International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Deep Learning Feature Extraction using Pre-Trained Alex Net Model for Indian Sign Language Recognition

Kruti J Dangarwala, Dilendra Hiran

Abstract*: Indian sign language is communicating language among deaf and dumb people of India. Hand gestures are broadly used as communication gestures among various forms of gesture. The real time classification of different signs is a challenging task due to the variation in shape and position of hands as well as due to the variation in the background which varies from person to person. There seems to be no availability of datasets resembling to Indian signs which poses a problem to the researcher. To address this problem, we design our own dataset which is formed by incorporating 1000 signs for the sign digits from 1 to 10 from 100 different people with varying backgrounds conditions by changing colour, and light illumination situations. The dataset comprises of the signs from left handed as well as right handed people. Feature extraction methodologies are studied and applied to recognition of Sign language. This paper focuses on deep learning CNN (convolution neural network) approach with pretrained model Alexnet for calculation of feature vector. Multiple SVM (Support Vector Machine) is applied to classify Indian sign language in real time surroundings. This paper also shows the comparative analysis between Deep learning feature extraction method with histogram of gradient, bag of feature and Speed up robust feature extraction method. The experimental results shown that Deep learning feature extraction using pretrained Alexnet model give accuracy of around 85% and above for the recognition of signed digit with the use of 60% training set and 40% testing set.

Keywords: Classification, Convolution neural network, Deep learning.

I. INTRODUCTION

According to the census 2011, there are around 50, 71,007 deaf people and 19, 98,535 dumb people in India. Sign Language Recognition is a multifaceted task that makes use of body expression detection such as hand and facial movement detection. This detection mechanism is based on various features such as shape and motion features which forms the essential structural blocks of the recognition system [1][2]. Among the various body movements and gestures, the hand gestures are most widely used as a medium of sign language based communication framework. Approximately 80% of ISL sign words [3] are composed of hand gestures. According to the press note released by Information Bureau, Government of India dated 23 march, 2018 11:55 IST, the Union Minister for Social Justice and Empowerment declared the ISL dictionary with 3000 words. Indian Sign Language Research & Training Centre has developed ISL dictionary to eliminate gap between deaf and non-deaf people. The problem that exists in this domain is the non-availability of sufficient datasets for researchers. In this work, our own sign datasets are developed by considering various background conditions of 100 different persons. Here all methodologies and analysis are done with our own creation number sign datasets. Feature extraction methods are used to get useful and relevant information form original image data and showing the information in low dimensional space. For larger image data processing, redundant data must be required. Feature extraction method is used to create feature vector which contain set of features with reduced representation. Histogram of gradient method [4][5] method is used to divide image into different blocks with cell & from cell feature extraction occurred. Magnitude of robust vector decide ensuing vector for a specific cell inside block. Speed up robust feature extraction method is based on hessian matrix which is depending on determinant of hessian matrix. Bag of feature extraction method [6][7][8] is used to examine images and convert into form of features group. It is very useful method for image classification, retrieval of images, texture recognition and many more. Bag of feature extraction method improve the efficiency compared to HOG method. Use of Deep learning [1][2][9][10] has tremendously increased in the area of image classification, natural language processing ,image recognition, speech recognition and many more. Several methodologies of image processing have been applied for classification of Indian sign language. Deep learning [10] has key progressive success for dealing with large complex image datasets. Deep learning [4] has variety of architectures based on methodologies. Deep neural network architecture has input layer, more than one hidden layer and output layer. Deep Auto Encoder has equal number of input layer nodes and output layer nodes with unsupervised learning method.

Revised Manuscript Received on 30 July 2019.

***** Correspondence Author
Kruti Dangarwala*, Ph.D Scholar, Faculty of Computer Engg. , Pacific Academy of Higher Education and Research University, Udaipur, Rajasthan, India.
Dr. Dilendra Hiran, Principal, Faculty of Computer Application, Pacific Academy of Higher Education and Research University, Udaipur, Rajasthan, India.

Deep Boltzmann machine has undirected connection between each sub layer. Deep belief network has undirected connection at top two layer with one visible layer and other hidden layers. Output is depending on previous state computation with same sharing weight across all networks in Recurrent Neural network. Deep learning convolution neural network [11][12] is applied for recognition of images by applying convolution which consider object edges to identify images. It is done with the help of sharing weights and pooling. CNN has different layers through which input is converted into output using differentiable function. Various pre-trained CNN DL models, Alexnet[13], VGGNet[14], GoogLeNet[15] ,LeNet and ResNet[16] are available which provide qualitative results in feature extraction as well as image classification. Alexnet pre-trained model is applied for image feature extraction. The paper is organized as follows. The literature survey is presented in section 2. Section 3 explains dataset development, feature extraction methods namely HOG,SURF and pre-trained alexnet CNN for finding feature vectors along with SVM classification. Section 4 shows the experimental results and analysis done with the help of confusion matrix to identify number sign.

