IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 July 2023, accepted 24 July 2023, date of publication 1 August 2023, date of current version 9 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3300712

RESEARCH ARTICLE

Hand Gesture Recognition for Characters Understanding Using Convex Hull Landmarks and Geometric Features

HIRA ANSAR[1], NAIF AL MUDAWI[2], SAUD S. ALOTAIBI[3], ABDULWAHAB ALAZEB[2], BAYAN IBRAHIMM ALABDULLAH[4], MOHAMMED ALONAZI[5], AND JEONGMIN PARK[6]

[1]Faculty of Computer Science and AI, Air University, Islamabad 44000, Pakistan
[2]Department of Computer Science, College of Computer Science and Information System, Najran University, Najran 55461, Saudi Arabia
[3]Information Systems Department, Umm Al-Qura University, Mecca 24382, Saudi Arabia
[4]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[5]Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj 16273, Saudi Arabia
[6]Department of Computer Engineering, Tech University of Korea, Siheung-si, Gyeonggi-do 15073, South Korea

Corresponding author: Jeongmin Park (jmpark@tukorea.ac.kr)

This work was supported by a grant of the Basic Science Research Program through the National Research Foundation (NRF) (2021R1F1A1063634) funded by the Ministry of Science and Information & Communications Technology (MSIT), Republic of Korea. The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work under the Research Group Funding program grant code (NU/RG/SERC/12/40). Also; Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R440), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This study is also supported via funding from Prince Satam bin Abdulaziz University project number (PSAU/2023/R/1444).

ABSTRACT With the latest advancements, hand gesture recognition is becoming an effective way of communication and gaining popularity from a research point of view. Hearing impaired people around the world need assistance, while sign language is only understood by a few people around the globe. It becomes challenging for untrained people to communicate easily, research community has tried to train systems with a variety of models to facilitate communication with hearing impaired people and also human-computer interaction. Researchers have detected gestures with numerous recognition rates; however, the recognition rate still needs improvement. As the images captured via cameras possess multiple issues, the light intensity variation makes it a challenging task to extract gestures from such images, extra information in captured images, such as noise hinders the computation time, and complex backgrounds make the extraction of gestures difficult. A novel approach is proposed in this paper for character detection and recognition. The proposed system is divided into five steps for hand gesture recognition. Firstly, images are pre-processed to reduce noise and intensity is adjusted. The pre-processed images region of interest is detected via directional images. After hand extraction, landmarks are extracted via a convex hull. Each gesture is used to extract geometric features for the proposed hand gesture recognition (HGR) system. The extracted features helped in gesture detection and recognition via the Convolutional Neural Network (CNN) classifier. The proposed approach experimentation result demonstrated over the MNIST dataset achieved a gesture recognition rate of 93.2% and 90.2% with one-third and two-third training validation systems, respectively. Also, the proposed system performance is validated on the ASL dataset, giving accuracy of 91.6% and 88.14% with one-third and two-third training validation systems, respectively. The proposed system is also compared with other conventional systems. Different emerging domains such as human-computer interaction (HCI), human-robot interaction (HRI), and virtual reality (VR) are applicable to the proposed system to fill the communication gap.

INDEX TERMS ASL sign language, character understanding, landmark identification, geometric feature, hand gesture recognition, CNN.

The associate editor coordinating the review of this manuscript and approving it for publication was Parikshit Sahatiya.

I. INTRODUCTION

Hand Gesture Recognition (HGR) is considered the best way of communication for hearing impaired people around the world, and it is also used as a second language [1]. Sign languages make communication easier between the hearing impaired and a normal-hearing person. Everyone cannot understand sign language, causing a communication gap between normal and hearing individuals [2]. The use of hand gestures helps people express their emotions in a variety of ways, as well as comprehend characters better and relieve tension and anxiety. Since we cannot communicate verbally with hearing impaired individuals, hand gestures are a great replacement because they allow us to understand what is happening within their thoughts. With the help of merging various fields, such as image processing along with pattern recognition, a system can easily understand these gestures made by special individuals.

HGR is primarily built utilizing a variety of techniques including sensors-based and vision-based [3]. In the sensor based approach, the sensors are built with the aid of arrays. They record the position, acceleration, and speed of the hand. These motion properties are then used to train and evaluate the recognition of hand gestures. Even though it has a high sensitivity, getting better accuracy is difficult. Since the sensor has to be well-configured and of high calibre to get better findings. The sensor system is pricey and has problems with longevity. As technology progresses, new sensors are often released on the market. While vision-based HGR, reduces the limitations of the sensors-based systems [4]. Using cameras, RGB pictures are gathered. Less money is required to buy the cameras, and a suitable arrangement is simple to create. For hand gesture identification, the RGB image's color, shape, orientation, contours, and locations are determined. The depth images add additional dimensions than RGB [5]. Either automated or empirical thresholding methods are available for depth. Empirical contains the ''error and trail'' approach, which excludes the search space and prioritizes manual localization above computing expense.

The vision-based systems provide challenges for researchers, the images contain variation in the light intensity, clutter sensitivity, and skin tone vary [6]. The localization of the hand is a crucial phase. The conventional procedures are broken down into several steps in order to increase accuracy while taking the challenges into consideration. From the acquired data, hand is detected. Few of the techniques are employed for hand detection such as, segmentation, and color-based. After that features are extracted and classified via algorithms [7]. HGR has worldwide practical applications such as sign language recognition, HCI, HRI, and VR [8]. Also, in hospitals, HGR is used to comprehend the symptoms of a hearing impaired patient through vision-based systems applied to Android phones. While some used gloves, wearables, or wires as a part of their systems [9], [10], [11], [12].

In this paper, a vision-based technique is applied and a novel method is used to extract hand gestures landmarks and features with all the constraints in consideration. The methodology of the proposed HGR undergoes the following series of steps. In the first step, preprocessing is used to detect hand position from the low-resolution grayscale images. In the second step, extreme landmark points are retrieved using a convex hull after the removal of convexity issues. In the third step, geometric features and Euclidean distance are applied to extract optimal feature values. Finally, character understanding through the CNN, is used over the MNIST dataset having 24 alphabets. Our approach detected all characters with compelling recognition accuracy over other state-of-the-art methods.

The paper's major contributions are summarized as follows:

- An efficient way to extract hands from a background in a low-resolution grayscale image is proposed.
- Geometric features are extracted for character understanding on hand.
- The proposed pre-processing model can separate the hand from the background in case of variation in light intensity.
- The proposed system is a generalized method for the detection of all the characters.
- The HGR system detected the maximum extreme points of all characters via the convex vertices points' mechanism.

The rest of the paper is divided in following sections: Section III presents the literature review. Section IV presents the methodology of the proposed system consisting of pre-processing techniques, hand detection, landmark detection, feature extraction and CNN classifier. Section V presents the experimental results, dataset description and performance evaluation after comparing them with other state-of-the-art methods. Section VI presents the discussion over algorithms merits and demerits. Finally, Section VI includes the conclusion of the paper.

II. LITERATURE REVIEW

This section presents the work exist in hand gesture recognition domain. Some of the researchers have detected hand gesture with the variety of techniques. Such as In [13], Stergiopoulou and Papamarkos proposed the SGONG network, it creates a mesh to detect the shape formed by neural gas. The system is color-based, it detects the color of the hand area However, there is a 10% error chance in detection as the distance between fingers differs in every gesture leading to false recognition. In [14], Zhao et al. discussed Histograms of oriented gradients (HOG) to define the hand images and discrimination via PCA-LDA. Although they detected the hand accurately, the shadow of the hand could not be tackled properly. In [15], Jani et al. developed a sensing glove hardware that recognized different hand gestures. In [16], Aly et al. worked on image gesture recognition using depth images and PCA net features. The litrature review is divided in two section hand gesture for character recognition and hand gesture for word generation as follows:

A. HAND GESTURE FOR CHARACTER GENERATION

In [17] Jalal et al. presented a new approach to recognize characters. The purpose of the research is to communicate with hearing impaired people. They used pattern recognition method for hand gesture recognition. Model was trained with multiple layers of CNN, and feature set was extracted for character recognition. The system attained an accuracy of 91.7%. Kale and Waghmare [18] also used the ASL dataset for character recognition with hearing impaired people. They used spatial feature identification and for classification convolution neural network was used. Xu [19] designed a system to recognize hand gestures with higher accuracy. In this paper, the Kalman filter is used to get the exact position of the hand. Then the gesture is passed through CNN for feature extraction and recognition. Abdulwahab and Raheem [20] presented a way to detect hand gestures, they first resized the image then it was converted it into a grayscale image. The pre-processed image was then passed through edge detection for feature collection. For recognition, they used deep neural learning methods for accuracy and error calculation. In [21] the proposed system the fine-tuned VGG19 model used the RGB and RGB-D static gesture recognition method. In this VGG19 model both RGB and RGB-D features layer concatenate to enhance the recognition accuracy of neural network. By using ASL dataset this model achieved the 94.8% accuracy with compared to other traditional and CNN based model. In this [22] model recognize the gesture without any preprocessing, filtration and image enhancement by using 2 light weight methods based on YOLO (You Only Look Once) v3 and DarkNet-53 convolutional neural network. The best part of this model is recognizing the gesture in complex background environment and in low light effect or resolution. The better result achieved from proposed YOLOv3 based model with accuracy of 97.68% and our model show real time detection both for static and dynamic gestures. In [23] proposed system start as a learning tool in hand gesture detection of sign language and model based on skin- color modelling just like explicit skin-color space thresholding. In explicit skin-color range is defined to extract the hand from background image and finally for the classification the convolutional neural network model used. In training of images Keras used which achieved accuracy of 93.67%. ASL alphabets achieved accuracy of 90.04%, Number recognition achieved 93.44% accuracy and static word recognition accuracy is 97.52%. In [24] a deep learning CNN model used to recognize the hand gestures based sign language, this model specialty is to achieve the better accuracy with few number of parameters. VGF-11 and VGG-16 used for training and testing, two datasets used in which one has large collection of Indian sign language gestures and the other one has American Sign language. The model achieved accuracy of 99.96% and 100% with the comparison of state-of-the-art other models. In [25] the proposed system recognize the real time hand gesture by inspiring the CNN performance by end-to-end fine-tuning of pre-trained CNN with fusion technique of score level with low number of gesture images. The proposed technique evaluated by leave-one-subject-out cross-validation regular CV test of two benchmark dataset MU and HUST-ASL.

B. HAND GESTURE FOR WORD GENERATION

In [26] Ansar et al. presented a way to recognize words first, the images were pre-processed then the hand was extracted through SSD-CNN. After that skeleton is detected and features are extracted. For classification KATH-Hashing algorithm was used. Shin et al. [27] used the media-pipe hands' algorithm for hand skeleton estimation. The features were then passed through SVM and a light gradient boosting machine (GBM). In [28] Vijaykumar et al. proposed a method in which gesture images were preprocessed and features were extracted via HOG. The extracted features were passed through CNN for recognition. In [29] the proposed recognition system enable recognition by artificial intelligence using sensing gloves. The sensing gloves pass from deep learning model with segmentation or without segmentation, in both techniques the recognition of words approach to 50 words and 20 sentences. The model treats the sentence signals as the words units and never seen sentences also recombine with new order combination with average recognition accuracy rate of 86.67%. In [30] the proposed system used deep learning methods apply on three datasets, NYU, First-Person and RKS-PERSIANSIGN to automatic recognize the sign words from input video streams. In CNN based model 2D input frames consider as a 3D hand key points, a midpoint algorithm used to construct the skeleton from these estimated key points and to show the more hands representation the 3D hand skeleton project into three surface images views. In detected key points apply the heatmap for stack fashion refinement, the 3DCNNs apply on stacked features of hand to extract the local spatio-temporal features and these features fed into LSTM model to recognize the dynamic hand gesture signs. In our model 3DCNN best recognize the dynamic hand gesture as compared to 2DCNN and outperform the state-of-the-art models in sign language recognition. In [31] proposed system two methodology used first in which used RGB+D data capture from real time scenario using Kinect Camera and segmented the data by using 3D reconstruction, affine transformation of depth map and RGB. Convolutional neural network used to segment the Indian sign language by achieved accuracy of 98.81% and showed good performance on ASL by giving accuracy of 97.71%. In combination of LSTM with convolution kernel achieved accuracy of 99.08%. This model not capable to maintain the bulk amount of data wit system computation effect the communication with people. In [32] the proposed deep transfer learning model reduce the number of observations from average of 10 to only 3 to predict the sign from different transferring schemes, vary sizes of vocabulary and labelled sentences. In this model data is collected from six degree of freedom inertial sensor wear on signer both hands and take into convolutional neural network, two bidirectional long short-term memory layers and connectionist temporal classification to recognize the sign without sign boundaries. In [33] real time hand gesture recognition proposed system capture RGB+D data by using computer vision techniques of 3D construction and

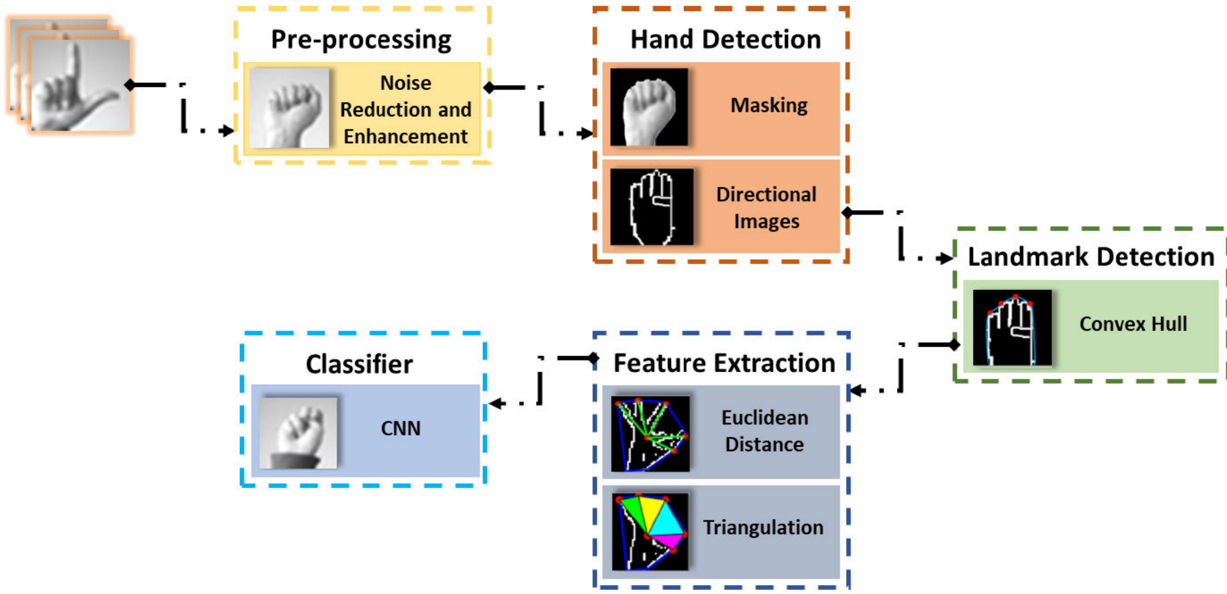


FIGURE 1. The proposed HGR systems architecture.

affine transformation. In one-to-one mapping of pixels, the date moves towards segmentation in which to remove the background noise segmentation of hand gesture done. At last, the Convolutional neural network trained on Indian sign language of numbers and alphabets and achieved accuracy of 98.81% on 45,000 RGB+D images, further convolutional LSTM trained on 10 ISL dynamic words to achieved accuracy of 99.08% on 1080 trained videos.

III. PROPOSED MODEL

In this section, proposed system methodology is explained in detail. The proposed HGR is divided into five phases, which makes it an efficient system to detect hand gesture. Firstly, a low-resolution grayscale image is pre-processed to adjust its size, then noise is reduced to remove extra information, after that image light intensity is adjusted by sharpening and enhancing image using filters, the image is then converted into binary and morphological operator are applied over it like dilation and erosion. Secondly, the directional region of hand gesture in the image is detected. Thirdly, extreme point localization is examined over the detected hand via a convex hull after convexity removal, resulting in landmark extraction. Fourthly, distance and geometric features are identified over extreme points on hand gestures. Lastly, landmark-based features are used to classify different hand gestures using CNN. The overview of proposed HGR system is represented in Fig. 1.

A. PRE-PROCESSING

In this section, hand gesture images are pre-processed to extract hand from low-resolution noisey grey-scale images, efficiently [34]. For the proposed HGR system two datasets MINIST and ASL are considered. The MINIST dataset consists of a 28∗28 resolution size of the grayscale image, while ASL consist of 200∗200 resolution. The ASL dataset are first converted to grey scale image. The images contains unsolicited information. In MINIST dataset Gaussian noise is removed. Gaussian noise is the random value added to the original pixel value [35]. Gaussian noise is represented as:

$$H(g) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(g-\mu)^2/2\sigma^2} \tag{1}$$

where g represents the gray level of the image, μ the mean grey value and σ is the standard deviation.

