Discrimination of Indian Tea Varieties using UV-VIS-NIR Spectrophotometer and Pattern Recognition Techniques

Santosh Kumar[1], P C Panchariya[2], Bhanu Prasad[3], AL Sharma[4]
[1,2,3] Digital Systems Group, CSIR-Central Electronics Engineering Research Institute, Pilani, India
[4] School of Instrumentation, Devi Ahilya University, Indore, India

***Abstract*: The food and water is an important and an extremely important beverage for all living beings. Inspite of food and water, there are other drinks in our daily life as soft drinks, cola, fruit juices, coffee, and tea etc. Tea is one of the popular beverages, not only for its taste but also for its beneficial medicinal properties for human beings. There are many different tea species or clones among the world and in India. India is one of the largest tea producers in the world and also the world's largest tea-drinking nation. There are lots of tea varieties available in the Indian market and all are having different taste, flavor and colors. The proposed paper gives the identification of most of the teas by means of UV-VIS-NIR spectrophotometer along with the clustering and classification techniques. UV-Vis-NIR spectroscopy is a reasonably fast and accurate technique that can be used as a replacement of conventional sensory evaluation methods and time-consuming chemical methods. UV-VIS-NIR spectra were recorded by measuring the absorption spectra in the range 240-2600nm. The spectral features of each category are reasonably differentiated by principal component analysis and the pattern recognition techniques for clustering and classification. Principal component analysis (PCA), a common chemometric method used for data reduction and visualization. Here nine varieties of teas are tested, and their spectral differences provide enough qualitative spectral information for the discrimination.**

***Keywords* : Spectroscopy, Discrimination, Principal Component Analysis, Pattern Recognition Techniques, Classification.**

INTRODUCTION

There is a large numbers of tea available in the Indian market. All are of same coloures and has some special tastes and flavours. This work will discriminate these teas and also based on these techniques we can develop the E-Tongue and E-Nose with different gas sensors and non specific sensors for the taste purposes and smell purposes. In this proposed research of work we have taken the commercial Indian tea as the sample for the discrimination by the UV-VIS-NIR spectroscopic analysis. We have used Spectrophotomer model number UV-3600 of Shimadzu instruments for the spectrum analysis. The entire data spectrums are stored in the absorbance mode with the spectrophotometer. We have captured the spectrum for the complete range as the wavelength from 240nm to 2600 nm. After recording the spectrum the data is first normalized in range of 0-1 and then the data is reduced with the standard data reduction technique, i.e., Principal Component Analysis PCA, we have done all the computations and algorithm development in MATLAB environment. We have used the MATLAB R2007 and Origin Lab 8.0 software for data normalization, data visualization and programming.[1-4] The latest research on various tea samples were done on different regions of teas of all over world, with different pattern recognition techniques. Linear Discriminant Analysis (LDA) and Artificial Neural Network (ANN) were compared to construct the identification models based on Principal Component Analysis (PCA).[12]. This work done is on the various Chinese teas samples discrimination.

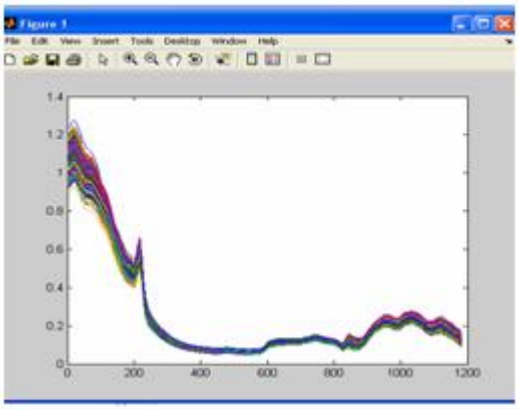

Fig 1 : Spectrum of Various Indian Tea.