II. RELATED WORK

Edge orientation histogram based method HOG is a robust technique to withstand variations in geometry and illumination change. Hog feature extraction, sift feature extraction and fusion of hog and sift methods are used [17] to identify sign alphabet datasets. Different methods for feature extraction and classification for Indian sign language recognition is highlighted by authors in [17]. Static hand gesture of Thomas Moeslund's gesture ASL database is recognized using deep learning architecture [18] namely deep convolution neural network and Stacked denoising auto encode (SADE). Three CNN's are used to train network with varying depth sizes. First CNN trained with two hidden layers. Second CNN is trained with three hidden layers and the third CNN is trained with four hidden layers. Hand sign images features are automatically extracted using [19] Deep Neural Networks. The features are scale and rotation invariant. Kinect sensor camera is used to capture images. Convolution neural network and Deep belief networks are applied for recognition of hand posture. DBN recognition accuracy is better than DNN. Image classification [20] is done using simple convolution neural network which is tested on MINIST and CIFAR-10 datasets of hand-written digits. Hybrid CNN-HMM approach[21] is used to recognize video sign databases. CNN with HMM framework in the context of sign language and gesture recognition. They apply their hybrid performance on three challenging standard benchmarks namely RWTH-PHOENIX-weather 2012, RWTH-PHOENIX-Weather Multisinger 2014 and SIGNUM single signer publicly available datasets. With RWTH-PHOENIX-Weather 2014 Multisigner, deep convolutional neural networks[22] have proved unbreakable accuracies for classification of images. Deep CNN architectures are developed to extract the spectral, spatial and combination of both deep features. 3D CNN is designed to extract spectral-spatial features for better classification accuracy. These methods are applied on three known hyper spectral datasets namely Indian Pines, University of Pavia and Kennedy space centre.

The pre-trained convolution neural network Alex-Net[22] is applied for training Image-Net Large-Scale Visual Recognition Challenge (ILSVRC). It is used to extract features for classifying 1.2 million images of image-Net database with 1000 different classes. AlexNet[23] is applied for scene classification work and gaining excellent success. Authors also proposed Alexnet–Scale pyramid pooling and Alexnet with side supervision for improving accuracy. 3D convolution neural network [24] is also applied for dynamic hand gesture recognition. Approach of 3D CNN is analysed with variety of methods like HOG, Spatial stream CNN,IDt-HOG,C3D with various modalities including color, depth, optical flow, color with depth, color with optical flow and many more. Further modification is done with 3D CNN model. Convolutional Neural Networks (CNN)[25] is class of deep learning (DL) models which assures about greater results in feature extraction and classification. DL techniques are widely used in medical science analysis now-a-days. To identify automatic recognition of malaria disease, DL methods are used by authors. Pre-trained CNNs are giving assurance for feature extraction which are proved by performing statistical analysis on the datasets. Feature extraction [26] with CNN achieve more and more success for image classification.

III. METHODOLOGIES

Indian sign language recognition system is developed by first gathering number sign datasets. Here we have used our own dataset creation. First section of methodology describes the development of datasets. Then we initiate pre-processing phase to find out interest of region. The hand segmentation technique is performed. The main part is feature extraction to find out feature vector. To achieve that objective, different methods are employed and compared with pre-trained model alexnet. The main goal is to achieve good accuracy. In all methods, multiple SVM classification is used to identify signs accurately.

A. Dataset creation: Indian Sign Language –The new number sign Database

As per survey/visit of various schools of Gujarat and Rajasthan state, Indian Sign Language dictionary is based on hand gestures signs. The taxonomy of hand gestures is categorized into single handed and double handed system. The students use double handed system to communicate with other people and for digits identification they make use of single handed system. There is not any computerized system for teaching and learning modules. There is lack of online dictionary of Indian Sign Language till date. So the need arises to develop easily available datasets/dictionaries for learning purposes. Datasets acquisition process is very difficult task as online Indian signs datasets for research is not sufficient.