For denoising, the weiner filter approach is applied that restore the maximum true values of the pixel. It is an adaptive filter for the removal of noise in signals which provides the nearby output value of the original image. Let the input original image is presented as $i(n)$ and the desired output denoised image as $di(n)$. The weight wv and of the input vectors are defined as follows:

$$wv = [wv_0, wv_1, \ldots., wv_{Y-1}]^T \tag{2}$$

$$i(n) = [i(n), i(n-1), \ldots., i(n-N+1)]^T \tag{3}$$

however the output weiner filter is presented as follows:

$$w = M^{-1}c \tag{4}$$

where M^{-1} autocorrelation matrix and c is the cross-correlation matrix. The autocorrelation matrix and cross-correlation matrix calculated using mean square error. The variance and error in the mean square root are minimized to a denoise image which resulted in soothing output [36]. The mean square error is defined as:

$$\varphi = E\left[e^2(n)\right] = E\left\{\left[di(n) - wv^T i(n)\right]\left[di(n) - i^T(n)\, wv\right]\right\} \tag{5}$$

where $e^2(n)$ is error in the denoised image and the output image, φ is the mean square error in denoised image and the input image. The mean square error is minimized to finally apply w on $i(n)$ to get the final noise free image.

After noise reduction the denoised image is passed through gamma correction [37] to adjust the image intensity for better extraction of hand gesture from image. Gamma correction consist of non-linear behaviour, the power law function is defined as:

$$gamma_i = Ck_y^{\gamma} \tag{6}$$

where k_y is the non-negative image raised to γ and multiplied with constant value C. $gamma_i$ presents the intensity adjustment lies between 0 and 1. Fig. 2 presents the preprocessed images.

FIGURE 2. Denoised and intensity adjusted characters of MINIST dataset: (a) D, (b).

B. HAND DETECTION

After the pre-processing method, hand region is extracted for further model process. The hand is detected using two way method. First the image is converted into binary using otsu's threshold method [38]. Then the binary image is passed through morphological operation such as dilation and erosion. In the case of dilation, the area of the hand region is examined which includes the sphere structuring element to get the desired output [39]. Let's assume, there is an image I that include structure element S in set form represented as:

$$I \oplus S = \left\{ h \mid (S)_h \cap I \neq \emptyset \right\} \tag{7}$$

The structuring element S is shifted by h and image I must have at least one common element with S. The dilation process enlarged the pixel value by 1 in the hand region and shrink the background having 0-pixel value.

While erosion is a complement of dilation, it reduced the size of the hand region. The image I with structuring element S is defined as [40]:

$$I \ominus S = \{ h \mid (S)_h \subseteq I \} \tag{8}$$

The structuring element S is translated by h which results in the loss of boundary pixel of the hand region and the background is enlarged. After the erosion and dilation process connected component is applied to get the exact boundary of hand and then the hand is masked over the binary image. Fig. 3 presents the F character step by step masking process.

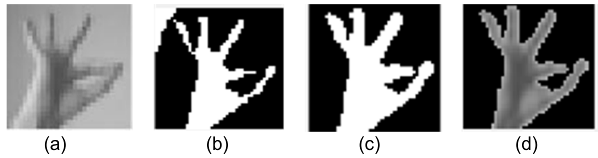


FIGURE 3. Hand detection via masking. (a) Grayscale image, (b) binarization, (c) connected component and morphological operations, and (d) masked image.

The second method used to detect the hand from gray scale image includes the detection of edges i.e. the outer and inner edge of the hand. Directional images are set of oriented segments. For the directional region of image, 3∗3 gradient vector-matrix $G_{(i,j)}$ is calculated [41]. For every pixel of an image I the gradient vector matrix is computed which is formulated as in (9), shown at the bottom of the page.

In the 3∗3 gradient vector matrix, pixel value in the second row and second column $I_{(2,2)}$ is compared with the distance calculated from the adjacent pixel distances. The adjacent pixel distance is calculated by subtracting the pixel values like pixel value $I_{(1,1)}$ is subtracted from $I_{(3,3)}$, $I_{(1,3)}$ is subtracted from $I_{(3,2)}$, $I_{(1,3)}$ is subtracted from $I_{(3,1)}$, $I_{(2,1)}$ is subtracted from $I_{(2,3)}$ and vice versa.

For every pixel value, distances dl is calculated after subtraction, the negative value will be converted into positive.

$$dl = dl^{*}(-1) \tag{10}$$

The distance values are compared via gradient vector with constant threshold. Thus, if the distance is greater than the threshold, pixel value 1 is assigned. While, if the distance is less than the threshold, it becomes a black pixel value of 0. The binary image is formed with outer and internal boundaries of the hand causing directional regions of hands. The resultant directional image is shown in Fig. 4.

FIGURE 4. Directional images of different characters over MINIST dataset.

Both of the methods were tested on ASL dataset as well. The directional images gave more promising result over masked images. To check the accuracy of hand detection from both methods, first the ground truth is computed. Then via geodesic distance both methods contour index pixel value are calculated. Table 1 shows the comparison of both methods

$$G_{(i,j)} = \begin{bmatrix} I_{(i,j)} - I_{(i,j-1)}, I_{(i,j)} - I_{(i,j+1)} & I_{(i,j)} - I_{(i-1,j)}, I_{(i,j)} - I_{(i+1,j)} \\ I_{(i,j)} - I_{(i-1,j-1)}, I_{(i,j)} - I_{(i+1,j+1)} & I_{(i,j)} - I_{(i-1,j+1)}, I_{(i,j)} - I_{(i+1,j-1)} \end{bmatrix} \tag{9}$$

over MINIST dataset. The comparison table shows the results got from directional images are better and these images are further selected for the next process.

TABLE 1. Comparison of detection accuracies over MINIST dataset.

Characters	Masked Images %	Directional Images %	Characters	Masked Images %	Directional Images %
A	69	91	N	69	91
B	68	92	O	72	91
C	70	92	P	76	92
D	69	90	Q	71	90
E	72	89	R	68	89
F	71	91	S	65	90
G	73	91	T	72	90
H	69	92	U	71	92
I	70	88	V	71	91
K	72	92	W	72	93
L	71	91	X	68	91
M	74	93	Y	70	91

C. LANDMARKS IDENTIFICATION

After attaining the directional regions of hands, the convex hull contour is drawn around the hand region. The convex hull contains all the set points in Euclidean space. It covers all the set points of the contour line forming an envelope [42]. Also, it includes minimum set points of the object and the hull is drawn by maintaining the property of convexity. The linear hull of the convexity is defined as:

$$lin\,(H) := \{x = \Sigma \lambda_i y_i \wedge \forall i : y_i \in H, \lambda_i \in \mathbb{R}\} \tag{11}$$

where $\boldsymbol{lin}\,(\boldsymbol{H})$ is the linear hull of the convexity having a set of linear combinations of $\boldsymbol{H}$ in the smallest linear subspace. If $\boldsymbol{H} = \{y\} \subset$, then $\boldsymbol{lin}\,(\boldsymbol{H})$ is the line through $\boldsymbol{y}$ and the center.

The landmarks are drawn using a convex hull and are directed to extra convex vertices points in the contour line. To set the convexity, the center of the hand is measured by dividing the hand region into equal planes, and a centroid is drawn on hand palms. The extra set points have been removed by maintaining the angle and distance between set points. Points are drawn from the start of convexity till the end of the contour line [43], [44]. Fig. 5 shows landmarks identification at top of the hand's figures.

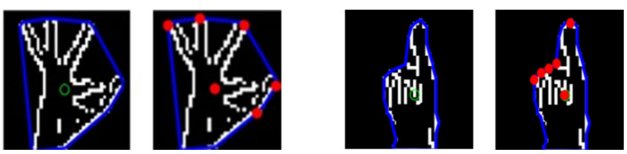


FIGURE 5. Landmarks identification using convex hull boundary line and key points on vertices of MINIST dataset classes i.e. (a) F (b) R.

The extracted landmarks are validated to check the error in the computation of landmark detection. The accuracy of landmarks are computed for all characters in MINIST dataset. For this the ground truth are first computed for all characters and

Algorithm 1 Landmark Identification

Input: segmented hand
Output: *LandmarkPoints(lp1,lp2,lp3,lp4,lp5)*
handcontour ← GetHandContour(hand)
Convex hull ← DrawConvexhull(handcontour)
%Extracting center point of hand %
Center ← GetContourcenter(contour)
LandmarkPoints.append(Center)
%detecting 5 extreme key points%
for *handkeypoint* on convexhull:
  if *handkeypoint<6* in contour:
    LandmarkPoints.append(point)
return *LandmarkPoints (lp1,lp2,lp3,lp4,lp5)*

then the euclidean distance is computed from both locations vector matrix of the characters. Table 2 shows the accuracy of landmark detected on MINIST dataset characters.

TABLE 2. Landmark detection accuracy over MINIST dataset.

Characters	Accuracy %	Characters	Accuracy %
A	99	N	99
B	99	O	100
C	100	P	99
D	100	Q	100
E	99	R	100
F	99	S	100
G	100	T	99
H	99	U	99
I	99	V	100
K	100	W	100
L	100	X	100
M	99	Y	100

D. FEATURE EXTRACTION

In this section valuable features are extracted such as euclidean distance and geometric features. Each directional region of an image is inspected with five extreme points on the fingertips and a center point of the hand. Features are extracted via two different methods i.e. the euclidean distance is calculated from the centroid to extreme points, and the cosine of angles is calculated between the sides of the four triangle formed by five extreme points and centroid.

