

**Laser Induced Breakdown Spectroscopy studies
with pharmaceuticals, high energy materials
and isomers towards unfolding the device
requirements**

A thesis submitted

to

University of Hyderabad

for the award of

Doctor of Philosophy in Physics



by

Ashwin Kumar M (09ACPA02)

Advanced Centre of Research in High Energy Materials (ACRHEM)

School of Physics, University of Hyderabad

Hyderabad 500046

Telangana, India.

November 2015.

DECLARATION



I, **Ashwin Kumar M**, hereby declare that this thesis entitled “**Laser Induced Breakdown Spectroscopy studies with pharmaceuticals, high energy materials and isomers towards unfolding the device requirements**” submitted by me under the guidance of **Dr. G. Manoj Kumar**, Assistant Professor, ACRHEM, School of Physics, University of Hyderabad, is a bonafide research work which is also plagiarism free. I also declare that it has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma. I hereby agree that my thesis can be deposited in Shodhganga/ INFLIBNET.

A report on plagiarism statistics from the University Librarian is enclosed.

Date:

Ashwin Kumar M

(09ACPA02)



**University of Hyderabad, Central University (P.O.),
Gachibowli, Hyderabad-500 046, India**

CERTIFICATE

This is to certify that the thesis entitled '**Laser Induced Breakdown Spectroscopy studies with pharmaceuticals, high energy materials and isomers towards unfolding the device requirements**' being submitted to the University of Hyderabad by **Ashwin Kumar M** (Reg. No: 09ACPA02), for the award of **Doctor of Philosophy in Physics**, is a bonafide work carried out by him under my supervision and guidance which is a plagiarism-free thesis.

The thesis has not been submitted previously in part or in full to this or any other University or Institution for the award of any degree or diploma.

Dr. G. Manoj Kumar
Thesis Supervisor
(ACRHEM)

Head of the Department/ Center

(ACRHEM)

Dean

(School of Physics)

To
Aayee, Baba
&
Nandu

Words of gratitude

First and foremost I offer my sincerest gratitude to my supervisor, Dr G. Manoj Kumar, Who has supported me throughout my PhD duration with his patience and knowledge whilst allowing me to work in my own way. His profound knowledge, kindness and personality have influenced me in many ways. Without his able guidance, this thesis would not be possible.

I would like to acknowledge present and past directors of ACRHEM, who supported during Ph.D work. I would like to express my gratitude towards the faculty of ACRHEM, Dr. P. Prem Kiran, Dr. A.K. Chaudhary, Dr. S. Venugopal Rao, Dr.G. S. Vaitheeswaran and also other non-teaching staff. I want to thank and acknowledge collaborators Ishan Barman, Narahari Chari Dingari, Ramachandra Rao Dasari and Rashkovsky for giving idea on neural network. I would like to greatly acknowledge DRDO and CSIR for funding.

The seed of my interest were planted at different stages of my education by Vijay Kumar, Jyothsna Sharma, Indumati, Laxmi Prasuna and Dr. Anjaneyulu; who motivated me to choose science as a career. Especially i feel happy to acknowledge my Post Graduation teacher, Mohammad. Najmuddin Khan, currently, working as Head of the Department of physics, Devarakonda, encouraged me to choose research as my career. I am happy to mention my PG batch mates Swarna, Narsimha rao, Narender, Sridhar, prasad sir and my child hood friend Balakrishna who encouraged me to do research.

I am very lucky to have jubilant research group members Tarun, Siva, Rajendhar, Nagendra and Raghavendra. I am blessed and fortunate to have a friends like Venkatesh, Krishna, Leela, Hamad,, Ramakanth, Nageswar Rao and Debasis for their sincere support during my research work. I specially appreciate my project students Yashaswi, Prakash

and Devi Varalakshmi for fruitful basic subject discussions, which enhanced my knowledge of understanding. I feel happy to mention all my friends Sreedhar, Nagendra, Anusha, Ganesh, Kondaiah, Vinoth, Sachin, Rakesh, Manikanta, Konda, Santanu, Mahendhar, Saishiva, Pinnoju, Yedukondalu, Narsimha, Ajay, Fahem, Nagaraju, Abdul, Chandu and Dharavath.

Next, I would like to thank my parents for their unconditional love and support over the years. Without their constant inspiration, this work would not have been possible as well. I would also like to thank all my family members Anand, Sangeetha, Prathibha, Vishnu, Nandini, Geetha, Aneel and Gayathri for their affection and my brother-in-law Bodigum Praveen for supporting me in my endeavors. I always in debt to my friend Nanda Kishore, where words are limited to express about him. I also like to thank from bottom my heart Nadimpalli Siva Kumar sir , Vijaya Lakshmi madam, uncle Trimurutulu for their moral support. At last I would like to thank all Indians, who paid taxes and because of them, the government able to provide funding.

Preface

Spectroscopy deals with the interaction of electromagnetic waves with matter. Atomic and molecular spectroscopic studies can reveal insights into properties of matter. The results of spectroscopy have great importance in various fields like chemistry, medicine, environmental studies, energy research etc. It is well known that atomic spectroscopy involves electronic transitions while molecular spectroscopy involves vibrational and rotational transitions. A major contribution to molecular spectroscopy was provided by Sir C.V Raman, who discovered the celebrated Raman Effect, and was awarded, with Nobel Prize in physics. Due to limitation in power of light sources in the early years of 20th century, it has not gained much attraction from researchers. The invention of laser not only dramatically revolutionized the Raman spectroscopy but also the entire field of spectroscopy. Particularly it has opened new avenues in the field of atomic spectroscopy. Laser Induced Breakdown Spectroscopy (LIBS) is one of such fields which was invented immediately following the invention of the laser.

The first experiments on laser induced plasma were started by Brech and Cross in 1962 [1]. The pioneering work on elemental analysis using laser induced plasma was initiated by David A. Cremers, Leon J. Radziemski at Los Alamos National Laboratory in the early 1980s. The name '*Laser-Induced Breakdown Spectroscopy*' was actually coined by them. It is an atomic emission based technique where a pulsed laser light is focused on to the sample producing a hot plasma [2-4]. It consists of electrons and ions, and emits radiation as it cools down. Typically, several peaks are observed in the LIBS spectrum in 200–1000 nm spectral window. The emission wavelengths are characteristic of atoms/ions present in the plasma and area under the specific curve(s) (i.e. intensity of the specific spectral features) depends on the concentration of the analyte(s) contributing to it. In some cases molecular emissions can also be observed. LIBS has received a great deal of attention from researchers in the past few years because of the many advantages it offers as a spectroscopic analytical tool. One of the most attractive features is the lack of substantial sample preparation requirements, which considerably reduces the typical time needed for the experiment. Samples in any form, be it solid, liquid or gas, can be used. Solid samples can be used in their original form directly for the experiment. Almost all atoms in the periodic table can be explored using this spectroscopy. Furthermore, the total

time needed to perform an experiment and obtain the results is very small. It has gained lot of attention from diverse fields of applications ranging from mars planetary surface exploration [5] to micro surface elemental mapping [6], table top experiments to portable instruments [7], geo-materials [8], explosives [9], biological compounds [10], polymeric materials [11], forensics [12], archaeological materials [13] and agriculture [14]. A fair comparison has been shown between existing analytical techniques and LIBS [15]. It has been shown to be advantageous in terms of less analysis time, low sample consumption, good discrimination, with low cost and least complexity. However, given the existence of many well established techniques for elemental analysis and combined with its disadvantages like ppm level detection limit, stochastic nature of the plasma and lesser precision, LIBS technique has yet not achieved its rightful place. More research work is still needed to translate this technique as an alternative to the existing techniques.

Motivation

Two distinct applications are possible with the use of LIBS spectrum a) Elemental analysis and b) Classification/identification. While the former deals with the identification and quantification of the elemental constituents, the latter classifies/detects a given material as belonging to a particular class. This thesis deals with the latter application in the context of organic materials like high energy materials (HEMs) and pharmaceuticals. There have been several research articles [16-20] in the area of LIBS based classification of materials and is rapidly growing into different fields of applications. The scope of LIBS is particularly more in those application fields where online, insitu and standoff monitoring is a primary requirement. There are few groups around the globe which are investigating and developing the solutions for sensing explosives. In the Indian context, explosive detection is a major issue for homeland security and developing smart solutions is the need of the hour. Specific to HEM detection, there are two principal research goals: first, to identify a specimen as a HEM and second, to classify the specimen as a specific HEM. Yet, HEMs typically are nitro-rich molecules and exhibit grossly similar spectral profiles with prominent emission lines corresponding to nitrogen, carbon, hydrogen and oxygen. Furthermore, the spectral interference emanating from trace quantities of grease, oils [21] and biomaterials [10] as well as the presence of oxygen and nitrogen in the ambient air, and hydrogen and oxygen in the atmospheric moisture impedes the reliable detection of the materials in the field.

Consequently, the basic spectroscopic approach of employing atomic (and in cases molecular) intensity ratios for these kinds of materials can yield very limited results. Also, the substrate effects for HEM residue detection, which have been detailed by Gottfried *et al.* [22], exacerbates the segmentation challenges due to the potential alterations to the light emission from the micro plasma. Taken together, it is not surprising that non-analyte specific peaks appear in the spectral profile due to the entrainment of air and the (variable) substrate information [23]. Approaches that partially alleviate these issues include noble gas purging and double pulse method. However, purging is not a viable option for on-field application and while the double pulse method shows some improvement in the signal enhancement [18], purely ambient air peaks (*e.g.* oxygen and nitrogen) cannot be avoided. There exists a wide scope of research to discover innovative solutions for tackling the problems such as size of the overall system, entrapment of the ambient air into the plasma, data collection protocols and data analysis approaches.

In this thesis work, wide classes of organic compounds viz., explosives, pharmaceutical compounds and isomers were selected. The selection of these particular compounds have their own objectives. Primitively the work was started with conventional data analysis i.e., exploiting the information of spectroscopy by calculating ratios of the peak intensities. As expected, due to complexities in plasma mechanism such as matrix effects and laser fluctuations, the calculated ratios were found to be not interpretable and no clear differentiation was observed using all the above organic compounds data. The main goal of work is to utilize the data of different types of spectrometers and feature inputs, thereby predicting the requirements of instrumentation for building a spectrometer and its optimization for experimentation. For this task, multivariate analysis methods are used. The thesis majorly concentrates to find the solutions towards miniaturization of LIBS detector based on analyzing two types of detectors available in the laboratory: gated and nongated spectrometers. Gated spectrometer ME 5000 is an echelle grating spectrometer equipped with an ICCD, whose output consists of intensities at 25499 wavelengths. It is bulkier with cooling electronics and requires skilled manpower for its operation. In contrast, Maya 2000 is a portable CCD spectrometer without any time gating and output consists of intensity values at 2080 wavelengths. Generally device manufacturer needs prior knowledge about the outcome of acquisition and sample analytics for specific

applications. Without the handshake of both aspects of device and experimental analysis, it is difficult to find the solutions for a research problem. The thesis is organised into six chapters as detailed in the following section.

Organization of Thesis

Chapter 1 discusses about LIBS theory and experimental details. Plasma diagnostics (temperature and electron density) techniques are explained with basic physics involved. The details of data analysis viz., ratiometrics, dendrogram construction using different distance metrics, Principal Component Analysis (PCA), Soft Independent Modelling Of Class Analogy (SIMCA), Partial Least Square Discriminant Analysis (PLSDA) and Artificial Neural Network (ANN) have been systematically illustrated using preliminary data of LIBS.

In chapter 2, a wide variety of organic compounds were chosen to evaluate classification based on LIBS data bank. All the spectra exhibited peaks corresponding to nitrogen, oxygen, hydrogen and carbon. Initial attempts were made at classification by calculating ratios of different elements O to N, H to O, C to O and H to N and compared with the stoichiometric line. But this approach nevertheless failed to find a linear correlation the stoichiometric line. To overcome this hurdle various multivariate analysis methods were applied to the data of five pharmaceutical compounds. In this chapter, multivariate analysis methods viz., PCA, SIMCA and PLSDA were used. An excellent classification rate of 94 % was obtained using SIMCA and 96% with PLSDA.

Chapter 3 is an application towards sensing of explosives. A set of HEM samples acquired from the High Energy Materials Research Laboratory, Pune, India were used for this study. Spectra for all the samples show the presence of emission peaks associated with carbon magnesium, calcium, hydrogen, nitrogen, oxygen and sodium. In addition, the CN violet bands and C₂ swan bands were also observed. The plasma temperature was estimated with Boltzman equation using the oxygen emission lines and found to be in the range of 8000-17000 K. The electron density was estimated to be of the order of 10^{15} - 10^{16} cm⁻³ using the stark broadening of the oxygen line at 844.4 nm. There are two main objectives involved: one is to identify explosive and other is to identify a specific explosive. LIBS data was collected from a set of secondary explosives and several non-

explosives using a gated spectrometer. Neural network analysis has been introduced as a classifier and results are compared with that of PLSDA. Judicious feature selection approaches - specific peaks, sub-spectral regions have been investigated. With specific peak of hydrogen as feature input, a highest rate of classification rate of ~ 97% was obtained among explosive and non-explosives. All the four sub-spectral regions and the full spectrum as feature input lead to almost similar classification rates ranging between 96-98%. For classification among the explosives, the full spectral input resulted in classification rate of *ca.*92%, and sub-spectra from carbon to hydrogen-specific features (~ 60% of the total intensity values) resulted in near identical performance.

Chapter 4 provides insights of application of non-gated LIBS for characterization and classification of organic materials viz., pharmaceutical tablets and HEMs. While use of such nongated system introduces substantive continuum background in the spectral dataset, appropriate treatment of the continuum and characteristic emission results in accurate discrimination of organic compounds of similar stoichiometry. Specifically, results suggest that near-perfect classification can be obtained by employing suitable multivariate analysis methods on the acquired spectra, without prior removal of the continuum background. Furthermore, this chapter even predicts that pre-processing in the form of background removal may introduce spurious features in the signal. Diverse classes of explosives and nonexplosives are classified with high degree accuracies and finally the chapter concludes with the possibility of a portable, non-gated LIBS system as a process analytical tool, for online, insitu and standoff monitoring; given its simple instrumentation needs, real-time capability and lack of sample preparation requirements.

Chapter 5 hints the possibility of using LIBS to identify isomer compounds which are more complicated due to their overwhelming similarity. A set of three home made pyrazoles were employed for this study. The spectra recorded with a non-gated spectrometer shows the presence of a broad featureless background emanating from the continuum emission overriding with sharp peaks characteristic elemental lines of hydrogen, unresolved triplets of oxygen, unresolved triplets of nitrogen, carbon, sodium and potassium. It was shown that a simple nongated LIBS device is sufficient for identification with the aid of suitable multivariate analysis. Despite the presence of continuum in the spectrum a high rate of correct classification has been achieved. Multivariate approaches like PCA and ANN have been utilized. The comparisons of

prediction capability with both gated and nongated have been shown. Using gated and nongated data, an average rate of identification accuracies of 85% and 97.3% respectively have been shown. The key mechanistic principles (causation) behind spectral variations have been investigated and confirmed their statistically significant nature that empowers the excellent classification performance. Further, utilizing the spectra recorded with gated spectrometer, the plasma temperature was estimated with Boltzman equation using the oxygen emission lines was found to be in the range of 10000 - 18000K. The electron density was estimated to be of the order of 10^{15} - 10^{16} using the stark broadening of the oxygen line at 844.4 nm. The plasma parameters showed an exponential decay as a function of time.

Chapter 6 describes the conclusions drawn from the present study and future scope. The future scope of the work is- Experiments can be conducted with explosives in trace quantities on substrates mimicking real time scenario. Further, the concept of LIBS system based on handheld CCD spectrometer can be combined with the above experiments. The isomer detection with LIBS has great potential and can be validated with more sets of isomer compounds and can also be extended to optical isomers. Additionally, a remote/standoff LIBS system is a natural step to take these laboratory results in to the field application. Exploring the possibility of multiplexing Raman spectroscopy with LIBS will be an additional milestone in this research. Furthermore, a large number of unexplored data analysis methods which can possibly lead to a design of a more robust LIBS system can also be examined.

References

- [1] Brech, F., Cross, L. Optical microemission stimulated by a ruby laser. Appl. Spectrosc. 1962, 16, 59.
- [2] Miziolek, A.W., Palleschi, V., Schechter, I. Laser Induced Breakdown Spectroscopy, Cambridge University Press, 2006.
- [3] Cremers, D.A., Radziemski, L.J. Handbook of Laser-Induced Breakdown Spectroscopy, Wiley, 2013.
- [4] Singh, J.P., Thakur, S.N. Laser-Induced Breakdown Spectroscopy, Elsevier Science, 2007.

