FULL PAPER

Macromolecular Symposia

www.ms-journal.de

Indian Sign Language Recognition Using Machine Learning Techniques

Ashok Kumar Sahoo

Automatic conversion of sign language to text or speech is indeed helpful for interaction between deaf or mute people with people who even do not have knowledge of sign language. This is the demand of current times to develop an automatic system to convert ISL signs to normal text and vice versa. This will be beneficial for both communities to express their fillings to one another in accessing publicly available facilities like ticketing, banking services, traffic signals, etc. A new feature extraction and selection technique using structural features and some of the best available classifiers are proposed to recognize ISL signs for better communication for computer-human interface. This paper narrates a system for automatic recognition of ISL immobile numeric signs, in which a standard digital camera was only used to acquire the signs, no wearable devices are required to capture electrical signals. The system is intended to convert isolated digit signs into text, that is, each entered sign image should contain precisely one numeric sign. To recognize ISL sign images in real time environment, a sign database containing ISL digits is created which contains 5000 images, 500 images for each numeral sign (0-9). Among two classifiers, the k-Nearest Neighbor outperforms the Naive Bayes classifier in stipulations of classification accuracy.

1. Introduction

To convey messages to unblessed community with hearing loss and to receive communication from them, sign language[1,2] is used. It primarily uses manual communication and body movements to convey information. A sign language contains a set of structured gestures, which are used habitually as a medium of communication among hearing weekend and hard hearing community. Sign language gestures involve movements of hands and body parts (also includes facial expressions). Different gesture has its own meaning as compared to normal languages. Sign language is a complete language in its own which has its syntax and grammar. Sign language is different for each country/subcontinent. Many sign languages developed for the deaf people but only few are having legal recognition. A lot of work has already been conducted by researchers for recognizing American (ASL),[3] Japanese (JSL),[4] and Australian (Auslan)[5] sign languages. The sign language of India and Indian subcontinent is often referred as Indian Sign Language (or ISL).[6] Most of the elements of ISL are derived from British Sign Language (BSL).[7] Only Devanagari-based finger spelling system is isolated from BSL. Sign Language signals can be categorized into two types: dynamic signs and static signs. Static signs have fixed posture of hand, whereas dynamic signs have maneuver of hands, body parts also including some kind of facial expressions.

Gesture recognition process involves two approaches: device based and vision based. In pattern recognition, the vision-based sign language recognition is generally used. The recognition of sign language is not an easy task; hence, a system for recognition of sign language is the demand of current times. In India, as of date, no automatic translator or machine is available to convert sign language into text or speech.

A. K. Sahoo
Professor, Department of Computer Science & Engineering
Graphic Era Hill University
Dehradun, Uttarakhand 248006, India
E-mail: ashoksahoo2000@yahoo.com

DOI: 10.1002/masy.202000241

2. Related Work

Subha Rajam, P., Balakrishnan, G., et al.[8] in their work, proposed a new method for all 32 combinations of fingers ($2^5 = 32$), where down fingers represented by "0" and UP fingers are represented by "1." The up and down positions of fingers are identified by canny edge detection technique.[9] They used a new idea for extracting features from sign images. Images are scanned from left to right and also from right to left to locate the finger tips and also to locate the bottom of the palm. The up position of a finger can be identified by the help of "least" height. From five fingers which represent a sign, they extract the binary equivalent of finger positions. Two set of databases are used in experiments: one set is without angular movements and other is with angular movements. This information is used for classification of sign images. The experimental result reported for Tamil signs was 96.87% without angular movements and 98.75% with angular movements.

An ISL system was developed by Deora and Bajaj [10] for 25 English characters and nine single digits. The users of the proposed system need to wear red and blue gloves. The researchers applied figure tip and segmentation algorithms for extracting features and experimented on Principal Component Analysis (PCA) to classify ISL signs. They have achieved an accuracy of 94% as reported in their paper.

