IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Sign Language Recognition Using Multiple Kernel Learning: A Case Study of Pakistan Sign Language

Farman Shah[1], Muhammad Saqlain Shah[1], Waseem Akram[2], Awais Manzoor[*,2], Rasha Orban [3], Diaa Salama AbdElminaam[3,4]

[1]International Islamic University Islamabad, Islamabad 44000, Pakistan
[2]COMSATS University Islamabad (CUI), Islamabad 45550, Pakistan
[3]Faculty of Computers and Artificial Intelligence, Benha University, Egypt.
[4]Faculty of Computer Science , Misr International University, Egypt.

*Corresponding author: Awais Manzoor (e-mail:e-mail: malik.awaismanzoor@gmail.com).

ABSTRACT All over the world, deaf people use sign language as the only reliable source of communication with each other as well as with normal people. These communicating signs are made up of the shape of the hand and movement. In Pakistan, deaf people use Pakistan sign language (PSL) as a means of communication with people. In scientific literature, many studies have been done on PSL recognition and classification. Most of these work focused on colored-based hands while some others are sensors and Kinect-based approaches. These techniques are costly and also avoid user-friendliness. In this paper, a technique is proposed for the recognition of thirty-six static alphabets of PSL using bare hands. The dataset is obtained from the sign language videos. At a later step, four vision-based features are extracted i.e. local binary patterns, a histogram of oriented gradients, edge-oriented histogram, and speeded up robust features. The extracted features are individually classified using Multiple kernel learning (MKL) in support vector machine (SVM). We employed a one-to-all approach for the implementation of basic binary SVM into the multi-class SVM. A voting scheme is adopted for the final recognition of PSL. The performance of the proposed technique is measured in terms of accuracy, precision, recall, and F-score. The simulation results are promising as compared with existing approaches.

INDEX TERMS Sign Language, Image recognition, Machine learning, and Features extraction.

I. INTRODUCTION

Sign language is the way of communication and interaction for deaf people all around the world. This kind of communication is accomplished over some hand gestures, facial expressions, or movement of arm/body. The sign language recognition system aims to enable the deaf community to communicate with normal society appropriately. It is a highly structured symbolic set that provides the human-computer interaction (HCI). Sign language is very beneficial as a communication tool, and every day millions of deaf people around the world use sign language to communicate and express their ideas. This facilitation and assistance to deaf persons enable and encourage them to be a healthy part of society and integrate them into society. As you move from one country/ region to another country/region, sign language changes like American, Japanese, Chinese, and Arabic sign language. Pakistan has the sign language known as Pakistan sign language (PSL), and the alphabets of PSL are the alphabets of the national language of Pakistan 'Urdu'.

Sign language has two broad categories which include static and dynamic sign language. In static, a fixed pose of the hand is considered while dynamic sign language includes motion or movement of the hand. In past, many researchers presented several approaches for the recognition of different sign languages, used all around the world. Most of the work has been done for American sign language, Chinese sign language, Arabic sign language. Unfortunately, PSL recognition has still a gap in recognition. Aleem Khalid Alvi [1] presented a statistical template matching technique. In this work, the input data is collected from a sensor device and its mean and standard deviation is calculated. The approach used by Sumaira Kausar [2] is based on colored gloves for recognition of PSL. Halim, Zahid [3] used a dynamic time warping (DTW) algorithm based on recognition of Pakistan

sign language. Although PSL recognition got the significant attention of the researchers in the recent decade, still this topic requires attention.

In this paper, we present a technique for the recognition of one-handed static alphabets of PSL based on vision-based features. This technique employs the idea of multiple kernel learning utilized in the very well-known binary classifier named Support Vector Machine (SVM) [4]–[6]. Each kernel learning method is used in SVM for recognition of Alphabets and their performance is analyzed. Once the performance of each kernel learning method is measured, then the best kernel learning method for recognition of alphabets is selected as the optimal kernel method for that feature.

The next section consists of the state of the art of sign language recognition techniques. Section II presents the proposed methodology, while the performance of the proposed methodology along with the complete description of the PSL dataset is given in section III. The final section of the paper concluded the remarks of the proposed technique.

II. RELATED WORK

Sign language recognition is a broad area of research and researchers in the past have proposed many techniques for sign language recognition. All of these approaches are well-developed according to their scenarios. Although these techniques have a recognition rate, each of these is bounded to their specific scenario, and has some restrictions and issues.

Mostly the researchers put their effort into accurately recognizing the sign language, which involves different approaches i.e., using colored hands, colored markers and gloves, Microsoft Kinect devices, and gyroscopes. The advantage of these approaches is comparatively equal to the other approach which is developed on some geometric based. The latter approach is considered costly.

Saba Jadooki et al. [7] and Bauer, Britta et al. [8] used the Kinect device and colored gloves respectively. Jitcharoenport, Rujira et al. [9] used flex sensors and gyroscopes for Thai sign language recognition. Nada B Ibrahim et al. [10] recognized the Arabic sign language by extracting multiple hand features. Fused features mining is presented et al. [7]. They extracted the hand features which are classified by the artificial neural network (ANN) algorithm. Although this approach claims an error rate of 0.8, their dictionary size is very limited with only 8 alphabets of ASL.

Sumaira Kausar et al. [2] recognized Pakistan sign language by fuzzy classifier approach. They have identified the different positions of fingers and orientation of the hand which is extracted using colored marked gloves. Aleem Khalid Alvi et al. [1] used a statistical template matching technique for Pakistan sign language recognition. The mean value and standard deviation of the sensors for a gesture are identified through which PSL is recognized up to 78.2%, the accuracy is affected by environmental changes. This approach also utilized sensors and gloves. Image-based recognition of Pakistan sign language is done by Muhammad Raees et al. [11]. They recognized the position of fingers in 2-D. although this approach is robust for PSL recognition, it gives 16% gesture failure from hand orientation. Ahmed et al. [12] have used SVM based approach for PSL recognition. In this work, region-based and boundary-based features are extracted. Experiments are conducted for only 10 static alphabets of PSL with a dictionary size of 60 samples and an accuracy rate of 83% is achieved. Halim, Zahid et al. [3] recognized Pakistan sign language using DTW algorithm-based approach. This work considers body parts including head, right wrist, left wrist, right hand, left hand, spine, hp bone, left shoulder, center shoulder, and right shoulder. Their accuracy rate varies base on the distance of the signer from the camera which shows that this approach does not scale-invariant and keeping a specific distance of the signer from the camera they achieved an accuracy rate of 91%. Tauseef et al. [13] proposed a new approach for recognition static PSL alphabets which achieved a high accuracy result of 97.4%. They used a color segmentation approach which can affect the accuracy rate in some situations. Syed Saqlain et al. [14] presented a new approach for categorization of static alphabets of PSL. The approach calculates the histogram of local binary patterns (LBP) of the input images which are further processed for extraction of different features including standard deviation, skewness, kurtosis, variance, entropy, and energy of the LBP histogram. For classification, multi-class SVM is applied which reported an accuracy rate of 78.18% for the dictionary size of over 3400 samples.