Ultrasonic testing is also used for the non destructive analysis of the foods and beverages. Other technique i.e.; Ultrasonic measurement, is also used for the various teas discriminations. The Ultrasonic properties are very sensitive to composition of liquid food products. [13]. Another Technique is also used that is with electronic Tongue (E-Tongue). In this paper the voltametric method is used. The voltametric method is a destructive type of measurement technique. In this research work the PCA-LDA Model is used on the data captured with E-Tongue.[14] and so on.. There is various research work on other teas but there is no data and work available on the commercial Indian teas. Therefore there is the need of this research for the discrimination of commercially available Indian teas. This

paper is gives the discrimination results of various Indian Teas, which are commercially available in every corner of the country. The data spectrum of all the tea samples used in this work are shown in figure 1.

PRINCIPLES OF SPECTROPHOTOMETRY

A spectrophotometer, one of the most useful physics lab equipments is the combination of two devices, a spectrometer and a photometer. Spectrometer is used for producing light of any selected wavelength or color while a photometer is used for measuring the intensity of light. The two devices are placed at either side of a cuvette filled with a liquid. Spectrometer produces the light of desired wavelength and it passes through the tube and reaches photometer that measures its intensity. Then the photometer produces a voltage signal to a display device, usually a galvanometer. As the amount of light absorbed by the liquid changes the signal also changes. The concentration of a substance in solution can be measured by calculating the amount of absorption of light at the appropriate wavelength or a particular color. The basic block diagram of spectrometer is shown in figure 2.

Spectrometry setup and arrangement:

Light source, monochromator, sample, detector (PMT), data acquisition and data handling unit (PC). (PMT - Photoelectron Multiplier Tube)

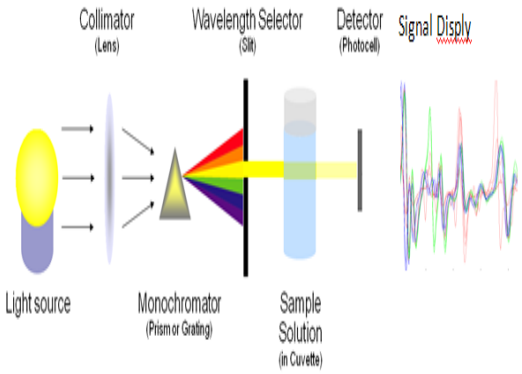

Figure 2: Spectrophotometer System.

Pattern Recognition Techniques:

Automatic (machine) recognition, description, classification, and grouping of patterns are important problems in a variety of engineering and scientific disciplines such as biology, psychology, medicine, marketing, computer vision, artificial intelligence, and remote sensing all the humans, birds, plants, animals, images, signals has different properties with each other and also different types of signals has different patterns and the same classes gives same pattern if we take the data with the same techniques. There are two main pattern recognition techniques.

1. **Unsupervised Techniques :** In this technique, there is no primary information about the sample classes. i.e.; known as Clustering. Here we have used K-Means Clustering algorithms on the data samples and the clustering result is shown.

2. **Supervised Techniques :** In this technique, we have the prior information with us about the sample classes . i.e.; known as Classification. In this paper we have used the PNN Classification technique. We have got very good result with this techniques. The Complete Block diagram of the Pattern Recognition Technique can be given in step by step as in following figure-3.[4,8]