We create our own datasets of 1000 images for 1-10 digits with 100 different persons at different places like playground, classes, college labs etc. Android MI phone is used to capture images by varying the backgrounds effects like colors, left handed people, right handed people, light illumination conditions etc.. The sizes of images are not fixed. So we resize each image with 227 by 227 pixels. Here image labelling is prepared for training set and testing set. Ideally, 60 % of images are taken for training and 40 % of images are taken for testing.

B. Feature extraction methodologies

Feature extraction methods Histogram of Gradient[4],Speed up robust feature[27], bag of feature extraction and deep learning pre-trained Alexnet model of feature extraction methods are applied to number sign datasets. Feature descriptors are found through these three methods. Multiple SVM classification is applied to classify Indian sign with digits 1-10. Following Fig 1 highlights all methodologies applied to number sign datasets.

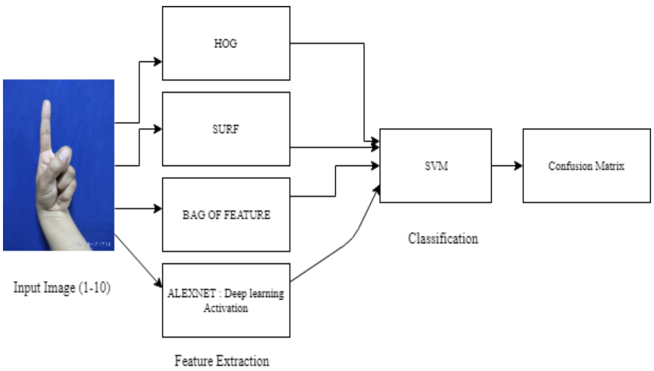


Fig 1. Feature extraction methods with classification method(s)

B. 1 Histogram of Gradient with SVM classification:

HOG is a feature extraction method based on orientation of the gradient. It is an edge orientation histogram[28]. HOG[4] calculation is divided into three parts namely gradient, histogram &Normalization. We apply HOG feature extraction with cell size [2, 2]. Output of feature extraction in an image is shown in fig 2.

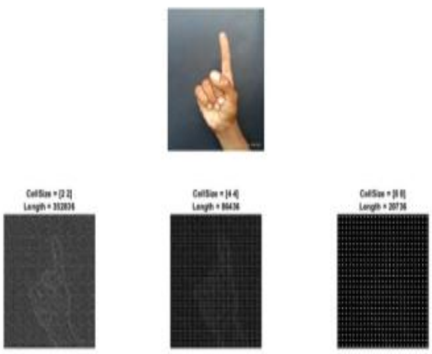


Fig 2. HOG feature extraction

Algorithm 1 : HOG Feature Extraction Algorithm [4]

1. At each pixel, calculate the illumination gradient. Illumination gradient is a vector with two parameters namely magnitude m and orientation φ.

Magnitude m =

$$\sqrt{(P(x+1, y)-P(x-1, y))^{2} + (P(x, y+1)-P(x, y-1))^{2}}$$

Orientation φ =

$$\tan^{-1}((P(x+1, y)-P(x-1, y)) / ((P(x, y+1)-P(x, y-1)))$$

Where P(x, y): Pixel Position

2. At each cell, generate histogram of gradient orientation
3. Perform normalization for feature measurement until it becomes robust to variations in illumination.

HOG-SVM Approach:

1. Resize RGB color images into 227 by 227 pixel size then convert into grayscale image
2. Apply hand region segmentation.
3. Split Digit Sign datasets into training and testing sets with 60:40 ratio.
4. Now Find out HOG feature descriptor from training set and testing set.
5. Apply SVM multiple classification to recognize class between 1-10.
6. Find out confusion matrix to see result.

B. 2 SURF (Speed up robust feature) with SVM classification:

SURF [29] method is based on hessian matrix for feature extraction and selection. This method [27][30] is divided into four parts. In first part, searching of all image locations and scales. The method find out interest points which are not vary to scale and orientation. Second part is used to select key points which are depending on stability measures.
Third part is related to orientation assignment, in which each key point locations assign one or more orientations. All operations are invariance to transformations. Fourth part is related to local image gradients calculations which are measured at each selected scale of key point regions. Transformations occur into significant level of local shape. This approach converts images into scale invariant coordinates which is like local features. Output of feature extraction is given in fig 3.

Fig 3. SURF feature extraction

SURF-SVM Approach:

1. Apply pre-processing step to resize and hand region segmentation.
2. Split Digit Sign datasets into training and testing sets with 60:40 ratios.

