Pakistan Journal of Engineering and Technology, PakJET
Multidisciplinary & Peer Reviewed
Volume: 04, Number: 01, Pages: 60- 65, Year: 2021

A Novel Machine Learning based Multiple-user Hand Gesture Recognition Approach

Anam Abid[1], Aaisha Abid[2], Zo-Afshan[3], and Shahzad Anwar[1]

[1]Department of Mechatronics Engineering, University of Engineering and Technology, Peshawar, Pakistan
[2]Department of Electrical Engineering, University of Engineering and Technology, Peshawar, Pakistan
[3]Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering, National University of Science and Technology, Rawalpindi, Pakistan
Corresponding author: Anam Abid (anam.abid@uetpeshawar.edu.pk).

Abstract- **Machines are generally designed offering attributes, such as, of usability, accuracy, affordability, and scalability. This study aims to facilitate value addition Human-Computer Interaction employing hand gesture recognition, introducing a user-friendly and natural interaction for conveying useful information. Initially, machine vision-based techniques are applied for hand segmentation and detection, subsequently, deep learning classifier is trained on various hand gestures. In this study, shape-based hand features, which offer less variation under various lighting conditions are presented in contrast to the other means of hand gesture recognition (such as texture and skin colour). Furthermore, experiments were conducted under varying light conditions and the recognition performance of the developed algorithm for multiple-user hand gesture is investigated. The developed method results in achieving a classification accuracy of 94.6% and 93.% for single and multiple-user hand gestures respectively. The developed hand gesture-based non-verbal communication would assist handicap and physically challenged personals for non-invasive machine interaction.**

***Index Terms*—Convolutional neural network, Hand gesture, Machine Vision, Segmentation.**

I. INTRODUCTION

Gesture recognition is streamed from Human-Computer Interaction (HCI), a user-friendly interface for providing real-time data to a computer. Gesture express both semantic meaning and contextual information are an effective medium of communicating [1]. Hand gestures convey reasonable information and provide better understanding compared to other body gestures such as facial and body movements [2], [3]. Gesture recognition exhibits promising performance for analyzing and extracting the user movement characteristics. Gesture control is a skill which involves human body movement recognition followed by their interpretation, thus resulting in interaction and controlling a real-world system without any direct physical contact. Human gesture control has numerous applications in different fields, for instance, virtual reality, appliance control, robotics, and computer games. In general, the hand gesture recognition is primarily categorized into two branches, namely dynamic and static. Dynamic gesture recognition tries to investigate spatial-temporal characteristics, while static recognition focuses on the single input image analysis [4]–[6]. This paper focuses on static hand gesture recognition. The analysis of static hand-gesture recognition is significant because different hand shapes provide specific information without motion cues. Besides, it could help in reducing redundant frames in dynamic problems. Few limitations of static hand gesture recognition based applications include noise susceptibility and environmental constraints. For some applications, users are limited to place their hand at the nearest place to the sensor which makes it non-user-friendly. However, in many gesture control applications, the input sensor is generally installed in a vicinity closed to the devices which are desired to be controlled. Therefore, static hand gesture recognition is a suitable choice for such applications. The main contributions are summarized as followings.

- The development of a shape-based hand features recognition algorithm offering less variation under various lighting conditions.
- The development of a multiple-user hand gesture database for multiple gestures

The remaining of this manuscript is organised as follows: Section II describes the Literature review, followed by Materials and Methodology in Section III. Section IV discusses the Implementation and Results are presented in section V. Finally, Section VI conclude this study.

II. LITERATURE REVIEW

The research related to hand gesture recognition deals with real-time data acquisition, hand detection (i.e., segmentation) and classification of the gesture (i.e., recognition). In this article, a literature review is organized addressing above mentioned three sections.