In their proposed ISL recognition system, Rekha et al.[11] consider static and dynamic ISL double handed character signs.

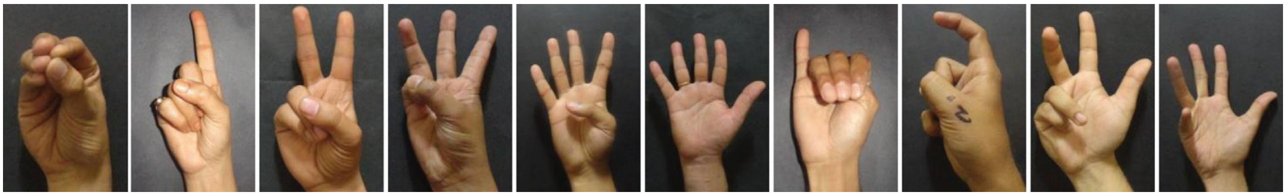
(0) (1) (2) (3) (4) (5) (6) (7) (8) (9)

Figure 1. ISL static gestures for isolated numerals.

Training samples were collected from 40 persons with 22 static ISL gestures and 22 dynamic gestures. With help of principle curvature-based region detector, shape features were extracted from input signs, wavelet packet decomposition technique has been used to extract texture features from signs, and complexity defects algorithm has been used to extract feature vectors from fingers. For classifying input signs, three different classification techniques were used, namely, multiclass-Support Vector Machine, Dynamic Time Wrapping, and k-Nearest Neighbor. The accuracy rates reported by researchers are 91.3% for steady signs and 86.3% for video signs.

Ghotkar et al.[12] have experimented their work on ISL one-handed and two-handed characters. They have used four different modules towards recognition of ISL characters. First module was on a system to track hand positions by help of Continuously Adaptive Mean Shift Algorithm, second module was segmentation of hand with Hue-Saturation-Value color model in association with neural network, the third module used was for feature extraction using Generic Fourier Descriptor, and finally the last module was used for gesture recognition where genetic algorithms were used to classification of input ISL signs. In the entire work, the authors not reported anything about dataset.

In a limited experimental environment, the authors, Goyal et al.[13] developed an innovative model for ISL recognition for only eight one-handed characters. The authors used scale invariant feature transform technique to extricate features from input signs and key point matching technique for recognition of characters. The recognition accuracy reported by the research is 95%.

3. Getting Started

As discussed above, the works described by various researchers are not complete in terms of dataset, feature extraction techniques, and classification techniques used in conduction of experiments. It is reported and admitted by the authors that they have developed some of the experimental data for their own and conducted research on those datasets only. Not a single author claims to have dataset in terms of adequate number of images and in terms of complete set of ISL alphabets, digits, etc. The experiments were conducted on extreme lab conditions and no consideration for factors like, noise induced, differing in illumination, etc. There are also a number of other feature extraction techniques, classification techniques that may produce higher results and can be used for experimentation. Also, the authors did not discuss any algorithmic aspects of their experiments. Most of the current applications are based on mobile apps, this factor cannot be neglected. Considering these factors, some experiments are adopted in this paper. The process of pattern recognition is followed strictly in conduction of experiments, like, dataset collection, pre-processing of input images, extraction of useful information from input ISL signs, classifier selection. At the end results, conclusion and future research direction in this area is discussed in detail.

As usual, the process of pattern recognition is followed in conduction of experiments. The following steps are strictly followed in experimenting with digit ISL signs. First, ISL sign digit images are collected with help of a digital camera. Then, the steps of image pre-processing, feature extraction, application of classifiers, and finally analysis of results were carried out in this research.