The extreme landmark i_{xy} and center point c_{xy} of hand is used to measure the euclidean distance feature d of the directional image [45] is formulated as:

$$\left\| \sqrt{(xi_2 - yc_2)^2 + (xi_1 - yc_1)^2} \right\| \tag{12}$$

where d represents the euclidean distance between two points, xi_1 and $xi_2$are x-coordinates, while yc_1 and yc_2 are the y-coordinates of the first extreme point and center point of the hand. Fig. 6 shows the euclidean distance from the centroid to extreme points of different characters.

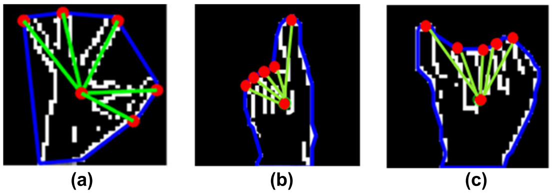


FIGURE 6. Features extraction via Euclidian distance between the extreme points and center point of hand on character classes (a) F (b) R (c) I.

The cosine of angles $(i.e., \alpha, \beta, \gamma)$ are extracted through geometric features that measure the directional angle of two extreme points, adjacent side and centroid forming a triangle as shown in Fig. 7.

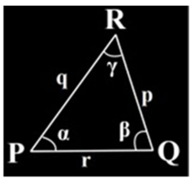


FIGURE 7. Angle references of triangle formulation.

Also, particular triangle have vertices i.e. P,Q and R. p, q and r are the sides of triangles shown in Fig. 8 respectively.

$$\alpha = cos^{-1}q^2 + r^2 - p^2/2qr \tag{13}$$
$$\beta = cos^{-1}p^2 + r^2 - q^2/2pr \tag{14}$$
$$\gamma = cos^{-1}p^2 + q^2 - r^2/2pq \tag{15}$$

where as α, β, γ are angle of the triangle, calculated between two adjacent sides $q \leftrightarrow r, p \leftrightarrow r$, and $p \leftrightarrow q$ respectively. These features are passed through classifier for character recognition.

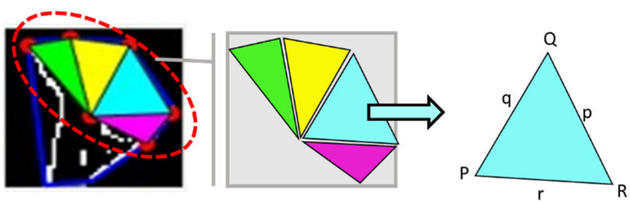


FIGURE 8. Features extraction via triangulation over MINIST dataset F character with angle references.

E. CLASSIFICATION VIA CNN

The extracted features are passed through CNN for character recognition on both datasets. As compared to other conventional classifiers of deep learning, CNN presents promising results for hand gesture recognition [46], [47], [48], [49], [50], [51]. Fig. 9 presents the overall architecture of CNN.

The CNN is composed of three convolution layers, three pooling layers and one fully connected layer. All of the layers possess its specific function to recognize the hand gesture. First layer is the convolution layer CL_1, and also it contains input matrix. The layer is set with 32 kernels containing size of 1 × 13. The input matrix contains 24,000 image sample for training with 10, 488 feature sets. After passing through first layer the matrix of 24, 000 × 10, 488 × 32 is formed. The convolution layer is defined as:

Algorithm 2 Hand Gestures Feature Extraction

Input: HKP: Extreme hand key points
Output: combined *FeatureVector* (*f1,f2,f3...fn*)
FeatureVector←[]
For HKP in [x,y] **do**
Euclidean_distance, ←Extract_ Euclidean_distance, (Hand_Feature)
Triangulation ←Extract_ Triangulation (Hand_Feature)
feature_vectors←GetfeatureVectors(Euclidean_distance, Triangulation)
feature_vectors.append(feature_vectors)
end for
feature_vectors←Normalize(feature_vectors)
return *FeatureVector* (*f1,f2,f3...fn*)
LandmarkPoints.append(point)
return *LandmarkPoints (lp1,lp2,lp3,lp4,lp5)*

$$CL_y^{x-1}(j,k) = ReLU(a) \tag{16}$$
$$ReLU\ (a) = \sum\nolimits_{q=1}^{p} \Omega \left(j\left(k - q + \frac{p+1}{2}\right)\right) M_y^x\ (q) + \alpha_y^x \tag{17}$$

where $CL_y^{x-1}(j,k)$ is produced at (j,k) coordinates of convolution layer $x-1$ with the y^{th} map. Size of the kernel is represented with a and Ω is previous map layer. M_y^x represents x^{th} convolution kernel over y^{th} layer. α_y^x presents the x^{th} bias on y^{th} layer.

The first layer results are passed to the first pooling layer after passing through ReLU. The ReLU activation function possess sum of weights and previous layer bias. The sliding window of size 1 × 2 is used in pooling layer to downsample the results obtained from previous convolution layer. Hence the pooling layer result is obtained using:

$$pool_y^{x-1}\ (a,b) = \max\left(CL_y^x\ (a\ ((b-1) \times (u,v)))\right) \tag{18}$$

where $pool_y^{x-1}\ (a,b)$ is pooling results of the $(u-1)^{th}$ layer on (a,b) coordinates, v is the kernel. $1 \leq u \leq v$ presents the window size. The results obtained using pooling layer is passed through CL_2, with 64 kernels. Then again the second convolution layer is passed through second pooling layer and the procedure is repeated third time with 128 kernels. At last, the fully connected layer gives result as follows:

$$F_r^{x+1} = \text{ReLU}\left(\sum_n a_t^x M_{xr}^x + \alpha_r^x\right) \tag{19}$$

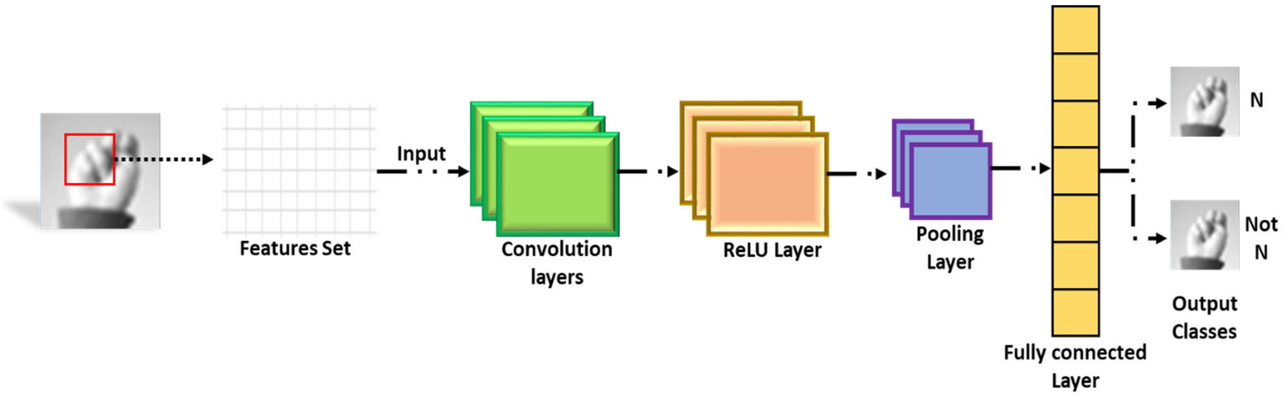


FIGURE 9. Overall structure of CNN used for proposed HGR.

where M^{x}_{xr} shows the matrix containing weights with n^{th} node on $(x+1)$ layer. a^{x}_{t} represents content of the layer. After fully connected layer the gesture is classified with the character recognized.

IV. EXPERIMENTAL SETTINGS AND ANALYSIS

This section consists of dataset description, recognition rate, performance evaluation, and comparison of experimental results with the other state-of-the-art algorithms and methods:

A. DATASET DESCRIPTION

1) MINIST DATASET

MNIST dataset [52] contains twenty-four characters from A to Y, excluding J and Z. These two alphabets need a motion to be expressed. The overall dataset is made in a static mode which causes the expulsion of these two alphabets. The total dataset consists of 24000 images of which 1000 grayscale images per alphabet have 28∗28 resolution sizes. Hand extraction from the background was challenging due to differences in light intensity in every image of the dataset. Fig. 10 shows a few examples of different alphabets from the dataset.

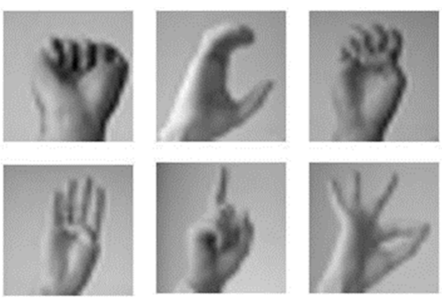


FIGURE 10. Examples of six different alphabets of the MNIST dataset.