A. DATA ACQUISITION

In literature, a variety of techniques are listed employing machine vision-based data capture devices, such as depth, IR camera, stereoscopic camera [7], [8] etc. In vision-based based approaches, depth images are either sensed directly with depth cameras or extracted from a stereo video. The stereoscopic image capturing device offers an alternative in several situations where video cameras capabilities are limited, such as, in low or unpredictable lighting, and in environments with skin coloured objects other than the hands (for instance a face). Other works include incorporating a wearable device which captures the information through different sensors such as gloves/rings with position sensors [9], accelerometers [10], [11] attached to gesture wrists, gesture pads, and wireless finger-ring. Few limitations include wearing of unusual finger rings/ gloves and their associated power requirements. In another approach, muscle tension is measured incorporating electromyogram, or EMG [12], [13]. Though this approach is highly suitable for humans with physical disabilities, however at least two electrodes are required to be attached to the skin which is placed at specific body points. Besides, Neverova [14] researched on multi-modal data, which comprised of audio streams, skeletal data and RGB-D images. Each modality was processed independently initially and then effectively merged to improve recognition performance.

B. HAND SEGMENTATION

After the acquisition of sensor data/image, hand localization is performed on the acquired information using different segmentation methods. The hand detection methods incorporate the principle of segmenting the hand from other parts via threshold configuration, which is generally estimated using depth cues or sometimes manually adjusted for different light conditions. The skeleton-based approaches [6], [15] exploit the geometric shape of the hand to extract an effective descriptor from hand skeleton connected joints returned by the depth camera. Sometimes finger detection and segmentation stages are also included for the determination of hand shapes. Different ways for finger detection include the utilization of knowledge concerning geometrical aspects of hand shapes, contour retrieval analysis, and hand's topological information. Since these methods depend on the sensitivity of the threshold; therefore, it may introduce segmentation errors. In another investigation, near-convex shape decomposition scheme was introduced in [8] which formulates the shape decomposition as an integer optimization.

In contrast, steps involved in the image-based segmentation process include background subtraction and foreground detection using thresholding techniques. In the context of hand gesture recognition, the hand posture/gesture is the foreground which needs to be separated from the background before further processing. The performance of the segmentation process is dependent on the quality of the extracted foreground which consequently depends on suitable thresholding method and light conditions. Few basic Thresholding techniques [16] include threshold binary, threshold binary inverse, threshold to zero, threshold to zero inverse, truncate threshold. Other thresholding methods include mean value, Gaussian, adaptive and OTSU binarization thresholding methods. In contrast to traditional video cameras whose performance is dependent on light conditions, depth camera and laser triangulation methods provide information about the depth of the object in the depth image. The intensity values in an image represent object distance from a viewpoint which could be colour-coded to visually represent the depth of an object.

C. GESTURE CLASSIFICATION

Finally, the segmented hand images and/or their tracked trajectories are classified as a particular gesture or pose, incorporating suitable classification techniques. Many techniques are listed in the literature for the classification of hand gestures such as support vector machines (SVM), Markov models (MM), and artificial neural network (ANN). Hofmann presented a recognition technique which is based on discrete hidden (dHMM) (Hidden Markov Model) to detect dynamic gestures [17]. Likewise, some researchers focused on the application of feature fusion and feature selection, such as fusion of the feature extracted from inertial and depth sensor, Principle Component Analysis (PCA) and hybrid features combining short-time energy with Fast Fourier Transform (FFT) [18]. The machine learning techniques that are widely used for hand gesture recognition in wearable devices based approaches such as wrist accelerometers and position sensors are SVM (Support Vector Machine), RF (Random Forest), KNN (k-nearest neighbours), Naive Bayes.

Recently, deep learning methods such as convolution neural network (CNN), have demonstrated excellent performance in many vision tasks, such as object detection, image classification, and segmentation [19]–[22]. An example of hand gesture recognition through CNNs was discussed in [23] where CNNs are applied on the RGB images for classification. Molchanov [24] also introduced a CNN architecture for RGB-D images in which the classifier is made of two CNN networks (a high-resolution network and a low-resolution network) whose class-membership outputs are combined with an element to element multiplication. Smedt in [21] proposed a deep learning method for hand gesture recognition on the dynamic hand gesture (DHG) dataset. The Guerry method is based on the sequencing the Red, Green, Blue and Depth channels of each RGB-D image. An already pre-trained VGG image classification model is applied to the sequences of five concatenated images successively in time. Generally, extracting features and detecting hand gesture from inputted colour videos is more difficult because of the huge variation in the hands. To address this, an effective HGR system was introduced in [25] for low-cost colour video. In their model, Deep Convolutional Neural Network (DCNN) was deployed for efficient hand features extraction, recognizing American Sign Language (ASL), incorporating a Multi-class Support Vector Machine (MCSVM) for hand sign identification. Shin et al [26] developed a dynamic hand gesture recognition technique utilizing a recurrent neural network (RNN) algorithm, which was assessed based on the gesture database captured by a SmartWatch. The following section would describe the methodology in details.