- [5] Maurice, S., Wiens, R.C., Saccoccio, M., Barraclough, B., Gasnault, O., Forni, O., et al. The ChemCam Instrument Suite on the Mars Science Laboratory (MSL) Rover: Science Objectives and Mast Unit Description. *Space Sci Rev.* 2012, 170, 95-166.
- [6] Menut, D., Fichet, P., Lacour, J.-L., Rivoallan, A., Mauchien, P. Micro-laser-induced breakdown spectroscopy technique: a powerful method for performing quantitative surface mapping on conductive and nonconductive samples. *Appl. Opt.* 2003, 42, 6063-71.
- [7] Guirado, S., Fortes, F.J., Lazic, V., Laserna, J.J. Chemical analysis of archeological materials in submarine environments using laser-induced breakdown spectroscopy. On-site trials in the Mediterranean Sea. *Spectrochim. Acta. Part B.* 2012, 74–75, 137-43.
- [8] De Lucia Jr, F.C., Gottfried, J.L. Rapid analysis of energetic and geo-materials using LIBS. *Materials Today.* 2011, 14, 274-81.
- [9] Gottfried, J., De Lucia, F., Jr., Munson, C., Miziolek, A. Laser-induced breakdown spectroscopy for detection of explosives residues: a review of recent advances, challenges, and future prospects. *Anal. Bioanal. Chem.* 2009, 395, 283-300.
- [10] Assion, A., Wollenhaupt, M., Haag, L., Mayorov, F., Sarpe-Tudoran, C., Winter, M., et al. Femtosecond laser-induced-breakdown spectrometry for Ca²⁺ analysis of biological samples with high spatial resolution. *Appl Phys B.* 2003, 77, 391-7.
- [11] Aquino, F.W.B., Pereira-Filho, E.R. Analysis of the polymeric fractions of scrap from mobile phones using laser-induced breakdown spectroscopy: Chemometric applications for better data interpretation. *Talanta.* 2015, 134, 65-73.
- [12] Rodriguez-Celis, E.M., Gornushkin, I.B., Heitmann, U.M., Almirall, J.R., Smith, B.W., Winefordner, J.D., et al. Laser induced breakdown spectroscopy as a tool for discrimination of glass for forensic applications. *Anal. Bioanal. Chem.* 2008, 391, 1961-8.
- [13] Giakoumaki, A., Melessanaki, K., Anglos, D. Laser-induced breakdown spectroscopy (LIBS) in archaeological science—applications and prospects. *Anal. Bioanal. Chem.* 2007, 387, 749-60.
- [14] Sirven, J.B., Bousquet, B., Canioni, L., Sarger, L. Laser-Induced Breakdown Spectroscopy of Composite Samples: Comparison of Advanced Chemometrics Methods. *Anal. Chem.* 2006, 78, 1462-9.
- [15] Kearton, B., Mattley, Y. Laser-induced breakdown spectroscopy: Sparking new applications. *Nat Photon.* 2008, 2, 537-40.
- [16] Hybl, J.D., Lithgow, G.A., Buckley, S.G. Laser-induced breakdown spectroscopy detection and classification of biological aerosols. *Appl. Spectrosc.* 2003, 57, 1207-15.

- [17] Clegg, S.M., Sklute, E., Dyar, M.D., Barefield, J.E., Wiens, R.C. Multivariate analysis of remote laser-induced breakdown spectroscopy spectra using partial least squares, principal component analysis, and related techniques. *Spectrochim. Acta. Part B.* 2009, 64, 79-88.
- [18] Gottfried, J.L., De Lucia, F.C., Munson, C.A., Miziolek, A.W. Double-pulse standoff laser-induced breakdown spectroscopy for versatile hazardous materials detection. *Spectrochim. Acta. Part B.* 2007, 62, 1405-11.
- [19] Sirven, J.-B., Salle, B., Mauchien, P., Lacour, J.-L., Maurice, S., Manhes, G. Feasibility study of rock identification at the surface of Mars by remote laser-induced breakdown spectroscopy and three chemometric methods. *J. Anal. At. Spectrom.* 2007, 22, 1471-80.
- [20] Myakalwar, A.K., Sreedhar, S., Barman, I., Dingari, N.C., Venugopal Rao, S., Prem Kiran, P., et al. Laser-induced breakdown spectroscopy-based investigation and classification of pharmaceutical tablets using multivariate chemometric analysis. *Talanta.* 2011, 87, 53-9.
- [21] Fortes, F.J., Ctvrtnícková, T., Mateo, M.P., Cabalín, L.M., Nicolas, G., Laserna, J.J. Spectrochemical study for the in situ detection of oil spill residues using laser-induced breakdown spectroscopy. *Anal. Chim. Acta.* 2010, 683, 52-7.
- [22] Gottfried, J.L., De Lucia Jr, F.C., Munson, C.A., Miziolek, A.W. Strategies for residue explosives detection using laser-induced breakdown spectroscopy. *J. Anal. At. Spectrom.* 2008, 23, 205-16.
- [23] De Lucia, F.C., Gottfried, J.L., Munson, C.A., Miziolek, A.W. Multivariate analysis of standoff laser-induced breakdown spectroscopy spectra for classification of explosive-containing residues. *Appl. Opt.* 2008, 47, G112-G21.

Table of Contents

Motivation	2
Organization of Thesis	4
References	6
1. Laser Induced Breakdown Spectroscopy	19
1.1 Introduction	19
1.2 Mechanism involving production of LIBS signal	20
1.3 Physics of LIBS plasma	20
1.4 Plasma diagnostics using Oxygen lines.....	23
1.5 Experimental arrangement	25
1.5.1 Data acquisition.....	25
1.5.2 Sample preparation	27
1.5.3 Optimization of experimental parameters.....	27
1.6 Data analyzing methods	31
1.7 References	39
2. Ratiometrics or Chemometrics?	41
2.1 Introduction	41
2.2 Materials - Experiment.....	43
2.2.1 Sample preparation	45
2.2.2 spectral observations.....	45
2.3 Results	48
2.3.1 Relative intensity of elements	48
2.3.2 Ratiometrics	51
2.3.3 Classification using multivariate methods	57
2.4 Conclusions	65
2.5 References	65

3. Explosive detection using LIBS	67
3.1 Introduction.....	67
3.2 Materials	72
3.3 Plasma parameters	79
3.4 Results and Discussion	81
3.4.1 Explosive and non explosive identification.....	81
3.4.2 Specific Explosive Identification.....	99
3.5 Conclusions.....	104
3.6 References.....	106
4. Nongated LIBS towards field applications	111
4.1 Introduction.....	111
4.2 Materials and methods	114
4.3 Instrument interface by LABVIEW	116
4.4 Spectral description.....	120
4.5 Results.....	123
4.5.1 Compound specific classification.....	123
4.5.2 Explosive/Non Explosive Classification	133
4.6 Discussion.....	134
4.7 Conclusion	136
4.8 References.....	137
5. Discrimination of organic isomer compounds.....	141
5.1 Introduction.....	141
5.2 Materials and methods	143
5.3 Spectral observations	144
5.4 Results and discussion	147
5.4.1 Gated spectrograph configuration	147
5.4.2 Nongated spectrograph configuration.....	154

5.5	Investigation of Plasma parameters.....	159
5.6	Conclusions	162
5.7	References	162
6.	Conclusion and future scope	165
6.1	Conclusion.....	165
6.2	Future Scope.....	167
7.	Publications	170
	Publications related to thesis	170
	Other Publications	170
	Awards	171

Table of figures

Figure 1. Mechanism involved producing LIBS signal on solids. The time shown in the figure corresponds to the time after the laser pulse interaction with sample.....	20
Figure 2. Oxygen lines used for calculation of temperature of plasma.....	24
Figure 3. Experimental arrangement used for LIBS studies.	25
Figure 4. Timing diagram representing laser pulse, continuum, plasma and detector.....	28
Figure 5. Influence of number of laser shots on the LIBS intensity a) 100 shots, (b) 75 shots, (c) 50 shots, (d) 25 shots, (e) 10 shots, (f) 5 shots. t_{100} , t_{75} , t_{50} , t_{25} , t_{10} and t_5 refers to the decay constants of 100, 75, 50, 25, 10 and 5 laser shots.	29
Figure 6. SEM images of craters developed on aluminum thick sheets using (a) 100 shots, (b) 75 shots, (c) 25 shots (d) 10 shots.....	30
Figure 7. Diameter of crater with increasing number of laser firings on the aluminum sample....	30
Figure 8. Organization of Data matrix: Preprocessing and mean centering of data matrix which contain intensities at different wavelengths of samples set.	32
Figure 9. Calculation of covariance matrix using mean centered data matrix.	32
Figure 10. Performing the eigen value problem.	32
Figure 11. The pictorial representation of relation between eigen matrix and eigen values with principle components	33
Figure 12. The schematic figure describing the process of neural network model for pattern recognition.....	35
Figure 13. Performance plot describing the best performance at 69 epoch.	37
Figure 14. Confusion matrix of 470 input spectra. Green colour diagonal indicates the correctly classified while offdiagonal elements depict the misclassification.....	38
Figure 15. LIBS spectra of the pharmaceutical samples used a) paracetamol b) vitamin C c) brufen d) coated brufen e) glucosamine f) glucosamine-coated. O(I) refers to neutral oxygen and Ca II refers to singly ionized calcium.	46
Figure 16. Representative LIBS spectra of organic compounds (a) HMX, (b) 4NP, (c) D- fructose (d) Glycine.....	47
Figure 17: Comparison of relative intensities of C, H and N elements present in organic compounds	50
Figure 18: a) Oxygen and b) Nitrogen LIBS peaks from the Neomycin spectrum. The dots represent the experimental points and the solid lines are the Lorentz fit.	52
Figure 19. Oxygen to nitrogen ratios of all organic compounds. The symbols correspond to measured O/N values from the spectra. Error bars denote the standard deviation of multiple	

spectral data. The diagonal is stoichiometric line deduced from empirical relation of oxygen and nitrogen from the formula.	53
Figure 20: Carbon to oxygen ratio of all organic compounds. The symbols correspond to measured C/O values from the spectra. Error bars denote the standard deviation of multiple spectral data. The diagonal is stoichiometric line.....	54
Figure 21: Carbon to Nitrogen ratio of all organic compounds. The symbols correspond to measured C/N values from the spectra. Error bars denote standard deviation of multiple spectral data. The diagonal is stoichiometric line.....	55
Figure 22: Hydrogen to oxygen ratio of all organic compounds. The symbols are corresponding to actual H/O values versus measured ones. Error bars denote the standard deviation of multiple spectral data.....	56
Figure 23: Hydrogen to Nitrogen ratio of all organic compounds. Actual H/N values are plotted against computed ones. Error bars are standard deviation of data. The diagonal line is stoichiometric line	56
Figure 24: Loadings plot of LIBS data of pharmaceutical compounds. The first three principal components corresponding to the entire spectral dataset acquired from the pharmaceutical samples. These three principal components, combined, explain 96.29% of the net variance in the dataset.	58
Figure 25: PCA scores plot of the first three principal components of the spectral dataset acquired from the six classes of pharmaceutical samples.	59
Figure 26: Bar plot of SIMCA classifications for a representative set of 30 test samples.	62
Figure 27: Classification of HEMs based on their performance, application and analytical importance	68
Figure 28. Organization of current detection technologies (Adopted from Explosives Detection Using Magnetic and Nuclear Resonance Techniques- NATO Science for Peace and Security Series)	69
Figure 29. Representative spectra of HEMs used in the LIBS study	74
Figure 30. Representative spectra of Dextrose, HMX, RDX, Cytosine and Glycine.	76
Figure 31. The Decay of LIBS signal of TNT compound with a function of time delay.....	77
Figure 32. The decay of elemental peaks C, H, N, O and molecular peak CN as a function of time delay. Upper quarter table shows the decay constants of the above peaks.....	77
Figure 33. Box plot illustrating the statistics of (a) plasma temperature and (b) electron density of different explosive compounds.....	79
Figure 34. The statistics of (a) plasma temperature, (b) electron density using non explosives.	80
Figure 35: Score plot of Non explosives and explosives.....	82
Figure 36: Loadings plot of explosives and non explosives.....	83

Figure 37 Score plot of non-explosives and explosives with data normalized to area under the curve.....	84
Figure 38. Loadings plot using normalised data of explosives and non-explosives	85
Figure 39. Full spectrum division into four sub spectral regions.....	86
Figure 40. Score plot using sub spectrum R1	87
Figure 41. Score plot using R2 sub spectrum.	88
Figure 42. Loadings plot using R2 sub spectrum.....	88
Figure 43. Score plot using sub spectrum R3	89
Figure 44. Loadings plot using subspectrum R3.....	90
Figure 45. Score plot using R4 sub spectrum	91
Figure 46. Loadings plot using sub spectrum R4.....	91
Figure 47. Score plot using carbon peak profile centered at 247.8nm.....	94
Figure 48. Score pot using CN peak profile	95
Figure 49. Score plot using hydrogen peak of non explosives and explosives at 656.6 nm	96
Figure 50. Score plot using nitrogen peaks of non explosives and explosives around 742, 744, 746 nm.	96
Figure 51. Score plot using oxygen peak of non explosives and explosives centred at 777.5 nm.	97
Figure 52: Dendrogram constructed between explosives and non explosives using euclidean distance as metrics.	99
Figure 53(a) Score plot, (b)loadings plot, (c)Scree plot using full spectrum of explosives.....	100
Figure 54. Score plots using (a) sub spectrum R1, (b)sub spectrum R2, (c)sub spectrum R3 and (d)sub spectrum R4.	101
Figure 55. Cumulative variances of PCs of four spectral regions R1, R2, R3 and R4.	102
Figure 56. Screenshot of the labview VI developed for non gated spectrometer(MAYA 2000).	117
Figure 57. Screenshot of the labview VI developed for measurement of laser power using power meter	118
Figure 58. Screenshot of the labview VI developed for stage controller. Pulse operating frequency and the size of the sample also considered for movement of the axis.	119
Figure 59. Representative LIBS spectra acquired from the pharmaceutical formulation (a) Cetirizine dihydrochloride; (b) Cipro pure; (c) Metformin hydrochloride; (d) Ciprofloxacin hydrochloride. Intensity on the y-axis is normalized with respect to the characteristic hydrogen emission peak at 656 nm.....	120
Figure 60. Representative LIBS spectra acquired from high energy material pellets (a) NTO; (b) RDX; (c) TATB; (d) TNT. Intensity on the y-axis is normalized with respect to the characteristic oxygen emission peak at 777 nm.	121

Figure 61. Nongated LIBS signal dominated with continuum . The atomic lines related to C,H,N and O are sitting on broad emission of continuum	123
Figure 62. Plot of the first three principal components corresponding to the entire spectral dataset acquired for all pharmaceutical samples. These three principal components, combined, explain 96.5% of the net variance in the dataset.	124
Figure 63. Discrimination of pharmaceutical samples based on their LIBS spectra: Scores plot corresponding to principal components 1, 2 and 3 for the spectral dataset acquired from the four samples. The data points corresponding to Cetirizine dihydrochloride, Cipro pure, Metformin hydrochloride and Ciprofloxacin hydrochloride are indicated by green squares, blue circles, black asterisks and yellow inverted triangles, respectively.....	125
Figure 64. Scree plot of HEMs used in the study	126
Figure 65. Plot of the first three principal components corresponding to the entire spectral dataset acquired for all samples of HEMs. These three principal components, combined, explain 99.5% of the net variance in the dataset.....	127
Figure 66. Discrimination of HEMs based on their LIBS spectra: Scores plot corresponding to principal components 1, 2 and 3 for the spectral dataset acquired from the four samples.	128
Figure 67. Hierarchical clustering using the dendrogram representation for LIBS spectra acquired from the 4 sets of pharmaceutical samples.....	129
Figure 68. Hierarchical clustering using the dendrogram representation for LIBS spectra acquired from the four sets of HEMsamples.....	129
Figure 69: Background corrected LIBS spectra acquired from the pharmaceutical formulations: (a) Cetirizine dihydrochloride; (b) Cipro pure; (c) Metformin hydrochloride; (d) Ciprofloxacin hydrochloride. Intensity on the y-axis is normalized with respect to the characteristic hydrogen emission peak at 656 nm. The continuum background is approximated and removed using a least squares-based polynomial curve-fitting algorithm.	135
Figure 70. The LIBS spectra of isomers 1Nitro pyrazole, 3-1H-Nitro pyrazole and 4-1H-Nitro pyrazole possessing formula $C_3H_3N_3O_2$	144
Figure 71. Time evolution of LIBS signal using 3NP with laser energy 92 mJ and step, gate width are fixed to 1 μ s and initial delay chosen 0.5 μ s.	146
Figure 72. Time evolution of LIBS signal using 4NP with laser energy 92mJ and step, gate width are fixed to 1 μ s and initial delay chosen 0.5 μ s	146
Figure 73. Time evolution of individual peaks and fit using exponential decay, j refers to decay constant.....	147
Figure 74. The score plot of 1NP, 3NP and 4NP with LIBS data collected by single scan	148

