Tea Classification using Fourier Transform Infrared Spectroscopy

Santosh Kumar, P. C. Panchariya, P. Bhanu Prasad
Digital Systems Group
CSIR-CEERI,
Pilani, Rajasthan, India
E-mail-ceerisk@gmail.com

A. L. Sharma
School of Instrumentation
Devi Ahilya University,
Indore, M.P. - 452 001
dralsharma@yahoo.com

***Abstract*— This paper deals with the classification of commercially available Indian teas using the Fourier Transform Infrared (FTIR) spectroscopy method in conjugation with pattern analysis techniques. In this work, two sample preparation approaches i.e. liquid as well as powdered solid sample were investigated. The FTIR spectra of 8 varieties of Indian teas were recorded in the range of 400cm^{-1} to 4000cm^{-1} in absorbance mode at room temperature (25ºC) using FTIR Spectrometer. The obtained spectra were preprocessed using standard preprocessing methods i.e. normalization and Standard Normal Variate (SNV). The feature extraction step was carried out using principal component analysis (PCA). The PCA integrated with Linear Discriminant analysis (LDA) was adopted for classification task. PCA-LDA was successfully classified all varieties of tea samples in liquid form as well as in powdered solid form. The FTIR measurements coupled with suitable pattern recognition methods can be used as a key measurement tool for analysis of quality of teas.**

Keywords— ***Infrared Spectra, Sample Process, Classification, Pattern Recognition.***

I. INTRODUCTION

Tea is a necessary part everybody's life. Researchers have been working on the tea for many years, for the quality assessment and also for the possibilities to improve the tea quality. Till now, it is very surprising to see that, after so many years of research on this, we get another surprising new with tea. It has a lot of health benefits like, Tea improves the alertness, reduces the risk from cancer like diseases etc. The major analysis on tea has been carried out using the high performance liquid chromatography (HPLC). HPLC analysis provides most of the tea components like, tea catechins, tea biochemical's etc. Electronic tongue (E-tongue) and Electronic nose (E-nose) are the other instruments, are used for the destructive and non-destructive analysis of teas for the tea analysis. E-tongue is based on voltammetric method, which gives tea-taster-like scores for tea quality discriminations [1]. E-nose provides the tea quality using the tea aroma. Ultraviolet (UV) and UHPLC are also used for the various tea analysis [2]. Gas chromatography-mass spectrometry (GC-MS) system is also used for the identification of volatile components in Kangra orthodox tea like, (E)-2-hexenal, 1-pentene-3-ol, and (Z)-3-hexenol among non-terpenoids and linalool, linalool oxides (furanoid), geraniol, methyl salicylate, 3,7-dimethyl-1,5,7-octatrien-3-ol etc [3]. FTIR is used for the tea components identification of powdered tea samples by Xiaowei Li, China [4]. Fourier Transform Infrared Spectroscopy analysis method is used by the researchers since 1960's afterwards from 1970s FTIR is popularly used for various applications. FTIR measurement is a modern spectroscopic analysis method works in the infrared (IR) region. FTIR spectrometer can choose many spectral collection parameters because of its wavelength selection nature [5]. FTIR measurement is very-very fast as it records the spectra in approximately 1 spectrum per second. By using the FTIR measurement methods one can analyze the samples chemicals, identifying an unknown compound in the samples, structures of the sample such as determine the chemical groups present in a specific compound and electronic information such as optical conductivity, band gaps etc[6]. The tea polyphenols has major impact on human health. The Fourier transform infrared spectroscopy analysis is used to analyze the cancer cells structures with the help of tea polyphenols such as curcumin, epigallocatechin gallate (EGCG) and quercetin [7]. The tea liquor is analyzed in this process. Flavonoids have some anti disease property. The tea catechins are found as the tea flavonoids. FTIR is also used to investigate the catechins from the tea aqueous samples [8]. The tea polyphenols and tea catechins can be analyzed by the peaks in the corresponding regions. In this analysis, statistical analysis such as Principal Component Analysis (PCA) and Linear Discriminant analysis (LDA) were applied for the discrimination between the commercial grades Indian teas. Fourier Transform Infrared (FTIR) Spectroscopy method were used as an experimental analysis such that, whether this method is applicable on our samples or not, and if this technique is applicable on our samples, then the results will be analyzed further using feature extraction and classification techniques. Therefore, in this research, FTIR experiments were done to analyze two different types of samples (Liquid Tea Sample and Powdered Tea Sample) was prepared for the FTIR experiments.