III. MATERIALS AND METHODOLOGY

In this manuscript, an intelligent hand gesture recognition is proposed for multiple-users hand gestures. For this purpose, machine vision techniques were applied for hand segmentation and detection. Subsequently, deep learning classifier was trained for the recognition of different hand gestures. The block diagram of the proposed methodology is given in FIGURE 1 and the steps involved are discussed as follows.

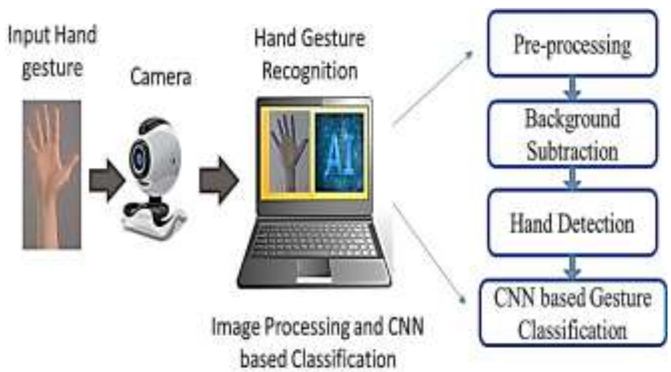

FIGURE 1. THE PROPOSED METHODOLOGY

Generally, pre-processing techniques are applied to transform raw data into understandable information in signal processing application. The processing steps involved here are image cropping, compression, despeckling, and resizing. Furthermore, a pre-defined rectangular-shape Region of Interest (ROI) is incorporated herewith a fixed-location camera setting. Afterwards, the image processing techniques are applied to ROI, therefore image frame is resized according to the defined height and width. In the proposed approach, the colour ROI is converted to grayscale before further processing. Image segmentation refers to the conversion of an image into a collection of regions of pixels that are represented by a mask or a label. The main steps involved in the segmentation process include background subtraction and foreground detection using threshold techniques. In the context of hand gesture recognition, the hand posture/gesture is the foreground which needs to be separated from the background before further processing. Background subtraction is a technique used for creating a foreground mask in static camera applications and to achieve this, background modelling is performed. The background modelling is a two-step process in which firstly a background initialization is carried out and later on background model is updated iteratively to accommodate for any changes in the scene. In this paper, running average technique is incorporated for the background modelling which is further used for separation of the foreground from the background. Initially, the image sequence is analyzed for a particular set of frames. During this sequence of frames, the running average over the current frame and the previous frames is computed using equation 1.

$$dst\left(x,y\right)=(1-\alpha).dst\left(x,y\right)+\alpha.src\left(x,y\right) \qquad (1)$$

where '*dst(x,y)*' and '*src(x,y)*' are the destination/accumulator and input/source image frames respectively and 'α' is the weight of the input image. The running average is iteratively computed and stored in the accumulator image *dst(x,y)*. 'α' decides the speed of updating such that for a lower value of this variable, the running average will be computed over a larger amount of previous frames and vice-versa.

The computation of the absolute difference between the background model (which is a function of time) and the current frame (which is newly introduced object) was performed for the foreground extraction. In the adopted background separation method, a pre-defined threshold is incorporated for foreground extraction. The threshold tuning level depends on the light conditions and the surroundings. Next, the Chain approximation technique is employed for contour approximation which stores only the end-points and for contour retrieval, the external mode has opted in which all the smaller contours are ignored and only the largest contour which is the hand posture is retained.

Traditionally, the segmentation stage is followed by feature extraction and classification stages in which feature vectors of the segmented image are extracted in different ways according to a particular application. In this study, a convolutional neural network (CNN) is employed for model training. The convolutional neural network (CNN) is the class of the deep neural network which is commonly used for the visualization, recognition, segmentation and classification of images. The developed CNN architecture is illustrated in FIGURE 2.