Figure.75: Scree plot shows the individual components versus corresponding variance associated with it, while the red colour circle denoted cumulative variance with increasing order with number of components.....	148
Figure 76. The score plot of 1NP, 3NP and 4NP with LIBS data acquired using accumulation of three pulses.....	149
Figure 77. Loadings plot of Principal components:This figure explains the correlation of original data with that of principal components	150
Figure 78. The score plot of isomer compounds using LIBS data obtained with 19 μ s gate width. The variation between 3NP and 4NP has found to be increased and cluster of all three compounds are closely packed.	151
Figure 79: Scree plot of number of components versus associated variance. Bar plot shows the individual components versus corresponding variance associated with it, while the red color circles denotes cumulative variance.....	152
Figure 80: Non gated LIBS spectra of 1NP, 3NP and 4NP	154
Figure 81. Plot of the first three principal component loading profiles corresponding to the dataset constituted by spectra recorded from the isomers 1NP, 3NP and 4 NP.....	155
Figure 82. Clustering of 1NP, 3NP and 4NP samples based on principal component analysis...	156
Figure 83. Plot of the spectral profiles of (A) hydrogen and (B) oxygen emission lines from the three classes of isomers, 1NP (red), 3NP (blue) and 4NP (green). The solid profile depicts the mean spectrum of each sample group and the shadow represents ± 1 S.D. Box plot of the FWHM of the corresponding Lorentzian fits of (C) hydrogen and (D) oxygen.	158
Figure 84: Electron temperature obtained using Boltzmann plot method The Electron temperature is calculated using oxygen peaks at 777.5, 794.8 and 844.4 nm.....	159
Figure 85. Box plot of plasma temperatures of 1NP, 3NP and 4NP. Small square box denotes mean and 'x' represents outliers.....	160
Figure 86. Box plot for electron density of 1NP, 3NP and 4NP.....	161
Figure 87. Temporal decay of electron temperature and electron density of 3NP and 4NP isomer compounds.	161

ACRONYMS

AES	Atomic Emission Spectroscopy
CCD	Charge Coupled Device
FTIR	Fourier Transform Infrared spectrometry
FWHM	Full Width at Half Maximum
ICCD	Intensified Charge Coupled Device
ICPAES	Inductively Coupled Plasma Atomic Emission Spectroscopy
ICPMS	Inductively Coupled Plasma Mass Spectrometry
LIBS	Laser Induced Breakdown Spectroscopy
LIPS	Laser Induced Plasma Spectroscopy
LOD	Limit of Detection
LTE	Local Thermodynamic Equilibrium
MCP	Micro Channel Plate
MS	Mass Spectroscopy
MVA	Multivariate Analysis/Multivariate Approach
NMR	Nuclear Magnetic Resonance
NQR	Nuclear Quadrupole Resonance
PC	Principal Component
PCA	Principal Component Analysis
Ppm	Part Per Million
Ppb	Part Per Billion
RSD	Relative Standard Deviation
SD	Standard Deviation
SEM	Scanning Electron Microscopy

SNR	Signal to Noise Ratio(S/N)
XRF	X-Ray Fluorescence
XRD	X-Ray diffraction
ns	Nanosecond
μs	Microsecond
ms	Millisecocond

1. LASER INDUCED BREAKDOWN SPECTROSCOPY

In this chapter, the characteristics of LIBS, explicitly the overview of spectroscopy, physics involved in laser produced plasma and plasma diagnostics methods are explained briefly. Additionally advantages and applications are also mentioned to illustrate the versatility of LIBS. Experimental details and some important optimization parameters are provided in detail. Finally, data analyzing methods- multivariate analysis methods viz., Principal Component Analysis (PCA), Soft Independent Modeling of Class Analogy (SIMCA) and Artificial Neural Network (ANN) are explained step by step elucidation.

1.1 Introduction

Atomic emission spectroscopy (AES) is related to the study of emission of radiation produced by atoms and ions which are in excited energy state. It is essentially an elemental analysis method which uses arc, flame, plasma or spark as a source of excitation. The emission at specified wavelength(s) can be considered as a fingerprint of the material. The number of atoms related to the elements are proportional to intensity of emitted light. The steps involved in AES are: atomization/ vaporization of the specimen, excitation of atoms and detection of light emitted. Unlike AES, LIBS does not employ external physical device for excitation rather simply uses plasma produced by focusing the laser itself. The light emitted by tiny plasma contains the information of atomic and molecular species. Winefordner et al., have compared the mostly used AES techniques with LIBS considering different figures of merits such as wide interest in research activity and application demand in various fields and considered it as "*a future super star*"[1].

When a pulsed laser focuses on a sample, (attaining power at focus typically for gases in the order of 10^{11} watts/cm²) it creates plasma on the surface. The light emitted from the plasma constitutes spectral information of the sample. The spectroscopic investigation involving this optical emission from plasma is called laser induced breakdown spectroscopy[2].

1.2 Mechanism involving production of LIBS signal

When a pulsed laser beam is focused on to the surface of the material, material absorbs the incident laser energy and material starts to evaporate at typically 1 ns time delay. At 5 ns time delay plasma is formed and emits continuum radiation typically till 1 μ s. Finally the sample specific information is emitted till 20 μ s depending upon laser energy and the material chosen for study. The following figure 1 gives the pictorial representation of plasma formation mechanism [3].

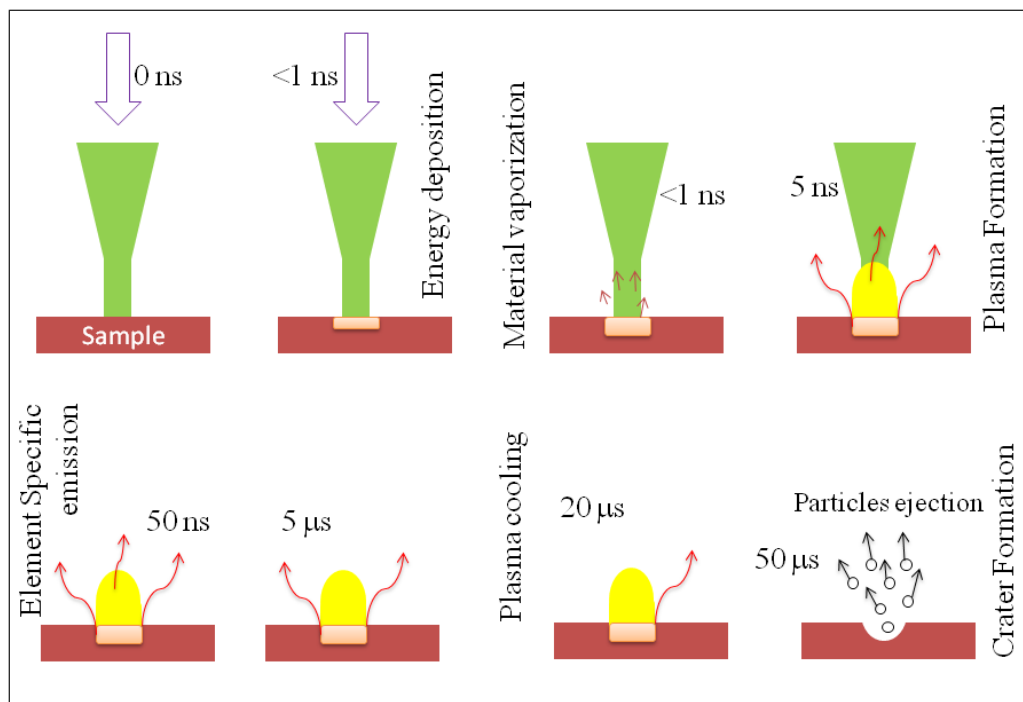


Figure 1. Mechanism involved producing LIBS signal on solids. The time shown in the figure corresponds to the time after the laser pulse interaction with sample.

1.3 Physics of LIBS plasma

The state of plasma and its physical parameters can be described by thermodynamic equilibrium i.e., the plasma system which comprises electrons, atoms, ions and radiation can be explained at the same temperature by the equilibrium distribution laws of statistical thermodynamics. In *complete thermodynamic equilibrium*, all processes pertaining plasma system are balanced and can be characterized by a single temperature.

The equilibrium between the processes can be explained by principle of detailed balance. If radiation law doesn't hold well where processes are mostly dominated by collisions, then the existed thermodynamic equilibrium is called as *local thermodynamic equilibrium*. The four laws are given as follows:

- a. **Atomic & ionic population density:** The ratio of a populations of two discrete energy level E_i and E_j of an atom/ion follow the Boltzmann distribution law

$$\frac{n_i}{n_j} = \frac{g_i}{g_j} \times e^{\frac{-(E_i-E_j)}{k_B T_e}} \quad (1)$$

Where n refers to number of atoms, g is statistical weights of an energy level, E_i , E_j are the energy of upper energy level 'i' and lower energy level 'j'. k_B is Boltzmann constant and T_e is the temperature of plasma (electron).

- b. **Particles velocity distribution:** All particles velocity distribution can be explained by Maxwellian distribution. The equation governing distribution is [4],

$$f(u)du = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left(-\frac{mu^2}{2kT}\right) u^2 du \quad (2)$$

Where fdu is the number of particles per cubic meter possessing velocity between u and $u+du$ and other symbols have their usual meanings.

- c. **Distribution of ionized stages:** The ratio of population densities for two ionization stages of one type of element/species is given by Saha-Eggert equation,

$$n_e \frac{n_+}{n_1} = 2 \frac{g_+}{g_1} \frac{(2\pi m k_B T)^{3/2}}{h^3} e^{\frac{\chi_1}{k_B T}} \quad (3)$$

Where n_e is the electron density, n_+ and n_1 are number densities of ionic and atomic species, χ_1 is the ionization energy which is (E_+-E_1) .

- d. **Radiation law:** The governing equation of intensity of radiation can be described by Planck function,

$$I_\nu = B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/k_B T} - 1} \quad (4)$$

LTE exist when only the first three laws hold good. LIBS data analysis relies on the assumption of LTE. The necessary but not sufficient condition for plasma to be in LTE is

given by a criteria called *Mc Whirter criterion* [5], which gives the threshold electron density for stationary, homogeneous and optical thin plasma as

$$N_e \geq 1.6 \times 10^{12} (T[K])^{\frac{1}{2}} (\Delta E[eV])^3 \quad (5)$$

however the review on LIBS part 1 made by Hahn and Omenetto, have suggested that McWhirter criterion is not alone a necessary condition to test whether the condition of LTE reached; additionally temporal and spatial variations should be calculated to confirm the existence of LTE [6]. At smaller time delays N_e is larger than the value obtained using equation 5 and while larger delays even the criterion may not be satisfied.

The integrated line intensity I_{ij} of a spectral line corresponding to the transition between the upper level i and the lower level j is given by [7]

$$I_{ij} = \frac{g_i A_{ij}}{U^s(T)} n^s \exp\left(-\frac{E_i}{k_B T}\right) \quad (6)$$

where $U^s(T)$ is the partition function, n^s is the number density of species in the plasma, g_i and A_{ij} are statistical weight and transition probability of the upper level respectively, and k_B is Boltzmann constant. The plasma temperature can be estimated by constructing the Boltzmann plot considering multiple peaks of a particular element which differ in their E_i values. Rewriting Eq.6 gives a straight line equation between with E_i as independent parameter and $\ln \frac{I_{ij}}{A_{ij} g_i}$ as dependent parameter.

$$\ln \frac{I_{ij}}{A_{ij} g_i} = \ln \frac{n^s}{U^s(T)} - \frac{E_i}{k_B T} \quad (7)$$

The slope of the line is related to the temperature of plasma and intercept can be related to concentration.

The ratio of intensity of two spectral lines from the same ionization stage, the excitation temperature is given by

$$\frac{I_1}{I_2} = \left(\frac{\lambda_{nm}}{\lambda_{ij}}\right) \left(\frac{A_{ij}}{A_{nm}}\right) \left(\frac{g_i}{g_n}\right) \exp\left(\frac{E_i - E_n}{k_B T}\right)$$

Where I_1 is the line intensity from the i - j transition and I_2 is that from the n - m transition.

When the upper energy level difference is more than 2eV, then the accuracy of temperature is good. Optical thin verification can be performed by calculating branching ratio of spectral lines from the same ionization stage as

$$\frac{I_1}{I_2} = \left(\frac{\lambda_{nm}}{\lambda_{ij}} \right) \left(\frac{A_{ij}}{A_{nm}} \right) \left(\frac{g_i}{g_n} \right) \quad (8)$$

The electron density can be calculated using stark broadened profile of the spectral line. The observed width of a line in the LIBS spectra is a result of Stark broadened [2, 8], because the contribution to spectral line width by Doppler and pressure broadening are negligible. The broadening of the spectral line due to above all mechanisms is given by,

$$\Delta\lambda = 2w \left(\frac{N_e}{10^{16}} \right) + 3.5 A \left(\frac{N_e}{10^{16}} \right)^{1/4} \left[1 - \frac{3}{4} N_D^{-1/3} \right] w \left(\frac{N_e}{10^{16}} \right)$$

Rearranging the above equation and only considering stark broadening, the electron density is given by

$$N_e = \frac{\Delta\lambda \times 10^{16}}{2w} \quad (9)$$

where $\Delta\lambda$ is the FWHM of the line after correcting for the instrument broadening, w is the electron impact width parameter [9].

1.4 Plasma diagnostics using Oxygen lines

Throughout the thesis, organic compounds such as pharmaceuticals, high energy materials and isomers are seldom used. The selection of spectral lines from C, H, N and O is a critical step for plasma diagnostics. Carbon peak has emission lines at the both ends of the spectrometers i.e., in the UV region peak at 247.89 nm and in the near infra red at 833.23 nm. These lines of carbon peak may not be a good choice due to the drastic change in quantum efficiency. Hydrogen peak emission comes in the centre of the spectrometer at 656.32 nm and no other emission wavelength related to hydrogen is observed explicitly. In the case of nitrogen all upper energy levels has similar/ closely matching values. It cannot be used for temperature measurements. Oxygen lines, shown in [figure 2](#), are the only elements in the organic compound which posses fairly different numerical values of upper energy level and are well separated. These are the lines concentrated in the same spectral region with almost same quantum efficiency. The

parameters associated with oxygen lines are drawn from NIST database and it is tabulated [table 1](#) [10]; where λ_{ij} (nm) is wavelengths of the oxygen lines, E_i (eV) is the upper energy level in eV, g_i is statistical weight and A_{ij} is the transition probability.

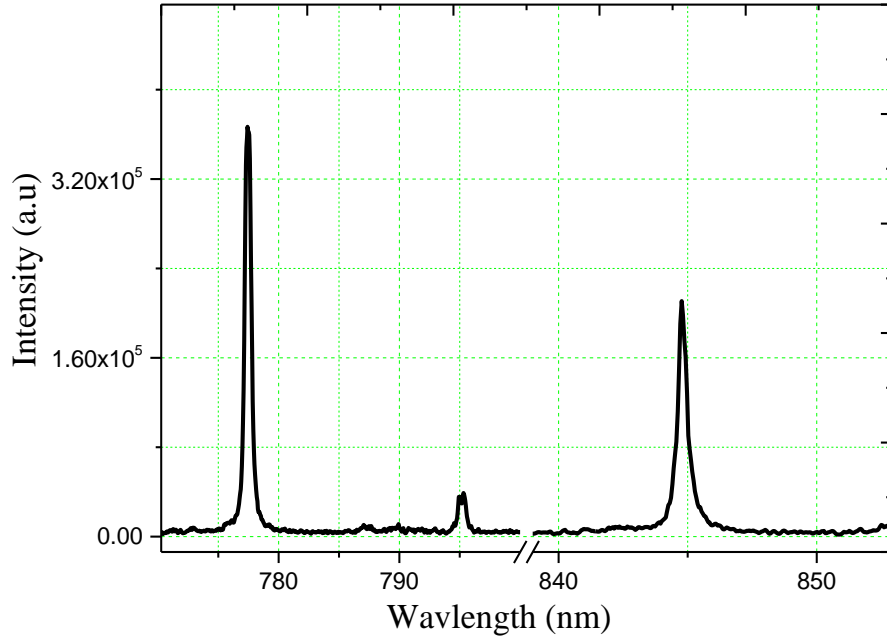


Figure 2. Oxygen lines used for calculation of temperature of plasma.

