IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue.4, April 2019 E-ISSN: 2347-2693

Deep Learning-based Hybridized LSTM model for Gesture Recognition

Sunil D. Kale

Computer Engineering Department, Pune Institute of Computer Technology, Pune, Maharashtra, India

Author's Mail Id: kalesunild@gmail.com, Tel.: +91-9921932670

DOI: https://doi.org/10.26438/ijcse/v7i4.12391246 | Available online at: www.ijcseonline.org

Accepted: 23/Apr/2019, Published: 30/Apr/2019

Abstract—In Human-Computer Interaction, gesture recognition is a prominent topic. Human-computer interaction (HCI) allows computers to recognize and interpret human gestures as commands. Gesture recognition is important for ease of use to operate computer machines. It has wider range of applications in the area like talking with machine, medical operation, computer game control, control of home appliances, car control driving and communication. In this research work A real-time Hand Gesture Recognition System is proposed with hybrid approach of Convolution Neural Network (CNN) and the Recurrent Neural Network (RNN). Moreover experimentation with pre-trained VGG16 is carried out with LSTM. Here LSTM is used to replace the final three layers of the VGG16 architecture, and a soft-max layer is used to produce the output. The integrated model is recognizing both static and dynamic hand motions. Proposed model has obtained training accuracy as 92.71% and validation accuracy is 87.50%.

Keywords— *Convolution Neural Network (CNN), Human-Computer Interaction (HCI), Recurrent Neural Network (RNN), LSTM (Long Short Term Memory).*

I. INTRODUCTION

People's interactions with computers are referred to as "human-computer interaction" (HCI). Making systems that are user-friendly for people is a goal of HCI. Human-computer interaction (HCI) aims to improve the usability and responsiveness of computers. It can be applied to many other fields, including gaming, education, medicine, disaster planning and response. Recognition of hand gestures has long been viewed as a helpful resource in human-computer interaction. Body movements are incorporated into gestures, which are nonverbal forms of communication that can be utilized to provide instructions to a system. Static gestures and dynamic gestures are the two categories of gestures. The thumbs-up gesture is an example of a static gesture that disregards movement.The thumbs-up gesture is an example of a static gesture that disregards movement. Dynamic movements, such as writing letters in the air, measure the angles formed by the fingers at their beginning and end points. Machine learning is used to address numerous real-time problems. It is frequently employed to address classification, detection, recognition, and prediction issues.

A method for automating data-driven tasks is machine learning. A model can provide new data or predictions based on previously collected data as its output. In order to extract more complex features from data, deep learning uses numerous layers of machine learning. "Deep" refers to the amount of levels through which the data is changed. The recognition of hand gestures was initially done via glove-based control. This system has uses not only in human-computer interaction but also in gaming, sign language interpretation, and other fields. Glove-based systems have advanced to the point where they can now perform vision-based recognition without any sensors. The Vision-Based Gesture Recognition system uses a camera to record the gesture as an image or video. Although more difficult to develop, this solution is more user-friendly and does not require the use of any additional gesture analyzing tools.

The vision-based method considers attributes including texture and color, and its limitations include lighting, location, and image noise. In glove- or sensor-based systems, the user must put on a glove that is attached to a sensor in order to detect a hand gesture. A bodysuit is also required for body gesture recognition.

Numerous sensors are required for this technology, along with cables to connect them. The sensors and gloves are pricey, which is the biggest disadvantage of this strategy. Another solid approach to gesture recognition is the depth-based approach. The 3D geometric information depth camera is used in this method. In this technique, the colored image reflects the depth image directly.

The input's illumination, shadow, and color have no impact on this approach. A Kinect device is used to measure the depth of the item in the most recent iteration of depth-based gesture recognition. The major drawback is that a Kinect camera is more expensive. Figure 1 shows the generalise framework of hand gesture recognition.

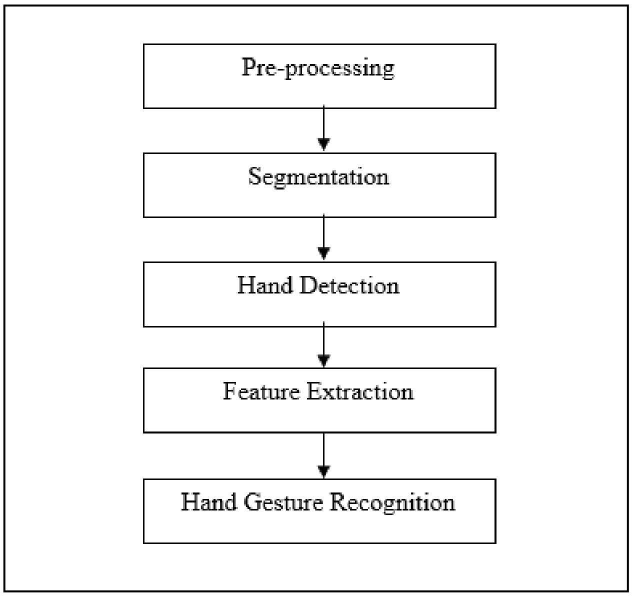

Figure 1. Framework for hand Gesture Recognition

Rest of the paper is structured as follows. In section 2, research done on Gesture recognition is provided. In section 3, proposed technique is shown. In section 4, experimentation and results are discussed and finally in section 5-conclusion and future scope is given.