3. Find out SURF features from training and testing sets.
4. Apply SVM classification to recognize classes between 1-10.
5. Find out confusion matrix to generate result

B3 Bag of Feature extraction method with SVM classification:

Bag of features [6] is a vector of occurrence counts of image features. Bag of Feature (BoF)[7] is used for different images and video classification problems. The Bag of Feature extraction method is performed with following steps (i) extraction feature detector and descriptor, (ii) Construction of visual vocabulary, (iii) Quantization of visual vocabulary, (iv) Generation of a training model for sign identification, (v) Evaluate Testing images.

Bag of feature extraction method:

1. Resize RGB color images into 227 by 227 then convert into grayscale image
2. Split Digit Sign datasets into training and testing sets with 60:40 ratio.
3. Extracting features using a custom feature extraction Function.
4. Keeping 80 percentages of the strongest features from each category. Balance the number of features across all image categories to improve clustering.
5. Train the category classifier.
6. Apply step 3 to step 4 on testing set.
7. Use evaluate function of matlab to test the classifier on a test set.
8. Find out the results from the confusion matrix

B4 Deep Convolution neural network with pre-trained Alexnet feature extraction with SVM classification:

Deep convolution neural network architecture is constructed with various convolutional layers, max-pooling layers, fully connected layers and softMax layers. Various models of deep convolution neural network exists namely LeNet, Alexnet, VGGNet, GoogLeNet, Residual Networks, DenseNet and Fractal Net. Among them Alexnet, VGG, GoogLenet, Dense CNN and FractalNet are most famous architectures for recognition of objects. Alexnet is powerful deep architecture with in advance research in deep learning. **A**lexNet[13]is a convolution neural network which is applied to large database for training and classifies more than 1000 categories of objects. Alexnet network is applied on image having size 227*227 pixels. Indian sign language classification is done with deep CNN alexnet architecture. Deep CNN[26] is made up with various convolution layers, max-pooling layers, fully connected layers and softMax layers. deep learning with alexnet architecture[12] The performance is based on extracting features from a pre-trained Alexnet CNN. Using activation function of matlab, image database is trained and features are extracted and then support vector machine is applied to classify signs based on extracted feature map. The input images of any sizes are resized with 227-by-227-by-3 .Then the whole image datasets is split into training and testing sets. Here we split into 60 % of training and 40% of testing. To get feature map, activation function with fully connected layer 'fc7' is applied. These extracted features from trained images are applied to multiclass support vector machine 'SVM'.

Algorithm 2: Feature Extraction using Pre-trained CNN Alexnet model with SVM Classification

---

Input: Number Sign Images (1000 images)
Output: Generation of Confusion Matrix (10 classes)

1. Load Input images with labels.
2. Split Image data set for training and testing in ratio of 60:40.
3. Load Pre-trained Alexnet Model.
4. Apply pre-processing to Image dataset for resizing image with 227 by 227 pixels size as suitable for alexnet model.
5. Extract features of training set with feature layer fc7 from alexnet model .
6. Apply training features to multiple SVM classifier.
7. Extract testing features from alexnet model.
8. Use the trained classifier to predict labelling of testing Set.
9. Generate confusion matrix for identify result.

IV. EXPERIMENT RESULT ANALYSIS

Experiment is done with 1000 number sign image datasets with 100 different persons. Datasets are split into 60 % training and 40% testing sets. Here 600 images are taken for training and 400 images for testing. The main goal is to identify correct hand signs shown by the person.

The implementation of our approach will lead to the generation of confusion matrix. Using confusion matrix, we easily identify number signs. Here we have shown four different feature extractions with classification analysis. Analysis is shown that pre-trained Alexnet model give better accuracy than HOG with SVM, SURF with SVM & Bag of feature extraction method. A confusion matrix is a table which find out quantitative relationship between correctly identify images and non-correctly identify images. Confusion matrix is a square matrix in which diagonal value indicate correctly classify number signs values. Others values in the matrix indicate that not correctly classify signs. Here confusion matrix is generated through implementation in all the approaches. It is also used to measure accuracy for image classification. Each row of confusion matrix indicates number sign 1-10. Different statistical analysis is done with confusion matrix result. Here matlab 2017 is used to implement all approaches. Accuracy is the most important consideration measure which indicates performance of model. Here we measure overall system accuracy, precision, recall and F1-score. Overall system accuracy is ratio of correctly identify signs to total signs data. Precision is the ratio of correctly identified sign data to total correctly positive data.

Recall is the ratio of correctly identified positive observations to the all observations in actual class - yes. Using confusion matrix we can easily find out precision and recall for each class. Diagonal value of confusion matrix is always true positive (TP). F1 Score is the weighted average of Precision and Recall. This measure considers false positives and false negatives both.