2) ASL DATASET

ASL dataset [53] consist of 29 classes containing 26 characters, space, delete and nothing. The dataset contains 87,000 images with 200∗200 pixel resolution. The images are captured in different scenarios. For proposed system evaluation 24 classes are selected A-Z excluding J and Z as they need motion. Fig 11 presents dataset images from the ASL dataset.

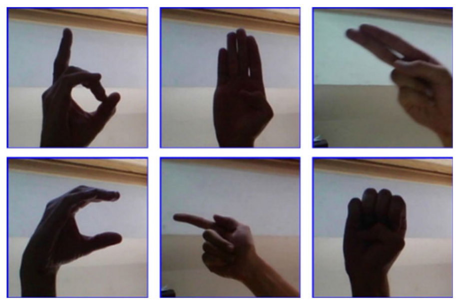


FIGURE 11. Examples of six different alphabets of the ASL dataset.

B. PERFORMANCE EVALUATION

The performance of the proposed solution is tested in MATLAB 2017a and Jupyter Notebook using core i7 (2.9 GH) with a RAM of 16 GB. Training and validation are done on the MNIST and ASL datasets. For system performance evaluation, metrics are calculated over both datasets. Precision, recall and f1-score are calculated for performance measures analysis.

1) EXPERIMENT: 1. ONE-THIRD TRAINING VALIDATION TEST

For the experimentation, one-third of the samples are used for testing and the rest are used for training. The experiment is performed on all alphabets of the MINIST and ASL dataset. Fig 12 and Fig 13 shows the confusion matrix of the experiment performed on both datasets with their mean accuracies. MINIST dataset depicted 93.2% accuracy and ASL presented 90.2% accuracy respectively.

2) EXPERIMENT: 2. TWO-THIRD TRAINING VALIDATION TEST

For this experimental setup, two-thirds of the dataset is used as training and one-third of the dataset is used as testing.

Fig 14 shows the confusion matrix of the two-third validation test performed on MINIST dataset with the mean accuracy of 91.6% using CNN.However, Fig 15 presents the confusion matrix over ASL dataset with 88.14% accuracy respectively.

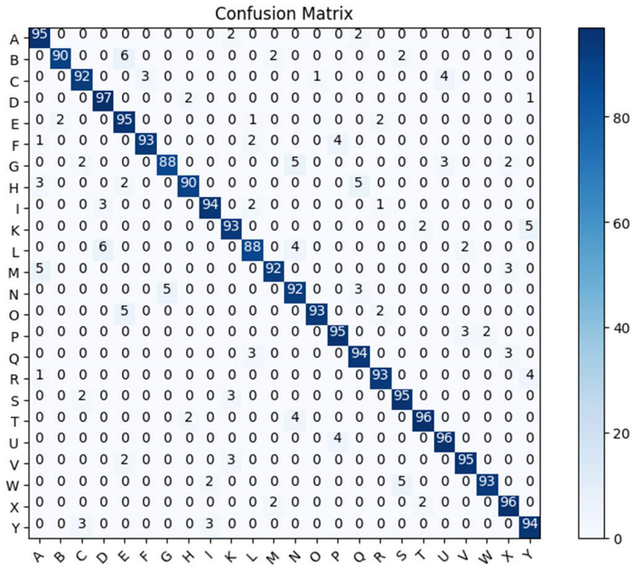


FIGURE 12. **Confusion matrix for one-third training validation test of MNIST dataset.**

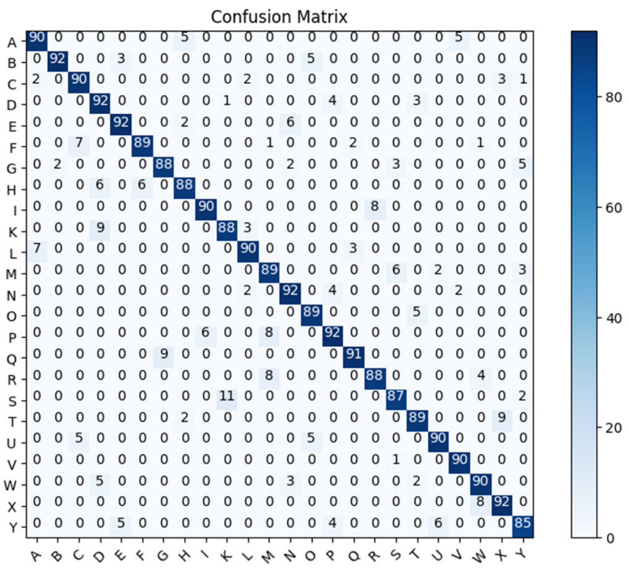


FIGURE 13. **Confusion matrix for one-third training validation test of ASL dataset.**

3) EXPERIMENT: 3. PRECISION, RECALL AND F1-SCORE

The precision, recall, and F1 scores are calculated for twenty-four classes of the MINIST and ASL dataset over one-third training validation test and two-thirds training validation. Table 3 shows performance metrics over one-third training validation test whereas, Table 4 presents evaluation metric on two-thirds training validation respectively. The proposed HGR system can recognize characters with high precision. Eq. (20), (21), and (22) are used to calculate the precision, sensitivity, and F1 scores of all the characters.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{20}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{21}$$

$$F1score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{22}$$

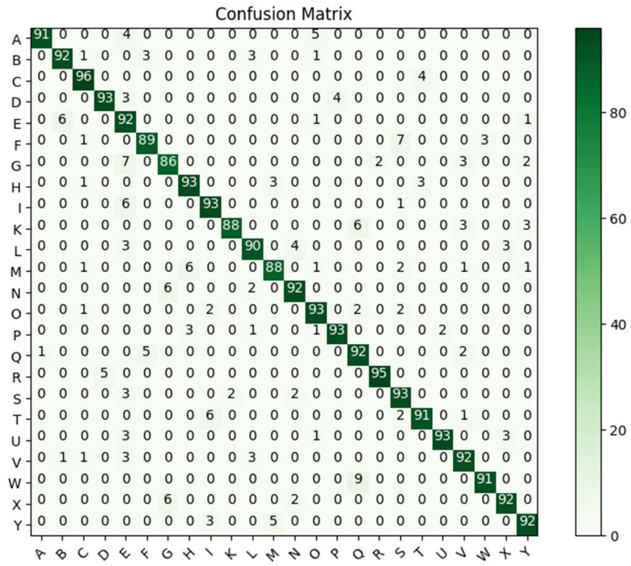


FIGURE 14. **Confusion matrix for two-third training validation test of MNIST dataset.**

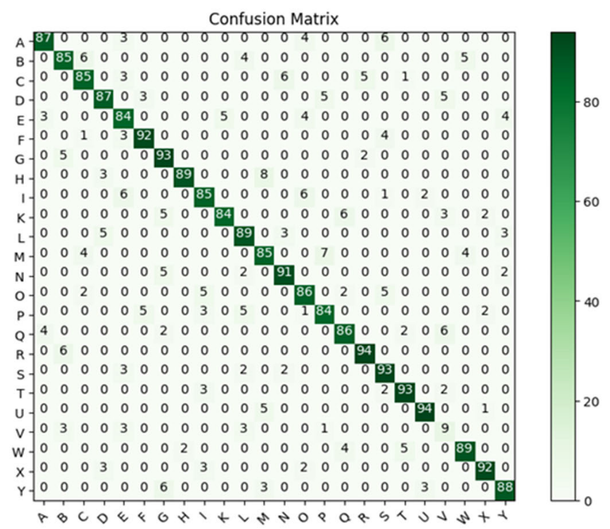


FIGURE 15. **Confusion matrix for two-third training validation test of ASL dataset.**

C. COMPARISON WITH OTHER CONVENTIONAL ALGORITHMS

The proposed system performance is evaluated with the other state of the art algorithms. The system is tested with SVM,

TABLE 3. Performance evaluation over MINIST and ASL dataset with one-third training validation test.