II. RELATED WORK

For the purpose of recognizing hand gestures, a variety of machine learning algorithms were employed. A few of them are mentioned below. [1] suggests a four-module system for real-time hand gesture detection, including data collecting, pre-processing, feature extraction, and gesture recognition. Using the k-nearest Neighbour (KNN) algorithm and the Histogram of Oriented Gradients (HOG) feature, real-time hand identification is carried out in MATLAB. [2] In this study, researchers examined the viability of identifying human hand motions using deep convolutional neural networks (DCNNs) and micro-Doppler signals obtained from Doppler radar. Radar-based hand gesture detection can be used to operate electronic equipment. When compared to an optical recognition system, radar can operate in any lighting situation and can be housed inside a case. Using solely micro-Doppler signatures on spectrograms devoid of range information, researchers categorize 10 different hand movements. Doppler radar was used to measure the ten gestures—swiping from left to right, right to left, spinning in a clockwise or counterclockwise direction, pushing, double pushing, holding, and double holding—and analyze their spectrograms. The spectrograms were classified using a DCNN, with 90% of the data used for training and 10% used for validation. The proposed method's classification accuracy was found to be 85.6% following five-fold validation. Seven gestures raised the accuracy to 93.1%.

Three crucial steps—hand form identification, tracing of detected hand, and tracing of detected hand—are proposed in [3] for a system for operating a computer utilizing six static and eight dynamic hand gestures. The pre-trained model used by the system to recognize the motions and convert the input into the required instruction is VGG16, a CNN architecture.

[4] presents a powerful Hand Gesture Recognition system that employs a Multi-class Support Vector Machine (MCSVM) for classification and a Deep Convolutional Neural Network (DCNN) for feature extraction. The AdaBoost classifier based on Haar and the CamShift algorithm for hand gesture tracking are used to partition hand gestures, and a convolution neural network is used to identify the hand gesture area.

[5] Presented a real time system for hand sign recognition on the basis of Detection of some meaningful shape based features like orientation, center of mass (centroid), status of fingers, thumb in terms of raised or folded fingers of hand and their respective location in image.

[6] Elaborated the basic concept of sign language recognition system and review of its existing techniques along with their comparison presented.

[7] Provides a multi-scale and multi-modal deep learning-based technique for gesture identification and localization. The entire system operates at three temporal scales, and each visual modality captures spatial information at a specific spatial scale (such as motion of the upper torso or a hand). The key to proposed method is a training strategy that makes use of two key techniques: (1) careful initialization of each modality; and (2) gradual fusion using random channel dropping (dubbed "ModDrop") to learn cross-modality correlations while preserving the individuality of each modality-specific representation. won first place out of 17 teams when submitted experiments on the ChaLearn 2014 Looking at People Challenge gesture recognition track. The model can correct for errors of the individual classifiers as well as noise in the individual channels by significantly increasing identification rates when multiple modalities are combined at various spatial and temporal dimensions. In order to make meaningful predictions from any number of accessible modalities, the proposed ModDrop training technique assures that the classifier is resistant to missing signals in one or more channels. Additionally, testing on the same dataset enhanced with audio showed how the suggested fusion scheme could be applied to modalities of any kind.

[8] The segmentation of hand gestures is accomplished in this study by developing a skin color model and an AdaBoost classifier based on haar to account for the specificities of skin color for hand motions. Additionally, hand movements are denaturized with one frame of video being cut for analysis. In this sense, the human hand is separated from the complex background, and the CamShift algorithm also makes it possible to track hand gestures in real-time. Convolutional neural network then recognizes the area of hand motions that have been detected in real-

time in order to achieve the recognition of 10 common digits. Research indicates 98.3% accuracy.

[9] The goal of this research is to create a model for gesture-based system control while adding continuous facial recognition to prevent unwanted access. To extract the foreground and minimize misclassifications, used a background subtraction model in this study. This method significantly enhances the facial and gesture detector's effectiveness in congested and complicated surroundings. On many databases, including the ORL, Caltech, and Faces96 database, the face detector's performance was evaluated. This study also shows how effective this method is in controlling a robot in real-time. In real-time applications, it offers an accuracy of 94.44% for detecting faces and more than 90.8% for recognizing motions.