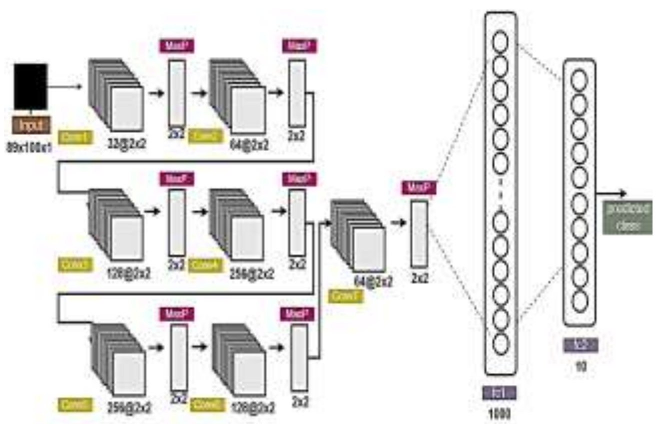

FIGURE 2: The developed CNN Architecture.

The architecture consists of a first convolutional layer having 32 filters of size 2x2 with RELU as its activation. Next, a convolutional layer having 2x2 pooling was introduced to summarize the extracted information from the feature maps. The second convolution layer consists of 64 filters of size 2x2 with the RELU as its activation, followed by another 2x2 pooling layer. Similarly, the next third to seventh convolution layers consists of 128, 256, 256, 128, 64 filters of size 2x2 with the RELU as its activation, followed by another 2x2 pooling layer. Afterwards, there are two fully connected layers in which the first fully connected layer has 1000 neurons. In RELU, all the neurons in a fully connected layer have connections to the previous activation functions. In literature, dropout is a technique employed to prevent model Overfitting, therefore, to avoid this a dropout rate of 0.75 was introduced in the developed model. Finally, there is a second fully connected layer which consists of 10 nodes/neurons and the activation function in this layer is Softmax, transforms input values (which is either positive, negative, zero, or greater than one) into values between 0 and 1 and interpreted as probabilities.

IV. IMPLEMENTATION

The implementation of hand gesture recognition is divided into including data collection and pre-processing, development of training and testing datasets, model training and recognition stages. The details of each stage are presented in the following section.

Firstly, a hand gesture dataset is developed with an image capturing device. The performance of deep learning method depends on the availability of the large database, therefore, a total of 21,000 images of static hand gestures are collected to obtain better accuracy in the recognition process, some examples are shown in Figure 4. The collection of image datasets and advanced image processing were implemented in Spyder platform using python language. Initially, a counter is set for 30 seconds and during this time the laptop camera is enabled to capture the live video feed and observe the background. In the meanwhile, the captured image frames are stored for the initialization of the background model, as discussed in section III. Subsequently to background initialization, the video feed window appears in which the pre-defined region of interest is indicated with green-boundary box. Afterwards, the user places his/her hand in front of camera range and particularly within the indicated green-box boundary which is the desired region of interest (ROI). Later, the threshold value is adjusted according to the light conditions. In this study, the effects of variation of the threshold value on the hand segmentation are studied and the optimal threshold value is determined which is discussed in the result section. Once, a threshold value is set then the data collection process is initiated. The background subtraction and hand segmentation algorithm are presented in TABLE I.

TABLE 1
HAND GESTURE SEGMENTATION

Algorithm
Input: Incoming frames from video data
Output: Extracted hand segment
Background Initialization:
$bg \leftarrow$ None, $\alpha \leftarrow 0.5$
for i = 1: 30
$src \leftarrow$ incoming frame
Update background: $bg = (1-a).bg + \alpha.src$
$i \leftarrow i+1$
end
Hand Segmentation:
$hd \leftarrow$ None, threshold$\leftarrow$30
for each "captured image, *img*"
$diff \leftarrow \lvert img-bg \rvert$
if $diff >$ threshold, **then**
$hd \leftarrow 255$ (white)
else $hd \leftarrow 0$ (black)
return contours in *hd* using chain approximation
end

The images were collected under different lighting conditions and with varying posture orientation. The training dataset consists of image/examples which are employed to fit and tune the parameters. For each gesture the collected training dataset consists of 2000 images, thus for ten gestures a total of 20,000 images were included in the training dataset. After training dataset collection data labels are assigned to the images such that each gesture class is presented by its particular label. The test dataset offers an unbiased evaluation of a final model fit on the training dataset. The test set is the unseen data for the model which has never been considered prior to the training stage. In this study, 100 images have been collected covering all gesture class thus the total test dataset is comprised of 1000 images. The developed approach is evaluated on a challenging hand gesture test dataset containing ten gestures, performed by various participants performing the same gesture with different orientations of hand and fingers. The proposed hand gesture recognition method uses CNN and for its implementation, a fixed size of the input image is required. Therefore, firstly image resizing is performed using PIL (Python Imaging Library) to a width of 100 pixels and a height proportional to the new width. The base width of 100 is suggested and the final resized image is 89 x 100 x 1.