III. PROPOSED METHOD

Usually, in an image-based recognition approach, the process starts from the image acquisition which ends with the recognition. A large number of existing approaches are based on supervised learning and unsupervised learning algorithms. Besides this, reinforcement learning, and deep learning approaches are also used. The recognition challenges are usually solved by the algorithms in the area of decision trees, naïve Bayes, k-nearest neighbor (KNN), logistic regression, SVM, dimensionality reduction, and Random forest (RF).

This paper suggests the use of SVM with multiple kernel learnings which include three well-known kernel functions i.e. Gaussian kernel, linear kernel, and polynomial kernel. The use of multiple kernel concepts is adopted and elaborated from the work presented in [4], [4], [6] where the use of multiple kernels is suggested over a single kernel. The basic idea behind the multiple kernels is to find out the best suitable kernel and use it for the understudied problem. In this work, the image frames are obtained from the PSL videos which are then segmented in the next step. The segmented image is converted to a grayscale image and four types of features are extracted separately. These features are further classified in the classification step. Figure 1 elaborates the flow of the proposed approach where the input image is further processed.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3077386, IEEE Access

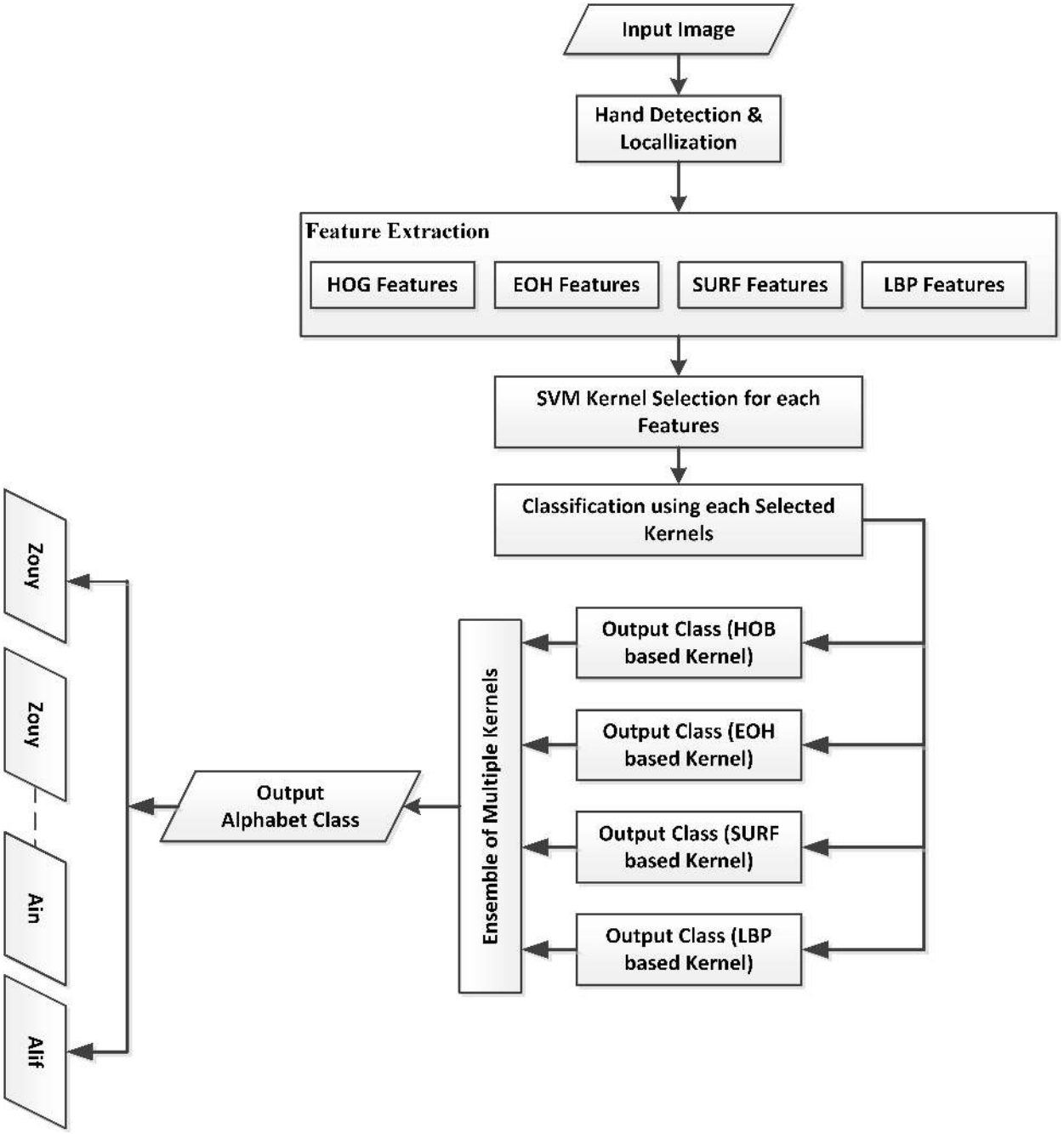


FIGURE 1: Working of the Proposed scheme

A. HAND SEGMENTATION AND CONVERSION TO GRAY SCALE

This domain contains a variety of algorithm which are presented for segmentation i.e. K-means clustering-based segmentation, thresholding, motion and interactive segmentation, compression-based methods, histogram-based methods. In this research K-means clustering-based segmentation is preferred due to its accuracy for segmenting the fingers and palm area of the image from the rest of the image [15]. Algorithm-1 illustrates the computation and working of K-means clustering.

Algorithm-I: Steps for K-means Clustering

1) Suppose $P = p_1, p_2, p_3....p_n$ is the set of data points and $C = c_1, c_2, c_3...c_n$ as the set of centres.
2) Randomly select centre of the cluster V.
3) Calculate the distance D between cluster centre and each data point P.
4) Assign the data point P to the cluster centre on the basis of minimum D.
5) Recalculate the new centre of the cluster using; $v_i = \frac{1}{c_i} \sum_1^{c_i} p_i$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3077386, IEEE Access

FIGURE 2: Segmentation of PSL signs with different scaling, illumination and rotation

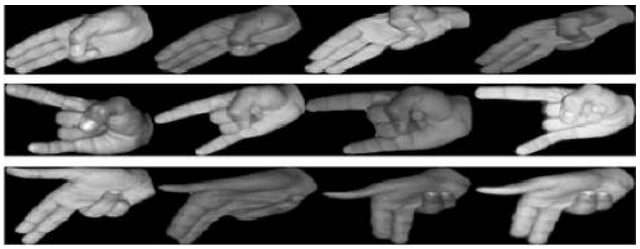


FIGURE 3: Top to Bottom: Grayscale Images of Class "Chay"," Dwad" and "Hamza" with different illumination and degree of rotation

6) Recalculate the Distance 'D' between new cluster centre and each data point P.
7) Stop, if no data point is assigned. Otherwise repeat step 3.