Classification's overall accuracy is measured using following equation:

Overall Accuracy = summation of all diagonal elements / Total no of elements (1)

Precision based on class (1-10) is measured using following equation:

Precision = Diagonal elements of row /other elements in row

=TP/TP+FP (2)

Recall is measured using following equation:

Recall = TP/TP+FN (3)

F1 score is measured using following equation:

F1 Score =

2*(Recall * Precision) / (Recall + Precision) (4)

Analysis is shown that pretrained Alexnet model give better accuracy than other three feature extraction methods. Diagonal value indicate correctly classify number value i.e. True positive value (TP) . The confusion matrix of various methodologies are discussed as per below.

B4.1 HOG with SVM:

Table I Confusion Matrix with HOG-SVM approach cell size [2 2]

	Target class										
Predicted class	Sign	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	1	20	3	3	0	3	0	0	6	3	2
	2	7	22	2	1	1	1	2	1	1	2
	3	8	0	27	2	3	0	0	0	0	0
	4	11	0	2	20	1	4	1	0	1	0
	5	12	0	2	1	18	1	3	0	0	3
	6	0	7	0	0	2	27	0	4	0	0
	7	7	3	2	4	2	1	19	0	2	0
	8	0	7	2	0	0	3	4	19	4	1
	9	2	2	4	4	1	1	3	4	18	1
	10	0	8	7	2	2	1	0	0	0	20

Overall system Accuracy =
(s11+s22+s33+s44+s55+s66+s77+s88+s99+s1010)
= (20+22+27+20+18+27+19+19+18+20)/400
=191/400 = 0.4775

Accuracy in percentage (%) =
Overall accuracy*100=0.4775*100=47.75%

Table II HOG-SVM Performance Metrics

Sign	Precision	Recall	F1-score
1	0.5	0.298507	0.373831413
2	0.55	0.423077	0.47826087
3	0.675	0.529412	0.593406593
4	0.5	0.588235	0.540540541
5	0.45	0.545455	0.493150685
6	0.675	0.692308	0.683544304
7	0.475	0.59375	0.527777778
8	0.475	0.558824	0.513513514
9	0.45	0.62069	0.52173913
10	0.5	0.689655	0.579710145

B4.2 SURF with SVM approach:

Table III Confusion Matrix SURF-SVM

	Target class										
Predicted class	Sign	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	1	15	4	4	4	7	0	0	0	0	6
	2	7	10	5	0	0	0	7	1	6	4
	3	7	3	10	0	4	7	6	7	2	2
	4	8	2	5	9	1	2	4	4	8	0
	5	7	2	1	3	14	5	4	4	3	6
	6	7	2	3	2	4	20	4	8	1	3
	7	8	2	4	3	6	4	18	1	4	4
	8	9	4	8	1	7	0	6	9	2	6
	9	8	4	4	4	6	1	3	3	13	3
	10	11	0	4	6	2	4	3	0	0	17

Overall Accuracy =
(15+10+10+0+14+20+18+9+13+17)/400=135/400 =0.3375

Accuracy in percentage (%) = Overall Accuracy*100
=0.3375*100 = 33.75%

Table IV SURF-SVM Performance Metrics

Sign	Precision	Recall	F1-score
1	0.375	0.172414	0.236220472
2	0.25	0.30303	0.273972603
3	0.25	0.208333	0.227272727
4	0.225	0.28125	0.25
5	0.1	0.27451	0.146596859
6	0.5	0.465116	0.481927711
7	0.45	0.327273	0.378947368
8	0.225	0.243243	0.233766234
9	0.325	0.333333	0.329113924
10	0.425	0.333333	0.373626374

B4.3 BAG with SVM approach

Table V Confusion matrix bag with SVM

	Target class										
Predicted class	Sign	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	1	32	2	3	0	0	0	0	0	0	3
	2	1	34	1	0	1	0	0	0	0	3
	3	6	2	32	0	0	0	0	0	0	0
	4	0	7	2	34	0	0	0	0	0	1
	5	0	0	6	2	32	0	0	0	0	2
	6	0	0	2	1	3	34	0	0	0	0

Retrieval Number: B2142078219/2019©BEIESP
DOI: 10.35940/ijrte.B2142.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

	7	1	0	0	0	1	2	32	0	2	1
	8	0	0	0	0	2	3	0	31	3	1
	9	0	1	0	0	2	1	2	2	32	0
	10	1	0	1	3	2	0	0	0	0	33