Hand gesture recognition model training is the main step and the proposed CNN architecture (refer to Figure 2) is developed in python and during model training, the weights of connections between neurons are adjusted for gesture classification problem. The regression layer is introduced in the TFlearn model which utilizes Adam optimizer to minimize the loss function. Adam combines the best attributes of adaptive gradient algorithm and root mean square propagation and it performs well for CNN. For the model training, the learning rate (i.e., a hyper-parameter to control the model modifications in response to estimated error) was adjusted using a trial and error method which is 0.001. Next to model training, the subsequent phase is the real-time gesture recognition using the developed model in which classification results are displayed as the predicted class of the identified gesture which is from one of the 10 classes illustrated in FIGURE 3.

FIGURE 3. VARIOUS TYPES OF HAND GESTURES

V. RESULTS AND DISCUSSIONS

This section presents the results of the developed hand gesture recognition algorithm. The performance of the developed algorithm is evaluated using different testing datasets under multiple real-world scenarios, as mentioned in the following.

Scenario 1: Initially, the effects of the binarization threshold were investigated under different light conditions. FIGURE 4(a) illustrates the hand segmentation results with varying threshold values indoor environment with moderate light effects. It is clear from the results that the dilation effect becomes prominent for a smaller value of the threshold. In contrast, a comparatively higher value of threshold results in sharp finger edges due to high erosion effect. The optimal value of threshold under normal room environment in the day time was determined as 30. Similarly, FIGURE 4(b) shows the hand segmentation results with varying threshold values during night time indoor environment. The segmented image is stretched at small values of threshold and it becomes sharp with an increase in the threshold value. Again the optimal value of the threshold lies in the range of 20 to 30. The developed methodology works effectively indoor environment under moderate light conditions and it is independent of day or night time. In the next experiment, the effects of dim light are investigated. FIGURE 4(c) shows the hand segmentation results with varying threshold values indoor environment under dim light. Nonetheless, few patches in the retrieved contours are observed due to dim light effects. The results illustrate that the erosion and dilation effects at higher and lower values of the threshold are extremely dim or blurred. Also, with moderate threshold values in the range of 20 to 30, hand segmentation results are satisfactory.

Scenario 2: In the second parameter tuning experiment, we conducted experiments on the constructed hand gesture dataset and the gesture classification accuracy was used as a performance evaluation metric. Three parameters (learning rate, number epochs and optimizer) were investigated together for the performance evaluation of CNN based gesture classification. Based on the results of experimentation optimal parameter setting of the CNN model is optimizer RMS Prop, number of epochs 15 and learning rate 0.0001.

Scenario 3: In this experiment, one person hand gesture dataset consisting of 21,000 images is developed which is divided into training and testing datasets. For each gesture the collected training dataset consists of 2000 images, thus for 10 gestures a total of 20,000 images were included in the training dataset. In the testing 100 images are collected for every gesture class thus the total testing dataset is comprised of 1000 images. The gesture classification result of the developed CNN method (with optimal parameters) using single-user test images in the form of confusion matrix are given in TABLE II.

TABLE II
CONFUSION MATRIX

	I	II	III	IV	V	VI	VII	VIII	IX	X
I	98	0	0	0	0	2	2	0	3	4
II	0	95	2	0	0	0	0	0	0	0
III	0	5	98	1	0	0	0	0	0	0
IV	0	0	0	92	0	0	2	2	0	0
V	0	0	0	0	97	5	0	0	0	6
VI	0	0	0	0	3	90	0	0	2	2
VII	2	0	0	0	0	0	95	0	0	0
VIII	0	0	0	7	0	0	1	8	0	0
IX	0	0	0	0	0	0	0	0	95	0
X	0	0	0	0	0	3	0	0	0	88

Scenario 4: Moreover, unbiased evaluation of the developed CNN model is performed by another constructed test dataset which contains 10 gestures of five different persons. Every gesture class contains 100 images for each person. Therefore, there are 5,000 (5x10x100) images in this multiple-users dataset. The snapshots of the hand gestures and collected dataset for user 1 to 5 are illustrated in FIGURE 5.