Process Flow of Pattern Recognition Technique

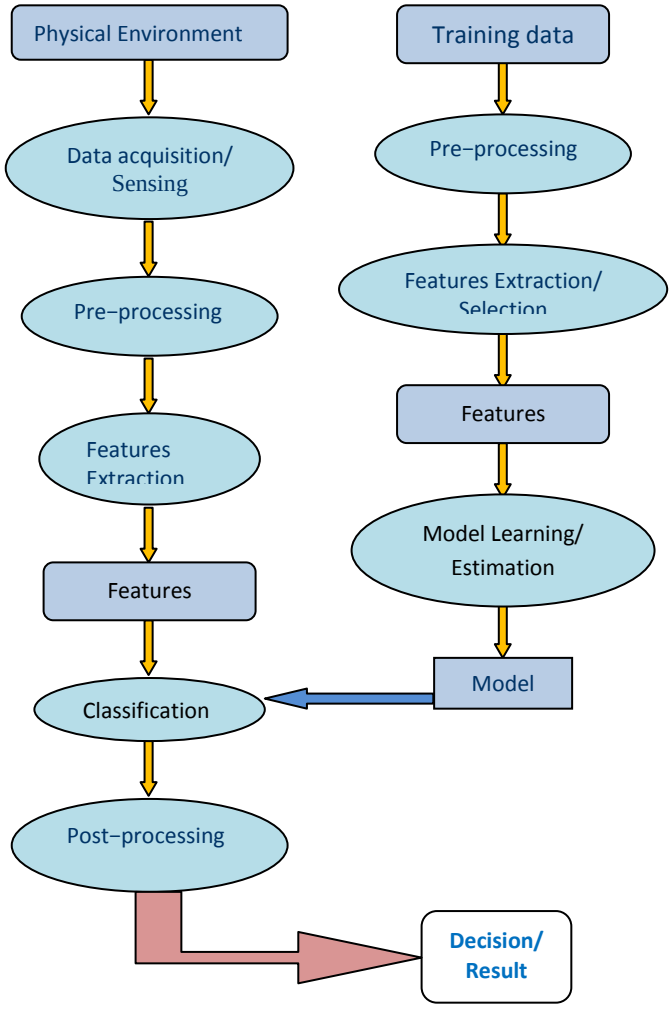

Fig 3: Process Flow Diagram of Pattern Recognition and Classification.

Tea Samples :

The tea samples are collected from the cooperative society (new and fresh stocks are always collected). Samples are also collected from the other general stores to analyzing and testing the receptivity in discrimination the tea samples. As the samples and data are collected at many times after one months. Therefore five times the samples are collected and data are recorded with the spectrophotometer and samples are prepared in same process.

Sample Preparation:

We have taken data of samples prepared with different methods. First we have taken 10 gram of samples and 100ml of boiled water, mixed them and wait for 5 minutes, 10 minutes for various ways of data recording. After the completion of the samples we have filtered it out and taken the sample in 1 mm and 2 mm cuvette and recorded the signal with the spectrophotometer after completion of the proper instruments initialization for different types of data analysis.

TRAINING AND TEST SAMPLES

After getting the data, complete data set needs to be partitioned into training and testing data sets. After separating the data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing. By using similar data for training and testing, we can minimize the effects of data discrepancies and better understand the characteristics of the model. After a model has been processed by using the training set, you test the model by making predictions against the test set. Because the data in the testing set already contains known values for the attribute that you want to predict, it is easy to determine whether the model's guesses are correct. We can divide data set such as the training set (seen data) to build the model (determine its parameters) and the test set (unseen data) to measure its performance (holding the parameters constant). Sometimes, we also need a validation set to tune the model (e.g., for pruning a decision tree). The validation set can't be used for testing (as it's not unseen). All three data set have to be representative samples of the data that the model will be applied.

Data Analysis :

The data is prepared for the data analysis. For the data analysis, the data is first preprocess the standard data reduction technique i.e; principal component analysis (PCA). And after the data reduction, applying the different classification and clustering techniques for the discrimination of the tea samples as it is given in next sections.

K-Means Clustering:

This is an unsupervised technique of pattern recognition. The initial step is the choosing a set of K instances as centres of the clusters. Often chosen such that the points are mutually "farthest apart", in some way. Next, the algorithm considers each instance and assigns it to the cluster which is closest. The cluster centroids are recalculated either after each instance assignment, or after the whole cycle of re-assignments. Initially, the number of clusters must be known, or chosen, to be K say. This process is iterated.