4. Database Collection/Image Acquisition

The major concern as described in this paper is to develop a standard database for ISL gestures also. A standard database is created, which is used in proposed ISL recognition system and in future researches also. A standard database for 10 static signs of numerals (**Figure 1**), ISL alphabets (single-handed and double-handed signs representing A-Z), and a limited number of dynamic signs in the form of words has been created. In this paper, only numeric signs are taken into the consideration. A total of 5000 static signs were captured from 100 different signers. Each signer contributed 50 signs, with a repetition factor of five signs for each digit. The sign database which is collected also have mixed variations, such as age (18 years up to 50 years), gender (male or female), and class of signer (actual deaf people or normal hearing people).

To simplify the process of feature extraction and classification, the images are captured in uniform and dark background. Gradually, the selection of different types of background will be included in subsequent research. A standard digital camera with resolution of 16 megapixel and 10X optical zoom is used in capturing images. Although a standard digital camera is sufficient for taking images for research purpose, flashlight has been used in capturing ISL signs. The images are accumulated in “JPEG” format, as it is considered as a usual format, requires less storage space, and providing high-quality images for processing. Original images are of 5.5 MB size and for processing these signs are resized to 200×300 pixels and need only 25 KB storage space. This is a requirement for conduction of experiments because in the experiments 5000 images are used, with high resolution images it is nearly impossible to carry out research due to limitations of primary memory and processing power. Images are captured from 100 persons, 69 persons are male and rest is female. The dataset contains only digit signs (0–9) and hence 500 signs per digit is available for research. The dataset is disjoined into training and testing sets in the proportion of 70% and 30% separately.

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	1	1	0	0	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(a)

Left Image | Right Image

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	1	1	1	1	0	0	0	0	0	0	1	1	1	0
3	0	0	0	0	1	1	0	0	0	0	0	0	1	1	1	1
4	0	0	0	1	1	1	1	0	0	0	0	0	1	1	1	1
5	0	0	0	1	1	1	1	0	0	0	0	0	1	1	1	1
6	0	0	0	0	1	1	1	1	0	0	0	1	1	1	1	0
7	0	0	0	0	1	1	1	1	0	0	0	1	1	1	1	0
8	0	0	0	0	0	1	0	0	0	1	1	1	1	1	0	0
9	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
11	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
12	1	1	1	0	0	0	1	1	1	0	1	1	1	1	0	0
13	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

(b)

2	**5**	**7**	**7**	**11**	**10**	**8**	**10**	**4**	**4**	**5**	**7**	**11**	**11**	**9**	P_0

(c)

0	**2**	**0**	**5**	**0**	**7**	**0**	**7**	**0**	**11**	**0**	**10**	**0**	**8**	**0**	**10**	**0**	**4**	**0**	**4**	**0**	**5**	**0**	**7**	**0**	**11**	**0**	**11**	**0**	**9**	**0**	**3**	P_1

(d)

Figure 2. a) The black and white resized (20 × 30) image of ISL static gesture "2." b) The pre-processed image. c) The vertical projection P_0, 2. d) creation of P_1 from P_0.

4.1. Image Pre-Processing

After size reduction of dataset, each individual image is processed through a step called image pre-processing. In this phase, the input images are transformed into monochrome images and further reduced in size of 20 × 30 pixels as shown in **Figure 2**a. Important statistical features are extracted from the reduced images, which in fact is a matrix containing binary values. The unwanted information still present in the feature vector are also removed by deletion of rows/columns from the matrix containing only background information, as shown in Figure 2b.

4.2. Feature Extraction

Feature extraction in image processing is a technique to identify, select and capture important information present in the dataset. Structural features[14] are extracted from input images in the proposed ISL recognition system. The detail of feature extraction process in the experiments is presented below.

Consider an image is presented by a tuple (a, b) and contains only binary information, “1” for foreground pixels and ‘0’ for background pixels. Let a_{max} and b_{max} are the width and height of an input image as shown in Figure 2b. The structural method segregates the image into two different sub-images horizontally (left and right sub-images) based on the summation of foreground pixel values. Similarly, the image also divided into two different sub-images vertically. The combination of vertical and horizontal division results into four sub-images.