[10] Introduction to gesture recognition, its various applications, the workings of the gesture recognition system, applications, the idea that has been proposed, along with tools for its hardware and software implementation, benefits of the idea, drawbacks of the idea, and the future potential of the gesture recognition system are all covered in this study. [11] This suggested study employs a MATLAB-based approach. This tool is used in the planned work to recognize and process hand gestures. The object detection algorithm was employed for the proposed study. The household appliances are controlled when the image is first taken by the camera and processed by MATLAB. If the preloaded gesture matches the already-existing gesture, the data is delivered to the microcontroller. Other programs, like media players, robotics, and virtual objects, could be operated by a gesture. The camera, PIC microcontroller, fan, light, power supply, LED, and GSM module make up the hardware module. A USB to serial converter bus is used to interface between this hardware component and simulation software.

[12] The purpose of this study is to use depth data and CNN to extract crucial spatial data from depth photos. A series of movements can be more precisely recognized when CNN and RNN work together. Additionally, numerous fusion techniques are investigated to merge both depth and skeleton data in order to retrieve temporal-spatial data. On the dynamic hand gesture-14/28 dataset, an overall accuracy of 85.46% is attained. [13] The purpose of this article is to build a man-machine interface and to identify hand postures. The number of active fingers is calculated after the hand region in the image is identified. This method uses an image or video frame as input, which can be from a web camera or any other camera. This color image is preprocessed, transformed to binary, and then the number of fingers is counted in MATLAB using the scanning method. This strategy is straightforward but effective. The primary goal of the scanning method is to make the code for counting fingers independent of hand size and rotation. [14] This research introduced a depth camera-based dynamic hand gesture recognition system for controlling home appliances. Static hand postures and hand trajectories are used to identify the dynamic hand gesture. Seven frequently used dynamic hand motions can be recognized by the suggested approach. According to experimental findings, the technology is efficient for controlling residential appliances.

[15] In this work, presented a reliable system that may be applied in real-world scenarios with complex backgrounds. Wristbands and long sleeves are not necessary for users to wear. By combining the Viola-Jones object detection framework with TRS moment invariants with operating a pedestal fan, shows how the system performs in various complicated environments.

[16] In this research, authors proposed a system for hand gesture recognition that can decode Indian Sign Language alphabets. Real-time hand tracking, hand segmentation, feature extraction, and gesture recognition are the four modules that make up suggested solution. Hand tracking and segmentation are performed using the Camshift method and the Hue, Saturation, and Intensity (HSV) color model. The Genetic Algorithm is used to recognize gestures. Moreover suggested a simple, low-cost method for properly identifying both single-handed and double-handed motions. Millions of deaf people can communicate with hearing people because to this technique.

[17] In this research An enhanced feature extraction method for hand gestures is proposed that combines the histograms of directed gradients with the skin similarity in order to lessen the impact of background edge information. Each image pixel's gradient now has weight based on the similarity of the skin tones. The hand's features can be improved by these new gradients. Because different cell sizes represent distinct local properties, histograms of directed gradients with various cell sizes are used to categorize the hand motions. The experiment's findings show that the size of the cell has a significant impact on identification rates and that the hand gesture elements can be accurately portrayed by combining histograms of directed gradients with two suitable distinct cells.

[18] This research addresses the creation of a user interface that follows and detects hand motions in real time using depth information gathered by a Kinect sensor. Based on the presumption that the user's hand is the object in the scene that is nearest to the camera, the interest space corresponding to the hands is initially divided. To decrease the scanning time and find the initial pixel on the hand contour in this area, a novel algorithm is suggested.

A directional search technique enables the identification of the full hand contour starting from this pixel. The average recognition rate across 55 static and dynamic gestures is 92.4%. The discussion and evaluation of two potential uses for this study include one for the intuitive control of a software interface and the other for the interpretation of sign digits and gestures for a friendlier human-machine connection.

[19] presents a universal remote that functions like a wristwatch and can control many different gadgets. There are a total of seven specified hand motion-based gestures. The wearable wristwatch-type device is compatible with these seven gestures. These motions can be applied in the same manner to numerous systems. Instead than being designed in discrete hand motions and then being analyzed in gesture commands, gestures are designed in continuous hand motions. A technique for modeling a virtual menu using a consumer electronic device's menu is also presented. The user can operate a number of devices with gestures by using the virtual menu.

The virtual menu must depict the functionality of electronic equipment and must incorporate hand motion features. Presented how a user's hand motions can be employed in a quick and efficient method to use a virtual menu. Finally, conducted user testing to assess the efficiency and comfort of wearable remote control compared to the conventional remote control.

[20] The gesture segmentation and feature extraction steps in conventional algorithms can be removed using the RGB-D static gesture recognition method that is provided in this research. To fine-tune the model, the authors use a two-stage training technique as opposed to typical CNN algorithms. This approach uses depth information to enhance the performance of gesture recognition by setting a feature concatenate layer of RGB and depth images in the CNN structure. The authors also compared their approach to other conventional machine learning techniques, CNN algorithms, and the RGB input alone approach using the American Sign Language (ASL) Recognition dataset. The authors' method achieved the highest accuracy of 91.35% among three groups of comparative tests, exceeding the present state-of-the-art on the ASL dataset.