	MINIST Dataset					*ASL Dataset*				
Character Classes	Precision	1-Precision	Recall	1-Recall	F1-Score	Precision	1-Precision	Recall	1-Recall	F1-Score
A	0.95	0.05	0.91	0.10	0.93	0.90	0.10	0.91	0.09	0.90
B	0.90	0.10	0.98	0.02	0.94	0.92	0.08	0.98	0.02	0.95
C	0.92	0.08	0.93	0.07	0.92	0.90	0.10	0.88	0.12	0.89
D	0.97	0.03	0.92	0.09	0.94	0.92	0.08	0.82	0.18	0.87
E	0.95	0.05	0.86	0.14	0.90	0.92	0.08	0.92	0.08	0.92
F	0.93	0.07	0.97	0.03	0.95	0.89	0.11	0.86	0.14	0.88
G	0.88	0.12	0.95	0.05	0.91	0.88	0.12	0.91	0.09	0.89
H	0.90	0.10	0.96	0.04	0.93	0.88	0.12	0.91	0.09	0.89
I	0.94	0.06	0.95	0.05	0.94	0.90	0.10	0.92	0.08	0.91
K	0.93	0.07	0.92	0.08	0.93	0.88	0.12	0.88	0.12	0.88
L	0.88	0.12	0.92	0.08	0.90	0.90	0.10	0.93	0.07	0.91
M	0.92	0.08	0.96	0.04	0.94	0.89	0.11	0.84	0.16	0.86
N	0.92	0.08	0.88	0.12	0.90	0.92	0.08	0.89	0.11	0.91
O	0.93	0.07	0.99	0.01	0.96	0.89	0.11	0.90	0.10	0.89
P	0.95	0.05	0.92	0.08	0.94	0.92	0.08	0.89	0.12	0.90
Q	0.94	0.06	0.90	0.10	0.92	0.91	0.09	0.95	0.05	0.93
R	0.93	0.07	0.95	0.05	0.94	0.88	0.12	0.92	0.08	0.90
S	0.95	0.05	0.93	0.07	0.94	0.87	0.13	0.90	0.10	0.88
T	0.96	0.04	0.96	0.04	0.96	0.89	0.11	0.90	0.10	0.89
U	0.96	0.04	0.93	0.07	0.95	0.90	0.10	0.90	0.10	0.90
V	0.95	0.05	0.95	0.05	0.95	0.91	0.09	0.93	0.07	0.92
W	0.93	0.07	0.98	0.02	0.95	0.92	0.08	0.87	0.13	0.90
X	0.96	0.04	0.91	0.09	0.94	0.92	0.08	0.89	0.12	0.90
Y	0.94	0.06	0.90	0.10	0.92	0.85	0.15	0.89	0.12	0.87
	Mean Precision=0.93 **Mean Recall**=0.93 **Mean F1 Score**=0.93					**Mean Precision**=0.89 **Mean Recall**=0.90 **Mean F1 Score**=0.90				

random forest and multiple perceptron. The accuracies graph is presented in Fig. 16 shows the accuracy attained via CNN gave the maximum achieved accuracy.

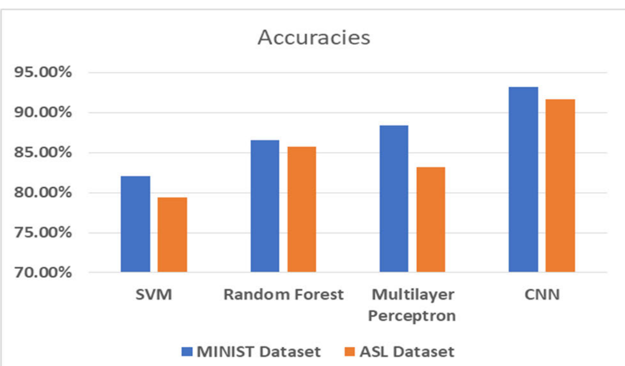


FIGURE 16. Accuracy comparison with other conventional algorithm.

D. COMPARISON WITH OTHER CONVENTIONAL SYSTEMS

This section presents that our proposed method recognizes characters with more accuracy than other state of the art methods. Table 5 discusses the comparison of the proposed method with the other state-of-the-art methods over the MNIST and ASL dataset.

V. DISCUSSION

In the proposed HGR, we have used optimized features and CNN to get higher accuracy. For the proposed system, we have used an adaptive median filter for pre-processing images to remove noise. As noise hinders the processing of any kind of extraction or detection. The system successfully identifies characters. Other conventional systems focus on one type of gesture recognition at a time. The system has a proper architecture as compared with other state-of-the art methods. To get better detection results, we first

TABLE 4. Performance evaluation over MINIST and ASL dataset with two-third training validation test.

	MINIST Dataset					*ASL Dataset*				
Character Classes	Precision	1-Precision	Recall	1-Recall	F1-Score	Precision	1-Precision	Recall	1-Recall	F1-Score
A	0.89	0.11	0.80	0.20	0.84	0.87	0.13	0.93	0.07	0.9
B	0.86	0.14	0.83	0.17	0.84	0.84	0.16	0.89	0.11	0.86
C	0.82	0.18	0.77	0.23	0.79	0.92	0.08	0.8	0.2	0.86
D	0.81	0.19	0.76	0.24	0.78	0.93	0.07	0.92	0.08	0.92
E	0.84	0.16	0.93	0.07	0.88	0.89	0.11	0.84	0.16	0.86
F	0.85	0.15	0.79	0.21	0.81	0.85	0.15	0.98	0.02	0.91
G	0.80	0.20	0.80	0.20	0.80	0.84	0.16	0.86	0.14	0.85
H	0.83	0.17	0.85	0.15	0.83	0.89	0.11	0.94	0.06	0.91
I	0.79	0.21	0.79	0.21	0.79	0.85	0.15	0.85	0.15	0.85
K	0.81	0.19	0.79	0.21	0.79	0.91	0.09	0.84	0.16	0.87
L	0.80	0.20	0.83	0.17	0.81	0.86	0.14	0.89	0.11	0.87
M	0.82	0.18	0.71	0.29	0.76	0.84	0.16	0.87	0.13	0.85
N	0.83	0.17	0.84	0.16	0.83	0.86	0.14	0.9	0.1	0.88
O	0.86	0.14	0.84	0.16	0.84	0.94	0.06	0.93	0.07	0.93
P	0.84	0.16	0.84	0.16	0.84	0.93	0.07	0.82	0.18	0.87
Q	0.79	0.21	0.79	0.21	0.79	0.93	0.07	0.92	0.08	0.92
R	0.79	0.21	0.84	0.16	0.81	0.94	0.06	0.95	0.05	0.94
S	0.81	0.19	0.84	0.16	0.82	0.9	0.1	0.85	0.15	0.87
T	0.79	0.21	0.81	0.19	0.79	0.89	0.11	0.91	0.09	0.9
U	0.81	0.19	0.92	0.08	0.86	0.92	0.08	0.95	0.05	0.93
V	0.82	0.18	0.82	0.18	0.82	0.88	0.12	0.91	0.09	0.89
W	0.80	0.20	0.84	0.16	0.81	0.87	0.13	0.93	0.07	0.9
X	0.83	0.17	0.84	0.16	0.83	0.84	0.16	0.89	0.11	0.86
Y	0.79	0.21	0.81	0.19	0.79	0.92	0.08	0.8	0.2	0.86
	Mean Precision=0.89 **Mean Recall**=0.89 **Mean F1 Score**=0.89					**Mean Precision**=0.88 **Mean Recall**=0.89 **Mean F1 Score**=0.89				

TABLE 5. Comparison of recognition accuracy of the proposed method with other state-of-the-art methods over the mnist dataset.

Methods	Recognition Accuracy (%)	
	MINIST dataset	ASL dataset
CovNet [54]	82.0	--
SVM [55]	--	79.83
LeNet [56]	82.19	--
SVM[57]	--	83.36
CNN[58]	90	--
SVM [59]	--	80.3
Proposed Method	**93.2**	**91.7**

pre-processed the images. Then the interest of the region is extracted from the pre-processed images. Usually, in any conventional system, this is the point where background complexity reduces the detection accuracy results, but our proposed method proved that the detection result for all characters performed well. After the successful accomplishment of detecting hands, we extracted landmarks. In conventional systems, detected hands are passed directly through deep learning methods. Our proposed method introduces landmark extraction through a convex hull. Then the feature set is passed through a deep learning algorithm to make the result more accurate. We have used CNN for our proposed system, and CNN has proven itself to be a better classifier than other deep learning algorithms for our system.

The proposed system has limitations, as described below:

- In images, different types of noises are found, but our system removes the impulse noise efficiently, if noises exist, they can affect our detection.
- When images are not particularly identical, convex hulls are susceptible to inaccuracies. In order to focus feature points on the same area of the image landmark,

techniques that employ affine invariant and noise-insensitive convex hulls with correlation criteria must be developed.
- The results for feature extraction and optimization can be improved more.
- Our system detects a limited number of gestures, but a lot of substantial gestures can help train our model.

VI. CONCLUSION

In this paper, novel approach for hand gesture recognition is proposed by applying five different stages i.e. preprocessing, hand detection, landmark detection, feature extraction, and classification, using geometric features for character recognition. The proposed HGR system targets the domain of HCI, HRI, and VR. The performance of the proposed geometric features approach is evaluated on the MNIST dataset on static gestures, also system performance is tested on ASL dataset for system validation. Experimental results show that our system competes with other state-of-the-art methods.

In the future, we will target more feature extraction techniques for recognition in different domains i.e. healthcare, interactive gaming and home appliances. Also gestures with both hands will be considered for recognition.

REFERENCES

[1] F. A. Farid, N. Hashim, J. Abdullah, M. R. Bhuiyan, W. N. S. M. Isa, J. Uddin, M. A. Haque, and M. N. Husen, "A structured and methodological review on vision-based hand gesture recognition system," *J. Imag.*, vol. 8, no. 6, p. 153, May 2022.