II. EXPERIMENTAL

1. Sample Preparation:

The eight varieties of commercial Indian tea were used for this research work, as they are-

1-	Taj Mahal	-	T 1
2-	Red Label	-	T 2
3-	Taaza	-	T 3

4-	Tata Tea Gold	-	T 4
5-	Marvell Red Tea	-	T 5
6-	Duncan Double Diamond	-	T 6
7-	Tetley	-	T 7
8-	Tez	-	T 8

(i) *Liquid Tea Sample preparation:*

The commercial Indian tea were purchased from the standard supermarket (Akshay cooperative society), from Pilani, Rajasthan, India. Eight variety of standard tea were chosen for the analysis. Five gram of tea samples were taken from each tea variety. The weighing machine is used to take the equal amount of tea. Mineral water of 100ºC is used to make the tea liquid samples. 5 gm of tea leaves and 100ml of water were mixed. After mixing the tea samples in the water, wait for 10 minutes for proper dissolving and extraction of tea in the liquor. After proper extraction of tea in water for 10 minutes, it will be filtered out. The liquid samples of 2 ml were used for the FTIR analysis. Five replica of eight variety of tea were prepared for the data measurements. Total 40 samples were prepared for the analysis.

(ii) *Powdered Tea Sample preparation:*

Equal amount of tea samples were taken and milled until fine powdered form and are filtered with same mesh after proper cleaning and drying. Fine tea powders were used for the measurements and further data analysis. The same amount of tea powder was taken for the infrared observation in the wave number ranging from of 400 cm^{-1} to 4000 cm^{-1} using TENSOR37 Fourier Infrared Spectrometer. The tea powder samples were not mixed with the KBr powder as in the case of Xiaowei Li et al [4].

(iii) *Fourier Transform Infrared (FTIR) Spectroscope*

The Tensor-37 of Bruker optics instrument were used as FTIR platform for all measurements. The measurements were taken between absorbance (a.u.) vs. wave numbers (cm^{-1}). The destructive tea spectrum data were recorded with Fourier Transform Infrared Spectroscopy (FTIR). The OPUS software is interfaced with this instrument and the user can operate the instrument and monitors the status.

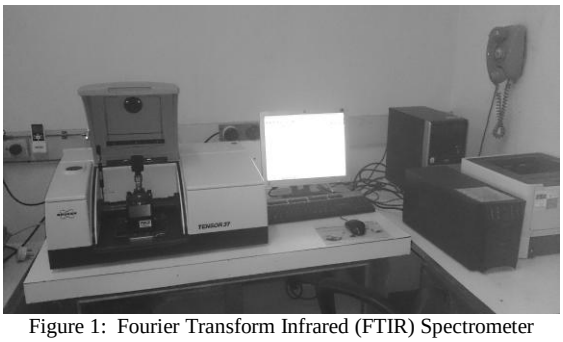


Figure 1: Fourier Transform Infrared (FTIR) Spectrometer Experimental Setup
(CSIR-CEERI, Pilani - India)

Sampling capabilities of the TENSOR series can be expanded by adding an optional external beam port [9]. The FTIR spectrum is analyzed after suitable pattern recognition techniques, here the clear separation between different samples have found and resulting that the samples are easily discriminable. The FTIR spectrometer setup (figure 1) at CSIR-CEERI were used for the measurements and the FTIR Liquid Sample Attachment Model no. - Miracle ATR is used for the experiments with liquid samples and Solid Sample attachment Model no. – Easi Diff is used to analyze the tea powdered samples.

2. Data Analysis :

(i) *Liquid Tea :*

The liquid tea data were first normalized in the range 0 to 1as the spectra are shown in figure 5a. Principal component analysis is applied to visualize the data samples. PCA Score plot of Normalized data is shown in the figure (5b). The normalized data is further used to find out the first derivative of the data to get most refined data with no noise content and applied the PCA algorithm on the derivative data. The derivative data plot and the PCA score plot are shown in the figure 5c and 5d. Tea samples were clearly discriminated by the principal component analysis combined with the Linear Discriminant analysis. 100 % Classification accuracy is achieved.