Overall Accuracy = (32+34+32+34+32+34+32+31+32+33)/400=326/400 =0.815

Accuracy in percentage (%) = Overall Accuracy*100 = 0.815*100 = 81.5%

Table VI BAG-SVM Performance Metrics

Sign	Precision	Recall	F1-score
1	0.8	0.780488	0.790123457
2	0.85	0.73913	0.790697674
3	0.8	0.680851	0.735632184
4	0.85	0.85	0.85
5	0.8	0.744186	0.771084337
6	0.85	0.85	0.85
7	0.8	0.941176	0.864864865
8	0.775	0.939394	0.849315068
9	0.8	0.864865	0.831168831
10	0.825	0.75	0.785714286

B4.4 Deep Lerning Using Alexnet With Svm

Table VII Confusion Matrix pretrained Alexnet with SVM

	Target class										
	Sign	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Predicted class	**1**	38	0	2	0	0	0	0	0	0	0
	2	0	37	0	3	0	0	0	0	0	0
	3	0	0	35	0	0	0	2	2	0	1
	4	0	0	0	38	0	0	2	4	3	2
	5	0	0	0	0	37	0	1	1	0	1
	6	0	2	0	0	4	34	0	0	0	4
	7	0	0	1	0	1	0	39	1	0	0
	8	0	4	0	1	2	0	0	35	3	0
	9	0	0	0	0	0	0	3	5	34	0
	10	0	0	0	0	0	2	1	2	0	35

Overall Accuracy = 38+37+35+38+37+34+39+35+34+35)/400 =362/400 =0.905

Accuracy in percentage (%) = Overall Accuracy*100 =0.905*100 = 90.5%

Table VIII DL-SVM Performance Metrics

Sign	Precision	Recall	F1-score
1	0.95	1	0.948717949
2	0.925	0.860465	0.891566265
3	0.875	0.921053	0.897435897
4	0.95	0.904762	0.926829268
5	0.925	0.840909	0.880952381
6	0.85	0.944444	0.894736842
7	0.975	0.8125	0.886363636
8	0.875	0.7	0.777777778
9	0.85	0.85	0.85
10	0.875	0.813953	0.843373494

B4.6 Analysis based on all methodologies

Here we have shown graphical representation of performance metrics with all approaches. Graphical view is shown that deep learning pretrained Alexnet model approach achieves almost 85% above result with all performance metrics.

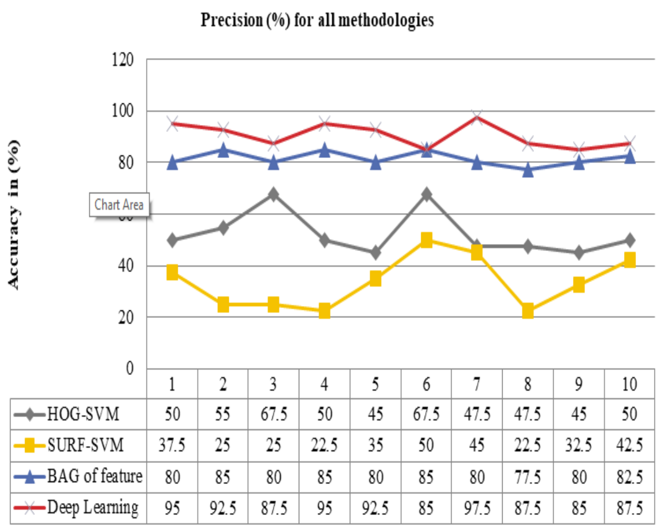


Fig 4. Summary chart for Precision with different methodologies

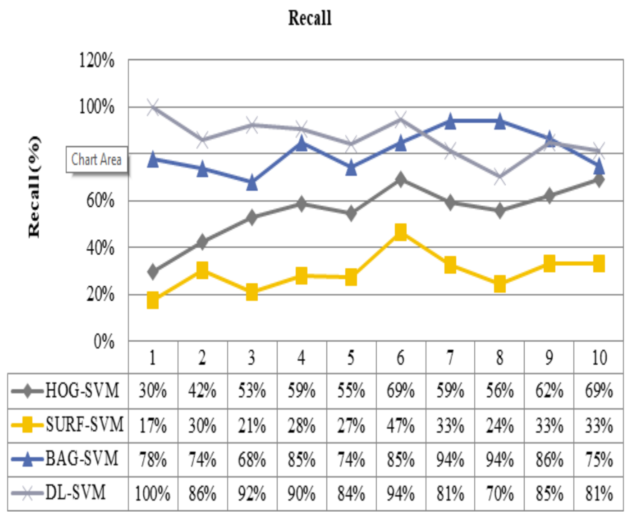
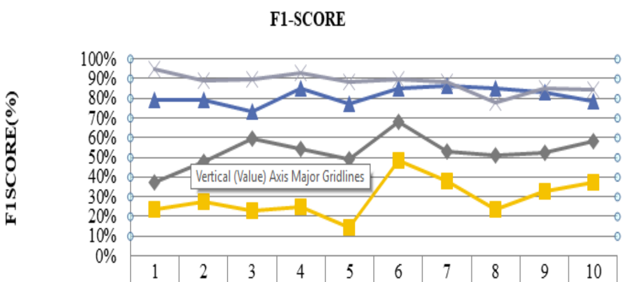