[2] Y. Si, S. Chen, M. Li, S. Li, Y. Pei, and X. Guo, "Flexible strain sensors for wearable hand gesture recognition: From devices to systems," *Adv. Intell. Syst.*, vol. 4, no. 2, Feb. 2022, Art. no. 2100046.

[3] S. Wang, A. Wang, M. Ran, L. Liu, Y. Peng, M. Liu, G. Su, A. Alhudhaif, F. Alenezi, and N. Alnaim, "Hand gesture recognition framework using a lie group based spatio-temporal recurrent network with multiple hand-worn motion sensors," *Inf. Sci.*, vol. 606, pp. 722–741, Aug. 2022.

[4] J. Singha, A. Roy, and R. H. Laskar, "Dynamic hand gesture recognition using vision-based approach for human–computer interaction," *Neural Comput. Appl.*, vol. 29, no. 4, pp. 1129–1141, 2018.

[5] T. R. Gadekallu, G. Srivastava, M. Liyanage, C. L. Chowdhary, S. Koppu, and P. K. R. Maddikunta, "Hand gesture recognition based on a Harris hawks optimized convolution neural network," *Comput. Electr. Eng.*, vol. 100, May 2022, Art. no. 107836.

[6] R. Jain, R. K. Karsh, and A. A. Barbhuiya, "Literature review of vision-based dynamic gesture recognition using deep learning techniques," *Concurrency Comput., Pract. Exper.*, vol. 34, no. 22, Oct. 2022.

[7] M. M. Damaneh, F. Mohanna, and P. Jafari, "Static hand gesture recognition in sign language based on convolutional neural network with feature extraction method using ORB descriptor and Gabor filter," *Exp. Syst. Appl.*, vol. 211, Jan. 2023, Art. no. 118559.

[8] J. Qi, G. Jiang, G. Li, Y. Sun, and B. Tao, "Surface EMG hand gesture recognition system based on PCA and GRNN," *Neural Comput. Appl.*, vol. 32, no. 10, pp. 6343–6351, May 2020.

[9] M. H. Syu, Y. J. Guan, W. C. Lo, and Y. K. Fuh, "Biomimetic and porous nanofiber-based hybrid sensor for multifunctional pressure sensing and human gesture identification via deep learning method," *Nano Energy*, vol. 76, Oct. 2020, Art. no. 105029.

[10] L. Chen, J. Fu, Y. Wu, H. Li, and B. Zheng, "Hand gesture recognition using compact CNN via surface electromyography signals," *Sensors*, vol. 20, no. 3, p. 672, Jan. 2020.

[11] Y. Sun, Y. Weng, B. Luo, G. Li, B. Tao, D. Jiang, and D. Chen, "Gesture recognition algorithm based on multi-scale feature fusion in RGB-D images," *IET Image Process.*, vol. 17, no. 4, pp. 1280–1290, Mar. 2023.

[12] S. Kamal and A. Jalal, "A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors," *Arabian J. Sci. Eng.*, vol. 41, no. 3, pp. 1043–1051, Mar. 2016.

[13] E. Stergiopoulou and N. Papamarkos, "Hand gesture recognition using a neural network shape fitting technique," *Eng. Appl. Artif. Intell.*, vol. 22, no. 8, pp. 1141–1158, Dec. 2009.

[14] Y. Zhao, W. Wang, and Y. Wang, "A real-time hand gesture recognition method," in *Proc. Int. Conf. Electron., Commun. Control (ICECC)*, Sep. 2011, pp. 2475–2478.

[15] A. B. Jani, N. A. Kotak, and A. K. Roy, "Sensor based hand gesture recognition system for English alphabets used in sign language of deaf-mute people," in *Proc. IEEE SENSORS*, Oct. 2018, pp. 1–4.

[16] W. Aly, S. Aly, and S. Almotairi, "User-independent American sign language alphabet recognition based on depth image and PCANet features," *IEEE Access*, vol. 7, pp. 123138–123150, 2019.

[17] A. Jalal, Y.-H. Kim, Y.-J. Kim, S. Kamal, and D. Kim, "Robust human activity recognition from depth video using spatiotemporal multi-fused features," *Pattern Recognit.*, vol. 61, pp. 295–308, Jan. 2017.

[18] K. S. Kale and M. B. Waghmare, "Hand gesture alphabet recognition for American sign language using deep learning," *Int. J. Sci. Res. Sci., Eng. Technol.*, vol. 8, no. 5, pp. 213–220, Sep. 2021.

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

[20] A. Abdulhussein and F. Raheem, "Hand gesture recognition of static letters American sign language (ASL) using deep learning," *Eng. Technol. J.*, vol. 38, no. 6, pp. 926–937, Jun. 2021.

[21] M. Khari, A. K. Garg, R. G. Crespo, and E. Verdú, "Gesture recognition of RGB and RGB-D static images using convolutional neural networks," *Int. J. Interact. Multim. Artif. Intell.*, vol. 5, no. 7, pp. 22–27, 2019.

[22] A. Mujahid, M. J. Awan, A. Yasin, M. A. Mohammed, R. Damaševičius, R. Maskeliūnas, and K. H. Abdulkareem, "Real-time hand gesture recognition based on deep learning YOLOv3 model," *Appl. Sci.*, vol. 11, no. 9, p. 4164, May 2021.

[23] L. K. S. Tolentino, R. O. S. Juan, A. C. Thio-Ac, M. A. B. Pamahoy, J. R. R. Forteza, and X. J. O. Garcia, "Static sign language recognition using deep learning," *Int. J. Mach. Learn. Comput.*, vol. 9, no. 6, pp. 821–827, Dec. 2019.

[24] S. Sharma and S. Singh, "Vision-based hand gesture recognition using deep learning for the interpretation of sign language," *Exp. Syst. Appl.*, vol. 182, Nov. 2021, Art. no. 115657.

[25] 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, p. 706, Jan. 2022.

[26] H. Ansar, A. Ksibi, A. Jalal, M. Shorfuzzaman, A. Alsufyani, S. A. Alsuhibany, and J. Park, "Dynamic hand gesture recognition for smart lifecare routines via K-ary tree hashing classifier," *Appl. Sci.*, vol. 12, no. 13, p. 6481, Jun. 2022.

[27] J. Shin, A. Matsuoka, M. A. M. Hasan, and A. Y. Srizon, "American sign language alphabet recognition by extracting feature from hand pose estimation," *Sensors*, vol. 21, no. 17, p. 5856, Aug. 2021.

[28] M. V. Vijaykumar, H. J. Vidyarani, A. C. Shekhar, A. Metre, N. M. Kavya, and P. Prakruthi, "Real time hand gesture recognition," *Int. Res. J. Modernization Eng. Technol. Sci.*, vol. 4, pp. 1–6, Sep. 2022.

[29] F. Wen, Z. Zhang, T. He, and C. Lee, "AI enabled sign language recognition and VR space bidirectional communication using triboelectric smart glove," *Nature Commun.*, vol. 12, no. 1, p. 5378, Sep. 2021.

[30] R. Rastgoo, K. Kiani, and S. Escalera, "Hand sign language recognition using multi-view hand skeleton," *Exp. Syst. Appl.*, vol. 150, Jul. 2020, Art. no. 113336.

[31] P. Likhar, N. K. Bhagat, and G. N. Rathna, "Deep learning methods for Indian sign language recognition," in *Proc. IEEE 10th Int. Conf. Consum. Electron. (ICCE-Berlin)*, Nov. 2020, pp. 1–6.

[32] S. Sharma, R. Gupta, and A. Kumar, "Continuous sign language recognition using isolated signs data and deep transfer learning," *J. Ambient Intell. Humanized Comput.*, vol. 14, pp. 1–12, Aug. 2021.

[33] N. K. Bhagat, Y. Vishnusai, and G. N. Rathna, "Indian sign language gesture recognition using image processing and deep learning," in *Proc. Digit. Image Comput., Techn. Appl. (DICTA)*, 2019, pp. 1–8.

[34] T. Zhou, W. Wang, S. Qi, H. Ling, and J. Shen, "Cascaded human-object interaction recognition," in *Proc. CVPR*, 2020, pp. 4263–4272.

[35] A. Jalal, A. Nadeem, and S. Bobasu, "Human body parts estimation and detection for physical sports movements," in *Proc. 2nd Int. Conf. Commun., Comput. Digit. Syst. (C-CODE)*, Mar. 2019, pp. 104–109.

[36] A. A. Rafique, A. Jalal, and A. Ahmed, "Scene understanding and recognition: Statistical segmented model using geometrical features and Gaussian Naïve Bayes," in *Proc. Int. Conf. Appl. Eng. Math. (ICAEM)*, Aug. 2019, pp. 225–230.

