Methodology

It is really a challenging issue to develop a practical handwritten character recognition (CR) system which can maintain high recognition accuracy. A generic character recognition system is shown in Fig. 1.

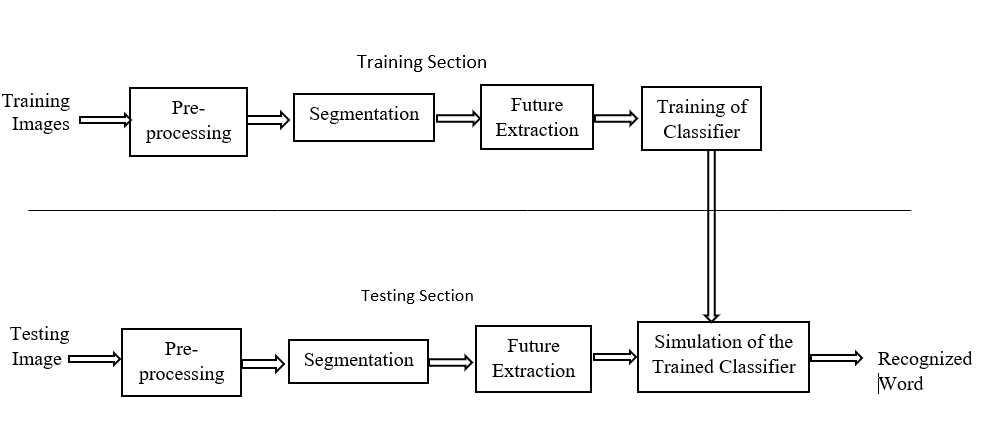
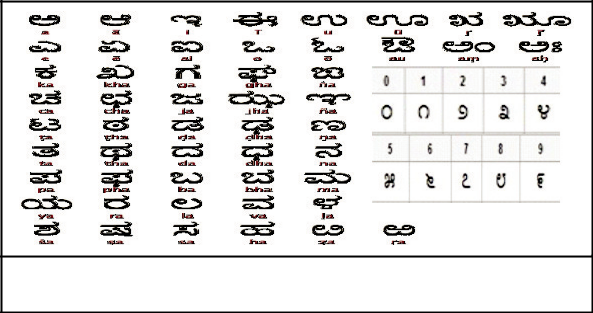


Fig 1: Generic CR System

In most of the existing systems recognition accuracy is heavily dependent on the quality of the input document. In handwritten text adjacent characters tend to be touched or overlapped. Therefore it is essential to segment a given string correctly into its character components. In most of the existing segmentation algorithms, human writing is evaluated empirically to deduce rules . But there is no guarantee for the optimum results of these heuristic rules in all styles of writing. Moreover handwriting varies from person to person and even for the same person it varies depending on mood, speed etc. This requires incorporating artiﬁcial neural networks, hidden Markov models and statistical classiﬁers to extract segmentation rules based on numerical data.

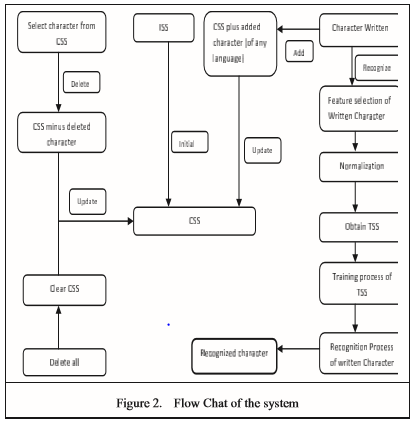
A precise definition of online HandWriting Recognition (HWR) is the task of transforming a language that is represented in its spatial form of graphical marks into its symbolic representation. This system includes datacollection, feature-selection and normalization, classification, training-sample creation and character recognition. In the proposed case the recognition is done for Kannada Characters. Kannada Language is extensively spoken in Southern parts of India. It has 49 characters and 10 numerals having 13 swaras(vowels), 34 vyanjanas(consonants) and 2 yogavahas(part vowels and part consonants) in its character set. Recognition mainly involves two stages say feature selection and classification. In the Feature selection stage, different features which form the basis for signature of each character are selected. Here, it is done using Direction based Stroke Density principle (DSD) which refers to recording the number of points in each stroke of a character thus determining the density of that stroke. Here a stroke is defined as any line greater than a predefined length drawn in a particular direction. The directions are considered based on 4 axes or 8 directions.

The Classification stage in turn involves many sub stages at the end of which the written character is classified to a particular set to obtain the character whose signature is closest to that of the written character. This is carried out by Kohonen Neural Network(KNN). The Kohonen system is an unsupervised neural system. It is simple and can be easily trained. It`s error is calculated based on how well it classifies the input provided to it. This approach was chosen because of its simplicity and efficiency. Even though Kohonen is used for OCR, its combination with DSD makes it a novel approach here. Very typical applications for on-line HWR are user interfaces of personal digital assistants (PDAs), smart phones or Microsoft's new tablet PCs.Kannada character set including numerals is as shown in Figure.



PROPOSED METHOD

Flow of the proposed method is as shown in Figure2:

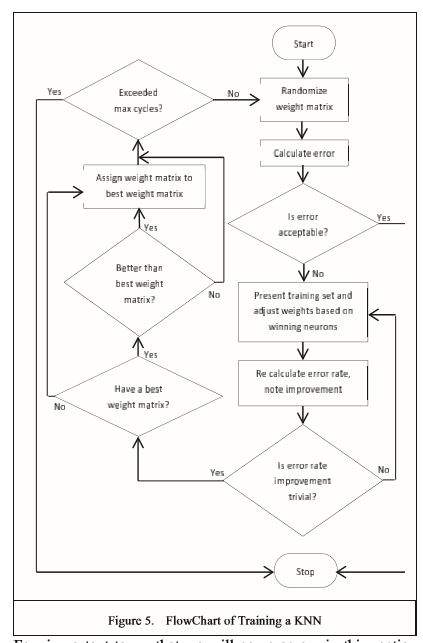


* Initial SampleSet(ISS): Sample set of all characters in Kannada Language(49characters + 10 numerals = 59)
* Current SampleSet(CSS): Contains the current sample set
* Training SampleSet(TSS): Contains sampleset for training which include only those characters which have the same click count feature as the written character.
* Focusing on the main recognition phase of the system, we come across three important processes namely feature selection, training and recognition processes.

Training process

Flowchart of the training process is as shown in Figure 5. In this stage the Training sample set is obtained i.e. the set which has same no.of click counts as the written character and then given to the Kohonen network for training. The Kohonen neural network differs from the feed forward back propagation network in several ways. The Kohonen neural network is trained in an unsupervised way which means the Kohonen neural network is given input data but no anticipated output. The Kohonen neural network then begins to map the training samples to each of its output neurons during training. In this case, each output neuron is associated with one character present in the training set

The basis of kohonen network is on two main aspects which are OutputNeuron weights and winning neuron. Unlike other neural networks, in kohonen there is only one winning neuron for every input pattern.



Few important terms that we will come across in this section are:

* Trainingset[][] -it contains input signatures of all the characters in Training sample set which are to be trained. If there are 59 characters then the training set will contain 59 sets with input values or signatures for each set
* InputNeuron[] -it is the value of each input pattern corresponding to each character.
* Outputneuron count - It is the no. of output neurons from which a winning neuron will be decided for each training set. This value will be equal to the no. of characters in the sample set.
* Inputneuron count – it is the no of input neurons given to the network which is the length of each training set given at a time.
* Outputweights[][] – it holds the weights between every input neuron and output neuron for each training set.