The atomic peaks are fitted using lorentzian function, because the line shape is due to stark broadening. The area under the curve is considered as intensity of spectral line rather considering height. Because the height involves large fluctuations, whereas area under is least affected. Boltzmann plot method is used for calculation of temperature of plasma as given by equation 7. The slope gives the temperature in eV.

λ_{ij} (nm)	E_i (eV)	g_i	$A_{ij}(10^7\text{s}^{-1})$
777.34	10.74	7	3.69
794.93	14.1	9	3.73
844.65	10.98	5	2.80

Table 1. Atomic data of oxygen spectral lines.

Electron density is calculated using stark broadened profile of the spectral line at 844.4 nm and instrumental profile correction also been considered. LIBS offer several advantages over other techniques.

Advantages of LIBS

- Minimal sample preparation
- Time consumption is less
- Sample in any form may be solid, gaseous or liquid
- Remote analysis
- Atomization and excitation are in one step
- Capable of simultaneous multi-element analysis
- Simple set-up

1.5 Experimental arrangement

1.5.1 DATA ACQUISITION

The experimental arrangement used for LIBS studies is shown [figure 3](#). Laser pulses from a second harmonic of Nd:YAG laser (InnoLas, Spitlite-1300) at 532 nm (7 ns pulse

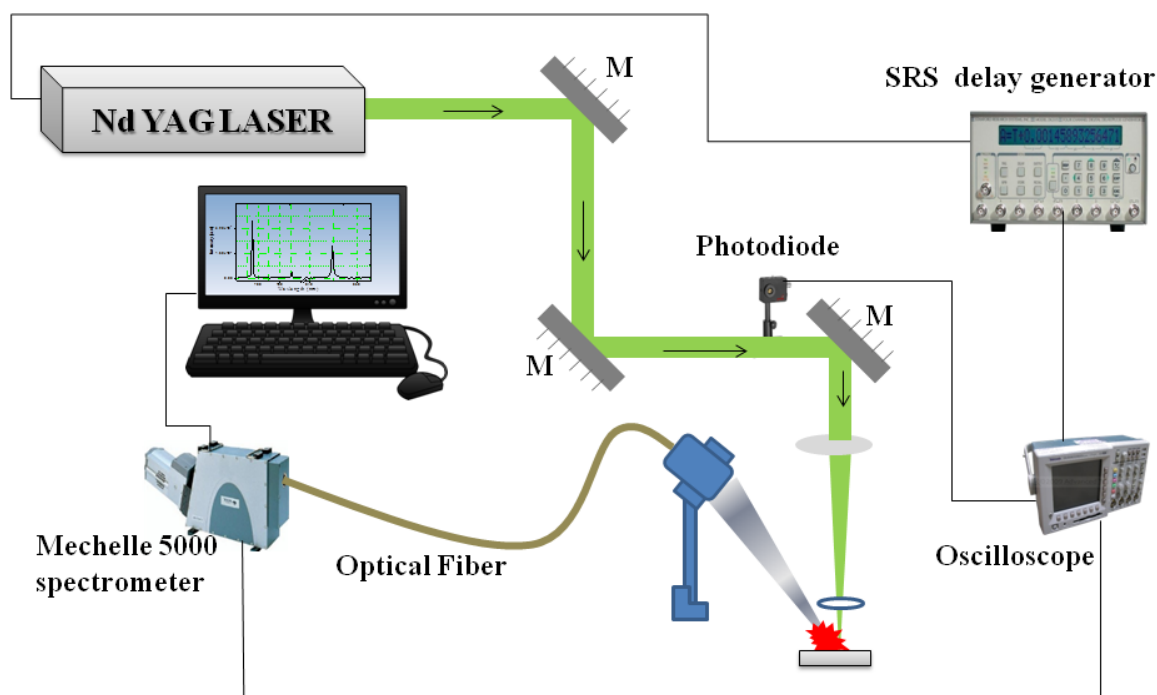


Figure 3. Experimental arrangement used for LIBS studies.

width, 10 Hz repetition rate) were focused onto the sample using an 80 mm convex lens. The signal was collected using a lens system and was coupled to the spectrograph.

Spectra were recorded with an integration time of 1-20 μ s and a delay of 0.5-1 μ s, the range defers for different investigations in the subsequent chapters. The delay refers to the time difference between the opening of the ICCD gate and the incidence of the laser pulse on the sample. The Pockels cell output of the laser triggered a SRS delay generator (Stanford Research Systems, Inc.), which in turn provided a TTL pulse to trigger the ICCD.

In the case of non-gated measurements Maya 2000(Ocean optics Inc.,) spectrometer was used. The spectrometer did not support gating electronics and it has no facility of gating. The minimum integration time possible with the spectrometer is 6 ms.

	Maya 2000	Mechelle
Spectral range Effective (nm)	200 – 1100 300-1044	200 – 975 200-850
Focal length (input)	101.6 mm	195 mm
Aperture	F/4	F/7
Pixel size (μ m square)	14	13
Number of pixels Active	2080 x 20 2048 x 14	1024 x 1024
Detector	Hamamatsu S9840, back-thinned FFT- CCD	i star DH734_18 mm- ICCD
Quantum efficiency	>90% peak	Peak up to 27%
Sensitivity	~0.45 counts/e-	-
Thermoelectric cooling	No	-35°C
Integration Time	6 ms – 10 s	Minimal optical gate width 2 ns

optical resolution FWHM (nm) at 632.8	~1.3	~0.14
Grating type	Reflection grating	Echelle grating
Entrance Slit μm^2	25	50
Triggering mode	Quasi Real-time	External and internal
Dimensions	149 mm x 109.3 mm x 50.4 mm (0.96 kg)	350 mm x 100 mm x 200mm (10 kg)

Table 2. Comparison of gated and non-gated spectrometers used for LIBS measurements.

1.5.2 SAMPLE PREPARATION

Fine powder material was pressed into 10 mm diameter pellets by a die-hydraulic pressing machine by applying load of 3-4 tonne. The pelletization operation improves signal reproducibility, as the position of the focal spot is almost unchanged for all the laser pulses arriving at the pellet surface in sharp contrast to a powder specimen. The theoretical maximum density for all the pelleted samples was in the range of 91-94% for organic explosives which has been given chapter 3. The samples were placed on a manual X-Y-Z translation stage such that spectra could be collected from different spatial locations in the pellet to estimate measurement variability.

1.5.3 OPTIMIZATION OF EXPERIMENTAL PARAMETERS

1.5.3.1 INSTRUMENT SYNCHRONIZATION

The pulsed laser is used essentially have the pulse width 7 ns. It is already discussed that the plasma formation occurs typically 5-6 ns after the laser pulse interacts with the sample. Emission of continuum predominates at the early stage of plasma about 1 μs . The elemental specific information emitted after 1 μs and lasts for 20-25 μs . In order to get good measurement the time window of detector should be opened till 25 μs with initial delay 1 μs . The time delay to avoid continuum is generally called as gate delay and the total open time of detector is called as gate width. The Gate delay and gate width of the spectrometers are provided by gating electronic incorporated in the ICCD architecture.

The timing diagram for synchronizing detector for LIBS signal measurement as been shown in figure 4

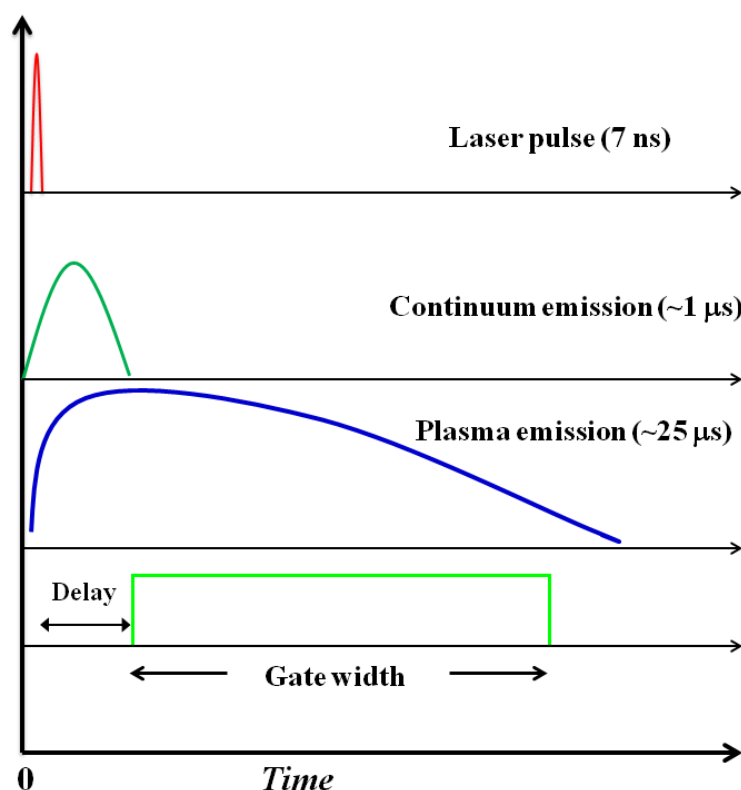


Figure 4. Timing diagram representing laser pulse, contiuum, plasma and detector.

1.5.3.2 DETERMINATION OF SAMPLE STEP SIZE

The motivation behind this work is twofold: first, to find the optimum number of laser shots, where the intensity change should be less than 30% of first laser shot. The second one is to determine minimum sample step size based on crater diameter formation due to number of laser shots.

The solution to the first one has importance in reducing the sample consumption and lack of sample scarcity. This is the same case with the organic compounds, especially in the case of explosives; the sample availability is a major problem; where utilization of explosives in bulk quantities is not appreciated/restricted/limited. The second one can help researchers to decide minimum step movement for each measurement. Aluminum sample was used for this study. LIBS signal was collected by ensuring experimental parameters constant throughout the investigation viz., laser energy, sample interaction

position and distance of optics collection. The same experiment was repeated with different laser energies 10 mJ, 20 mJ and 30 mJ.

Keeping the laser energy constant at 10 mJ consecutive shots of 5 pulses on the aluminum sheet, produces intensity variation for the 5th pulse is 72% of its first pulse intensity. Decay of intensity has been observed as a function of number of shots fired on the same place as shown in [figure 5](#) and the corresponding SEM images of the crater

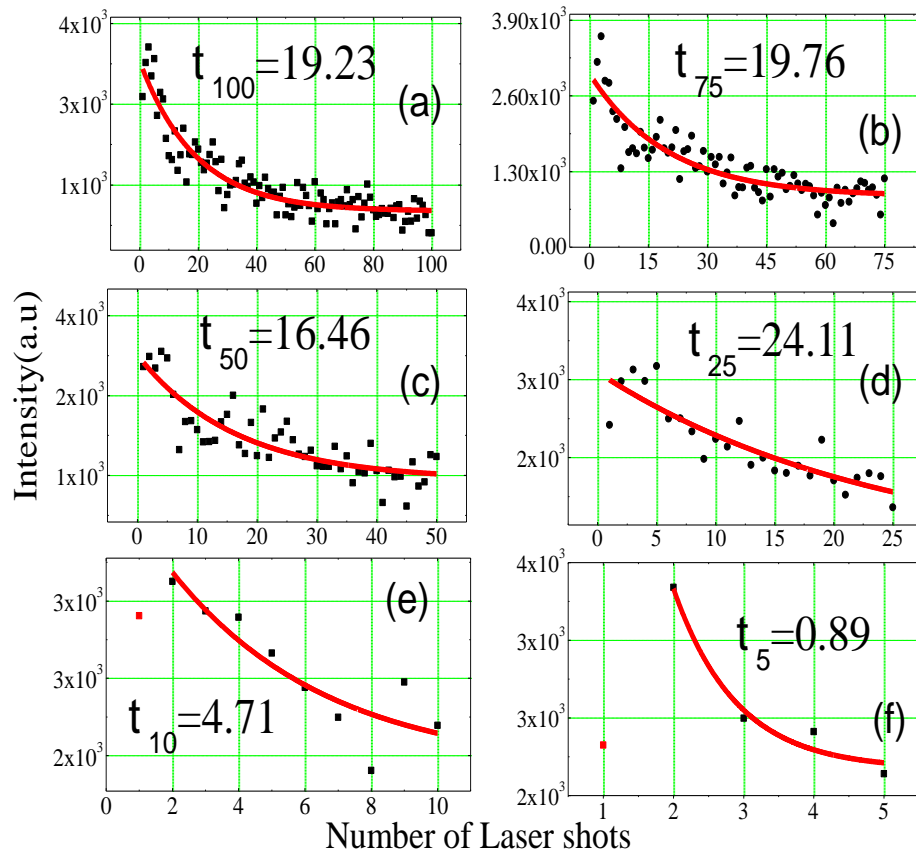


Figure 5. Influence of number of laser shots on the LIBS intensity a)100shots, (b)75 shots, (c)50 shots, (d)25 shots, (e)10 shots, (f)5 shots. t_{100} , t_{75} , t_{50} , t_{25} , t_{10} and t_5 refers to the decay constants of 100, 75, 50, 25, 10 and 5 laser shots.

produced has shown in [figure 6](#). Inference from visual inspection can be drawn as the number of shots are increased, the diameter and depth of the crater also increasing. The graphs in the [figure 5](#) are fitted using exponential decay and the decay constant is found to less for 5 shot investigation. This outcome gives an idea to opt five laser shots at a same place where sample scarcity exist and also multiple number of trials for user interests. The maximum diameter achieved using 30 mJ is about 100 μm , hinting the step

size should not below this value while moving the stage ([figure 7](#)). Although these are specific to the material under investigation, they suggest order typical number of shots.

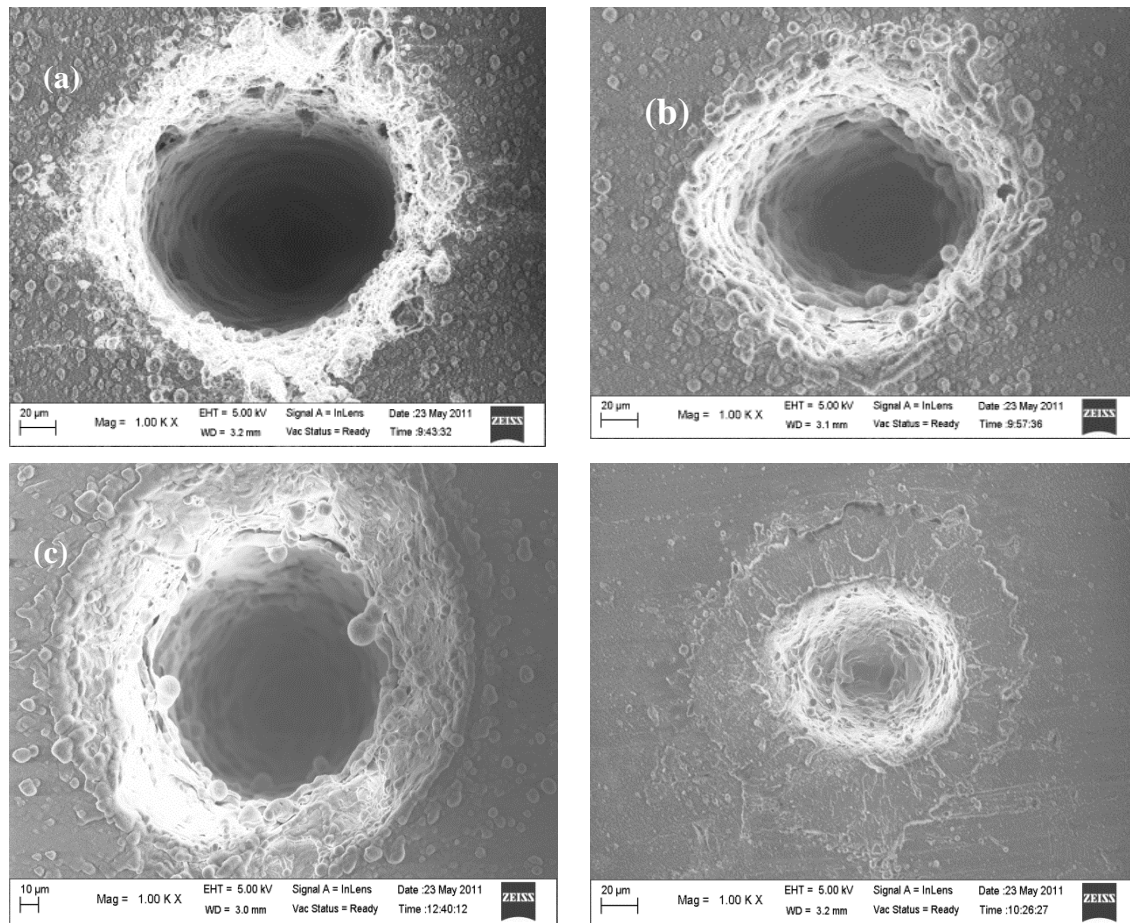


Figure 6. SEM images of craters developed on aluminum thick sheets using (a) 100 shots, (b) 75 shots, (c) 25 shots (d) 10 shots

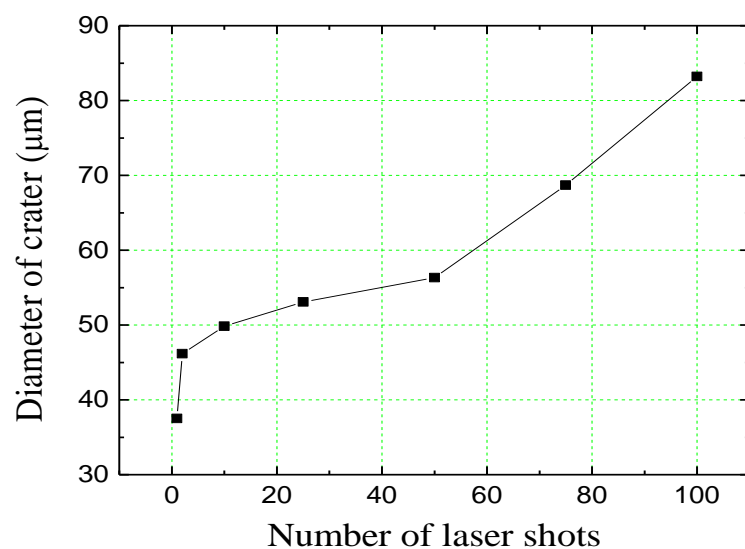


Figure 7. Diameter of crater with increasing number of laser firings on the aluminum sample.