K-mean Algorithm features:

Using cluster centroid to represent cluster. Assigning data elements to the closest cluster (centre). Goal: Minimise the sum of the within cluster variances Variations of K-Means. Initialisation (select the number of clusters, initial partitions).Updating of centre. [7] The K Means Clustering Result of the nine Indian teas can be given as in figure 4.

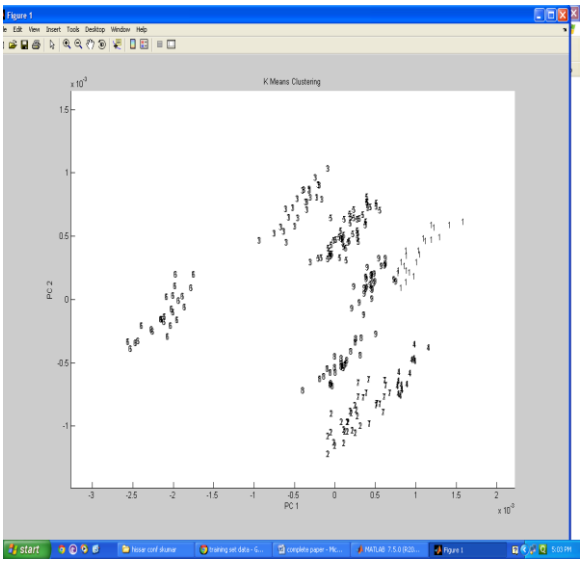

Fig 4 : K-Means Clustering Result.

PROBABILISTIC NEURAL NETWORK

A probabilistic neural network (PNN) has 3 layers of nodes. The figure below displays the architecture for a PNN that recognizes K = 2 classes, but it can be extended to any number K of classes. The input layer (on the left) contains N nodes: one for each of the N input features of a feature vector. These are fan-out nodes that branch at each feature input node to all nodes in the hidden (or middle) layer so that each hidden node receives the complete input feature vector x. The hidden nodes are collected into groups: one group for each of the K classes as shown in the figure 5. Each hidden node in the group for Class k corresponds to a Gaussian function centered on its associated feature vector

in the kth class (there is a Gaussian for each exemplar feature vector). All of the Gaussians in a class group feed their functional values to the same output layer node for that class, so there are K output nodes.[5-6]

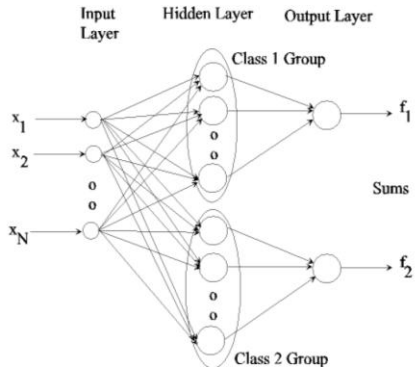

Fig 5 : The Architecture of Probabilistic Neural Networks.

The probabilistic neural network is an artificial neural network belongs to the supervised pattern recognition technique. The result of PNN Classification of the nine Indian teas can be given as in figure 6.[9-11]

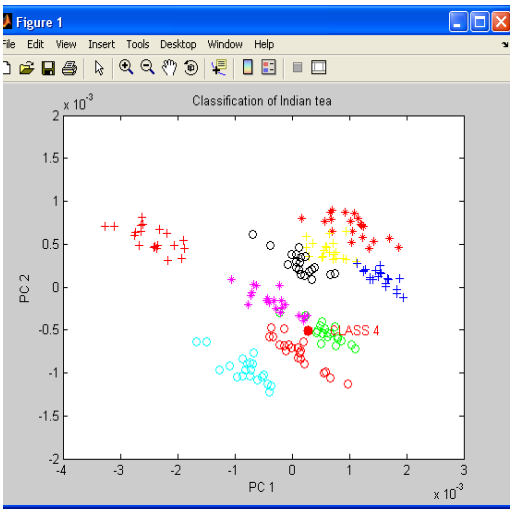

Fig 6 : PNN Classification Result of tea varieties.