(ii) *Powdered Tea:*

The powdered tea sample data were first preprocessed using the Standard Normal Variate (SNV) technique. This removes the slope variations of the data. Principal component analysis is done on SNV data for dimensionality reduction and data visualization. The reduced dataset i.e. Principal Components (PC) are used to create the linear Discriminant Analysis model. The 100% classification accuracy was achieved using the LDA model.

(iii) *Principal Component Analysis (PCA):*

PCA is one of the most popular data reduction techniques. The data obtained with the actual measurements are plotted in the principal directions. The data variation in principal directions gives that, the first principal components (PC) has the maximum data variation and the second principal components (PC) has the second largest data variation. PCA is the most suitable technique for the data visualization in statistical approach.

(iv) *Linear Discriminant Analysis (LDA):*

LDA, a well known method, is also used for the dimensionality reduction and samples classification. It keeps as much of the class discriminatory information. The distance between the data variations is not always a very good measure

for the prediction of data class because it does not take standard deviation into account within the classes. To classify the data class, mean and standard deviation has important contribution. LDA measures the within-class variability and between-class variability to classify the data in better way, therefore PCA-LDA approach is applied for the data reduction as well as for the data classification and prediction of the data classes. The software package "The Unscrambler" of ver. 10.2 from Camo is used for this analysis.

Ⅲ. RESULTS AND DISCUSSION:

(i) Liquid Teas Samples with Fourier Transform Infra Red (FTIR) Spectrometer

The liquid tea samples were prepared for this analysis. The absorbance spectra were recorded in the range of 400 cm^{-1} – 4000 cm^{-1} Wave Numbers. The spectra were first preprocessed using the normalization techniques and apply the principal component analysis on the normalized data. The FTIR spectra are shown by splitting the full range spectra (figure 2) into four ranges of wave numbers, as in figure 3, in expanded view, i.e. (1) 550 cm^{-1} to 800 cm^{-1} (2) 800cm^{-1} to 2000 cm^{-1}, (3) 2000cm^{-1} to 2500 cm^{-1} and (iv) 2500cm^{-1} to 4000 cm^{-1}.

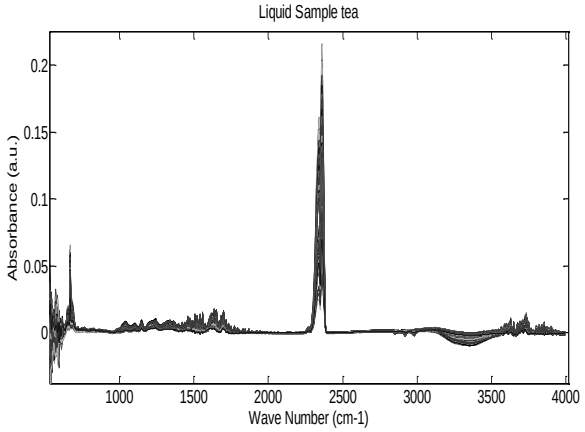
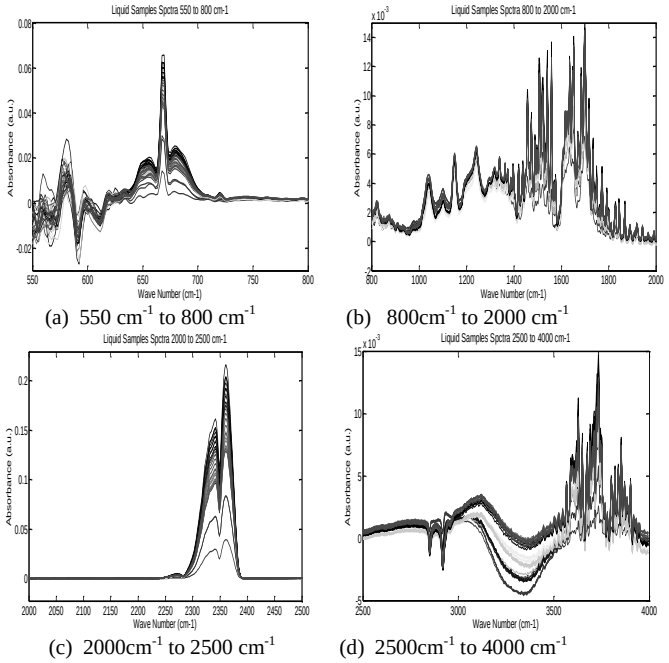