[21] In order to enable deaf and dumb individuals communicate with normal people and each other more effectively, software that can automatically understand sign language is presented in this work as a system prototype. The two emerging topics of study are pattern recognition and gesture recognition. Hand gestures play a crucial role in nonverbal communication and are essential to daily living. With the aid of a hand gesture detection system, provided access to a novel, comfortable, and user-friendly method of interacting with computers that is more suited to human needs. The software intends to propose a real-time system for hand gesture recognition on the basis of detection of some form-based features, keeping in mind the similarities of the human hand shape with four fingers and one thumb.

[22] In this study, developed a reliable marker-free hand gesture detection system that effectively tracks both static and dynamic hand gestures. System converts the detected motion into actions like opening websites and starting programs like PowerPoint and VLC Player. The presentation slides are cycled through using the dynamic gesture. Findings demonstrate that an intuitive HCI is feasible with the least amount of hardware.

[23] For the whole set of 24 hand motions taken from Thomas Moeslund's gesture recognition database, suggest using deep learning to address the hand gesture recognition problem. Demonstrated that deeper, more biologically inspired neural networks, including convolutional neural networks and stacked denoising autoencoders, can learn the challenging categorization job of hand gestures with lower error rates. The networks under consideration are trained and evaluated using data from the aforementioned public database, and the results are then compared to those of earlier studies that only took into account a tiny fraction of ASL hand gestures.

[24] In this study, provided a method for controlling a computer using a combination of six static and eight dynamic hand gestures. Hand form detection, tracing of the detected hand (if dynamic), and turning the data into the necessary command are the three basic phases. Studies reveal a 93.09% accuracy rate.

[25] This research offers a novel feature vector that can describe dynamic hand gestures and provides a workable method for dynamic hand gesture recognition using only a Leap Motion controller (LMC). Other newspapers have not covered these. To identify dynamic hand motions, the feature vector with depth information is calculated and input into the Hidden Conditional Neural Field (HCNF) classifier. The suggested method's systematic structure consists of two primary steps: feature extraction and classification using the HCNF classifier. On two dynamic hand gesture datasets containing frames obtained using an LMC, the suggested approach is assessed. The LeapMotion-Gesture3D dataset has a recognition accuracy of 89.5%, whereas the Handicraft-Gesture dataset has a recognition accuracy of 95.0%. The proposed technique is suitable for some dynamic hand gesture recognition tasks, according to experimental findings.

[26] In this paper, a brand-new skeleton-based method for 3D hand motion identification is proposed. To be more precise, use the hand's geometric shape to extract an efficient description from the connected joints of the hand skeleton that the Intel RealSense depth camera has returned. Then, using a Gaussian Mixture Model to produce a Fisher Vector representation, each descriptor is encoded. A temporal pyramid ensures that Fisher Vectors and other skeleton-based geometric features are represented on several levels in the final feature vector, which is then employed by a linear SVM classifier to do the classification. The suggested method is tested using a difficult hand gesture dataset made up of 14 movements that were executed by 20 subjects using two different numbers of fingers for the same gesture. According to experimental findings, skeleton-based technique routinely outperforms a depth-based approach in terms of performance.

[27] In this project, created a real-time hand gesture-based human-computer interaction system. The system as a whole is made up of three parts: hand detection, gesture

recognition, and human-computer interaction (HCI) based on recognition. It achieves robust management of mouse and keyboard events with a greater level of gesture recognition accuracy. The convolutional neural network (CNN) is specifically used by us to recognize gestures and makes it possible to identify rather complicated gestures with just one inexpensive monocular camera. To estimate the hand location, on which a steady and fluid mouse cursor control is realized, introduced the Kalman filter. During the HCI stage, created a straightforward technique to prevent the mistaken recognition brought on by noises—mostly fleeting, misleading gestures—and so to increase the interaction's dependability. With more complex command formats than just mouse and keyboard events, the proposed system is extremely extensible and can be utilized in human-robotic or other human-machine interaction scenarios.

[28] The segmentation of hand gestures is accomplished in this study by creating a skin color model and an AdaBoost classifier based on haar to account for the specificities of skin color for hand motions. Additionally, hand gestures are denaturized with one frame of video being cut for analysis. In this sense, the human hand is separated from the complex background, and the CamShift algorithm also makes it possible to track hand gestures in real-time. Convolutional neural network then recognizes the area of hand motions that have been detected in real-time in order to achieve the recognition of 10 common digits. Research indicates 98.3% accuracy.