1.6 Data analyzing methods

In the thesis, two types of multivariate analysis methods are used: Unsupervised and supervised classification methods.

(A) PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is one of the most extensively used multivariate statistical techniques and represents a powerful tool for exploratory data analysis and for making predictive models [11]. The linear multivariate PCA models are developed using orthogonal basis vectors (eigenvectors), which are called principal components. Linear combinations, as used in PCA models, are especially attractive because of their ease of computation and analytical tractability. The principal components model the statistically significant variation in the data set as well as the random measurement error. Thus, PCA can be viewed as a projection method that transforms the high-dimensional data onto a lower dimensional space by combining the otherwise distinct features in the acquired spectroscopic data. Since PCA seeks a projection that best represents the data in a least-squares sense, it can be employed to eliminate the principal components associated with noise, thereby reducing the dimensionality of complex problems and minimizing the effects of measurement error.

In PCA one performs a linear mathematical transformation of the data into a new coordinate system such that the largest variance lies on the first axis and decreases thereafter for each successive axis. For the analysis presented in this article, PCA can be thought of as an unsupervised classification technique that separates the samples into clusters based on the variance of their corresponding LIBS spectra. The coordinates along the principal components (also known as scores) are plotted to visualize the similarities between respective spectra thereby linking similar spectra and also demarcating separate classes of spectra. Ideally, the separation obtained based on PCA analysis of LIBS spectra would correspond to the separate classes of over-the-counter pharmaceutical drugs tested in this study. The Pictorial step-wise implementation of PCA algorithm is as shown in figures 8-11.

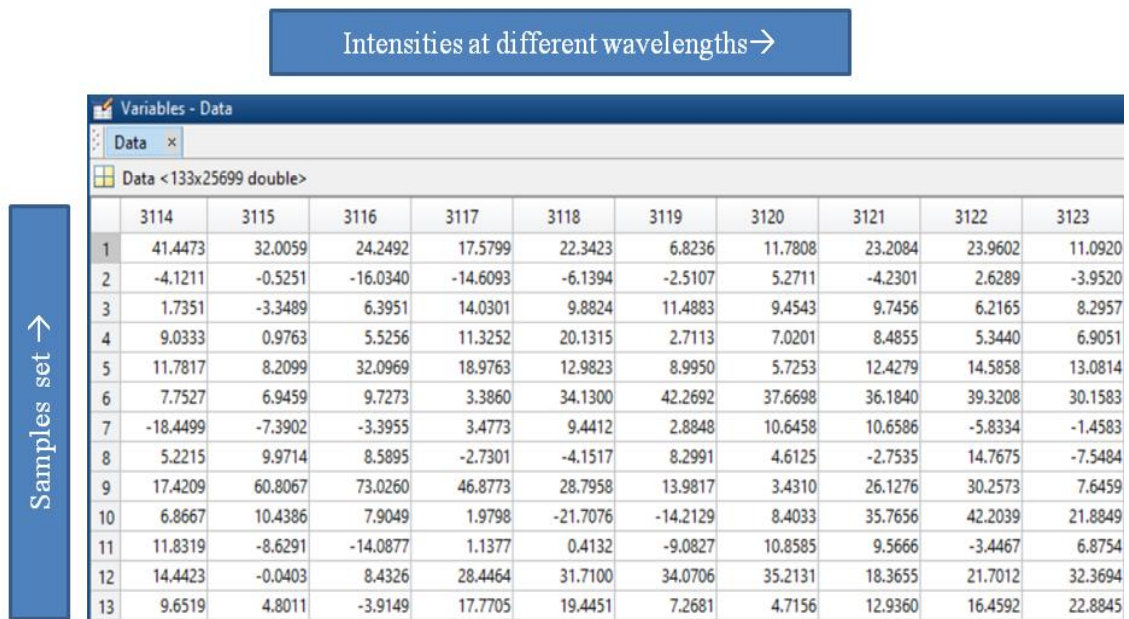


Figure 8. Organization of Data matrix: Preprocessing and mean centering of data matrix which contain intensities at different wavelengths of samples set.



Figure 9. Calculation of covariance matrix using mean centered data matrix.

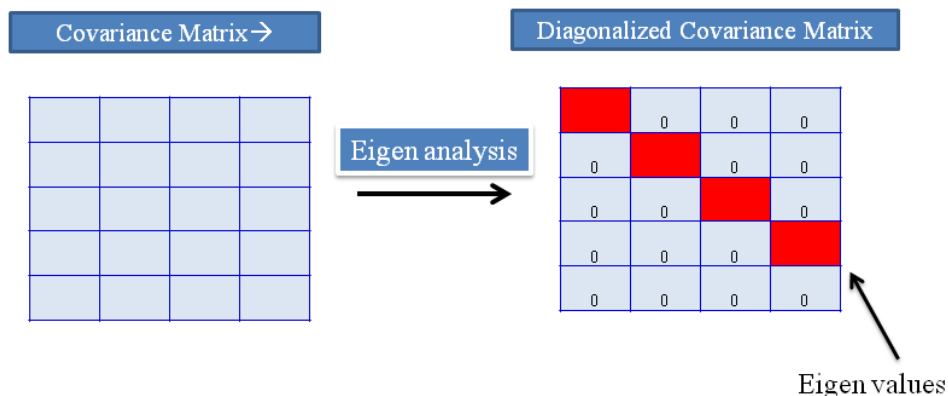


Figure 10. Performing the eigen value problem.

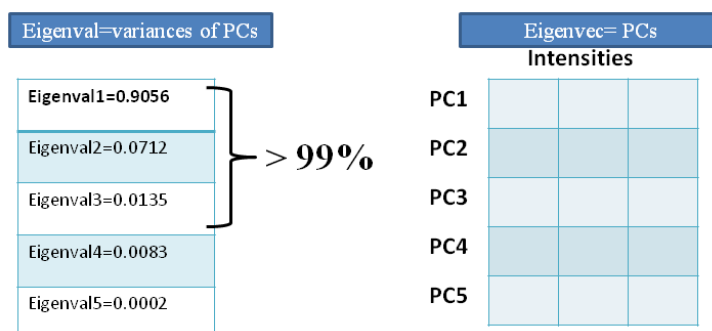


Figure 11. The pictorial representation of relation between eigen matrix and eigen values with principle components

(B) SIMCA

SIMCA is a supervised classification technique, which incorporates the application of PCA for dimensionality reduction thereby rendering it suitable for classification of high-dimensional observations [12, 13]. Because of its supervised nature, it necessitates a training data set consisting of samples with a set of attributes (*e.g.* LIBS spectra) and, importantly, their class membership (*e.g.* type of pharmaceutical tablet). The primary idea of soft modeling refers to the fact that the classifier can identify samples as belonging to multiple (overlapping) classes and is not constrained to producing a classification of samples into strictly discrete (non-overlapping) classes. Classically, this represented a major shift in understanding from hard modeling (*e.g.* conventional discriminant analysis) - although it is possible to have class distances derived from hard models that are proximal to two or more groups.

Of greater significance to our present analysis is the fact that SIMCA enables independent modeling of the classes as opposed to an overall variance modeling as performed in PCA. In SIMCA, each group (as defined by the class membership information) is independently modeled using its characteristic number of principal components. (The optimization of number of principal components retained in our models is further detailed in the ensuing Data Analysis section.) For each class, the resulting model then describes a line (for one PC), plane (for two PCs) or hyper-plane (for more than two PCs). The class distance is estimated as the geometric distance (*e.g.* mean orthogonal distance) from the respective PC models. SIMCA-based predictive classification is subsequently performed by comparing the residual variance of the prospective sample with the mean residual variance of the training samples belonging to

the specific class. A F -test is often performed to determine the similarity (or lack thereof) of the sample residual variance to the mean residual variance of the class. Such significance testing can be extended to compute the statistical probability (*i.e.* the likelihood) of the prospective sample belonging to the particular class.

The SIMCA approach has been widely adopted for a wide range of applications as it offers a multiple advantages over other classification techniques. First, SIMCA inherits the strengths of PCA-based dimensionality reduction, which as mentioned before reduces the chances of spurious correlations and noise corruption. Second, the prospective sample is only assigned to a class with a sufficiently high probability (above a pre-assigned threshold value) rather than to "the nearest class". This reduces the chances for misclassification as for any sample there always exists a nearest class, even if the sample is an outlier or belongs to a class which is not represented in the training set. Third, SIMCA provides a robust classification method in the sense that it does not require a large training dataset nor does it constrain the number of measurement variables. This represents a significant improvement over standard discrimination techniques which do not work under such scenarios due to well-characterized problems of collinearity and chance classification [14]. This feature makes SIMCA a particularly attractive classification tool for spectroscopic characterization, where a large number of measurement variables is almost always encountered.

(C) **PLSDA**

Although it was originally developed as a multivariate regression method, PLS [15] has been extensively used to solve classification problems by encoding the class membership of the measured samples in the target matrix. [16, 17] An important feature of this supervised technique is that it is specifically suited to deal with problems in which the number of variables is large (compared to the number of observations) and collinear, two major challenges encountered when LIBS spectra are used. Furthermore, PLS-DA provides a powerful tool for discrimination even when the variability within a class is similar to the inter-class variability by virtue of establishing the maximal separation between each class via fitting one global model to the entire dataset. In PLS-DA, a PLS regression model is calculated that relates the independent variables (spectral data) to a binary response encoded as: [1,0,0,0,0] for sample belonging to class 1, [0,1,0,0,0] for

sample belonging to class 2, and so on until class 5. Typically, a threshold value based on the Bayesian method,[18] is defined between 0 and 1, and calculated on the basis of the values predicted during the classification process, so an object is assigned to a particular class if its prediction is larger than the threshold value for this class. As in the PLS regression model, the optimal number of latent variables (LVs) retained are chosen before the modeling process and this is determined by using the fractional misclassification error rate in cross validation [19].

(D) ARTIFICIAL NEURAL NETWORK

In this section Neural network is elucidated by an example of pharmaceutical nongated LIBS data

Description of selection of Input for Neural Network:

Consider Total spectra 470 samples of LIBS spectra have 2080 wavelength. Input is a 2080×470 matrix representing 470 samples of 2080 elements. Target is a 5×470 matrix representing 470 samples of 5 elements.

ANN model:

In pattern recognition problems, NN used for classify inputs in to a set of target categories. A two-layer feed-forward network, with sigmoid hidden and output neurons, classifies vectors arbitrarily well, given 10 neurons in its hidden layer. The schematic picture shown in figure (9), it describes the 2080×470 elements as input, hidden layer neurons 10. The network was trained with scaled conjugate gradient back propagation method1 [20].

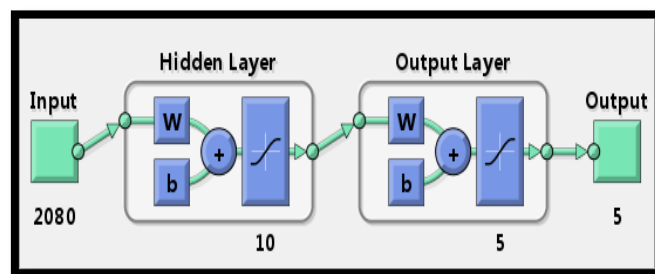


Figure 12. The schematic figure describing the process of neural network model for pattern recognition.

For the classification task, neural network pattern recognition tool box, Matlab 2010a, Mathworks Inc., was used. This toolbox helped selecting data, creating and training a network and evaluating its performance using a mean squared error and confusion matrices.

Testing and Validation of data:

Training: These are presented to the network during training, and the network is adjusted according to its error.

Validation: These are used to measure network generalization, and to halt training when generalization stops improving.

Testing: These have no effect on training and so provide an independent measure of network performance during and after training.

The total input randomly divided up in to three parts, 1.training, 2.validation, 3.testing. For training 70%, validation 15% and for testing 15% of the inputs were used randomly. Training will stop according to mean squared error reduces to zero or approaches to zero.

To determine when the training should be stopped, an stopping criteria based on the validation set was used [21]. The number of epochs was not relevant in this case. To avoid an over fitting of the NN model, the learning process was repeated while the verification mean square error (MSE), defined in Eq. (1), was decreased:

$$MSE = \frac{1}{N} \sum_k^N (r_k - y_k)^2$$

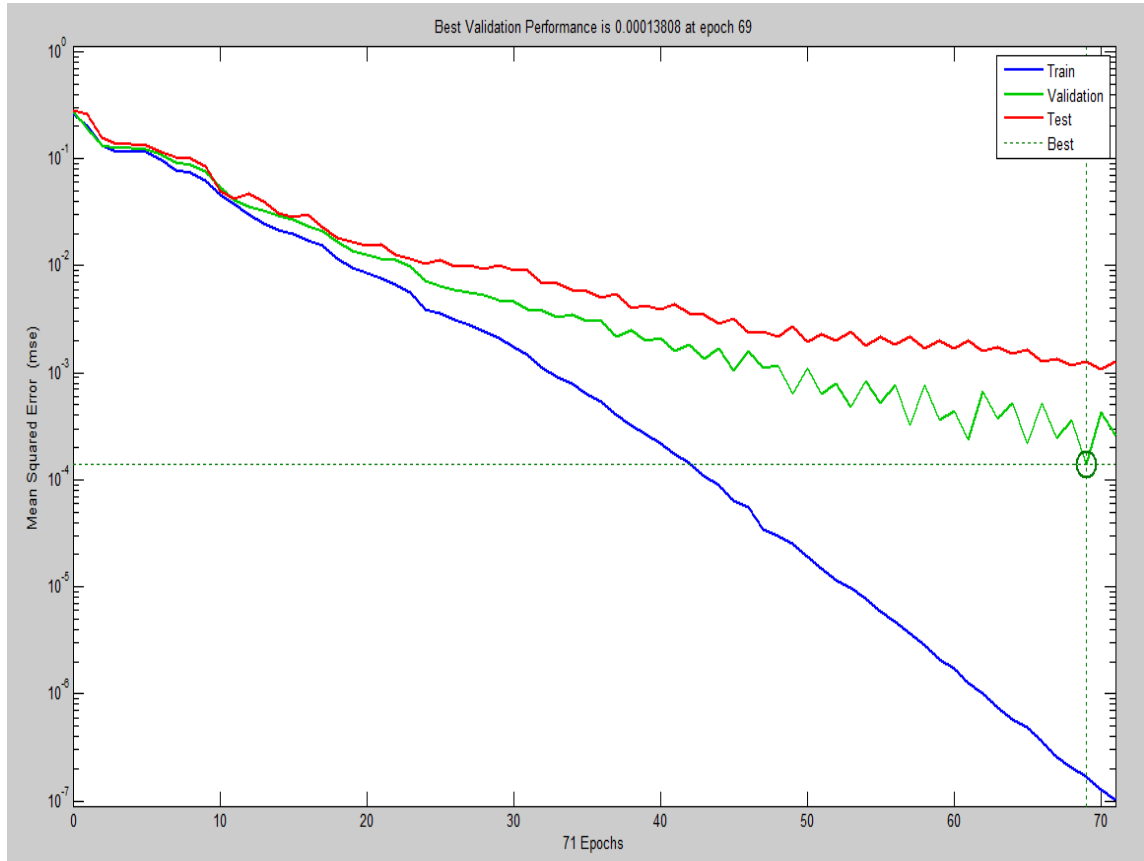


Figure 13. Performance plot describing the best performance at 69 epoch.