Figure 2: FTIR Spectrum of Commercial Indian Tea (Liquid Sample)

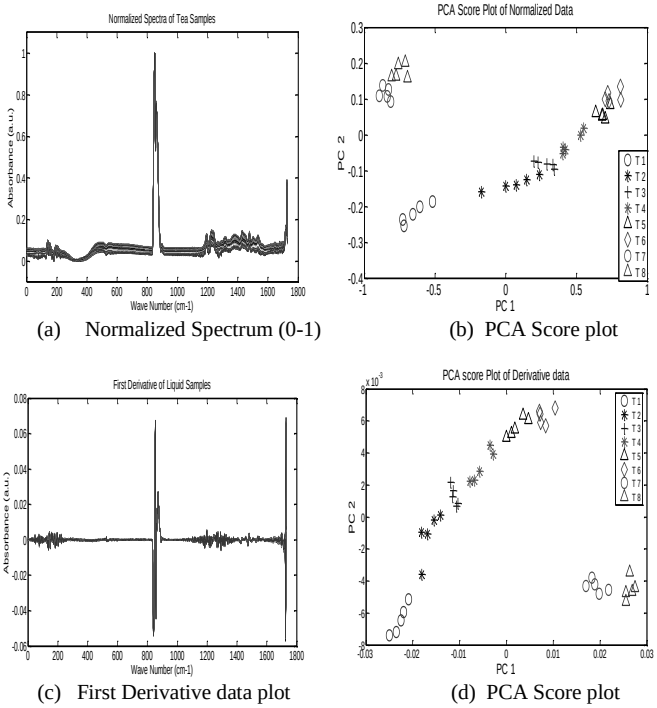
All these spitted wave forms are analyzed separately and found that the spectrum between wave number 2000cm^{-1} to 2500 cm^{-1} are giving the best classification result and this range can be used for the tea variety identification and new sample prediction. This spectrum can be seen in figure 3d. The normalized and derivative spectra with their PCA score plot are shown in figure 4.The liquid tea data samples were classified by applying the linear discriminant analysis. The classification accuracy is improved by applying various preprocessing methods. Two principal components are used to make the LDA model. The LDA classification accuracy was 90% of raw data, 97.5% accuracy were achieved by applying the data normalization and after taking the derivative of the normalized data and applying the LDA classification, 100% accuracy were achieved. The LDA plot and their corresponding confusion matrices are given in the figure 5 and in Table 1, 2 and 3 for raw data, normalized data and derivative data.



(a) 550 cm^{-1} to 800 cm^{-1} (b) 800cm^{-1} to 2000 cm^{-1}

(c) 2000cm^{-1} to 2500 cm^{-1} (d) 2500cm^{-1} to 4000 cm^{-1}

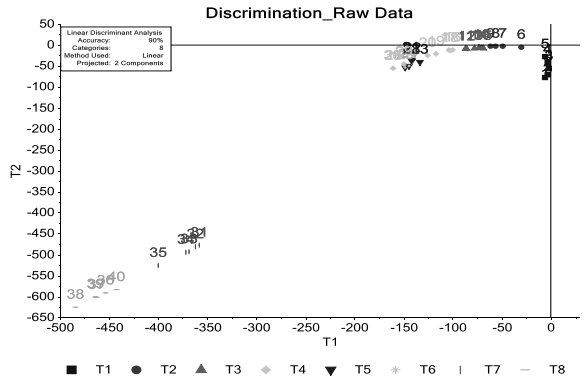
Figure 3: FTIR Spectrum of Commercial Indian Tea (Liquid Sample)



(a) Normalized Spectrum (0-1) (b) PCA Score plot

(c) First Derivative data plot (d) PCA Score plot

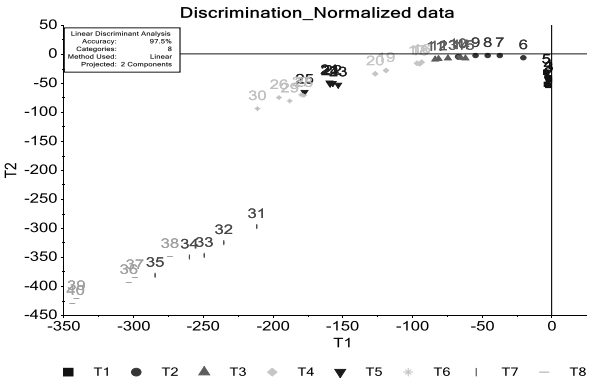
Fig 4 : Liquid Samples Analysis (a) Normalized data, (b) PCA of Normalized data, (c) First Derivative data and (d) PCA of First derivative data