FIGURE 5. MULTIPLE-USERS HAND GESTURE DATASET

The performance of the developed hand gesture recognition algorithm is evaluated on this challenging multiple-user hand gesture test dataset which has never been used before for training. The classification results are summarized in TABLE III.

(A) DAY WITH ILLUMINATION (B) NIGHT WITH ILLUMINATION (C) DAY WITHOUT ILLUMINATION
FIGURE 4. THE EFFECT OF THRESHOLD VARIATION

TABLE III
CLASSIFICATION ACCURACY (IN % AGE) FOR MULTIPLE-USERS

User / Gesture	1	2	3	4	5	Overall
Swing	98	96	98	98	96	97.2
Palm	95	94	95	95	90	93.8
Stop	98	96	98	98	94	96.8
Shoot	90	89	91	92	90	90.4
Best	97	96	97	97	90	95.4
Fist	90	89	89	90	86	88.8
Allgood	95	92	95	95	94	94.2
Victory	98	96	96	98	98	97.2
Rock	92	90	93	95	90	92.0
Exit	86	86	87	88	84	86.2

The results show promising recognition performance of all the gestures for multiple persons. The developed algorithm has multiple applications such as appliance control, healthcare sector, handicap special personel, and security, etc.

VI. CONCLUSION

In this study, a hand gesture recognition based method for Human-Machine Interface is presented, incorporating machine vision and machine learning. An image capturing device was introduced for data acquisition, subsequently, intelligent computation was performed for hand gesture segmentation and classification, employing threshold and deep learning techniques. Initially, a single-user hand gesture database consisting of 21,000 images was developed which achieved an overall accuracy of 94.6%. Next, hand gesture identification performance of the developed algorithm was assessed for multiple-users. A challenging dataset, consisting of 5000 images, for ten different gestures of five different humans, was developed with varying hand gesture orientation and postures. The classification accuracy of 93.2% was achieved.

Experimentation on an increasing number of a dataset with the inclusion of both static and dynamic features is in-progress.

REFERENCES

[1] P. Bernardis and M. Gentilucci, "Speech and gesture share the same communication system," *Neuropsychologia*, vol. 44, no. 2, pp. 178–190, 2006.

[2] L. Cabrera-quiros, D. M. J. Tax, and H. Hung, "Gestures In-The-Wild : Detecting Conversational Hand Gestures in Crowded Scenes Using a Multimodal Fusion of Bags of Video Trajectories and Body Worn Acceleration," *IEEE Trans. Multimed.*, vol. 22, no. 1, pp. 138–147, 2020.

[3] H. Li, L. Wu, H. Wang, C. Han, W. Quan, and J. Zhao, "Hand Gesture Recognition Enhancement Based on Spatial Fuzzy Matching in Leap Motion," *IEEE Trans. Ind. Informatics*, vol. 16, no. 3, pp. 1885–1894, 2020.

[4] C. Liu, Y. Yang, X. Liu, L. Fang, and W. Kang, "Dynamic-Hand-Gesture Authentication Dataset and Benchmark," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 1550–1562, 2021.

[5] O. Köpüklü and A. Gunduz, "Online Dynamic Hand Gesture Recognition Including Efficiency Analysis," *IEEE Trans. Biometrics, Behav. Identity Sci.*, vol. 2, no. 2, pp. 85–97, 2020.

[6] Q. De Smedt, H. Wannous, J. Vandeborre, and U. M. R. Cristal, "Skeleton-based Dynamic hand gesture recognition," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV.*, 2016, pp. 1206–1214.

[7] J. Suarez and R. R. Murphy, "Hand Gesture Recognition with Depth Images : A Review," in *21st IEEE International Symposium on Robot and Human Interactive Communication*, 2012, pp. 9–13.

[8] Z. Ren, J. Meng, and J. Yuan, "Depth Camera Based Hand Gesture Recognition and its Applications in Human-Computer-Interaction," in *ACM on Multimedia Conference*, 2016, pp. 102–116.