[29] In this research, provided a novel deep learning-based method for 3D hand gesture detection. Suggested a brand-new Convolutional Neural Network (CNN) in which parallel convolutions are used to handle sequences of hand-skeletal joint locations, and then assessed how well this model performs on classification tasks involving hand gesture sequences. Model doesn't use any depth images; it solely uses hand-skeletal data. When compared to previous published algorithms, experimental findings demonstrate that methodology achieves state-of-the-art performance on a difficult dataset (the DHG dataset from the SHREC 2017 3D Shape Retrieval Contest). For the 14 gesture classes scenario, model achieves a classification accuracy of 91.28%, and for the 28 gesture classes case, it achieves a classification accuracy of 84.35%.

[30] In this paper, outlined the most recent machine learning techniques for recognizing hand gestures using depth information from time-of-flight sensors. This study highlight the accomplishments made in a particular area of research at the Ruhr West University of Applied Sciences' Computational Neuroscience lab. This study confirm that Long Short-Term Memory and Convolutional Neural Networks produce the most trustworthy results by comparing findings to other researchers' work in this area. To validate system in real-world scenarios, conducted user surveys while researching various sensor data fusion approaches within a deep learning framework. Authors collected and published data during the study process in a novel benchmark dataset (REHAP), which contains more than a million distinct three-dimensional hand posture samples.

[31] The usage of existing gesture-recognition systems in low-end devices is restricted by their considerable power and computing resource requirements. Presented AllSee, the first battery-free gesture recognition technology that can be used with a variety of computing devices. AllSee can provide always-on gesture detection for smartphones and tablets while using three to four orders of magnitude less power than cutting-edge solutions. It collects gesture data from wireless signals that are already in use (like TV transmissions), but without the power and computational costs of earlier wireless methods. AllSee prototypes that can detect gestures on RFID tags and power-harvesting sensors. Additionally, combined technology with a generic Nexus S phone to show impressive gesture recognition in through-the-pocket settings. Findings demonstrate that AllSee succeeds.

[32] Presented WiGest is a technology that detects in-air hand gestures near a user's mobile device by using variations in WiFi signal strength. WiGest differs from prior research in that it makes use of common WiFi hardware without any modifications or gesture recognition training. Prepared mutually independent gesture families from the system's identification of various signal change primitives. It is possible to link these families to distinct application actions. Deal with issues such signal cleaning, gesture type and attribute identification, lowering false positives from human interference, and signal polarity adaptation. a proof-of-concept prototype was implemented using readily available computers, and the system was thoroughly tested in both an office setting and a typical apartment with common WiFi access points. According to the results, WiGest is able to detect the fundamental primitives with an accuracy of 87.5% when employing a single AP, even in through-the-wall non-line-of-sight situations. Three overheard APs boost this accuracy to 96%. Additionally, the system's categorization accuracy was 96% when tested utilizing a multi-media player application. This accuracy is unaffected by the presence of other interfering humans, underscoring WiGest's potential to enable widespread hands-free gesture-based mobile device engagement in the future.

[33] A reliable real-time hand gesture recognition system is presented in this study. In this method, the hand is first detected using a specific gesture, followed by tracking; the hand is then segmented using motion and color cues; and finally, scale-space feature detection is integrated with gesture recognition to overcome the aspect ratio limitation present in the majority of learning-based hand gesture methods. The proposed strategy performs satisfactorily when used to picture browsing navigation, according to experimental findings.

[34] The goal of this study is to develop a reliable Kinect sensor-based hand gesture detection system. Proposed a

unique distance metric, Finger-Earth Mover's Distance (FEMD), to evaluate the dissimilarity of hand shapes in order to handle the noisy hand shapes collected from the Kinect sensor. It can more easily discern between hand movements with minor variations because it only matches the finger portions and not the entire hand. Extensive testing shows that hand gesture recognition system is reliable (averaging 93.2% mean accuracy on a difficult 10-gesture dataset), quick (averaging 0.0750 s per frame), resistant to hand articulations, distortions, and changes in scale or orientation, and capable of operating in unfavorable conditions (cluttered backgrounds and poor lighting). Two real-life HCI projects further illustrate the excellence of proposed technology.

[35] A convolution neural network (CNN) method to identify hand gestures of human task activities from a camera image is proposed in this research. The skin model and calibration of hand position and orientation are employed to get the training and testing data for the CNN in order to accomplish the robustness performance. Employed a Gaussian Mixture model (GMM) to train the skin model, which is used to effectively filter out non-skin colors in a picture because the light condition has a significant impact on skin color. By translating and rotating the hand image to a neutral stance, the calibration of hand position and orientation seeks to achieve. After that, the CNN is trained using the calibrated images. An experiment that validated the suggested method for identifying human gestures produced reliable findings using a range of hand positions, orientations, and lighting situations. The viability and dependability of the suggested method are demonstrated through experimental evaluation of seven participants doing seven different hand gestures, with average identification accuracies of roughly 95.96%.