(a)

Table 1 - Confusion Matrix: Raw data

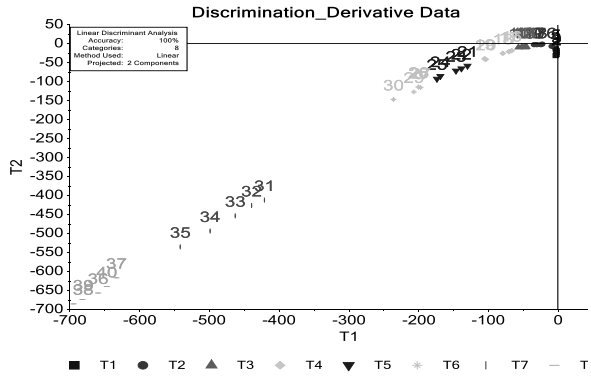
Sample name		Predicted Sample							
		T1	T2	T3	T4	T5	T6	T7	T8
Actual Sample	T1	5	0	0	0	0	0	0	0
	T2	0	4	0	0	0	0	0	0
	T3	0	1	5	0	0	0	0	0
	T4	0	0	0	5	0	0	0	0
	T5	0	0	0	0	3	1	0	0
	T6	0	0	0	0	2	4	0	0
	T7	0	0	0	0	0	0	5	0
	T8	0	0	0	0	0	0	0	5



(b)

Table 2 - Confusion Matrix: Normalized data

Sample name		Predicted Sample							
		T1	T2	T3	T4	T5	T6	T7	T8
Actual Sample	T1	5	0	0	0	0	0	0	0
	T2	0	4	0	0	0	0	0	0
	T3	0	1	5	0	0	0	0	0
	T4	0	0	0	5	0	0	0	0
	T5	0	0	0	0	5	0	0	0
	T6	0	0	0	0	0	5	0	0
	T7	0	0	0	0	0	0	5	0
	T8	0	0	0	0	0	0	0	5



(c)

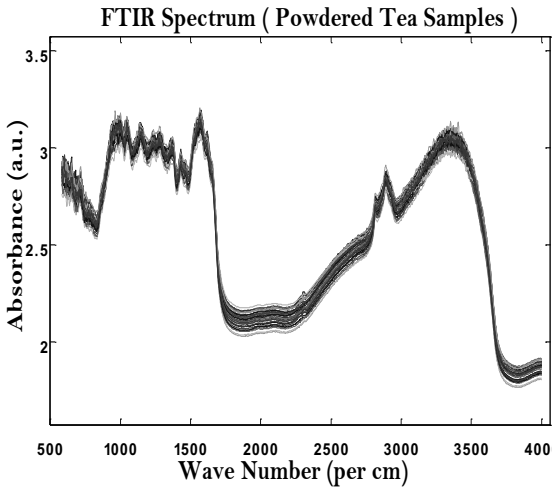
Table 3 - Confusion Matrix: Derivative Data

Sample name		Predicted Sample							
		T1	T2	T3	T4	T5	T6	T7	T8
Actual Sample	T1	5	0	0	0	0	0	0	0
	T2	0	5	0	0	0	0	0	0
	T3	0	0	5	0	0	0	0	0
	T4	0	0	0	5	0	0	0	0
	T5	0	0	0	0	5	0	0	0
	T6	0	0	0	0	0	5	0	0
	T7	0	0	0	0	0	0	5	0
	T8	0	0	0	0	0	0	0	5

Figure 5: LDA Plot and Confusion Matrix of liquid tea samples - (a) Raw data (b) Normalized data (c) Derivative data

(ii) Powdered Teas Samples with Fourier Transform Infra Red (FTIR) Spectrometer

In this analysis, the powdered tea samples were prepared for the FTIR analysis. Xiaowei Li et al. used FTIR analysis to identify the Zymic Pu-er Tea with the tea powder samples and analyzed the data with the vector normalization, derivative and smoothing preprocessing methods integrated with the feature extraction algorithm to give the best discrimination results on the tea powder samples [4]. The tea property is affected by adding the milk to it is analyzed and shown that FTIR is well suited to analyze the α- and β-caseins with tea polyphenols [**10**]. The powdered tea sample data for commercial Indian tea were recorded using FTIR. Xiaowei Li et al captured FTIR spectra and analyzed on the data for the components composition. The commercial Indian tea FTIR spectra by our methods are given in figure 6. The FTIR spectra of Xiaowei and our spectra's are very similar having some slope variations.