[37] M. Batool, A. Jalal, and K. Kim, "Sensors technologies for human activity analysis based on SVM optimized by PSO algorithm," in *Proc. Int. Conf. Appl. Eng. Math. (ICAEM)*, Aug. 2019, pp. 145–150.

[38] K. Kim, A. Jalal, and M. Mahmood, "Vision-based human activity recognition system using depth silhouettes: A smart home system for monitoring the residents," in *J. Elect. Eng. Technol.*, vol. 14, pp. 2567–2573, Sep. 2019.

[39] A. Ahmed, A. Jalal, and K. Kim, "Region and decision tree-based segmentations for multi-objects detection and classification in outdoor scenes," in *Proc. Int. Conf. Frontiers Inf. Technol. (FIT)*, Dec. 2019, pp. 209–2095.

[40] A. A. Rafique, A. Jalal, and K. Kim, "Statistical multi-objects segmentation for indoor/outdoor scene detection and classification via depth images," in *Proc. 17th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2020, pp. 271–276.

[41] A. Ahmed, A. Jalal, and K. Kim, "RGB-D images for object segmentation, localization and recognition in indoor scenes using feature descriptor and Hough voting," in *Proc. 17th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2020, pp. 290–295.

[42] M. Mahmood, A. Jalal, and K. Kim, "WHITE STAG model: Wise human interaction tracking and estimation (WHITE) using spatio-temporal and angular-geometric (STAG) descriptors," *Multimedia Tools Appl.*, vol. 79, nos. 11–12, pp. 6919–6950, Mar. 2020.

[43] M. A. K. Quaid and A. Jalal, "Wearable sensors based human behavioral pattern recognition using statistical features and reweighted genetic algorithm," *Multimedia Tools Appl.*, vol. 79, nos. 9–10, pp. 6061–6083, Mar. 2020.

[44] S. Badar ud din Tahir, A. Jalal, and M. Batool, "Wearable sensors for activity analysis using SMO-based random forest over smart home and sports datasets," in *Proc. 3rd Int. Conf. Advancements Comput. Sci. (ICACS)*, Feb. 2020, pp. 1–6.

[45] S. A. Rizwan, A. Jalal, and K. Kim, "An accurate facial expression detector using multi-landmarks selection and local transform features," in *Proc. 3rd Int. Conf. Advancements Comput. Sci. (ICACS)*, Feb. 2020, pp. 1–6.

[46] S. Badar, A. Jalal, and K. Kim, "Wearable inertial sensors for daily activity analysis based on Adam optimization and the maximum entropy Markov model," *Entropy*, vol. 22, no. 5, p. 579, 2020.

[47] X. Chen, Y. Li, R. Hu, X. Zhang, and X. Chen, "Hand gesture recognition based on surface electromyography using convolutional neural network with transfer learning method," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 1292–1304, Apr. 2021.

[48] 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. Electr. Comput. Eng.*, vol. 2019, pp. 1–12, Oct. 2019.

[49] S. Ahmed, K. D. Kallu, S. Ahmed, and S. H. Cho, "Hand gestures recognition using radar sensors for human–computer-interaction: A review," *Remote Sens.*, vol. 13, no. 3, p. 527, Feb. 2021.

[50] J.-W. Choi, S.-J. Ryu, and J.-H. Kim, "Short-range radar based real-time hand gesture recognition using LSTM encoder," *IEEE Access*, vol. 7, pp. 33610–33618, 2019.

[51] M. Abavisani, H. R. V. Joze, and V. M. Patel, "Improving the performance of unimodal dynamic hand-gesture recognition with multimodal training," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1165–1174.

[52] M. Kumar, P. Gupta, R. K. Jha, A. Bhatia, K. Jha, and B. K. Shah, "Sign language alphabet recognition using convolution neural network," in *Proc. 5th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, May 2021, pp. 1859–1865.

[53] M. M. Rahman, M. S. Islam, M. H. Rahman, R. Sassi, M. W. Rivolta, and M. Aktaruzzaman, "A new benchmark on American sign language recognition using convolutional neural network," in *Proc. Int. Conf. Sustain. Technol. Ind. 4.0 (STI)*, 2019, pp. 1–6.

[54] S. Ameen and S. Vadera, "A convolutional neural network to classify American sign language fingerspelling from depth and colour images," *Exp. Syst.*, vol. 34, no. 3, Jun. 2017, Art. no. e12197.

[55] C.-H. Chuan, E. Regina, and C. Guardino, "American sign language recognition using leap motion sensor," in *Proc. 13th Int. Conf. Mach. Learn. Appl.*, Dec. 2014, pp. 541–544.

[56] M. Bilgin and K. Mutludogan, "American sign language character recognition with capsule networks," in *Proc. 3rd Int. Symp. Multidisciplinary Stud. Innov. Technol. (ISMSIT)*, Oct. 2019, pp. 1–6.

[57] Y. Du, S. Liu, L. Feng, M. Chen, and J. Wu, "Hand gesture recognition with leap motion," 2017, *arXiv:1711.04293*.

[58] Y. Zhao and L. Wang, "The application of convolution neural networks in sign language recognition," in *Proc. 9th Int. Conf. Intell. Control Inf. Process. (ICICIP)*, 2018, pp. 269–272.

[59] T.-W. Chong and B.-G. Lee, "American sign language recognition using leap motion controller with machine learning approach," *Sensors*, vol. 18, no. 10, p. 3554, Oct. 2018.

HIRA ANSAR received the B.S. degree (Hons.) in computer science from Fatima Jinnah Women, Rawalpindi, Pakistan. She is currently pursuing the M.S. degree in computer science with Air University, Islamabad, Pakistan. Her research interests include artificial intelligence, machine learning algorithms, deep learning classification, and hand gesture recognition.

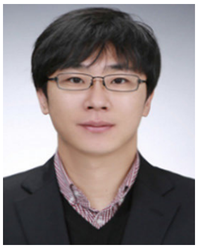
NAIF AL MUDAWI received the master's degree in computer science from Australian La Trobe University, in 2011, and the Ph.D. degree from the College of Engineering and Informatics, University of Sussex, Brighton, U.K., in 2018. He is currently an Assistant Professor with the Department of Computer Science and Information System, Najran University. He has many published research and scientific papers in many prestigious journals in various disciplines of computer science. He was a member of the Australian Computer Science Committee.

SAUD S. ALOTAIBI received the B.Sc. degree from King Abdul Aziz University, in 2000, the master's degree in computer science from King Fahd University, Dhahran, in May 2008, and the Ph.D. degree in computer science from Colorado State University, Fort Collins, CO, USA, in August 2015, under the supervision of Dr. Charles Anderson. He started his career as an Assistant Lecturer with Umm Al-Qura University, Mecca, Saudi Arabia, in July 2001. He was the Deputy of the IT-Center for EGovernment and Application Services, Umm Al-Qura University, in January 2009. From 2015 to 2018, he was with the Deanship of Information Technology to improve the IT services that are provided to the Umm Al-Qura University. He is currently with the Computer and Information College, as the Vice Dean, for academic affairs. He is an Assistant Professor of computer science with Umm Al-Qura University. His research interests include AI, machine learning, natural language processing, neural computing IoT, knowledge representation, smart cities, wireless, and sensors.

ABDULWAHAB ALAZEB received the B.S. degree in computer science from King Khalid University, Abha, Saudi Arabia, in 2007, the M.S. degree in computer science from the Department of Computer Science, University of Colorado Denver, Denver, CO, USA, in 2014, and the Ph.D. degree in cybersecurity from the University of Arkansas, USA, in 2021. He is currently an Assistant Professor with the Department of Computer Science and Information System, Najran University. His research interests include cybersecurity, cloud and edge computing security, machine learning, and the Internet of Things.

BAYAN IBRAHIMM ALABDULLAH received the Ph.D. degree in informatics from the University of Sussex, Brighton, U.K., in May 2022. She was an Assistant Professor with the Department of Information System, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University. She teaches several courses with the Information System Department, such as data governance, system security, and database system. Her research interests include machine learning, data science, privacy, and security.

MOHAMMED ALONAZI received the B.Sc. degree in computer science from King Saud University, Saudi Arabia, in 2008, the M.Sc. degree in computer science from the Florida Institute of Technology, Melbourne, USA, in 2015, and the Ph.D. degree in informatics from the University of Sussex, U.K., in 2019. He is currently an Assistant Professor with the Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia. His research interests include human–computer interaction, UX/UI, digital transformation, cyber security, and machine learning.



JEONGMIN PARK received the Ph.D. degree from the College of Information and Communication Engineering from Sungkyunkwan University, in 2009. He is currently an Associate Professor with the Department of Computer Engineering, Tech University of Korea, South Korea. Before joining the Tech University of Korea, in 2014, he was a Senior Researcher with the Electronics and Telecommunications Research Institute (ETRI) and a Research Professor with Sungkyunkwan University, South Korea. His research interests include high-reliable autonomic computing mechanism and human-oriented interaction system.