The vertical and horizontal division of images can be either forming two disjoint sub-images or sub-images share the pixels of common division row or column based on the number of foreground pixels. The intersection of horizontal line and vertical line, whether synchronous or asynchronous, is referred as common area (CA). The process is repeated recursively up to seventh level to drag out more intersection points. The collection of all points constitutes the feature vector representing the image for processing.

The point of intersection of an input image (known as common area, or CA) is calculated as follows.

Let X_{max} is the number of columns of the image matrix as shown in Figure 2c containing at least one non-zero element. Similarly, let P_0 is the vertical projection elements in an array. The value stored in P_0 (in a position of the array) is the sum of binary information in that particular column of the image array. Another array, P_1, is created which contains two times the elements of P_0. The elements of P_0 are also part of P_1, but before and after each element of P_0, zeros are inserted in P_1 which can be seen in Figure 2d. The element (a_q) which divides the image matrix vertically is identified. The coordinate (a_0) is calculated as $a_q/2$ and is part of coordinate (a_0, b_0). Similarly, the coordinate point (b_0) is calculated from the element (b_q).

The array is divided vertically on the basis of the value of (a_0), and is divided horizontally on the basis of (b_0). Depending upon the values of a_q and b_q, it will be decided that whether the image array is divided as separate images or share the division line among them. The points (a_0, b_0) constitute the zero-level granularity. As explained earlier, the process is repeated recursively for seven levels and a number of such points (a_i, b_i) are extracted from the image array. The value of i is in the range of 0–7.

Suppose T is the current level of granularity, total number of sub-images possible at this level is given by the formula $4^{(T+1)}$. The number of common areas at this level is equals to $4 \times T$ depending upon the value of T, the coordinate points (a_i, b_i) constitute the feature vector of the image. The number of elements of feature vector at level T is $8 \times T$ and is normalized in the range [0, 1] which is obtained by dividing the original elements of feature vector by a factor N, where N is the maximum value present in the feature vector.

4.3. Classification

Classification[1,15] groups input vectors into separate classes based on the theory of probability or distance measures. In the proposed research, 10 different classes are involved. As explained earlier, out of 5000 input images, feature vectors extracted from 3500 images are provided to classifier as training samples. By help of these training samples, the classifiers have developed some model for prediction. The rest of the feature vectors extracted from 1500 images are used as testing samples. Based on the developed model and testing samples, the classifiers are able to predict the class labels of testing vectors. From the confusion matrix produced by the classifiers as a result of prediction, one can determine various metrics, like, recognition rate, precision, specificity, sensitivity, and F1_Score. From these metrics one can determine the efficiency of classifiers. In conducting experiments, two classifiers have been used, namely, Naive Bayes, and k-Nearest Neighbour.

4.3.1. The Naive Bayes Classifier

The classification task performed by this classifier[1,16] is based on supervised learning technique and theory of probability. Its performance is at par with other popular classifiers, like, neural network and support vector machine, etc. This classifier works on input vectors, say for instance X, which is expressed as a combination of many attributes and the target function f(x) is selecting a subset of values from the input vector. A dataset containing training vectors are given to the classifier and for testing a new instance is inputted to it. The attribute values of the testing instance is (a_1, a_2...a_n). The learner algorithm of the classifier predicts the target value of the testing instance based on the concept of prior, posterior, likelihood factors of probability theory.