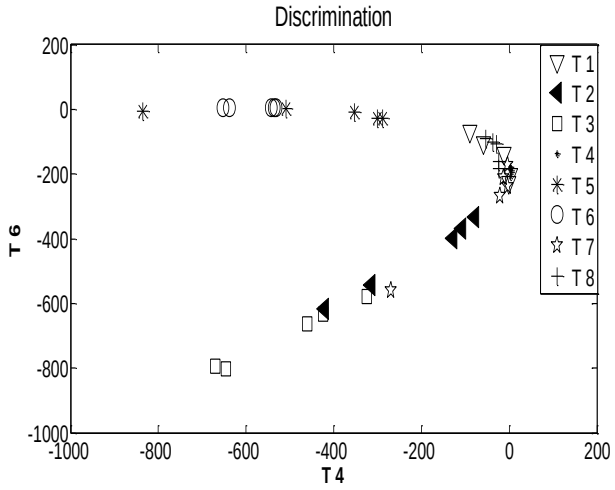


Figure 7: LDA Plot of powdered tea samples

Therefore, it is proved by this method, that the powder samples could be used to discriminate the Tea samples. Xiaowei Li et al. gives the idea about the tea compositions identification of various samples of Chinese tea, while in this paper the statistical methods were applied to identify the different commercial Indian tea varieties.

The data samples recorded using FTIR were preprocessed using standard normal variate (SNV) method, and further analyzed by PCA-LDA with three methods as (i) Linear LDA analysis, (ii) Mahalanobis LDA method and (iii) Quadratic LDA method. The Linear Discriminant Analysis classification is tested with projected components 2, 3, 4 and 5.

The Linear method gives 87.5% classification accuracy with 2 projected components, giving 95% classification accuracy with 3 projected components, 92.5% classification accuracy with 4 projected components and 90% classification accuracy with 5 projected components,, Mahalanobis method gives 90% classification accuracy with 2 projected components, 97.5% classification accuracy with 3 projected components, 97.5% classification accuracy with 4 projected components and giving 100% classification accuracy with 5 projected components.

The Quadratic LDA method gives 90% classification accuracy with 2 projected components and giving 100% classification accuracy with 3 projected components 97.5% classification accuracy with 4 projected components and 100% classification accuracy with 5 projected components. Therefore, the LDA has the power to discriminate the tea samples taken by FTIR spectrometer. The best classification result, i.e.; 100% accuracy is achieved with Quadratic linear Discriminant method with 3 numbers of projected score. LDA plot is shown in the figure 7, and the Confusion Matrix is shown in table 4.

Table4: Confusion Matrix

Sample Name		Predicted Samples							
		T1	T2	T3	T4	T5	T6	T7	T8
Actual Samples	T1	5	0	0	0	0	0	0	0
	T2	0	5	0	0	0	0	0	0
	T3	0	0	5	0	0	0	0	0
	T4	0	0	0	5	0	0	0	0
	T5	0	0	0	0	5	0	0	0
	T6	0	0	0	0	0	5	0	0
	T7	0	0	0	0	0	0	5	0
	T8	0	0	0	0	0	0	0	5

IV. CONCLUSION:

This Research gives the overall idea about the commercially available Indian tea varieties and its experiments with the Fourier Transform Infrared spectrometer for food quality analysis. The eight varieties of tea samples were prepared in liquid sample and powdered samples. The data were preprocessed using the standard methods as normalization and standard normal variate. The principal component analysis was applied for the data reduction and data visualization. The linear discriminant analysis was applied for the sample classification. In both the cases, 100% classification accuracy is achieved. Therefore, Fourier Infrared spectrometer has the high potential to discriminate the Liquid tea samples as well as the Powdered Tea samples with suitable pattern recognition techniques. This method of infrared spectrometer may be considered as a powerful tool for food quality assessment.