Fig 5. Recall (%) for all methodologies



	1	2	3	4	5	6	7	8	9	10
HOG-SVM	37%	48%	59%	54%	49%	68%	53%	51%	52%	58%
SURF-SVM	24%	27%	23%	25%	15%	48%	38%	23%	33%	37%
BAG-SVM	79%	79%	74%	85%	77%	85%	86%	85%	83%	79%
DL-SVM	95%	89%	90%	93%	88%	89%	89%	78%	85%	84%

Fig 6. F1-score (%) for all methodologies

V. CONCLUSION

Enhancement in accuracy is found with pre-trained alexnet CNN model compared to other methods such as HOG, SURF & bag of feature extraction methods.

We have calculated Precision, Recall and F1-score matrices for finding out performance of our model. We have achieved almost 85 % and above accuracy with all performance matrices. This work can be applicable to all pre-trained deep learning CNN models with the aim of improvement in future. Enhancement of datasets will occur with alphabet sign and daily conversation videos as we have used only number sign. This work can be extended by using 3D CNN model.

REFERENCES

1. M. A. Jalal, R. Chen, R. K. Moore, and L. Mihaylova, "American Sign Language Posture Understanding with Deep Neural Networks," in 21st International Conference on Information Fusion, FUSION 2018, 2018, pp. 573–579.
2. S. Gattupalli, A. Ghaderi, and V. Athitsos, "Evaluation of Deep Learning based Pose Estimation for Sign Language Recognition," in Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments, 2016, p. 12.
3. A. Ghotkar and G. Kharate, "Dynamic Hand Gesture Recognition for Sign Words and Novel Sentence Interpretation Algorithm for Indian Sign Language Using Microsoft Kinect Sensor," J. Pattern Recognit. Res., vol. 10, no. 1, pp. 24–38, 2015.
4. Y. Li and G. Su, "Simplified histograms of oriented gradient features extraction algorithm for the hardware implementation," in International Conference on Computers, Communications and Systems, ICCCS 2015, 2015, pp. 192–195.
5. K. M. Lim, A. W. C. Tan, and S. C. Tan, "Block-based histogram of optical flow for isolated sign language recognition," J. Vis. Commun. Image Represent., vol. 40, no. Part B, 2016.
6. A. Singh, R. Kumar, and R. P. Tripathi, "Study and Analysis of Different Kinds of Data Samples of Vehicles for Classification by Bag of Feature Technique," in Soft Computing: Theories and Applications, Springer Singapore, 2018, pp. 53–61.
7. N. Nida, M. Yousaf, A. Irtaza, and S. Velastin, "Bag of Deep Features for Instructor Activity Recognition in Lecture Room," in International Conference on Multimidea Modelling(MMM), 2019, vol. 281, pp. 481–492.
8. H. Mittal and M. Saraswat, "Classification of Histopathological Images Through Bag-of-Visual-Words and Gravitational Search Algorithm," in Soft Computing for Problem Solving, vol. 817, Springer Singapore, 2019, pp. 231–241.
9. H. G. LeCun Yann, Bengio Yoshua, "Deep learning," Nat. Methods, vol. 13, no. 1, p. 35, 2015.
10. J. Schmidhuber, "Deep Learning in neural networks: An overview," Neural Networks, vol. 61, pp. 85–117, 2015.
11. P. J. Jino, K. Balakrishnan, and U. Bhattacharya, "Offline Handwritten Malayalam Word Recognition Using a Deep Architecture," in Soft Computing for Problem Solving, vol. 816, Springer Singapore, 2019, pp. 913–925.
12. Y. Chen, "Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolution Nerual NetworkC," IEEE Trans. Geosci. Remote Sens., vol. 54, no. 10, pp. 6232–6251, 2016.
13. H. G. Krizhevsky Alex , Sutskever Ilya , "ImageNet Classification with Deep Convolutional Neural Networks," in International conference on Advances in neural information processing systems, 2012, pp. 1097–1105.
14. S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," in 3rd IAPR Asian Conference on Pattern Recognition, ACPR 2015, 2016, pp. 730–734.