Let V_{map} is the most probable target for the attribute set (a_1, a_2...a_n), therefore

$$V_{MAP} = \arg\max_{\nu_j \in V} P(\nu_j | \mathrm{a}_1, \mathrm{a}_2 \ldots \ldots . \mathrm{a}_n) \tag{1}$$

Using help of Bayes theorem, the Equation (1) can be rewritten as

$$V_{MAP} = \arg\max_{\nu_j \in V} \frac{P(\nu_j | \mathrm{a}_1, \mathrm{a}_2, \ldots \ldots ., \mathrm{a}_n)\mathrm{P}(\mathrm{v}_j)}{P\left(a_1,\ a_2, \ldots ., a_n\right)} \tag{2}$$

$$= \arg\max_{\mathrm{v}_j \in \mathrm{V}} \frac{\mathrm{P}(\mathrm{v}|\mathrm{a}_1,\mathrm{a}_2 \ldots \ldots \mathrm{a}_n|\mathrm{v}_j)\mathrm{P}\left(\mathrm{v}_j\right)}{\mathrm{P}\left(\mathrm{a}_1,\mathrm{a}_2, \ldots .,\mathrm{a}_n\right)}$$

$$= \arg\max_{\nu_j \in V} P\left(a_1,\ a_2,\ \ldots .a_n | \nu_j\right)\ P\left(\nu_j\right) \tag{3}$$

The terms of Equation (3) are dependent on training samples. Estimation of values of $P(v_j)$ are dependent on the frequencies of target values (ν_j) present in the training samples. Assessment of various $P(a_l,\ a_2 ... a_n | \nu_j)$ terms is infeasible if a small training samples is provided.

Given target value of an instance, the probability of discovering the attribute conjunction (a_1, a_2...a_n) is nothing but the product of the probabilities of independent attributes. Therefore, one can write, P ($a_1 a_2 \ldots \ldots . a_n$ vj) = $\prod_i P(a_i|\nu_j)$.

Now the Equation (3) becomes,

$$\text{Naive Bayes classifier (VNB)} = \underset{\nu_j \in V}{\mathrm{argmax}}\, P\left(\nu_j\right) \prod_i P(a_i|\nu_j) \tag{4}$$

The variable V_{NB} is the target value of the input testing samples as predicted by Naive Bayes classifier. In this classification task, estimation of differing $P(a_i|v_j)$ terms must be from training samples and is the count of distinct attribute vales multiplied by distinct number of target vales.

The learning step in Naive Bayes classifier is nothing but only the estimation of $P(v_j)$ and $P(a_i\ |v_j)$ which is available in training samples. This leads to a learning hypothesis and it is the primary source of classification by help of the Equation (4).

4.3.2. k-Nearest Neighbour (k-NN) Classifier

This classification technique[17–19] is based upon the concept of distance metric and known as one of the instance-based technique. The training vectors are belongs to points in a n-dimensional space R^n. Consider x is an instance of input vector to the algorithm and, $x = \{a_1(x), a_2(x), ..., a_n(x)\}$ where $a_k(x)$ is the k-th attribute of x.

The distance between two testing vectors x_i and x_j is given as $d(x_i, x_j)$, is given by the Equation (5) as below

$$d\ \left(x_i, x_j\right) = \sqrt{\sum_{r=1}^{n} \left(a_r(x_i\ - x_j\ )\right)^2} \tag{5}$$

The target function is dependent upon the number of outputs or targets and hence can be discrete or real valued. It is of the form, $f: R^n \rightarrow V$, the set V is the finite set containing the target values $\{v_1, v_2, ...,v_s\}$.An algorithm for k-NN is given below, the value of $f(x_q)$ is returned by the algorithm which is the most common value of the function "f" among all k training samples very near to x_q. The algorithm can assign most appropriate value of nearest neighbor training samples if k is very large. In the proposed experimentation, the value of k is 10.

Training algorithm:

- For training sample (x, f (x)), append the sample to the list "*training_samples*"

Classification algorithm:
Given a query test vector x_q inputted for classification,

- Let x_l ...x_k represent the k training samples from training samples and are nearest to x_q
- Return

$$\hat{f}\ \left(x_q\right) \leftarrow \arg\max_{v \in V} \sum_{i\ =\ 1}^{k} \delta(v, f\ \left(x_i\right) \tag{6}$$

Where, $\delta(\mathrm{a},\ \mathrm{b})\ = \begin{cases} 1, & if\ a = b \\ 0, & otherwise \end{cases}$

The k-NN algorithm never forms an obvious rough hypothesis $\hat{f}$ for the target function f. It only predicts for each new query instance as and when required.