RESULT AND CONCLUSION

The output plot and results we have shown here in figure1 and in figure 7, are giving the discrimination of all the nine varieties of commercial Indian teas. The unsupervised technique K-means clustering is a powerful method of clustering in pattern recognition analysis of a data set. In the next step, we have used here a supervised technique i.e.; PNN classification to discriminate the tea varieties. Here an unknown sample is classified as of class – 4. the unknown sample is picked from the 4th class of tea of untrained samples. Therefore we have successfully discriminated and analyzed the nine varieties of commercial Indian teas using UV-VIS-NIR and the given Pattern Recognition Techniques.

ACKNOWLEDGMENT

The Author is higly acknowledged to the Council of Scientific and Industrial Research - Central Electronics Engineering Research Institute CSIR-CEERI, Pilani (Rajasthan) to giving me this excellent type of research work opportunity.

REFERENCES

[1] "Biomimetic Classification of Juices", 2012 Sixth International Conference on Sensing Technology, ICST-2012 at Kolkata, (AH Kiranmayee, P C Panchariya, P. Bhanu Prasad, A L Sharma)

[2] "Non Destructive Classification of Himalayan Orthodox Black Teas" Sensors & Transducers Journal, Vol. 145, Issue 10, October 2012, pp. 77-85, (S Kumar. PC. Panchariya, Bhanu Prasad P. and A. L. Sharma)

[3] "Acoustic signature based discrimination of drinking water" 2012 Sixth International Conference on Sensing Technology ICST-2012 at Kolkata, (RS Chouhan, AH Kiranmayee, PC Panchariya & P. Bhanu Prasad)

[4] "Introduction to Pattern Recognition" Selim Aksoy Department of Computer Engineering, Bilkent University,CS 551, Fall 2012.

[5] "An Improved Architecture for Probabilistic Neural Networks" Proceedings of International Joint Conference on Neural Networks, San Jose, California, USA, July 31 – August 5, 2011 (B. Chandra, K.V. Naresh Babu)

[6] "Probabilistic Neural-Network Structure Determination for Pattern Classification", IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 11, NO. 4, JULY 2000 (K. Z. Mao, K.-C. Tan, and W. Ser)

[7] "An Efficient k-Means Clustering Algorithm: Analysis and Implementation" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO.

7, JULY 2002 (Tapas Kanungo, Senior Member, IEEE, David M. Mount, Member, IEEE, Nathan S. Netanyahu, Member, IEEE, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu, Senior Member, IEEE)

[8] "Statistical Pattern Recognition: A Review", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 22, NO. 1, JANUARY 2000(Anil K. Jain, Fellow, IEEE, Robert P.W. Duin, and Jianchang Mao, Senior Member, IEEE).

[9] "A PROBABILISTIC NEURAL NETWORK APPROACH FOR PROTEIN SUPERFAMILY CLASSIFICATION", Journal of Theoretical and Applied Information Technology (PV NAGESWARA RAO, T UMA DEVI, DSVGK KALADHAR, GR SRIDHAR, ALLAM APPA RAO).

[10] "Maximum Likelihood Training of Probabilistic Neural Networks", IEEE Transactions on Nural Network, Vol. 5, No 5, September 1994 (Roy L Streit, Senior Member IEEE, and Tod E Liginbuhl, Member IEEE)

[11] "Adaptive Probabilistic Neural Networks for Pattern Classification in Time-Varying Environment" IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL.15 issue 4, Leszek Rutkowski, Senior Member, IEEE.

[12] Chen Q., Zhao J., Liu M., Cai J. (2008): Nondestructive identification of tea (Camellia sinensis L.) varieties using FT-NIR spectroscopy and pattern recognition. Czech J. Food Sci., 26: 360–367.

[13] Aba Priev and Yechezkel Barenholz, Ultrasonic Food Quality Analyzer Based on Cylindrical Standing Waves, PACS: 43.20.Hq, Ks; 43.25.Qp; 43.35.Yb, Zc