Where N , y_k , and r_k are the numbers of input data, the response from each output neuron, and the real output response, respectively. A detailed process of calculation is given in literature [22]. Mean Squared Error is the average squared difference between outputs and targets. Lower values are better. Zero means no error.

Results and Discussions for NN:

The NN was trained for spectral characteristics using the input data of 470 samples. To get the reproducibility, repeatability and for robustness NN was trained, validated and tested 100 times. While doing NN training, the training will stop when MSE approaches to zero. A typical graph shown in figure 11, describes the training, when it approaches 0.00013808 at epoch 69. This graph is called performance plot. A confusion matrix is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one, in our case neural network. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The figure 12 represents the confusion matrix of one of the

NN . In this confusion matrix, of the 110 of four pharmaceutical compounds and 30 air spectra, the system predicted that four were pharmaceutical compounds correctly, and of the 30 are air LIBS spectra. All correct guesses are located in the diagonal of the table, so it's easy to visually inspect the table for errors, as they will be represented by any non-zero values outside the diagonal.

Confusion Matrix

1	110 23.4%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	110 23.4%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	110 23.4%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	110 23.4%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	30 6.4%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%
	1	2	3	4	5	

Target Class

Figure 14. Confusion matrix of 470 input spectra. Green colour diagonal indicates the correctly classified while offdiagonal elements depict the misclassification

After applied neural network 100 times, the total resultant classification table was shown in table 2. This table describes correct classification, unclassification and misclassification. If the predicted sample same as that of actual it is said to be correctly classified, while if predicted is different from actual it is said to be misclassification. If predicted does not provide any of actual one it falls in to category of unclassification.

Average rate of ..	Correct classification	Misclassification	Unclassified
Cetirizine Di hydrochloride IP	98.76	1.24	0
Cipro pure	98.69	1.31	0
Metformin hydrochloride IP	98.17	1.83	0
Ciprofloxacin hydrochloride USP	99.2	0.8	0
Air	97	3	0
Total average	98.6	1.4	0

Table 3. Classification performance of neural network on the nongated LIBS data of pharmaceutical & air data combined. The results are shown are iterated over 100 runs.

1.7 References

- [1] Winefordner, J.D., Gornushkin, I.B., Correll, T., Gibb, E., Smith, B.W., Omenetto, N. Comparing several atomic spectrometric methods to the super stars: special emphasis on laser induced breakdown spectrometry, LIBS, a future super star. J. Anal. At. Spectrom. 2004, 19, 1061-83.
- [2] Singh, J.P., Thakur, S.N. Laser-Induced Breakdown Spectroscopy, Elsevier Science, 2007.
- [3] Noll, R. Laser-Induced Breakdown Spectroscopy: Fundamentals and Applications, Springer Berlin Heidelberg, 2012.
- [4] Athay, R.G. Radiation Transport in Spectral Lines, Springer Netherlands, 2012.
- [5] Huddleston, R.H., Leonard, S.L. Plasma diagnostic techniques, Academic Press, 1965.
- [6] Hahn, D.W., Omenetto, N. Laser-induced breakdown spectroscopy (LIBS), part I: review of basic diagnostics and plasma-particle interactions: still-challenging issues within the analytical plasma community. Appl. Spectrosc. 2010, 64, 335A-66A.
- [7] Miziolek, A.W., Palleschi, V., Schechter, I. Laser Induced Breakdown Spectroscopy, Cambridge University Press, 2006.

- [8] Cremers, D.A., Radziemski, L.J. Handbook of Laser-Induced Breakdown Spectroscopy, Wiley, 2013.
- [9] Griem, H. Spectral Line Broadening by Plasmas, Elsevier Science, 2012.
- [10] NIST data base http://physics.nist.gov/PhysRefData/ASD/lines_form.html.
- [11] Brereton, R.G. Applied Chemometrics for Scientists, Wiley, 2007.
- [12] Wold, S. Pattern recognition by means of disjoint principal components models. Pattern Recognition. 1976, 8, 127-39.
- [13] Blomquist, G., Johansson, E., Söderström, B., Wold, S. Classification of fungi by means of pyrolysis-gas chromatography-pattern recognition. Journal of Chromatography A. 1979, 173, 19-32.
- [14] Gemperline, P. Practical Guide To Chemometrics, Second Edition, CRC Press, 2006.
- [15] Suykens, J.A.K., Vandewalle, J. Least Squares Support Vector Machine Classifiers. Neural Processing Letters. 1999, 9, 293-300.
- [16] Shao, X., Bian, X., Liu, J., Zhang, M., Cai, W. Multivariate calibration methods in near infrared spectroscopic analysis. Analytical Methods. 2010, 2, 1662-6.
- [17] Balabin, R.M., Lomakina, E.I. Support vector machine regression (SVR/LS-SVM)-an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data. Analyst. 2011, 136, 1703-12.
- [18] Witten, I.H., Frank, E. Data Mining: Practical Machine Learning Tools and Techniques, Second Edition, Elsevier Science, 2005.
- [19] Verboven, S., Hubert, M. LIBRA: a MATLAB library for robust analysis. Chemometrics and Intelligent Laboratory Systems. 2005, 75, 127-36.
- [20] Møller, M.F. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks. 1993, 6, 525-33.
- [21] Christopher, B. Neural Networks. In: Handbook of Neural Computation, Taylor & Francis, 1996.
- [22] M.H. Beale H.B. Demuth, M.T. Hagan, "Neural Network Toolbox for Use with MATLAB: User's Guide 9th for Version 6.0 " (2008).

2. RATIOMETRICS OR CHEMOMETRICS?

A wide variety of organic compounds are chosen to evaluate the task of identification/classification based LIBS data. Initial attempts at classification are made by calculating different ratios O/N, H/O, C/O and H/N to match the stoichiometric line. But this approach nevertheless failed to find a linear correlation with that of the stoichiometry. This could be on account of stochastic nature of plasma and interference from ambient air. To overcome this hurdle various multivariate statistical methods (chemometrics) are applied to the data of pharmaceutical compounds. Multivariate statistical methods such as PCA, SIMCA, PLSDA are successfully applied for identification of pharmaceutical compounds.

2.1 Introduction

The goals of analytical methods are majorly two-fold: identifying the presence of elements in the compounds and the identification/classification of materials. Qualitative information of spectral features enable the discovery of the elemental constituents which may be part of compound or impurities. Development over last three decades in LIBS instrumentation and analysis made it applicable in most of the fields as discussed in chapter 1. Particularly the latter problem of classification/identification has shown promising applications. Identification of organic compounds which are CHNO type compounds are of particular interest and these are different classes of materials used for study such as pharmaceuticals, HEM, polymers, plastics. The discrimination based on their LIBS spectra is an arduous task due to similarity in their spectra. Some of the factors which pose challenges are: atmospheric contribution, matrix effect, self absorption and laser pulse fluctuations. The identification of materials based on LIBS have many applications: it can aid a pharmacologist to identify a compound at drug discovery stage [1]; assist detection of explosives on site in the field [2]; helpful in polymer testing and waste management [3]; assist law enforcement agencies in the identification of counterfeit drugs to identification of isomers in industrial process. LIBS is an ideal tool in such applications given in inherent advantages.

In this chapter : Two different methods have been studied towards correlating the spectra with the materials and also tries to find out a suitable methodology which is robust and it will be used widely in the subsequent chapters. In the First method , different element ratios have been calculated from the spectra and compared with the actual stoichiometric ratios. Similarly, one can also employ these data interpretation and analysis tools to classify among various classes of samples. However, precise quantitative analysis of LIBS datasets using conventional univariate model (where the intensity of a specific peak is correlated with the elemental concentration) is very difficult due to the uncontrollable fluctuations of the experimental parameters and the physical and chemical matrix effects [4]. Such fluctuations can distort the theoretical relationship between the intensity and concentration, presenting substantive challenges to the univariate approach. Further, such univariate modeling is extremely difficult to implement in multi-element samples, where inter-element interference due to overlap of the spectral signatures is unavoidable. Hence, there is need of using multi element based analysis to cope up the real world problems.

In this perspective, while LIBS investigations of several materials have been reported in the literature, the detection and compositional analysis of pharmaceutical samples have received much less attention [5-7]. In this chapter, other organic compounds of different categories have been selected: Carbohydrates, amino acids, explosives and pharmaceuticals' LIBS data are exploited for the feasibility of routine classification of compounds. First intensity of spectral lines of C, H, N and O are selected to calculate ratios, this approach will be referred as ratiometrics. Various ratios like O/N, C/O, C/N, H/O and H/N are calculated and compared with the stoichiometric line. Most of the compounds' ratios are not in good agreement with the line and the results are scattered. Though enough care has been taken to get repeatable spectra with less standard deviation of LIBS signals by enhancing signal to noise ratio, this approaches fails to get an acceptable classification. To overcome this bottleneck in the detection process, different multivariate methods, namely Principal Component Analysis (PCA) and Soft Independent Modeling of Class Analogy (SIMCA) and Partial Least Square Discriminant Analysis (PLSDA) and artificial neural network are implemented to exploit the multivariate nature of the LIBS data. These multivariate methods are generally referred to as 'chemometrics' in LIBS literature. This analysis is performed only on the pharmaceuticals data. The results on HEMs will be discussed extensively in chapter 4 and 6. The results show

excellent prospective classification accuracy using supervised classifications, demonstrating the usefulness of this methodology in process analytical technology, especially for fast on-line process control monitoring applications in the pharmaceutical industry and standoff detection of explosives in the warfare.

2.2 Materials - Experiment

Experimental studies are undertaken to accomplish the following objectives: (1) comparing the C, H, N and O signal relative intensity with the sample stoichiometry (2) comparing the classification performance of compounds using ratiometrics to multivariate methods. To accomplish these objectives, a total of 28 samples, from four categories of pharmaceuticals, carbohydrates, amino acids and explosives are chosen. Among them common tablets like brufen, glucosamine, paracetamol and vitamin c are pharmaceutical compounds were purchased over the counter from the local medical shop; D-fructose and Dextrose are carbohydrates; Glycine, Hypoxanthine, Ltryptophan, Methionine, Neomycin, Pyridine and thymine are amino acids; triethylamine, sodium lauryl sulphate and EDTA are chemicals; pyrazole, 1 nitro pyrazole, 3 nitro pyrazole, 4 nitro pyrazole, HMX, NTO, PETN, RDX and TNT are explosives. The details of compounds are listed in table 1. Second column of the table 4 shows the name of compound name/abbreviated name; second column is for chemical formulae and third is for number of spectra used for the analysis.

S.No	Compounds	Formulae	Number of spectra
1	Vitamin C	$C_6H_8O_6$	15
2	Paracetamol	$C_8H_9NO_2$	20
3	Glucosamine with coating	-----	10
4	Glucosamine with non coating	$C_6H_{13}NO_5$	10
5	Brufen with coating	-----	15
6	Brufen with non coating	$C_{13}H_8O_2$	15
7	Thymine	$C_5H_6N_2O_2$	32
8	TEA(Triethanolamine HCl)	$C_6H_{15}NO_3$	21

		HCl	
9	SLS(Sodium lauryl sulphate)	NaC ₁₂ H ₂₅ SO ₄	46
10	Pyridine	C ₅ H ₅ N	35
11	Neomycin	C ₂₃ H ₄₆ N ₆ O ₁₃	37
12	Methionine	C ₅ H ₁₁ NO ₂ S	43
13	Ltryptophan	C ₁₁ H ₁₂ N ₂ O ₂	61
14	Hypoxanthine	C ₅ H ₄ N ₄ O	31
15	Glycene	C ₂ H ₅ NO ₂	50
16	EDTA	C ₁₀ H ₁₆ N ₂ O ₈	37
17	Cytosine	C ₄ H ₅ N ₃ O	34
18	Dfructose	C ₆ H ₁₂ O ₆	30
19	Dextrose	C ₆ H ₁₂ O ₆	23
20	4 Niro pyrazole	C ₃ H ₃ N ₃ O ₂	101
21	3 Nitro pyrazole	C ₃ H ₃ N ₃ O ₂	88
22	1 Nitro pyrazole	C ₃ H ₃ N ₃ O ₂	76
23	Pyrazole	C ₃ H ₄ N ₂	83
24	TNT	C ₇ H ₅ N ₃ O ₆	143
25	RDX	C ₃ H ₆ N ₆ O ₆	61
26	PETN	C ₅ H ₈ N ₄ O ₁₂	60
27	NTO	C ₂ H ₂ N ₄ O ₃	75
28	HMX	C ₄ H ₈ N ₈ O ₈	133

Table 4: Details of organic samples used in the study, here spectra refers to usable spectra from the original recorded spectra

2.2.1 SAMPLE PREPARATION

Over-the-counter drug samples were purchased from a local pharmacy. The details of the samples are provided in table1 from s. no. 1-6. The tablets were used directly without any pre-processing. Laser pulses with the energy in the range of 25-35 mJ were used. Spectra were recorded with an integration time of 1 μ s and a delay of 0.5 μ s. Initially, twenty spectra from each sample are acquired after averaging over ten consecutive pulses. It is worth noting that the acquired spectra are screened for further analysis by subjecting them to a minimum signal-to-noise ratio threshold and performing spectral outlier detection using student's t-test and the Mahalanobis distance function [8]. The resultant spectra from each sample are averaged prior to the ensuing analysis. However, in order to prevent any spurious effects on our classification models, chemometric analysis was performed without the application of any additional pre-processing methods.

Carbohydrates and amino acids were purchased from Sigma Aldrich, Hyderabad; while pyrazole compounds are homemade chemicals. Explosives HMX, NTO, PETN, RDX and TNT were acquired from High Energy Materials Research Laboratory (HEMRL), Pune, India. These powders were pressed into pellets with a constant pressure of 5 tons for *ca.* 3 min using a hydraulic pressure machine. Pelletizing the powder is of significance as it typically causes a dramatic improvement in the reproducibility of the LIBS data [9]. Such an increase can be attributed to the improved rigidity of the pellets in comparison to the powders, which ensures that the position of the focal spot is almost unchanged for all the laser pulses arriving at the pellet surface during a data acquisition cycle. Furthermore, the pellet forms are more representative of the solid dosage forms available commercially and thus the variability encountered in the pellet experiments are expected to closely simulate real-world conditions. Laser energy of 20-25 mJ was used. Initial delay of 1 μ s, gate width of 2 μ s was chosen for all the spectral measurements.