III. PROPOSED SYSTEM

A Gesture Recognition System for operating electrical appliances is being developed to provide remote access without the use of switchboards. The system uses a deep neural network (DNN), a mix of long short-term memory (RNN) and convolutional neural networks (CNN). Combining CNN and LSTM, to recognize dynamic gestures by becoming better at recognizing a range of actions over time. The system is a type of vision-based gesture recognition system. Camera technology is used to record the input hand gesture.

In order to extract features and recognize gestures, the system uses the VGG16 architecture. The final three layers of the VGG model are replaced with the recurrent neural network, which is used to recognize dynamic motions. The Gesture Recognition System's architectural layout is shown in Figure 1. The last three layers of the Convolution Neural Network are initially swapped out for the Recurrent Neural Network to generate the RCNN model. The proposed system employs a pre-trained convolutional neural network with a VGG 16 architecture.

The model is tested with a fresh set of data to examine well it performs with fresh data after being trained with the training datasets, which contain a variety of hand movements in varied lighting and background conditions. The real-time hand motion is recorded using a camera, and it is then pre-processed, including the removal of the background and the scaling and normalization of the image. The pre-processed image is next supplied to the RCNN network in order to extract the features from the input gesture. The recovered features are compared to pre-trained datasets in order to recognize the gesture.

The real-time hand motion is recorded using a camera, and it is then pre-processed, including the removal of the background and the scaling and normalization of the image. The pre-processed image is next supplied to the RCNN network in order to extract the features from the input gesture. The recovered features are compared to pre-trained datasets in order to recognize the gesture. Each gesture is interpreted as a command, in this case, to turn on or off the electronic devices. The four elements that make up this Gesture Recognition system are Dataset collection and pre-processing, Training and Testing, Hand Detection, Feature extraction, and Gesture Recognition.

A. Gathering Datasets and Pre-Processing

The dataset, which consists of distinct hand gesture photographs taken under various circumstances, including lightning, varying backgrounds, dimensions, and so on, is gathered and pre-processed. The pre-processing of the image entails scaling, background removal, and RGB to greyscale conversion. Training and Testing

After pre-processing the dataset, the RCNN model is implemented to recognize the gestures and to interpret the gestures with a command. The model is trained using a training dataset so that the model learns the features of the hand gestures. After training a model, a new dataset is provided for testing. The testing phase helps to check the accuracy of the model.

B. Hand detection

First, The edges of the hand are identified to extract the features for a real-time hand gesture. The number of fingers in a gesture and the gesture's orientation are used to identify the hand.

C. Gesture recognition and feature extraction

The gesture's features are taken from the detected hand and compared to the available datasets to identify the gesture. Each recognized gesture has a command attached to it that the electrical appliances will obey.

IV. EXPERIMENTAL RESULT

A few of the tactics outlined in the literature review can be applied to develop a gesture recognition system. One such technology is the convolution neural network, which can recognize motions. The data set utilized for experimentation is collected from kaggle

In order to develop the system, VGG16, a CNN network that has already been trained, was used. The method makes use of pre-processing, feature extraction, and gesture recognition. Prior to passing through the CNN+LSTM architecture, which uses a hierarchical approach to classification and recognition, the input data is pre-processed. It uses a selective search technique to extract the data and recognize the motion.

FIG. 2 presents train loss versus validation loss. Figure 3 demonstrates train accuracy versus test validation accuracy over 30 epochs.

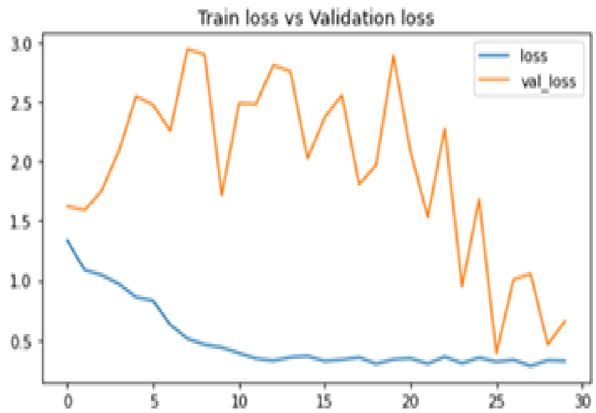

Figure 2. Train loss vs. Validation loss

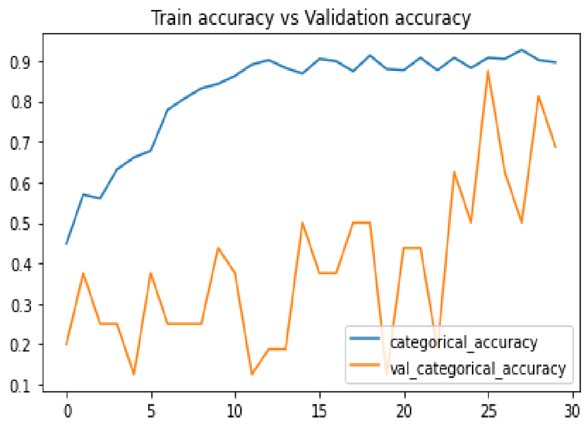

Figure 3. Train accuracy Vs. validation accuracy

The combined CNN and RNN proposed work is contrasted with the current model. LSTM is used to replace the final three layers of the VGG16 architecture, and a softmax layer is used to produce the output. In order to correctly recognize both dynamic and static motions, this is employed to uncover temporal patterns. The suggested model, outperforms the current model in terms of accuracy while using fewer epochs and losses.

