gesture Recognition using Convolutional Neural Network and Histogram of Oriented Gradients Features," CEUR Workshop Proc., vol. 2280, pp. 75–79, 2018.

[9] D. G. Leon et al., "Video Hand Gestures Recognition using Depth Camera and Lightweight CNN," IEEE Sens. J., vol. 22, no. 14, pp. 14610–14619, 2022. https://doi.org/10.1109/JSEN.2022.3181518

[10] T. T. Nguyen et al., "A Continuous Real-time Hand Gesture Recognition Method based on Skeleton," 2022 11th Int. Conf. Control. Autom. Inf. Sci. ICCAIS 2022, pp. 273–278, 2022. https://doi.org/10.1109/ICCAIS56082.2022.9990122

[11] J. P. Sahoo, S. Ari, and S. K. Patra, "Hand Gesture Recognition Using PCA Based Deep CNN Reduced Features and SVM Classifier," Proc. - 2019 IEEE Int. Symp. Smart Electron. Syst. iSES 2019, pp. 221–224, 2019. https://doi.org/10.1109/iSES47678.2019.00056

[12] G. Amrutkar, M. Shinde, K. Thattekar, and M. Dixit, "Hand Gesture Recognition Using Deep Learning," AIP Conf. Proc., vol. 2494, 2022. https://doi.org/10.1063/5.0110700

[13] A. Ikram and Y. Liu, "Skeleton-based dynamic hand gesture recognition using LSTM and CNN," ACM Int. Conf. Proceeding Ser., pp. 63–68, 2020. https://doi.org/10.1145/3421558.3421568

[14] A. Ikram and Y. Liu, "Real-Time Hand Gesture Recognition Using Leap Motion Controller Based on CNN-SVM Architecture," Int. Conf. Virtual Rehabil. ICVR, vol. 2021-May, pp. 5–9, 2021. https://doi.org/10.1109/ICVR51878.2021.9483844

[15] F. Khan, S. K. Leem, and S. H. Cho, "In-Air Continuous Writing using UWB Impulse Radar Sensors," IEEE Access, vol. 8, pp. 99302–99311, 2020. https://doi.org/10.1109/ACCESS.2020.2994281

[16] D. S. Korti and Z. Slimane, "Improving Dynamic Hand Gesture Recognition based IR-UWB using Offline Data Augmentation and Deep Learning," 4th IEEE Int. Conf. Artif. Intell. Eng. Technol. IICAIET 2022, 2022. https://doi.org/10.1109/IICAIET55139.2022.9936862

[17] J. H. Sun, T. T. Ji, S. Bin Zhang, J. K. Yang, and G. R. Ji, "Research on the Hand Gesture Recognition based on Deep Learning," 2018 12th Int. Symp. Antennas, Propag. EM Theory, ISAPE 2018 - Proc., 2019. https://doi.org/10.1109/ISAPE.2018.8634348

[18] N. Pugeault and R. Bowden, "Spelling It Out : Real–Time ASL Fingerspelling Recognition the University of Surrey," Hand Comput. Vis. Work. (ICCV Work. 2011 IEEE Int. Conf., pp. 1114–1119, 2011. https://doi.org/10.1109/ICCVW.2011.6130290

[19] M. Yasen and S. Jusoh, "A Systematic Review on Hand Gesture Recognition Techniques, Challenges, and Applications," PeerJ Comput. Sci., vol. 2019, no. 9, pp. 1–30, 2019. https://doi.org/10.7717/peerj-cs.218

[20] S. Ahmed, D. Wang, J. Park, and S. H. Cho, "UWB-Gestures, a Public Dataset of Dynamic Hand Gestures Acquired using Impulse Radar Sensors," Sci. Data, vol. 8, no. 1, pp. 1–10, 2021. https://doi.org/10.1038/s41597-021-00876-0

[21] E. Ohn-Bar and M. M. Trivedi, "Hand Gesture Recognition in Real Time for Automotive Interfaces: A Multimodal Vision-Based Approach and Evaluations," IEEE Trans. Intell. Transp. Syst., vol. 15, no. 6, pp. 2368–2377, 2014. https://doi.org/10.1109/TITS.2014.2337331

[22] O. Köpüklü, A. Gunduz, N. Kose, and G. Rigoll, "Real-time hand gesture detection and classification using convolutional neural networks," Proc. - 14th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2019, 2019. https://doi.org/10.1109/FG.2019.8756576

[23] P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree, and J. Kautz, "Online Detection and Classification of Dynamic Hand Gestures with Recurrent 3D Convolutional Neural Networks," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 4207–4215, 2016. https://doi.org/10.1109/CVPR.2016.456

[24] Z. Liu, X. Chai, Z. Liu, and X. Chen, "Continuous Gesture Recognition with Hand-Oriented Spatiotemporal Feature," Proc. - 2017 IEEE Int. Conf. Comput. Vis. Work. ICCVW 2017, vol. 2018-Janua, pp. 3056–3064, 2017. https://doi.org/10.1109/ICCVW.2017.361

[25] V. Adithya and R. Rajesh, "A Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition," Procedia Comput. Sci., vol. 171, no. 2019, pp. 2353–2361, 2020. https://doi.org/10.1016/j.procs.2020.04.255

8 AUTHORS

Parsipogu Glory Veronica received the B.Tech. degree (2020) in Electronics and Communication Engineering from the College of Hyderabad Institute of Technology and Management (HITAM), Hyderabad, and the M.Tech degree (2023) in Electronics and Communication Engineering (Automation and Robotics) at Koneru Lakshmaiah Education Foundation, Vijayawada. She can be contacted at email: gloryveronica99@gmail.com.

Ravi Kumar Mokkapati completed his Ph.D. from Koneru Lakshmaiah Education Foundation in the area of Communications. His significant research interests are ionospheric scintillation studies, satellite communications, and radio wave propagation. He is involved with the research work carried out at Koneru

Lakshmaiah Education Foundation using GPS and IRNSS/NAVIC receivers. He can be contacted at email: ravikumar@kluniversity.in.

Lakshmi Prasanna Jagupilla received her B.Tech. in Electronics and Computers Engineering from affiliated JNT University, Kakinada, India in 2013. She was awarded the M.Tech., Degree in VLSI Design from affiliated JNTU Kakinada, India in 2015, and she is Pursuing a Ph.D., Degree from the National Institute of Technology, Warangal in the area of photovoltaics for solar cell applications. Currently, she is in the position of Assistant Professor in the Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist., AP, India. Her research interests include VLSI, solar cells, photovoltaics, and IoT. She can be contacted at email: lakshmiprasannanewmail@kluniversity.in.

Chella Santhosh was born in Andhra Pradesh, India. He received his master's and Ph.D. degrees from VIT University, Vellore. He is presently working as an Associate Professor in the Microelectronics Research Group, Department of Electronics and Communication Engineering at Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India. His current research areas are Nanomaterials, Energy and Environmental Applications, MOSFET, FINFET, and VLSI Circuits. He can be contacted at email: raurisanthosh@gmail.com.