Once the segmentation is done, images are converted to gray scales to extract features. The images of PSL signs containing varying illumination, different scaling, and translation of hand are shown in Figure 2 along with the images obtained after segmentation.

A Cropping algorithm is applied to the original images in order to remove the unwanted black region of the image. Some classes of PSL dataset are given in Figure 3 which shows the images having no extra black region and also these images are converted to grayscale for extraction of features which is done in the next phase.

B. FEATURES EXTRACTION

In this phase, the segmented grayscale images are given as input which is further processed for features extraction. In this article, four different types of features are analyzed which are extracted from the grayscale images which include a histogram of oriented gradients (HOG), Edge orientation histogram (EOH), local binary patterns (LBP), and Speeded up Robust Features (SURF) for PSL recognition. The subsections define each feature briefly.

1) Speeded Up Robust Features (SURF)

SURF is a rotation and scale-invariant detector and is mostly used to calculate the region of interest in the image. In this method, a $9x9$ filter is used to estimate the blob response. Next, other filters of the size $15x15$,$21x21$,and $27x27$ are adopted. A $3x3$ filter is applied to localize the region of interest in the image. Herbert Bay et al. [16] done the blobs detection by SURF using the Hessian matrix. Equation 3 defines the Hessian matrix which is derived from the integral image for interest point detection.

Given an image I and $P = (p, q)$ is a point, then the hessian matrix $H(p, \sigma)$ in p having the scale σ can be defined as given in equation 1.

$$H(p,\sigma) = \begin{bmatrix} c_{pp}(p,\sigma) & c_{pp}(p,\sigma) \\ c_{pp}(p,\sigma) & c_{pp}(p,\sigma) \end{bmatrix} \quad (1)$$

In equation (3) $C_pp\ (p, \sigma)$ is called the convolution of the second order derivative $\frac{\partial^2}{\partial x^2} g(\sigma)$ of the Gaussian function having the image I in the point p.

In our approach we have extracted 10 strongest SURF points in X and Y directions and used the features set for PSL recognition.

2) Edge orientation histogram (EOH)

Previously, in [17], the author used edge orientation histogram features for sign language recognition. The technique detects borders and develops a histogram along the direction of the gradients. We have adopted the canny edge detector which recognizes the edges in five directions i.e. horizontal, vertical, two diagonals, and one non-directional. The oriented histogram is calculated using the following pseudo-code of the edge orientation histogram.

Algorithm-II. Edge orientation histogram

1) **Input:** Image_gray, Image_edge
2) **Initialize:** C, where $C = 0$ to $n - 1$. ($n = 16\ or\ 32\ or\ 64\ or\ 128$.)
3) For every edge pixel of Image_edge; Let $j = 1$ to k , where k = number of edge pixels
   a) Find corresponding gray level intensity.
   b) C=C+1
   c) **goto** Next edge, j = j + 1
4) Goto steps 3(a) & 3(b).
5) If j > k, Break
6) Print histogram

A feature vector of 80 dimensions is obtained which is the representation of edges in five different directions. The features set is further processed for classification in the next phase.

3) Local binary patterns (LBP)

LBP provides lamination's in-variance of the input image [18]. In this process, the threshold on each neighborhood pixel is applied and results in the binary number of the image. A comparison of the $3x3$ neighborhood is carried out with 8 pixels and a threshold for each value. If the value of a center pixel is greater compare to the surrounding value return 0 and otherwise return 1. This binary code is converted to decimal

for convenience. Equation 2 summarizes the working of the LBP operator.

$$LBP_{P,R} = \sum_{p=0}^{p-1} s(g_p, g_c)2^p; \\ s(x) = f(x) = \begin{cases} 0, if x < 0 \\ 1, if x \geq 0 \end{cases} \tag{2}$$

In this paper, we have analyzed LBP for PSL recognition because of the properties of rotation invariance and computational simplicity. The computation of binary code is done by using a binomial factor $2p$. The histogram of 8 neighborhood pixels is used as the features descriptor.

4) Histogram of oriented gradient (HOG)

HOG is used to extract all features from the image. It also refers to "dense features extraction". The author in et al. [19] have used HOG for the human detection and tracking problem. Most commonly, this method is used for object detection, face detection, body detection, and also vehicle detection. The working of the method is shown in Figure 4.

The image is divided into a block size of 2x2, 8x8, 16x16, or 64x64, which is further divided into the size of a cell. Usually, cell size is kept up to 4x4, 8x8, or 16x16. Then, use the sobel or canny operator to compute the vertical and horizontal gradients for each pixel in a cell as given below.

$$G_x(y,x) = Y(y, x+1) - Y(y, x-1) \\ G_y(y,x) = Y(y+1, x) - Y(y-1, x) \tag{3}$$

Where $G_x(y,x)$ is horizontal gradient and $G_y(y,x)$ is the vertical gradient of a pixel. The magnitude and phase of both the gradients are computed as,

$$G(y,x) = \sqrt{G(x)(y,x)^2 + G(y)(y,x)^2} \\ \theta(y,x) = arctan(\frac{G(y)(y,x)}{G(x)(y,x)}) \tag{4}$$

The next step is to find HOG for all cells, and select bins for an angle to calculate the histogram. HOG features descriptor is used with a block size of 2x2 and a cell size of 16x16 for HOG features extraction. The feature vector of the Histogram of oriented gradients is further processed for PSL recognition and classified through multiple kernel learning.

C. SUPPORT VECTOR MACHINE AND MULTIPLE KERNEL LEARNING FOR CLASSIFICATION

Support vector machine algorithm (SVM) is a supervised learner used for classification problems. In [20], the authors proposed an extended version of SVM consists of multiple kernel learning.