15. R. A. Szegedy Christian , Liu Wei , Jia Yangqing , Sermanet Pierre, Reed Scott , Anguelov Dragomir, Erhan Dumitru, Vanhoucke Vincent, "Going Deeper with Convolutions," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 95–111.
16. K. H. B, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in Lecture Notes in Computer Science (Computer Vision – ECCV), vol. 9908, Springer International Publishing AG, 2016, pp. 630–645.
17. K. J. Dangarwala and D. Hiran, "a Research Gap on Automatic Indian Sign Language Recognition Based on Hand Gesture Datasets and Abstract :," Int. J. Comput. Eng. Appl., vol. XII, no. Iii, pp. 46–54, 2018.
18. O. K. Oyedotun and A. Khashman, "Deep learning in vision-based static hand gesture recognition," Neural Comput. Appl., vol. 28, no. 12, pp. 3941–3951, 2017.
19. A. O. Tang, K. Lu, Y. Wang, J. Huang, and H. Li, "A Real-time Hand Posture Recognition System Using Deep Neural Networks," ACM Trans. Intell. Syst. Technol., vol. 9, no. 4, p. 39:1-39:23, 2013.
20. T. Guo, J. Dong, H. Li, and Y. Gao, "Simple convolutional neural network on image classification," 2017 IEEE 2nd Int. Conf. Big Data Anal. ICBDA 2017, pp. 721–724, 2017.
21. O. Koller, S. Zargaran, H. Ney, and R. Bowden, "Deep Sign: Hybrid CNN-HMM for Continuous Sign Language Recognition," in Proceedings of the British Machine Vision Conference (BMVC), 2017, p. 136.1-136.12.
22. M. G Alaslani and L. A. Elrefaei, "Convolutional Neural Network Based Feature Extraction for IRIS Recognition," Int. J. Comput. Sci. Inf. Technol., vol. 10, no. 2, pp. 65–78, 2018.
23. A. Sivaramakrishnan et al., "Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification," Remote Sens., vol. 9, no. 8, p. 848(1-22), 2017.
24. P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree, and J. Kautz, "Online Detection and Classification of Dynamic Hand Gestures with Recurrent 3D Convolutional Neural Networks," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, vol. 2016–Decem, pp. 4207–4215.
25. S. Rajaraman et al., "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images," PeerJ, vol. 6, p. e4568, 2018.
26. Y. Gong, L. Wang, R. Guo, and S. Lazebnik, "Multi-Scale Orderless Pooling of," in European Conference on Computer Vision(Eccv2014), 2014, pp. 1–17.
27. B. B. Swapnali and K. S. Vijay, "Feature Extraction Using Surf Algorithm for Object Recognition," Int. J. Tech. Res. Appl., vol. 2, no. 4, pp. 2320–8163, 2014.
28. B. Gupta, P. Shukla, and A. Mittal, "K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion," 2016 Int. Conf. Comput. Commun. Informatics, ICCCI 2016, pp. 7–11, 2016.
29. A. Chaturvedi and N. Sethi, "Automatic License Plate Recognition System using SURF Features and RBF Neural Network," Int. J. Comput. Appl., vol. 70, no. 27, pp. 37–41, 2013.
30. D. LOWE, "Distinctive Image Features from Scale-Invariant Keypoints," Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, 2004.

AUTHORS PROFILE

Kruti Dangarwala has completed her bachelor degree computer engineering from M.S. University, Baroda India in 2001. She has completed M.E. Computer Engg from S.P University, Vallabh vidhyanagar, India in 2006. Currently she is pursing Ph.D from pacific university, Rajasthan India. She is presently working as associate professor of computer Engg. Department since 2003. in Shri s'ad vidya mandal Institute of technology Bharuch, India. She has published more than 20 papers in international journals and conferences. She has specialization in computer graphics, compiler design, Image processing and computer vision.

Dr. Dilendra Hiran PH.D in Computer Science from Pacific University Udaipur in 2015. He completed his master of science in Mathematics and Computer Science 1999. Bachelor of Science in Maths from MLSU, Udaipur 1994. He is currently working as Principal, Faculty of Computer Application, Pacific University Udaipur.