The *training_samples* forms a form of convex polyhedra in the classification space for larger values of k. This is also known as Voronoi diagram. If the value of k = 1, only two classes involved, then the classification space is divided into two areas. One for a particular class and rest area is for the second class. Similarly for larger value of k, the classification space is divided among all classes. For a particular query point, x_q, its target function, $\hat{f}(x_q)$ will be computed, the target function will fit in a particular area in the classification space and hence belongs to a particular class. In this way, all query points are classified.

Table 1. Results obtained from Naive Bayes Classifier.

ISL Sign	TP	TN	FP	FN	Accuracy	Precision	Sensitivity	Specificity	F1_Score
0	144	1352	2	2	99.73%	98.63%	98.63%	99.85%	98.63%
1	140	1346	7	7	99.07%	95.24%	95.24%	99.48%	95.24%
2	138	1344	12	6	98.80%	92.00%	95.83%	99.12%	93.88%
3	132	1318	18	32	96.67%	88.00%	80.49%	98.65%	84.08%
4	114	1346	36	4	97.33%	76.00%	96.61%	97.40%	85.07%
5	117	1350	33	0	97.80%	78.00%	100.00%	97.61%	87.64%
6	140	1354	2	4	99.60%	98.59%	97.22%	99.85%	97.90%
7	140	1353	2	5	99.53%	98.59%	96.55%	99.85%	97.56%
8	126	1293	24	57	94.60%	84.00%	68.85%	98.18%	75.68%
9	120	1301	30	49	94.73%	80.00%	71.01%	97.75%	75.24%
Average					97.79%	88.91%	90.04%	99.78%	89.09%

Table 2. Results obtained from k-NN Classifier.

ISL Sign	TP	TN	FP	FN	Accuracy	Precision	Sensitivity	Specificity	F1_Score
0	140	1344	10	6	98.93%	93.33%	95.89%	99.26%	94.59%
1	140	1350	1	9	99.33%	99.29%	93.96%	99.93%	96.55%
2	139	1340	11	10	98.60%	92.67%	93.29%	99.19%	92.98%
3	132	1334	18	16	97.73%	88.00%	89.19%	98.67%	88.59%
4	139	1349	11	1	99.20%	92.67%	99.29%	99.19%	95.86%
5	139	1324	11	26	97.53%	92.67%	84.24%	99.18%	88.25%
6	140	1350	5	5	99.33%	96.55%	96.55%	99.63%	96.55%
7	142	1346	8	4	99.20%	94.67%	97.26%	99.41%	95.95%
8	145	1320	17	18	97.67%	89.51%	88.96%	98.73%	89.23%
9	119	1322	31	28	96.07%	79.33%	80.95%	97.71%	80.13%
				Average	98.36%	91.87%	91.96%	99.09%	91.87%

5. Results

As described in previous sections, experiments are conducted on hierarchical centroid feature extraction technique in association with Naive Bayes and k-NN classifiers. Results obtained from both classifiers in the form of confusion matrix are presented in **Tables 1 and 2**. The performance of classifiers in terms of various measures are shown in these tables, like, true positive (TP), true negative (TN), false positive (FP), and false negative (FN). From these measures, various metrics for analyzing the results, like, accuracy, precision, sensitivity, specificity and F1_Score for all ISL digits are also presented in Tables 1 and 2.