The basic SVM works as follows. First, it divides the data into binary classes. Then, finds a hyperplane that differentiates these classes. The data that lies near the hyperplane are called support vectors and mathematically represented as:

$$(x,y), ...(x_i, y_i), ...(x_n, y_n); \quad x_i \in R^N, y_j \in \{-1, 1\} \tag{5}$$

The hyperplane which classifies linearly separable a given set of data can be expressed by the following equation. This hyperplane is called Optimal Separating Hyperplane.

$$f(x) = \sum_{i=1}^{n} \alpha_i y_j (X_i^T X) + b \tag{6}$$

The representation of the optimal hyperplane along with the support vectors x_1, x_2, and x_3 which lies on the boundary of the hyperplane is given by:

$$g(\overrightarrow{X}) = \overrightarrow{W^T}\ \overrightarrow{X} + b \tag{7}$$

Where the decision rule is formulated as if $\overrightarrow{W^T}\ \overrightarrow{X}$ + b > 0 then C1 otherwise C2.

Multiple kernel multiple employs a set of machine learning methods. Each method includes a predefined set of kernels. These kernels estimate optimal linear and non-linear combinations of kernels from a large set of kernels. This approach reduces the bias effect in the learning process.

In our work, we have used three types of very well-known kernel functions in multi-class SVM and analyzed the different results given for PSL recognition by each kernel function by using Gaussian, linear, and polynomial kernels [4], [5]. These kernels are different to determine the best decision boundary. Each of these kernels is used when training the SVM and the best kernel is identified for prediction. SVM calculates the Posterior probabilities for each instance of each class to best divide the classes.

The requirements on a kernel function $k(x_i, x)$ is satisfying Mercer's theorem. There exist many possible inner product kernels. Equations 8 to 10 Illustrates the idea of different kernel functions.

D. GAUSSIAN KERNEL

A Gaussian kernel has the shape of a Gaussian (normal distribution) curve. The width of the Gaussian shape is defined in terms of σ by using the standard ways of statistics,

$$g(\overrightarrow{X}) = \overrightarrow{W^T}\ \overrightarrow{X} + b \tag{8}$$

E. LINEAR KERNEL

The linear kernel is the dot product of two feature filters, and linear distribution of features results into a linear decision boundary.

$$exp(-\frac{1}{2\sigma^2}||X_j - X|^2) \tag{9}$$

F. POLYNOMIAL KERNEL

A polynomial kernel can be represented as given in equation 10.

$$k_{lin}(x,z) = x^T z \tag{10}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3077386, IEEE Access

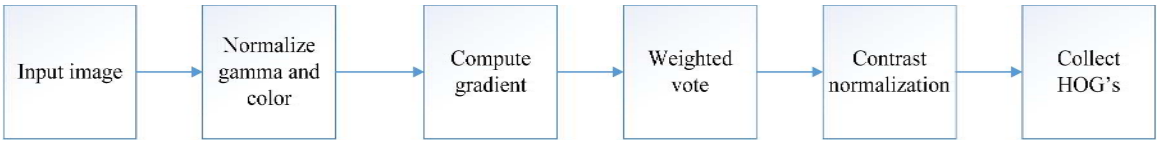


FIGURE 4: Steps for HOG features extraction

The performance of the Support vector machine is dependent on the optimal kernel selection and parameters construction.

G. ENSEMBLE OF MULTIPLE KERNELS FOR MULTIPLE FEATURE SPACES

Once the features have been extracted in four spaces i.e., HOG, EOH, LBP, and SURF, each of the feature space is subject to be used for classification purposes through the use of three kernels i.e., linear, Gaussian, and polynomial. The selection of the best performer SVM kernel, $KF(s, s_j)$, is based on the competitive process explained in the kernel selection algorithm(KSA).

The KSA takes, feature vector, fv_k, as input where it represents the extracted features using all the four spaces i.e., SURF, LBP, HOG, and EOH. The algorithm applies K-fold cross-validation, K=10, and computes average accuracy through each of the three kernels and over each of the feature space separately. The kernel function over a single feature space(among the three kernels) that shows the highest of the average accuracy is selected, fv_k, for that specific feature space.

The next step is to apply selected kernel SVM models over test data. The trained models predict output classes for the test data and the distances from decision boundaries, using $y = u.\Phi(s) + b$, for each of the classes. The probability estimates are computed from the computed decision boundaries using the sigmoid function [154,155]i.e.,

$$p(C_i|fv) = \frac{1}{(1 + e^{(-y_i)})} \tag{11}$$

where $p(C_i)$ represents the probability of i^th class using decision boundaries from the classifier. The output of the selected kernel models is ensembled to produce the final predicted class and the process is explained in the Ensemble of Feature-Based Kernels Algorithm(EoFKA).

Algorithm 1 : Kernel Selection Algorithm (KSA)

INPUT: feature vector set, $fv_k : k \in \{SURF, LBP, HOG, EOH\}$, representing {Dataset Minus TestData}

OUTPUT: selected kernel function, $SKF_u : u \in \{linear, Gaussian, polynomial\}$

initialize accuracy for each of the kernel as zero i.e., $acc\_KF_m = \{0\}$
for $(1 \leq i \leq 10)$ **do**
  Randomly Divide fv_k into training and validation subset i.e.,$fv_{k(train)}$, $fv_{k(valid)}$
  for (each $KF_m : m \in linear, Gaussian, polynomial$) **do**
    Train KF_m based multiclass SVM over $fv_{k(train)}$
    Validate the model using $fv_{k(valid)}$ and compute its accuracy i.e., $tempAcc\_KF_m$
    $acc\_KF_m = acc\_KF_m + tempAcc\_KF_m$
  end for
end for
Compute average validation accuracy for each of the kernel, $KF_m : m \in linear, Gaussian, polynomial$, after 10 iterations i.e., $avgAcc\_KF_m = acc\_KF_m/10$
Find kernel function corresponding to maximum average validation accuracy i.e., $SKF_u = MAX(avgAcc\_KF_m : \forall m \in \{linear, Gaussian, polynomial\})$
return SKF_u

IV. DATASET AND EVALUATION METRICS

We have developed a large dataset of Pakistan sign language for thirty-six static alphabets. The signs are performed by six individual signers which are recorded by a fixed positioned camera. The images are obtained from the sign language videos at 30 frames per second. The captured images contain a change in the position of the hand, rotation of hand in different angles, scaling of hand, and illumination variations. The distance of the signer from the camera is different i.e. of the 1 foot, 1.5 foot, and 2.0 foot. Rotation from 0 to 45 degrees of hand is included in the signing to make the approach more signer friendlier. Although the signs are performed by bare hands, there are also restrictions on background appearance and clothing, which are kept quite different from the skin color for correct and effective segmentation. The constraints which are defined for the signer are much easier to follow instead of using the complex and costly gadgetry of Kinect and data gloves, also the signer is not required to use color hands as this approach is based on bare hands.