[9] H S. Aziz, Z. Aslam, M. Rizwan, and S. Nawaz, "EARLY Heart Disease Prediction with Minimal Attributes using Machine Learning", Pakistan J Engg & Tech, vol. 3, no. 2, pp. 178-182, Oct. 2020.

[10] J. Rekimoto, "GestureWrist and GesturePad: Unobtrusive wearable interaction devices," in *5th International Symposium on Wearable Computers*, 2001, pp. 21–27.

[11] H. Gjoreski, J. Bizjak, M. Gjoreski, and M. Gams, "Comparing Deep and Classical Machine Learning Methods for Human Activity Recognition using Wrist Accelerometer," in *25th International Joint Conference on Artificial Intelligence*, 2016, pp. 1–7.

[12] K. Rhee and H. Shin, "Electromyogram-based hand gesture recognition robust to various arm postures," *Int. J. Distrib. Sens. Networks*, vol. 14, no. 7, pp. 1–15, 2018.

[13] S. Mehboob, "Sentiment Base Emotions Classification of Celebrity Tweets by Using R Language", Pakistan J Engg & Tech, vol. 3, no. 2, pp. 95-99, Sep. 2020.

[14] N. Neverova, C. Wolf, G. Taylor, and F. Nebout, "ModDrop : adaptive multi-modal gesture recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. December 2014, 2015.

[15] P. Wang, Z. Li, Y. Hou, and W. Li, "Action Recognition Based on Joint Trajectory Maps Using Convolutional Neural Networks," in *ACM on Multimedia Conference*, 2016, pp. 102–106.

[16] S. Jyothi, S. Padmavati, and M. Visvavidyalayam, "A Survey on Threshold Based Segmentation Technique in Image Processing," *Int. J. Innov. Res. Dev.*, vol. 3, no. 12, pp. 234–239, 2016.

[17] F. G. Hofmann, P. Heyer, and G. Homme, "Velocity profile based recognition of dynamic gestures with discrete hidden markov models," in *International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction*, 1998, pp. 81–95.

[18] T. Marasovi and V. Papi, "Accelerometer-Based Gesture Classification Using Principal Component Analysis," in *19th International Conference on Software, Telecommunications and Computer Networks*, 2011, pp. 1–5.

[19] S. K. Leem, G., and F. Khan, "Detecting Mid-Air Gestures for Digit Writing With Radio Sensors and a CNN," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1066–1081, 2020.

[20] S. Hazra and A. Santra, "Short-Range Radar-Based Gesture Recognition System Using 3D CNN With Triplet Loss," *IEEE Access*, vol. 7, pp. 125623–125633, 2019.

[21] S. Tam, M. Boukadoum, A. Campeau-lecours, and B. Gosselin, "A Fully Embedded Adaptive Real-Time Hand Gesture Classifier Leveraging HD-sEMG and Deep Learning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 2, pp. 232–243, 2020.

[22] C. Du, L. ZHANG, X. SUN, J. WANG, and J. SHENG, "Enhanced Multi-Channel Feature Synthesis for Hand Gesture Recognition Based on CNN With a Channel and Spatial Attention Mechanism," *IEEE Access*, vol. 8, pp. 144610–144620, 2020.

[23] G. Strezoski, D. Stojanovski, I. Dimitrovski, and G. Madjarov, "Hand Gesture Recognition using Deep Convolutional Neural Networks Hand," in *ICT Innovations*, 2016, no. September, pp. 1–10.

[24] P. Molchanov, S. Gupta, K. Kim, J. Kautz, and S. Clara, "Hand Gesture Recognition with 3D Convolutional Neural Networks," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 1–7.

[25] M. R. Islam, U. K. Mitu, R. A. Bhuiyan, and J. Shin, "Hand Gesture Feature Extraction Using Deep Convolutional Neural Network for Recognizing American Sign Language," in *4th International Conference on Frontiers of Signal Processing (ICFSP)*, 2018, pp. 115–119.

[26] S. Shin and W. Sung, "Dynamic Hand Gesture Recognition for Wearable Devices with Low Complexity Recurrent Neural Networks," in *International Symposium on Circuits and Systems*, 2016, pp. 2274–2277.