2.2.2 SPECTRAL OBSERVATIONS

The LIBS spectra of these are shown in Figure 15. All the spectra exhibited peaks corresponding to nitrogen, oxygen, hydrogen and carbon. Typical representative LIBS spectra of each carbohydrates, amino acids and explosives are been shown in Figure 16. These lines can be attributed to the primary components of the samples. Though the

primary component of brufen and vitamin C do not contain any nitrogen in them, yet their LIBS spectra show the nitrogen peaks. The possible reason for observing these nitrogen peaks in these samples could be the

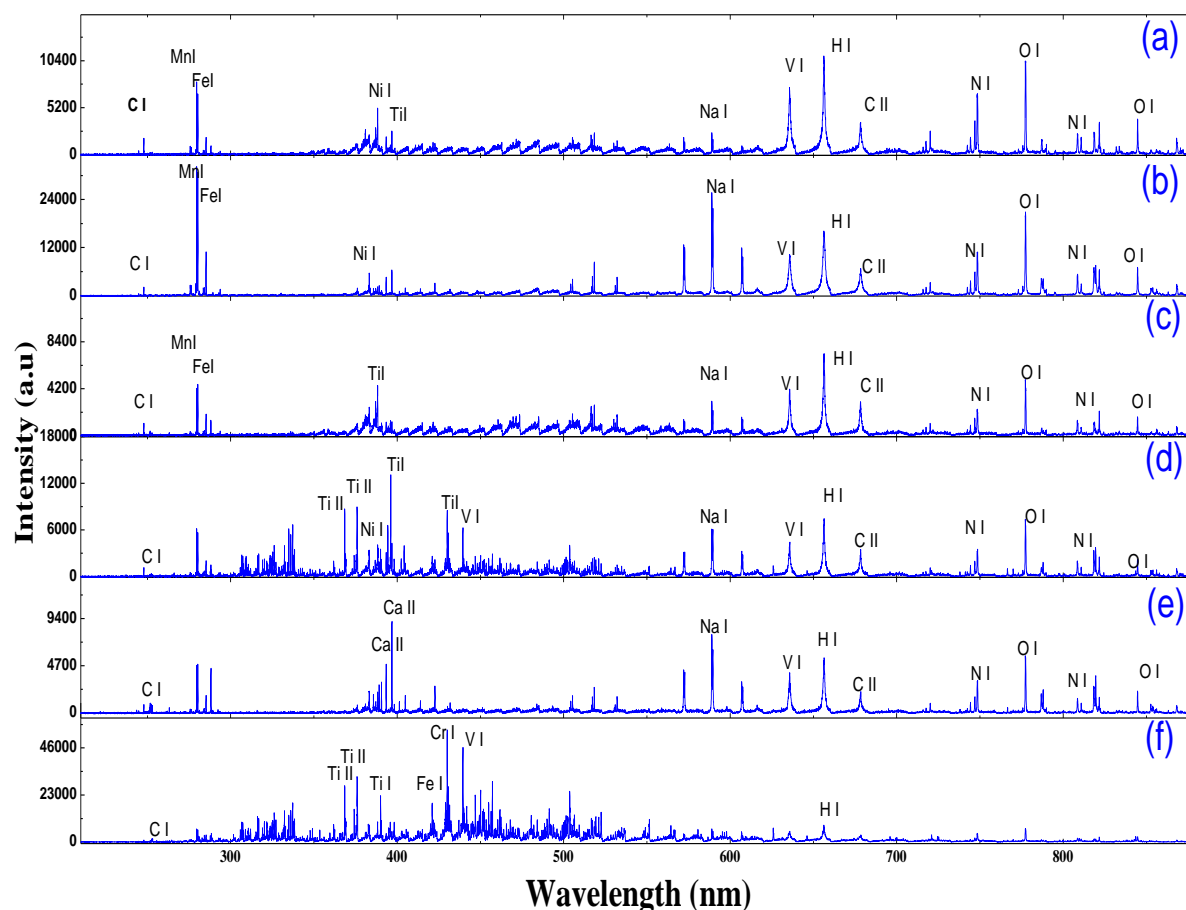


Figure 15. LIBS spectra of the pharmaceutical samples used a) paracetamol b) vitamin C c)brufen d) coated brufen e) glucosamine f) glucosamine-coated. O(I) refers to neutral oxygen and Ca II refers to singly ionized calcium.

presence of other ingredients such as flavoring and coloring agents that are added to the pharmaceutical tablets and entrapment from atmosphere. Differences were also observed with respect to coated and uncoated samples. (Here, coated samples refer to the tablets that are available with colored coatings on them.) For these coated samples, the spectra were first recorded directly with the coating. The protocol outlined in Missaghi and Fassihi [10] was used for removal of coating ensuring the relative flatness of sample. The spectra were again recorded. While both the recordings showed peaks corresponding to carbon, hydrogen, oxygen and nitrogen, the coated spectra showed strong peaks corresponding to titanium. The spectra also exhibited peaks corresponding to iron, manganese, sodium, vanadium, magnesium, titanium and calcium.

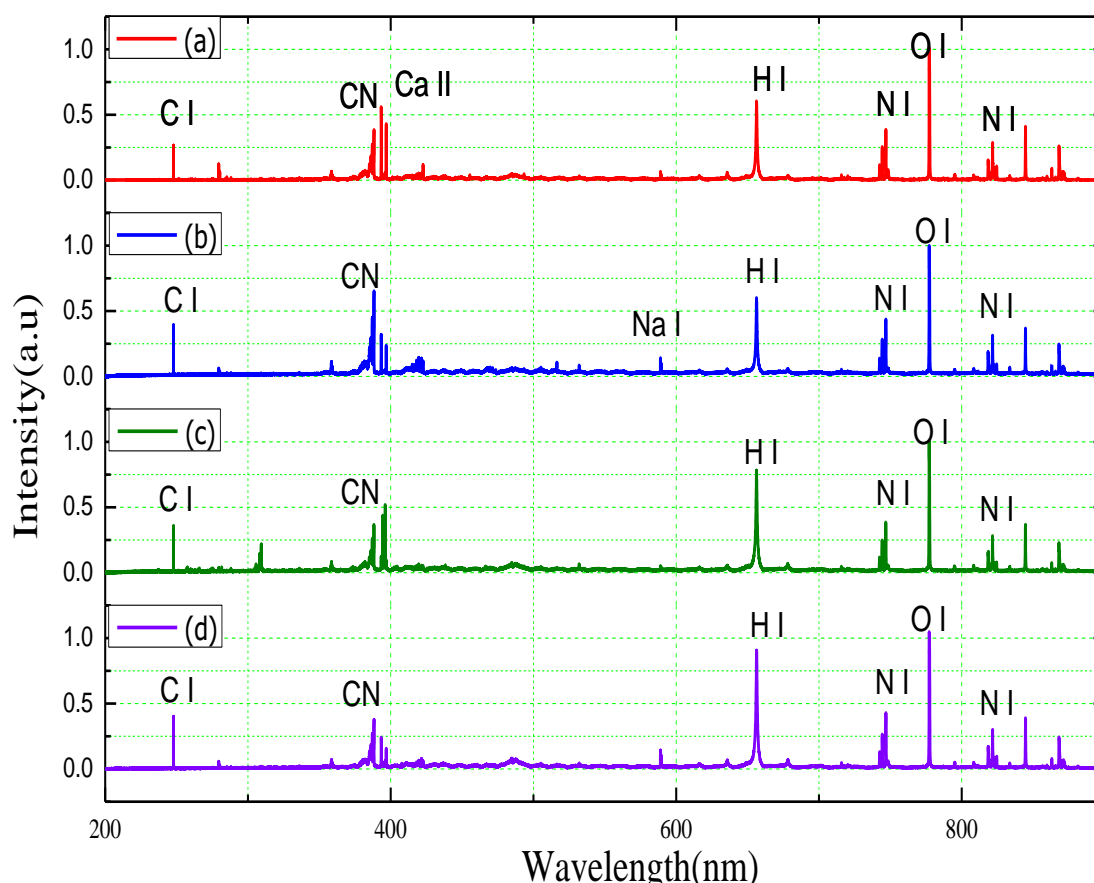


Figure 16. Representative LIBS spectra of organic compounds (a) HMX, (b) 4NP, (c) D-fructose (d) Glycine.

Except for vanadium, all the elements were detected when a separate measurement was performed using ICP-AES. While sodium and iron could be additives, the other metals are possibly contaminants. Similarly in the case of sodium lauryl sulphate (SLS), D fructose and Dextrose samples also there is no nitrogen in their chemical formulae but the presence of nitrogen in LIBS spectra clearly shows the entrainment of atmospheric nitrogen in the plasma. It is well known that atmosphere contains nitrogen about 78%, oxygen 21% and CO_2 about 0.03%. The chemical formulae of pyridine and pyrazole compounds do not contain oxygen but the presence of oxygen in LIBS spectra shows the complex nature of plasma. To observe the correlation between the stoichiometry of compound and the corresponding elemental emissions, relative intensities of carbon, hydrogen and nitrogen are compared, The elements identified in all the organic compounds are tabulated with the associated wavelengths shown in Table 5.

S. No.	Element	Wavelength (nm)
1	Carbon	247.85
2	Magnesium	279.78,283.59,285.18
3	Manganese	279.10,279.48,380.96
4	Sodium	589.0,589.60
5	Vanadium	251.16, 572.68, 635.70
6	Oxygen	777.19,777.41,777.53,822.18,822.76
7	Nitrogen	742.36,744.23,746.83,818.48,818.80,821.63,824.23
8	Hydrogen	656.27
9	Magnesium	518.36
10	Titanium	394.8, 395.6, 395.8, 399.8
11	Calcium	393.37, 396.86
12	CN	384.82,385.20,385.86,386.85,388.05
13	C ₂	469.65,471.40,473.61,512.94
14	Potassium	766.52,769.96

Table 5: Different peaks observed in the LIBS spectra and the corresponding atomic elements

2.3 Results

2.3.1 RELATIVE INTENSITY OF ELEMENTS

As seen from [figure 15](#) LIBS spectrum consists of sharp peaks corresponding to elements like C, H, N and O. Peak heights observation is the simplest way to interpret the spectra. It is well known phenomena that LIBS signal is resultant of several physical and chemical processes occurring in the plasma. The peaks which are main constituents of the compounds representing organic compounds carbon at 247.8 nm, hydrogen at 656.6 nm, nitrogen at 746.8 nm and oxygen at 777.2 nm peak intensities have been considered for

comparison to assess the correlation of emitted signal intensity with number of elements present in empirical formula of compounds. As seen from the [figure 17](#) there is no correlation of the intensities of the peaks to its abundance. Neomycin has large number of C, H, N and O with chemical formula $C_{23}H_{46}N_6O_{13}$; while on the other hand least number in pyrazole with $C_3H_4N_2$. In the case of compounds like D fructose and dextrose, nitrogen is not a constituent of sample but its entrapment from ambient atmosphere makes presence of nitrogen in the LIBS spectra. From the above two examples and observing the relative intensity bars in the [figure 17](#), there is no relation found in between the chemical formula and the relative intensities. All other types of compounds are also not following a trend with the formulae. In the case of explosives the relative intensities are similar in the vicinity of standard deviation but not have correlation with the empirical relation with that of formulae. Isomeric compounds like 1NP, 3NP, 4NP and carbohydrates do possess similar relative intensity bars, which are equal but differing in carbon intensities. Hence discriminating organic compounds on the basis of relative intensity measure may not be a viable option.



Figure 17: Comparison of relative intensities of C, H and N elements present in organic compounds
 (a) Carbon at 247.8 nm, (b) hydrogen at 656.6 nm and (c) nitrogen 746.8 nm are chosen to correlate the content of elements with the stoichiometry of chemical formulae of different organic compounds. Intensities are from the spectra normalized to 1.

2.3.2 RATIOMETRICS

The ratiometrics approach is explored to identify the organic compounds based on their respective oxygen to nitrogen, carbon to oxygen, hydrogen to oxygen, carbon to nitrogen, hydrogen to nitrogen intensity ratios. This approach is previously shown to yield reasonable identification of organic nitro-compounds, namely 4-nitroaniline and 4-nitrotoluene, by Rai *et al* [11]. Sreedhar and his coworkers using different type of detectors successfully classified inorganic compounds on the basis of O/N ratios.

The oxygen peak at 777 nm (O) and nitrogen peak at 746.83 nm (N) are used for evaluating the O/N ratios. The oxygen peak at 777 nm is a triplet and is not fully resolved in LIBS spectra of all organic compounds. It is better to recall the LIBS equation where the integrated line intensity I_{ij} of a spectral line corresponding to the transition between the upper level i and the lower level j is given by [12]

$$I_{ij} = \frac{g_i A_{ij}}{U^s(T)} n^s \exp\left(-\frac{E_i}{kT}\right) \quad (2)$$

where $U^s(T)$ is the partition function, n^s is the number density of species s in the plasma, g_i and A_{ij} are statistical weight and transition probability of the upper level respectively, and k is Boltzmann constant. From the equation (2) the number density of the species can be represented by I/Ag , where I is the observed area under the peak. As the observed oxygen peak at 777 nm is a triplet, it is deconvolved using a single Lorentz fit. Figure 18 shows a typical fit for the oxygen and nitrogen using a single Lorentz fit. Oxygen peak area is first divided in to three areas by multiplying the factors 870/2430, 810/2430 and 750/2430 where it denotes the relative intensities of oxygen triplet lines 777.2, 777.42 and 777.54 nm as given in NIST database [13] and termed as O1, O2 and O3. It is well known that collected spectral line emission intensity of oxygen is integrated collection of all triplets. Hence the partition of oxygen peak with the relative intensities is an important step while deducing O/N ratio. I/Ag is calculated for three peaks corresponding to 777.2 nm $(I/Ag)_1$, 777.42 nm $(I/Ag)_2$ and 777.54 nm $(I/Ag)_3$ for oxygen. The values of A and g are taken from NIST database. Table 6

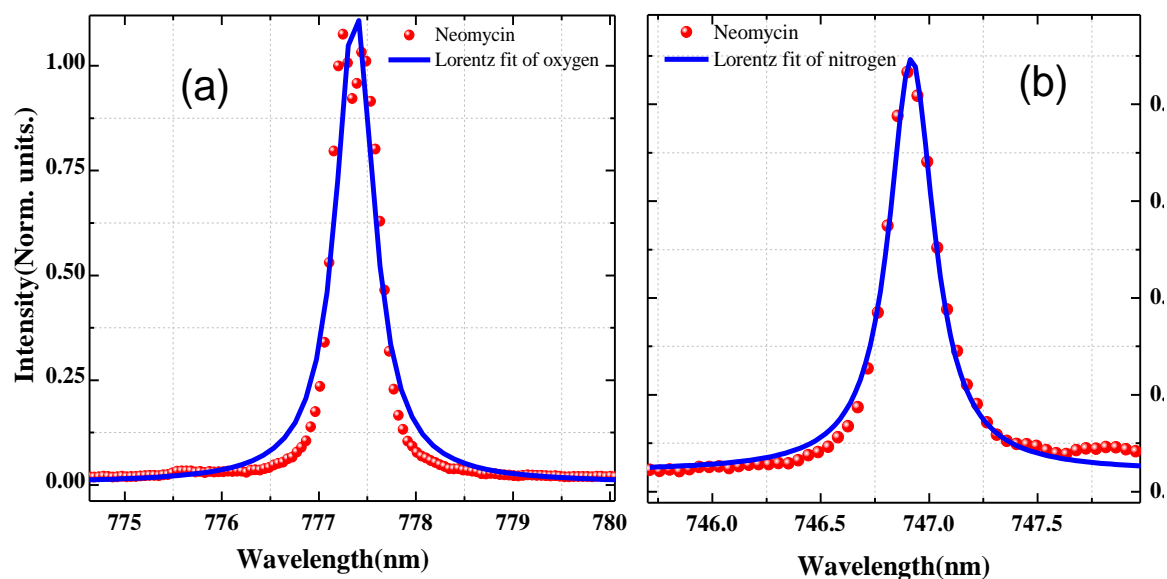


Figure 18: a) Oxygen and b) Nitrogen LIBS peaks from the Neomycin spectrum. The dots represent the experimental points and the solid lines are the Lorentz fit.