PSL dataset consists of a total of 6633 one-handed static alphabet images of Pakistan sing language. The dictionary size of the dataset is larger as compared to other methodologies used for PSL recognition. We have an average of 184 samples for each alphabet of Pakistan sign language. In the learning and classification process, every class is divided into

Algorithm 2 : Ensemble of Feature-Based Kernels Algorithm(EoFKA)

INPUT: Predicted Class by each Selected Trained Model($C_{ik} : 1 \leq i \leq 36, k \in \{SURF, LBP, HOG, EOH\}$)
Class Probability by each Selected Trained Model($p(C_{ik}) : 1 \leq i \leq 36, k \in \{SURF, LBP, HOG, EOH\}$)
OUTPUT: Predicted Class through Ensemble C_{ensem}

Vote among the predicted class labels i.e., C_{ik}
if one of the class,$C_p : 1 \leq p \leq 36$, has higher votes than any other **then**
  $C_{ensem} = C_p$
else if Two of the classes, $C_p, C_q : 1 \leq p, q \leq 36$,have two votes each **then**
  Compute commutative class confidence score(ccs) for p and q i.e.,
  $ccs_p = p(C_{ps}) + p(C_{pt}) : s, t \in \{SURF, LBP, HOG, EOH\}$ based selected kernels and predicting class p
  $ccs_q = p(C_{qu}) + p(C_{qv}) : u, v \in \{SURF, LBP, HOG, EOH\}$ based selected kernels and predicting class q
  if $(ccs_p > ccs_q)$ **then**
    assign C_{ensem}=p
  else if $(ccs_q > ccs_p)$ **then**
    assign C_{ensem}=q
  else if $(ccs_q == ccs_p)$ **then**
    assign C_{ensem}=class corresponding to Highest($p(C_{ps}), p(C_{pt}), p(C_{qu}), p(C_{qv})$)
  end if
else if (all the four trained models predicted distinct classes i.e., m,n,p,q) **then**
  C_{ensem}=class corresponding to Highest($p(C_{mSURF}), p(C_{nLBP}), p(C_{pHOG}), p(C_{qEOH})$)
end if
return C_{ensem}

a ratio of 70% for training,15% for validation, and 15% for testing purposes.

The performance of the proposed approach is evaluated using the four matrices i.e. accuracy, precision, recall, and F-score. Following is the mathematical representation of matrices.

$$Accuracy = \frac{T.N + T.P}{T.P + T.N + F.P + F.N} \tag{12}$$

$$Precision = \frac{T.P}{(T.P + F.P)} \tag{13}$$

$$Recall = \frac{(T.P}{(T.P + F.N)} \tag{14}$$

$$F - score = 2\frac{(Precision * Recall)}{(Precision + Recall)} \tag{15}$$

where, T.P denotes true positive, F.P is false negative and F.N is false negative.

V. EXPERIMENTAL RESULTS AND DISCUSSION

This section discusses the procedure of experiments along with the performance results of the proposed methodology. The technique of multiple kernel learning in SVM is used and three kernel methods are utilized for the performance evaluation. The accuracy of each kernel method for each feature set is calculated. The kernel function having the highest accuracy rate is selected as the optimal kernel for that feature set which is then tested with testing features.

Once the segmentation is done, the images are processed for features extraction i.e. Histogram of oriented gradients (HOG), edge orientation histogram (EOH), local binary patterns (LBP), and SURF. Features vectors are individually processed for classification through MKL. Only based on accuracy, the decision of selecting the optimal kernel function is made. The noticing point here is that once a kernel function is proved optimal for a given set of features, that kernel will be used in the testing phase.

A total of 16 experiments in four phases are conducted for the performance evaluation of four feature sets with three kernel functions. Three experiments are conducted for performance evaluation of each kernel function while one experiment is conducted for the optimal kernel among the three kernels by testing the optimal kernel function with testing features.

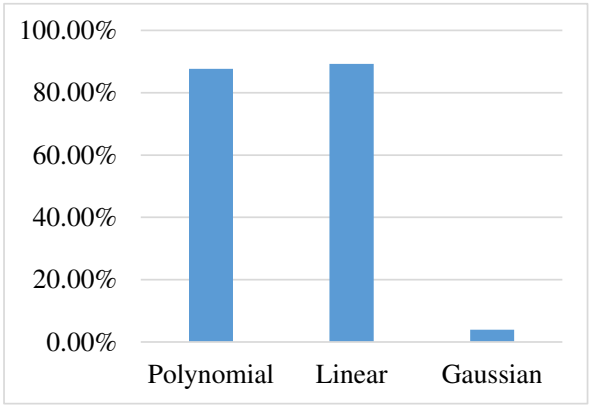
A. EXPERIMENTS FOR HOG FEATURES CLASSIFICATION

The first Phase's experiment is conducted for the classification of Histogram of Oriented Gradients (HOG) features. The recognition results are evaluated using the multi-class SVM and multiple kernel learning approach MKL. Three kernel functions, linear, polynomial, and Gaussian are tested. The graph in Figure 5a shows the performance of each kernel method in terms of accuracy which reveals that the linear kernel function outperforms with an accuracy rate of 89.24% then the non-linear kernel functions. The performance of the polynomial kernel function is high as compared to the Gaussian function which returns only an accuracy rate of 3.95%. Besides the highest recognition performance, the linear kernel function is also proven a computationally efficient function in both the training and testing phase. The linear kernel is further selected for the testing features which achieved an accuracy rate of 89.52%.

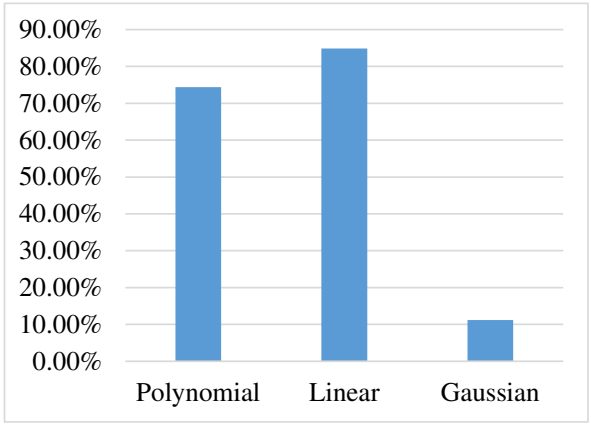
TABLE 1: Best selected kernels functions for each feature set

Dataset	Feature set	Highest average accuracy	Selected kernel function	Accuracy after testing
PSL Dataset	HOG	89.24%	Linear	89.52%
	EOH	84.87%	Linear	87.67%
	LBP	43.65%	Linear	45.71%
	SURF	20.68%	Polynomial	15.41%