The accuracy is defined as the ratio of (TP + TN) to (TP + TN + FP + FN). Accuracy is a statistical measurement used in classification task to test whether the classifier correctly identifies or it conditionally excludes. It often provides the efficiency of the classifier, but it also hides some other relevant information. The

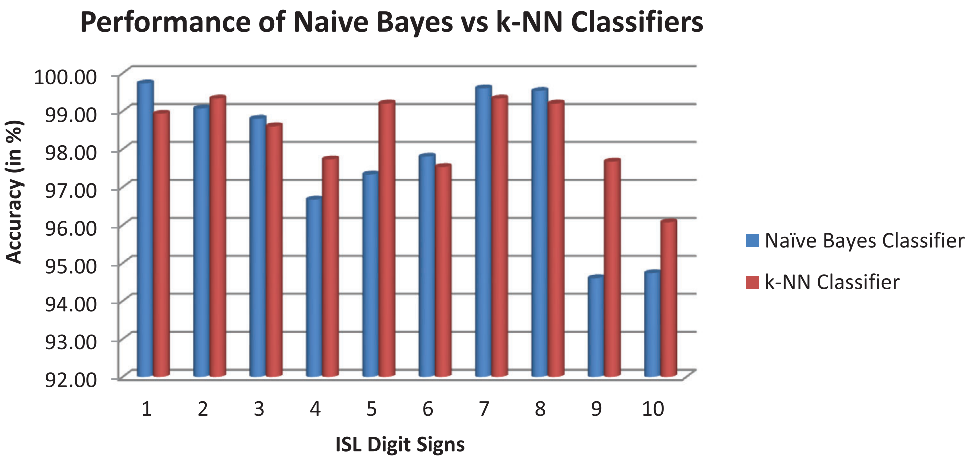

Figure 3. Performance of Naive-Bayes versus k-NN Classifiers.

Table 3. Comparison of Accuracy Rates (in %) of Naive Bayes and k-NN Classifiers.

ISL Sign	0	1	2	3	4	5	6	7	8	9
Naïve Bayes	99.73	99.07	98.80	96.67	97.33	97.80	99.60	99.53	94.60	94.73
k-NN	98.93	99.33	98.60	97.73	99.20	97.53	99.33	99.20	97.67	96.07

second measurement is precision and is defined as the ratio of (TP) to (TP + FP). It is a quality measurement and high precision indicates the classifier returns more relevant results than irreverent results. The third measurement is sensitivity or recall and is defined as the ratio of (TP) to (TP + FN). It is reliable when the result is negative whereas the sensitivity rate is high. Next measurement is specificity and is defined as the ratio of (TN) to (TN+FP). It specifies that if the specificity rate is high then the chance for type- I error, in statistics, is very low. The final measurement is F1_Score and is defined as the two multiplied by the ratio of (Precision × Sensitivity) to (Precision + Sensitivity). It is the harmonic mean of sensitivity and precision, its perfect value should be 1.0. It is helpful in analyzing results when uneven distribution of dataset per class is present, and in this case, its significance is more than accuracy. However, in the experiments on ISL digits, the dataset is distributed evenly.

Comparison of results of both classifiers is presented in Table 3. ISL digit wise result comparison graphically is shown in **Figure 3**. The result produced by k-NN classifier is higher than Naive Bayes classifier in terms of recognition accuracy. The average accuracy rate of k-NN is 98.36% whereas that of Naive Bayes is 97.79%. It is marginal victory for k-NN classifier. As one can verify that all metrics in Tables 1 and 2 produces good results for both of the classifiers. This is a very positive sign in the research direction of ISL recognition. This is possible due to the contribution of hierarchical centroid feature vector technique. In the case of ISL digit classification, the positions of figures are very important. Therefore, with help of the proposed feature extraction technique, it is possible to extract various important features from the input images. The proposed feature extraction technique is able to distinguish clearly between all digits. Interestingly, it can be verified from the results that the minimum misclassification rate reported by classifiers is 94.60%.