V. CONCLUSION

As a result, the Convolution Neural Network and the Recurrent Neural Network are combined to form the Gesture Recognition System, which is used to control electrical appliances. The proposed model recognize both static and dynamic hand gesture with training accuracy 92.71% and validation accuracy 87.50. Here the CNN model will extract useful features from hand gestures to recognize dynamic motions, while the LSTM (RNN) model will recognize temporal patterns. Future scope for gesture recognition to improve the result of accuracy by improving image quality of gesture images. Future devices will be able to be control by a wider variety of gestures, improving the quality of gesture images.

REFERENCES

[1] Sanmukh Kaur, Anuranjana, "Electronic Device Control Using Hand Gesture Recognition System for Differently Abled", 8th International Conference on Cloud Computing, Data Science & Engineering, **2018.**

[2] Youngwook Kim And Brian Toomajian, "Hand Gesture Recognition Using Micro-Doppler Signatures With Convolutional Neural Network", IEEE Access Vol.**4, 2016.**

[3] Soeb Hussain, Rupal Saxena, Xie Han, Jameel Ahmed Khan, Hyunchul Shin, "Hand Gesture Recognition Using Deep Learning", International SoC Design Conference, **2017.**

[4] Md Rashedul Islam, Rasel Ahmed Bhuiyan, Ummey Kulsum Mitu, Jungpil Shin, "Hand Gesture Feature Extraction Using Deep Convolutional Neural Network for Recognizing American Sign Language", 4th International Conference on Frontiers of Signal Processing, **2018.**

[5] Panwar, M., February. Hand gesture recognition based on shape parameters. In 2012 international conference on computing, communication and applications, pp.**1-6, 2012.** IEEE.

[6] Zhang, L.G., Chen, Y., Fang, G., Chen, X. and Gao, W., October. A vision-based sign language recognition system using tied-mixture density HMM. In *Proceedings of the 6th international conference on Multimodal interfaces*, pp.**198-204, 2004.**

[7] Molchanov, P., Gupta, S., Kim, K. and Pulli, K., May. Multi-sensor system for driver's hand-gesture recognition. In 2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG), Vol.**1**, pp.**1-8, 2015.** IEEE.

[8] Jing-Hao Sun, Shu-Bin Zhang, "Research on the Hand Gesture Recognition Based on Deep Learning", 12th International Symposium on Antennas, Propagation and EM Theory, **2018.**

[9] Anush Ananthakumar, "Efficient Face and Gesture Recognition for Time Sensitive Application", IEEE Southwest Symposium on Image Analysis and Interpretation, **2018.**

[10] Mohd. Baqir Khan, Kavya Mishra, Mohammed Abdul Qadeer, "Gesture Recognition using Open-CV", 7th International Conference on Communication Systems and Network Technologies, **2017.**

[11] Arathi P.N, S.Arthika, S.Ponmithra, K.Srinivasan, V.Rukkumani, "Gesture-Based Home Automation System", 2017 International Conference on Nextgen Electronic Technologies: Silicon to Software, **2017.**

[12] Kenneth Lai and Svetlana N. Yanushkevich, "CNN+RNN Depth and Skeleton based Dynamic Hand Gesture Recognition", 24th International Conference on Pattern Recognition, **2018.**

[13]. P Raghu Veera Chowdary, M Nagendra Babu, Thadigotla Venkata Subbareddy, Bommepalli Madhava Reddy, V Elamaran, "Image Processing Algorithms for Gesture Recognition using Communication Control and Computing Technologies, **2014.**

[14]. Chih-Hung Wu and Chang Hong Lin, "Depth-based hand gesture recognition for home appliance control", IEEE 17th International Symposium on Consumer Electronics, **2013.**

[15] Ransalu Senanayake, Sisil Kumarawadu, "A Robust Vision-based Hand Gesture Recognition System for Appliance Control in Smart Homes", IEEE International Conference on Signal Processing, Communication and Computing, **2012.**

[16] Archana S. Ghotkar, Rucha Khatal, Sanjana Khupase, Surbhi Asati &Mithila Hadap, "Hand Gesture Recognition for Indian

Sign Language", International Conference on Computer Communication and Informatics, **2012.**

[17] Lan Tiantian, Shen Jinyuan, Liu Runjie, Guo Yingying, "Hand Gesture Recognition Based on Improved Histograms of Oriented Gradients", The 27th Chinese Control and Decision Conference, **2015.**