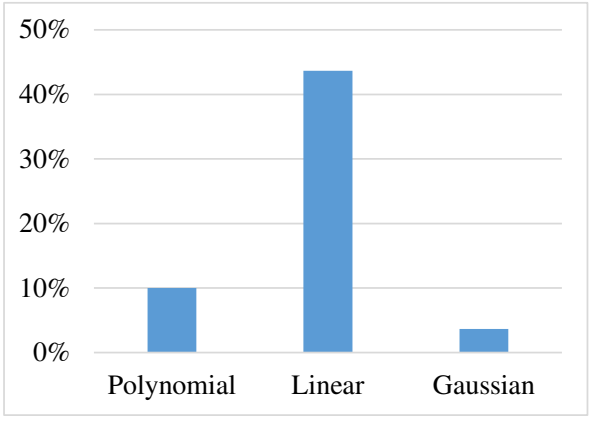
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3077386, IEEE Access



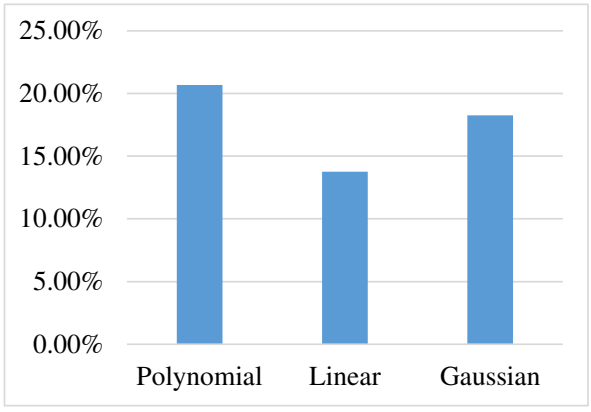
(a) HOG features classification



(b) Edge oriented histogram



(c) LBP Features



(d) SURF

FIGURE 5: Performance of each kernel

B. EXPERIMENTS FOR EOH FEATURES CLASSIFICATION

The second Phase experiments are conducted for the performance evaluation of edge orientation histogram features. Here the linear kernel function is found to be more accurate for the recognition of PSL. Figure 5b shows the accuracy results of each kernel function along with the testing features results for the optimal selected linear kernel function. The Gaussian kernel function returned the lowest accuracy rate.

C. EXPERIMENTS FOR LBP FEATURES CLASSIFICATION

The phase three experiments are conducted for Local binary patterns (LBP) features classification. The local binary patterns have comparatively fewer recognition results for PSL. The linear kernel function is again proven optimal with an accuracy rate of 43.65% in the recognition process while testing the linear kernel on testing features achieved an accuracy rate of 45.71%. The result of each kernel method is given in Figure 5c.

D. EXPERIMENTS FOR SURF FEATURES CLASSIFICATION

Phase four experiments are conducted for the classification of SURF features. SURF is not proven accurate for PSL recognition. The accuracy decrease of SURF features is due to the non-rigid surface. The polynomial kernel function is proven optimal in recognition with an only accuracy rate of 20.68% for validation and 15.41% for testing features.

E. KERNAL SELECTION

The accuracy results of the optimal kernel functions are gathered for all features and carried out for final recognition of PSL. Three feature sets are best recognized through the linear kernel function in SVM while for SURF features classification the polynomial kernel is proven accurate. Table 1 shows the best-selected kernel method for the PSL dataset along with the testing accuracy of the selected kernel function.

F. VOTING SCHEME/ENSEMBLING FOR PSL RECOGNITION

The final accuracy of the proposed methodology is measured using the best-selected kernel method for every feature set classification. we compared the predicted labels of four features classification and selected the label with the highest votes as the true label of that class. Four possible cases are dealt with in the selection of labels.

TABLE 2: labels predicted by all features classification methods

S. No	EOH	HOG	LBP	SURF	Predicted class
1	8	8	11	3	8
2	25	25	25	25	25
3	35	6	6	18	6
4	8	8	8	5	8
5	17	17	18	3	17
6	33	33	33	36	33
7	11	11	34	3	11
8	25	25	25	30	25
9	11	10	3	3	10
10	6	6	18	28	6
11	31	31	17	17	31
12	25	25	25	25	25
13	6	6	19	13	6
14	14	25	25	25	25
15	13	28	25	16	28
16	2	2	2	2	2
17	28	34	2	34	34
18	21	21	21	7	21
19	13	13	13	33	13
20	10	10	36	11	10

Case-1 HOG feature classification predicted the label as "Alif", EOH features classification predicted the same label as "Chay" while the LBP and SURF classification predicted it as "Noon" and "Alif" respectively. Class "Alif" is having more votes so it is selected.

Case-2 The number of votes is the same i.e. a tie situation, in this case, the pre-calculated posterior probabilities are utilized. For example, LBP and SURF identified the class as the class "Seen", while the HOG and EOH recognized the same class as "Zay" so the posterior probability of LBP and SURF are summed up i.e (PPLBP + PPSURF) and also the posterior probability of HOG and EOH are summed i.e (PPHOG + PPEOH). The set with the highest probability values is given the vote.

If PPLBP + PPSURF > PPHOG + PPEOH = "Seen " and "Zay" otherwise.

Case-3 The predicted class for each feature classifier may be different i.e., the class is recognized as "Meem" by HOG features classifier, "Fay" by EOH, and "Hamza" and "Wow" by LBP and SURF classifiers, respectively. Now, in this case, the posterior probability of each predicted class is checked, and the class with high probability is selected as the true predicted class label.

Case-4 The class which is predicted by all features is assigned as the same.

This process is implemented and the predicted class labels are compared with the actual label of the class to measure the performance of our proposed methodology. The different labels for classification of different features and the predicted class label for an overview of only twenty labels are given in Table 3. At location 11 it can be seen that for classification of edge orientation histogram (EOH) and HOG both recognized the class label 31 while LBP and SURF recognized the class label of 17. In such a condition where the number of votes is the same, the posterior probability for each feature classification is calculated and the decision of class selection is made on the highest probability.

The classes having the highest votes is selected as a recognized class. The posterior probability calculated for labels 31 for EOH and labels 31 for HOG is given in the figure. Taking sum of both probability values i.e. 0.2623+0.9679 = 1.2302. The sum of the probabilities of LBP and SURF features for label 17 is 0.1371+0.0724= 0.2095. The higher probabilities labels are selected as shown in Table 2. In the case where each predicted label is different for all features classification (see Table 1, s.no 15), the posterior probability is calculated in the same way and the label with the highest probability is selected.

According to the description given in Table 3, the performance of the proposed methodology is evaluated in terms of accuracy, precision, recall, and F-score. The recognition accuracy rate of 91.93% is achieved. Figure 6 shows the precision, recall, and F-score of each of the alphabet.