In the literature review, researchers have reported a maximum of 97% recognition rates, however, in this research a marginal increment of about 1% in terms of recognition rate is achieved. Therefore, in all aspects of results obtained, the proposed method is better than what is presented in the literature. The results of proposed method are complete in all respects including dataset, feature extraction, classification, and analysis of results.

6. Conclusions and Future Work

The research presented in this paper is efficient in terms of recognition of ISL digits with utmost accuracy. The experimental results can be enhanced with some graphical user interface to recognize ISL digits, where the system will capture images of hand posture of users and display the text, which in fact, is the meaning of the hand gesture. If the user has contributed his signs to the system, the system is capable of prediction of 100% accurate results, as the system can learn while working. The researchers of this paper are working continuously to develop a system in which ISL digits, characters (single handed and double handed), words, and sentence-level images or videos in real time can be interpreted with high accuracy. In next phase of this research, the researchers working in this area can try various other feature extraction techniques, classifiers, even combination of different classifiers to develop a complete system to recognize ISL signs. Other techniques can be experimented in future for development of a complete system which can be truly termed as ISL recognition system.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

Indian sign language, k-NN classifier, naïve bayes classifier, structural features, support vector machine

[1] A. K. Sahoo, G. S. Mishra, K. K. Ravulakollu, *ARPN J. Eng. Appl. Sci.* **2014**, *9*, 116.
[2] U. Zeshan, *Sign Language in Indo-Pakistan: A Description of a Signed Language*, John Benjamins **2000**.
[3] C. Valli, C. Lucas, *Linguistics of American Sign Language: An introduction*, Gallaudet University Press **2000**.
[4] S.-I Ito, M. Ito, M. Fukumi, *International Journal of Advances in Intelligent Informatics* **2019**, *5*, 243.
[5] T. Johnston, A. Schembri, *Australian Sign Language: An Introduction to Sign Language Linguistics*, Cambridge University Press **2007**.
[6] A. K. Sahoo, G. S. Mishra, P. Ahmed, *International Journal of Computer Application* **2012**, *2*, 158.
[7] K. D. Rogers, M. Pilling, L. Davies, R. Belk, C. Nassimi-Green, A. Young, *Quality of Life Research* **2016**, *25*, 1825.
[8] P. S. Rajam, G. Balakrishnan, in IEEE International Conference on Communication Technology (ICCT) **2011**, pp.737-742.
[9] D. Sangeetha, P. Deepa, *Journal of Real-Time Image Processing 16*, 957, **2019**.
[10] D. Deora, N. Bajaj, in International Conference on Emerging Technology Trends in Electronics Communication and Networking **2012**, pp. 1-5.
[11] J. Rekha, J. Bhattacharya, S. Majumder, in International Conference on Trendz in Information Sciences and Computing **2011**, pp. 30–35.
[12] A. S. Ghotkar, R. Khatal, S. Khupase, S. Asati, M. Hadap, in International Conference on Computer Communication and Informatics **2012**, pp. 1–4.
[13] S. Goyal, I. Sharma, S. Sharma, *Int. J. Eng.* **2013**, *2*.
[14] G. Vamvakas, B. Gatos, S. J. Perantonis, *Pattern Recognit.* **2010**, *43*, 2807.
[15] T. M. Mitchell, *Machine Learning*, McGraw-Hill Science/Engineering/Math **1997**, pp. 177–178, 231–233.
[16] K. A. K. Niazi, W. Akhtar, H. A. Khan, Y. Yang, S. Athar, *Sol. Energy* **2019**, *190*, 34.
[17] A. K. Sahoo, K. K. Ravulakollu, *Int. J. Appl. Eng. Res.* **2014**, *9*, 7347.
[18] M. Sharma, R. Pal, A. K. Sahoo, *ARPN J. Eng. Appl. Sci.* **2014**, *9*, 1255.
[19] Mir Aman Sheheryar, P. K. Mishra, A. K. Sahoo, *J. Eng. Appl. Sci.* **2016**, *11*, 5800.