References:

1. "Classification of Black Tea Taste and Correlation With Tea Taster's Mark Using Voltammetric Electronic Tongue", Mousumi PALIT *et al* , IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, VOL. 59, NO. 8, AUGUST 2010.
2. "Quantitative and qualitative analysis of flavonoids in leaves of *Adinandra nitida* by high performance liquid chromatography with UV and electrospray ionization tandem mass spectrometry detection", JieZhanga et al. Analytica Chimica Acta 532 (2005) 97–104.
3. "Characterization of volatile components of Kangra orthodox black tea by gas chromatography-mass spectrometry", Renu Rawat et al, food chemistry, Volume 105, Issue 1, 2007, Pages 229–235.
4. "A New Method for Identification of Zymic Pu-er TeaBased on Fourier Transform Infrared Spectroscopy", Xiaowei Li, Yanwei Zhang, Rongxiang Zhang, Xiaohui Zhao, Lianshui Zhang, 2010 3rd International Congress on Image and Signal Processing (CISP2010) from IEEE Explore.
5. "Real Time Emission Measurements Using FTIR Spectroscopy (EPA Method 320) ", Jeffrey LaCosse , Spectral Insights LLC , December 8, 2010 , www.spectralinsights.com
6. "Fourier Transform Infrared Spectroscopy FTIR Michael C. Martin, Advanced Light Source Division, MCMartin@lbl.gov"
7. "FTIR spectroscopy: A new valuable tool to classify the effects of polyphenolic compounds on cancer cells.", Allison Derenne, Vincent Van Hemelryck , Delphine Lamoral-Theys, Robert Kiss, Erik Goormaghtigh, Biochimica et Biophysica Acta 1832 (2013) 46–56, Elsevier.
8. "Investigation of alumina/(+)-catechin system properties. Part I: a study of the system by FTIR-UV–Vis spectroscopy", M.M. Ramos-Tejada, J.D.G. Dura´n, A. Ontiveros-Ortega, M.Espinosa-Jimenez, R. Perea-Carpio, E. Chibowski, Colloids and Surfaces B: Biointerfaces 24 (2002) 297–308, Elsevier.
9. "FT-IR spectrometer, for today's and tomorrow's laboratories Bruker Optik"
10. "Interaction of milk α- and β-caseins with tea polyphenols", Imed Hasni, Philippe Bourassa, Saber Hamdani, Guy Samson, Robert Carpentier, Heidar-Ali Tajmir-Riahi, Food Chemistry, Volume 126, Issue 2, 15 May 2011, Pages 630–639, Elsevier.

Biographies

Santosh Kumar is **B.E** in Electronics and communication engineering (2000) from Govind Ballabh Pant Engineering College Pauri Garhwal, UP India and **M. Tech** in Electronics Design and Technology (2003) from UP Technical University, Lucknow, India. He is currently working as a Scientist at CSIR-CEERI, Pilani, Rajasthan - India. He is doing the PhD on food quality assessments. His research interests are instrumentation, signal processing, process automation, machine Intelligence and soft computing.

PC Panchariya received his M. Sc. in Physics with specialization in Electronics and Ph.D. degree in instrumentation engineering sciences from the Devi Ahilya University, Indore, India, in 1990, 1994 and 2003 respectively. Since 1994, he is working as Sr. principal scientist at the Central Electronics Engineering Research Institute (CEERI), Pilani, India. His research interests include neural networks, fuzzy logic, multivariate data analysis, pattern recognition techniques, and intelligent measurement systems.

P. Bhanu Prasad is professor ACSIR, divisional head of digital systems group CSIR-CEERI, Pilani, Rajasthan, India. Presently he is working as Chief Scientist at CSIR-CEERI. His research interest includes instrumentation, process control automation and signal processing systems for industrial applications.

AL Sharma is professor and head at school of Instrumentation, Devi Ahilya University, Indore, India since 1986. He obtained his M.Sc. degree, M. Phil. and Ph. D. in physics from the Aligarh Muslim University, Aligarh, India in 1971, 1975 and 1978 respectively. His research interest includes fuzzy logics, neural networks and pattern recognition algorithms applied to food quality measurements.