The proposed methodology is easier to adopt for Pakistan sign language recognition in many aspects. As described earlier, this is a bare-handed approach and it does not require any external gadgetry. Previously, many other approaches developed for Pakistan sign language recognition have claimed a better recognition accuracy also. The features used in this paper are individually used previously for sign language recognition. A comparison of our work with these approaches is given in Table 3.

VI. CONCLUSION

In the presented method, Pakistan sign language (PSL) is recognized using multiple kernel learning techniques in SVM from a set of vision-based features. The proposed technique takes the bare hand images as input and segments them in the next phase using K-means clustering segmentation. The images of the PSL dataset include the change in the position of the hand, containing the rotation of the hand in different

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3077386, IEEE Access

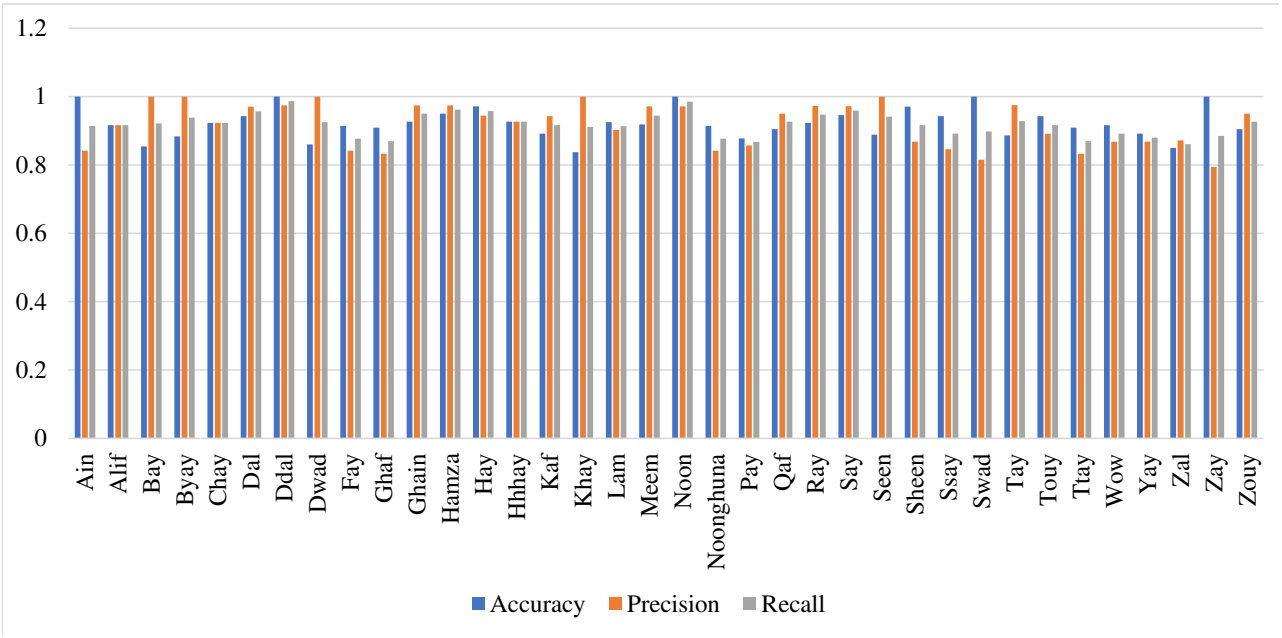


FIGURE 6: Precision, recall and F-score of Each PSL alphabet Class

TABLE 3: Comparison of other techniques with proposed technique

Ref.	Dataset	Features				Dataset properties				Classifier	Performance
		SURF	LBH	EOH	HOG	Rotating	Scaling	Illumination variation	Require <external Gadgetry		
[1]	PSL Alphabets						✓	✓	✓	Fuzzy classifier	86%
[2]	PSL Alphabets								✓	Template matching	78.20%
[10]	PSL Alphabets									Template matching	84%
[11]	Only 10 PSL Alphabets									SVM	83%
Proposed method	PSL	✓	✓	✓	✓	✓	✓	✓		SVM with MKL	91.98%

angles, scaling of hand, and illumination variations. Four vision-based features i.e. histogram of oriented gradients (HOG), edge orientation histogram, local binary patterns (LBP), and speeded up robust features (SURF) are extracted in the next phase, individually. These features are classified using the multiple kernel learning (MKL) technique in support vector machines (SVM) classifier. Each feature set is classified by three kernel methods i.e. Gaussian, linear and polynomial kernel, and the results are compared. The kernel function with the highest accuracy is selected as the best kernel method and final accuracy is calculated based on all the selected kernels for each feature set. The proposed methodology achieved considerably good recognition results for PSL recognition. Graphs are given for average accuracy with precision, recall, and F-score of each of the PSL alphabet classes.

In our method, three kernel functions are utilized for each feature set classification. By observing the accuracy of each feature set it is clear that the accuracy of SURF features is only 15%. In the future, we suggest the use of other features e.g. maximally stable extremal regions (MSER) features instead of SURF. Also, there is a lot of room to further improve the dataset. The segmentation phase of our methodology follows some constraints which are also needed to be user-friendlier.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest. Non-financial competing interests.

CREDIT AUTHOR STATEMENT

Farman Shah: Methodology, Software, Writing - Original draft. **Muhammad Saqlain Shah:** Conceptualization, Writing - Original draft, Data curation, Formal analysis. **Waseem Akram:** Methodology, Resources, Formal verification, Writing - review & editing. **Awais Manzoor**: Conceptualization, Methodology, Software, Writing - review & editing. **Rasha Orban:** Formal analysis, Resources, Writing - review & editing. **Diaa Salama AbdElminaam:** Formal analysis, Validation, Resources, Writing - review & editing. All authors

read and approved the final paper.

REFERENCES

[1] A. K. Alvi, M. Y. B. Azhar, M. Usman, S. Mumtaz, S. Rafiq, R. U. Rehman, and I. Ahmed, "Pakistan sign language recognition using statistical template matching," International Journal of Information Technology, vol. 1, no. 1, pp. 1–12, 2004.

[2] S. Kausar, M. Y. Javed, and S. Sohail, "Recognition of gestures in pakistani sign language using fuzzy classifier," in Proceedings of the 8th conference on Signal processing, computational geometry and artificial vision, pp. 101–105, World Scientific and Engineering Academy and Society (WSEAS), 2008.

[3] Z. Halim and G. Abbas, "A kinect-based sign language hand gesture recognition system for hearing-and speech-impaired: a pilot study of pakistani sign language," Assistive Technology, vol. 27, no. 1, pp. 34–43, 2015.