[18] Guillaume Plouffe and Ana-Maria Cretu, "Static and Dynamic Hand Gesture Recognition in Depth Data Using Dynamic Time Warping", IEEE Transactions on Instrumentation and Measurement, Vol.**65**, Issue.**2, 2016.**

[19] Dong-Woo Lee, Jeong-Mook Lim, John Sunwoo, Il-Yeon Cho and Cheol-Hoon Lee, "Actual Remote Control: A Universal Remote Control using Hand Motions on a Virtual Menu", IEEE Transactions on Consumer Electronics, Vol.**55**, Issue.**3, 2009.**

[20] Bin Xie1, Xiaoyu He1, Yi Li1, "RGB-D static gesture recognition based on convolutional neural network", The Journal of Engineering, Vol. **2018**, Issue.**16, 2018.**

[21] Nikam, A.S. and Ambekar, A.G., November. Sign language recognition using image based hand gesture recognition techniques. In *2016 online international conference on green engineering and technologies (IC-GET)*, pp.**1-5, 2016.** IEEE.

[22] Haria, A., Subramanian, A., Asokkumar, N., Poddar, S. and Nayak, J.S., Hand gesture recognition for human computer interaction. *Procedia computer science*, **115,** pp.**367-374, 2017.**

[23] Oyedotun, O.K. and Khashman, A., 2017. Deep learning in vision-based static hand gesture recognition. *Neural Computing and Applications*, Vol.**28**, Issue.**12**, pp.**3941-3951, 2017.**

[24] Hussain, S., Saxena, R., Han, X., Khan, J.A. and Shin, H., 2017, November. Hand gesture recognition using deep learning. In *2017 International SoC design conference (ISOCC)*, pp.**48-49, 2017**. IEEE.

[25] Lu, W., Tong, Z. and Chu, J., 2016. Dynamic hand gesture recognition with leap motion controller. *IEEE Signal Processing Letters*, Vol.**23**, Issue.**9**, pp.**1188-1192, 2016.**

[26] De Smedt, Q., Wannous, H. and Vandeborre, J.P., 2016. Skeleton-based dynamic hand gesture recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.**1-9, 2016.**

[27] Xu, P., 2017. A real-time hand gesture recognition and human-computer interaction system. *arXiv preprint arXiv:1704.07296*.

[28] Sun, J.H., Ji, T.T., Zhang, S.B., Yang, J.K. and Ji, G.R., 2018, December. Research on the hand gesture recognition based on deep learning. In *2018 12th International symposium on antennas, propagation and EM theory (ISAPE)*, pp.**1-4, 2018.** IEEE.

[29] Devineau, G., Moutarde, F., Xi, W. and Yang, J., 2018, May. Deep learning for hand gesture recognition on skeletal data. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pp.**106-113, 2018.** IEEE.

[30] Zengeler, N., Kopinski, T. and Handmann, U., 2018. Hand gesture recognition in automotive human–machine interaction using depth cameras. *Sensors*, Vol.**19**, Issue.**1**, p.**59, 2018.**

[31] Kellogg, B., Talla, V. and Gollakota, S., 2014. Bringing gesture recognition to all devices. In 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14), pp.**303-316, 2014.**

[32] Abdelnasser, H., Youssef, M. and Harras, K.A., 2015, April. Wigest: A ubiquitous wifi-based gesture recognition system. In 2015 IEEE conference on computer communications (INFOCOM), pp.**1472-1480, 2015.** IEEE.

[33] Fang, Y., Wang, K., Cheng, J. and Lu, H., 2007, July. A real-time hand gesture recognition method. In 2007 IEEE International Conference on Multimedia and Expo, pp.**995-998, 2007.** IEEE.

[34] Ren, Z., Yuan, J., Meng, J. and Zhang, Z., 2013. Robust part-based hand gesture recognition using kinect sensor. IEEE transactions on multimedia, Vol.**15**, Issue.**5**, pp.**1110-1120, 2013.**

[35] Lin, H.I., Hsu, M.H. and Chen, W.K., 2014, August. Human hand gesture recognition using a convolution neural network. In 2014 IEEE International Conference on Automation Science and Engineering (CASE), pp.**1038-1043, 2014.** IEEE.

Authors Profile

Sunil D. Kale has received his graduate degree from Government College of Engineering, Aurangabad, Maharashtra, India and master's degree M. Tech. (CSE) from Visvesvaraya Technological University, Karnataka, India. He is a Research Scholar in Computer Engineering Department of Smt. Kashibai Navale College of Engineering, Vadgaon (Bk) and working as Assistant Professor in Computer Engineering Department of Pune Institute of Compter Technoloy, Pune, India. His area of interest is text analytics and pattern recognition. He has published 17 papers in national and international journals and conferences. He is a life member of Indian Society of Technical Education.