[4] Z. Ren, H. Li, C. Yang, and Q. Sun, "Multiple kernel subspace clustering with local structural graph and low-rank consensus kernel learning," Knowledge-Based Systems, vol. 188, p. 105040, 2020.

[5] I. Lauriola, C. Gallicchio, and F. Aiolli, "Enhancing deep neural networks via multiple kernel learning," Pattern Recognition, vol. 101, p. 107194, 2020.

[6] Z. Ren, H. Li, C. Yang, and Q. Sun, "Multiple kernel subspace clustering with local structural graph and low-rank consensus kernel learning," Knowledge-Based Systems, vol. 188, p. 105040, 2020.

[7] S. Jadooki, D. Mohamad, T. Saba, A. S. Almazyad, and A. Rehman, "Fused features mining for depth-based hand gesture recognition to classify blind human communication," Neural Computing and Applications, vol. 28, no. 11, pp. 3285–3294, 2017.

[8] B. Bauer and H. Hienz, "Relevant features for video-based continuous sign language recognition," in Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), pp. 440–445, IEEE, 2000.

[9] R. Jitcharoenpory, P. Senechakr, M. Dahlan, A. Suchato, E. Chuangsuwanich, and P. Punyabukkana, "Recognizing words in thai sign language using flex sensors and gyroscopes," i-CREATe2017, vol. 4, 2017.

[10] N. B. Ibrahim, M. M. Selim, and H. H. Zayed, "An automatic arabic sign language recognition system (arslrs)," Journal of King Saud University-Computer and Information Sciences, vol. 30, no. 4, pp. 470–477, 2018.

[11] M. Raees, S. Ullah, S. U. Rahman, and I. Rabbi, "Image based recognition of pakistan sign language," Journal of Engineering Research, vol. 1, no. 4, pp. 1–21, 2016.

[12] H. Ahmed, S. O. Gilani, M. Jamil, Y. Ayaz, and S. I. A. Shah, "Monocular vision-based signer-independent pakistani sign-language recognition system using supervised learning," Indian J. Sci. Technol, vol. 9, no. 25, pp. 1–16, 2016.

[13] H. Tauseef, M. A. Fahiem, and S. Farhan, "Recognition and translation of hand gestures to urdu alphabets using a geometrical classification," in 2009 Second International Conference in Visualisation, pp. 213–217, IEEE, 2009.

[14] S. M. S. Shah, H. A. Naqvi, J. I. Khan, M. Ramzan, H. U. Khan, et al., "Shape based pakistan sign language categorization using statistical features and support vector machines," IEEE Access, vol. 6, pp. 59242–59252, 2018.

[15] A. Behura, "The cluster analysis and feature selection: Perspective of machine learning and image processing," Data Analytics in Bioinformatics: A Machine Learning Perspective, pp. 249–280, 2021.

[16] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," Computer vision and image understanding, vol. 110, no. 3, pp. 346–359, 2008.

[17] S. Nagarajan and T. Subashini, "Static hand gesture recognition for sign language alphabets using edge oriented histogram and multi class svm," International Journal of Computer Applications, vol. 82, no. 4, 2013.

[18] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," Pattern recognition, vol. 29, no. 1, pp. 51–59, 1996.

[19] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol. 1, pp. 886–893, IEEE, 2005.

[20] M. Athoillah, M. I. Irawan, and E. M. Imah, "Support vector machine with multiple kernel learning for image retrieval," in 2015 International Conference on Information & Communication Technology and Systems (ICTS), pp. 17–22, IEEE, 2015.

FARMAN SHAH received the master's degree in Computer Science from the International Islamic University, Islamabad in 2019. His research interest lies in computer vision, machine learning, and deep learning applications used in the recognition systems.

MUHAMMAD SAQLAIN SHAH received the bachelor's and master's degrees from the International Islamic University, Islamabad, where he is currently pursuing the Ph.D. degree in computer science. He has over 15 years of experience in research, development, and teaching. His research interests include image processing, computer vision, activity analysis, sign language recognition, pattern classification, and data mining. He has research profile with over 20 research articles in national and international journals. He has participated in international conferences and workshops of computer science and related fields.

WASEEM AKRAM is a research student of COSMOSE Lab at COMSATS University Isalamabad, Pakistan. He got a BCS (Hons) and MSCS degree from Hazara University Mansehra and COMSATS University Islamabad, Pakistan respectively. During research work, he has published one conference and one journal paper. He has also authored two book chapters and some papers- currently under review. He is also a reviewer for international journals and conferences including Complex Adaptive System Modeling, IEEE Transaction on Emerging Topics of Computing, International Conference on Sensor Networks and Signal Processing, International Conference on Fuzzy System and Data Mining. His research interest includes the formal methods, complex adaptive systems, agent-based modeling and simulation, smart grid, and the Internet of Things.

AWAIS MANZOOR received the Master's degree in computer science from COMSATS University, Islamabad, Pakistan, in 2017. He has previously worked as a Lecturer with the Department of Computer Science, The University of Lahore Pakistan. He has authored several journal articles and books chapters, and a reviewer for journals such as IEEE Systems Journal. His research interests include medical image analysis using deep learning & machine learning techniques, explainable deep learning for medical image analysis, smart grids, and autonomous intelligence of industrial robotics with a special interest in robot manipulation, optimal grasp and manipulation of non-rigid objects.

RASHA O. MAHMOUD is an Assistant Professor at Benha University, Faculty of Computers and Artificial intelligence, Computer Science Department, Egypt since 2011. Assistant Prof. 2010-2011 Nile Higher Institute for Commercial Science and Computer Technology- Mansoura-Egypt. Assistant Prof. 2001-2005 Delta Academy for Computers- Talkha- Egypt. Assistant Lecturer. 1997-2001 Electronic Engineering- Computer and Control Systems Dep. Faculty of Engineering-Mansoura University- Egypt. Assistant Researcher. Her areas of interest are image/ video processing, biometrics, and face recognition and watermarking. (Mobile: +201001910819) ; E-mail: rasha.abdelkreem@fci.bu.edu.eg)

DIAA SALAMA ABDELMINAAM received his Ph.D. degree in information system from the faculty of computers and information, menufia university, Egypt in 2015. He is currently an associate Professor in the data science department, Faculty of Computers ,Misr International University , Egypt on leave benha university Egypt, since Feb 2020. He has worked on several research topics. Diaa has contributed more than 80 technical papers in the areas of optimization , meta heuristics ,Feature Selection , Opinion mining , Machine and Deep learning , wireless networks, wireless network security, Information security and Internet applications, Cloud Computing, Mobile Cloud Computing, Internet of Things, and Machine learning in international journals, international conferences, local journals, and local conferences. He majors in Optimization , Cryptography, Network Security, IoT, Big Data, Cloud Computing, deep learning.