Table 3 summarizes the transitions and the constants that are used for oxygen & nitrogen. I/Ag for oxygen is taken as the sum of these three ratios $(I/Ag)_{\text{oxygen}} = (I/Ag)_1 + (I/Ag)_2 + (I/Ag)_3$. The ratio of $(I/Ag)_{\text{oxygen}} / (I/Ag)_{\text{nitrogen}}$ is considered as O/N ratio. [figure 19](#) shows the O/N ratios of all the samples. The calculated O/N ratios of all

Element	Wave-length (nm)	Transition Probablity (10^7 S^{-1})	E_i (cm^{-1})	E_k (cm^{-1})	g_i	g_k	Configurations
Carbon	247.8	2.8	21648.01	61981.82	1	3	$2s^2 2p^2 1s - 2s^2 2p 3s$
Hydrogen	656.6	4.41	82259.16	97492.3	-	-	2-3
Nitrogen	746.83	1.93	83364.62	96750.84	6	4	$2s^2 2p^2 (^3P) 3s - 2s^2 2p^2 (^3P) 3p$
Oxygen I	777.19	3.69	73768.2	86631.45	5	7	$2s^2 2p^3 (^4S^\circ) 3s - 2s^2 2p^3 (^4S^\circ) 3p$
Oxygen II	777.41	3.69	73768.2	86627.78	5	5	$2s^2 2p^3 (^4S^\circ) 3s - 2s^2 2p^3 (^4S^\circ) 3p$
Oxygen III	777.53	3.69	73768.2	86625.76	5	3	$2s^2 2p^3 (^4S^\circ) 3s - 2s^2 2p^3 (^4S^\circ) 3p$

Table 6: Transition probability, energy levels and statistical weights of the levels corresponding oxygen and nitrogen peaks used for the ratiometrics calculations

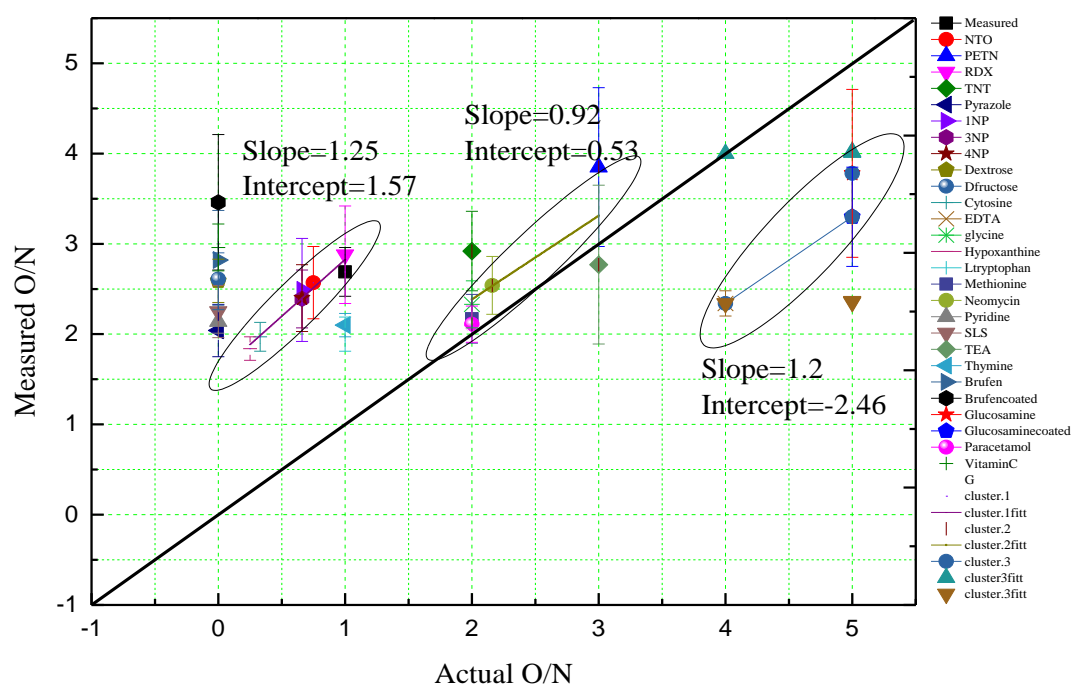


Figure 19. Oxygen to nitrogen ratios of all organic compounds. The symbols correspond to measured O/N values from the spectra. Error bars denote the standard deviation of multiple spectral data. The diagonal is stoichiometric line deduced from empirical relation of oxygen and nitrogen from the formula.

samples are plotted against actual O/N (i.e., deduced from the formula). The standard deviations of all the O/N ratios are computed and the average of all measurements of each organic compound calculated. The plot contains standard deviation as error bars. The diagonal line, which is called stoichiometric line is the principally derived from the formula.

The actual O/N empirical ratios of Methionine, paracetamol and TEA are 2, 2 and 3; the computed average O/N using LIBS are 2.54, 2.11 and 2.77, the corresponding standard deviation are 0.32, 0.20 and 0.88 respectively. Only these three compounds have values nearer to diagonal line. For Other compounds, the deviation from stoichiometric line is large and have larger standard deviations.. This could be a result of inhomogeneity in the sample consisting of various additives including the matrix which is true in the case of tablets. While the reason for the other compounds like pyridine and pyrazole compounds do not contain oxygen in their chemical formula but their spectrum contain oxygen peak.

Similarly active content of brufen, SLS, D fructose and Dextrose samples also do not have nitrogen in chemical formulae but they possess nitrogen in LIBS spectra. This clearly shows the entrainment of atmospheric oxygen and nitrogen involvement in the plasma. As a result, atomic emission from plasma is not reflecting proportionate distribution of O/N with that of formula. Three different clusters have been modeled and slopes/intercepts neither match with each other nor with stoichiometric line.

C/O and C/N ratios are computed for all organic compounds, considering the same calculation protocol with that of O/N. The atomic transition probability and statistical weight pertaining to C, H, N and O are given table 3 (above). Actual C/O and C/N ratios versus computed ratios are plotted and compared with stoichiometric line as shown in Figure 20 & Figure 21 . It can be observed that C/O ratios not reflected the stoichiometric line, it has formed three different clusters which are fitted with straight line and the slopes are neither equal nor equal to the stoichiometric line slope. Similarly also for C/N, where two different clusters and stoichiometric line nowhere close.

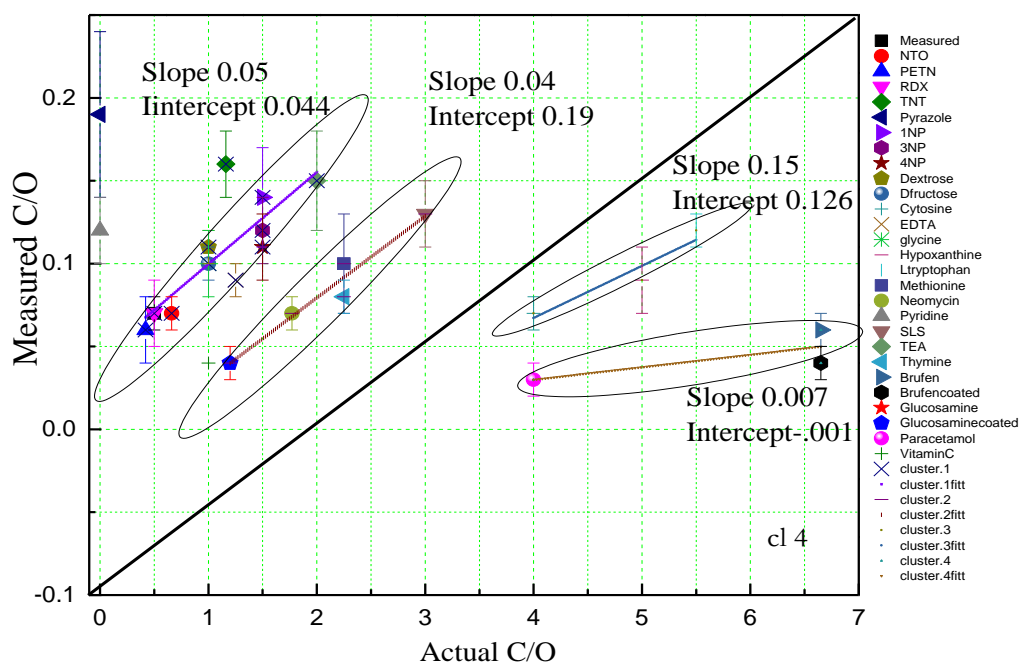


Figure 20: Carbon to oxygen ratio of all organic compounds. The symbols correspond to measured C/O values from the spectra. Error bars denote the standard deviation of multiple spectral data. The diagonal is stoichiometric line

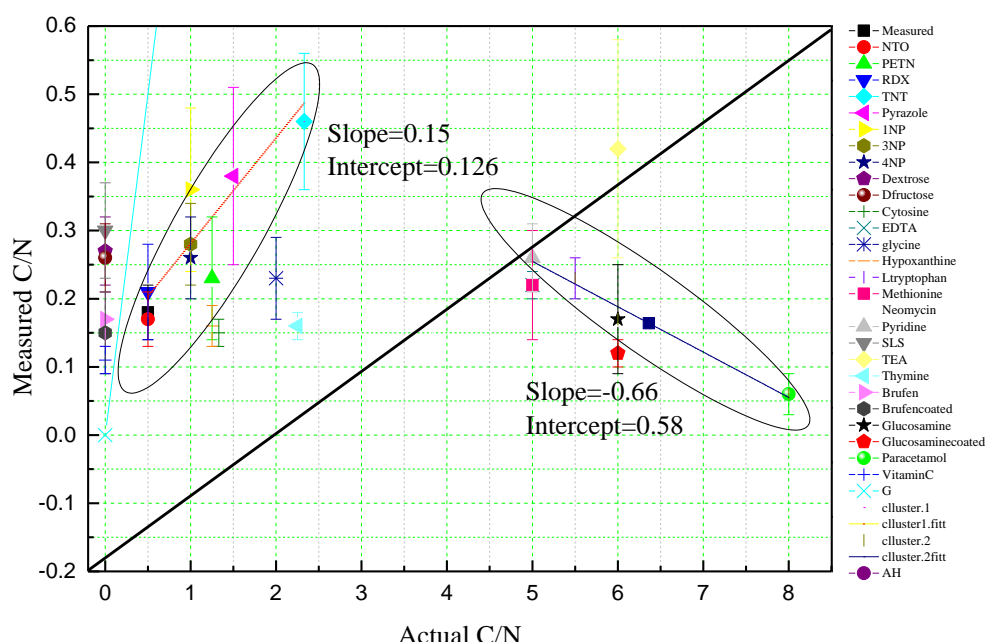


Figure 21: Carbon to Nitrogen ratio of all organic compounds. The symbols correspond to measured C/N values from the spectra. Error bars denote standard deviation of multiple spectral data. The diagonal is stoichiometric line

Since C lies in the ultraviolet region, it may better to choose the peaks that lie close to each other in the spectrum. The same method of calculation elements number density I/Ag calculated as discussed in the O/N ratio approach section. Actual H/N & H/O ratios plotted against computed ratios as shown in Figure 22 & Figure 23. It is interesting to note that from the figures H/N & H/O ratios of Neomycin and TEA are scattered nearer to stoichiometric line which could be purely coincidental. However the remaining samples are scattered either side of diagonal line.

The ratios O/N, C/O, C/N, H/O and H/N are not consistent even for one sample listed in table 1. The slope corresponds to different clusters are also not in good agreement with stoichiometry. This hints at the failure of ratiometrics in discriminating the organic compounds. Finally, in order to overcome the inherent constraints of the ratiometrics approach for accurate classification, multivariate statistical methods are employed to exploit the multi-channel spectral dataset of pharmaceutical tablets have been incorporated.

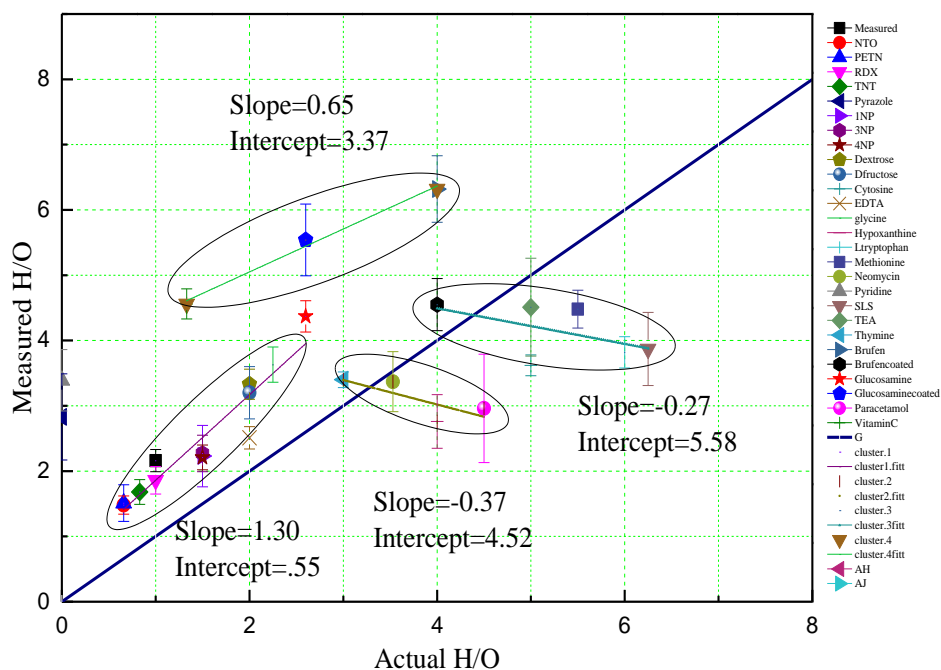


Figure 22: Hydrogen to oxygen ratio of all organic compounds. The symbols are corresponding to actual H/O values versus measured ones. Error bars denote the standard deviation of multiple spectral data.

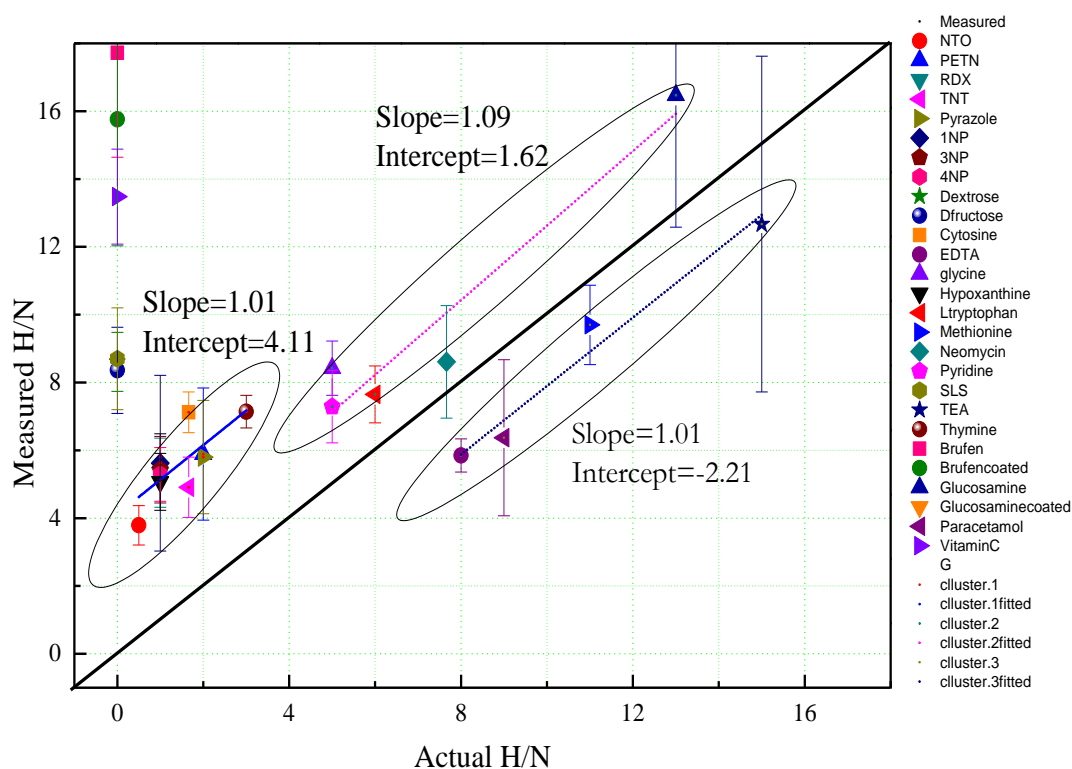


Figure 23: Hydrogen to Nitrogen ratio of all organic compounds. Actual H/N values are plotted against computed ones. Error bars are standard deviation of data. The diagonal line is stoichiometric line

2.3.3 CLASSIFICATION USING MULTIVARIATE METHODS

This section explores the effectiveness of Multivariate analysis for the identification a sample from given set of samples. The analysis framework is referred to as chemometrics. According to the International Chemometrics Society, definition of chemometrics is “the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods.” As a typical identification problem involves a set of materials of interest and the task is to identify a given unknown based on the training model. Pharmaceutical samples have been considered for case study in this section. It is worth noting that the complex chemical composition as well as the turbid media of the typical pharmaceutical samples significantly hinders the employment of other comparable optical and spectroscopic techniques. For example, most capsules and tablets consist of a mixture of organic active substances as active pharmaceutical ingredients (API) and excipients, in addition to possible impurities. Evidently, LIBS provides a powerful tool to detect these impurities and monitor the quality of products with potential extension to detection of counterfeit drugs. The outcome of previous section clearly demonstrated that ratiometrics are impractical for identification. There different methods principal component analysis (PCA), Soft independent modeling of class analogy (SIMCA) and partial least square discriminant analysis (PLSDA), have been used for the analysis.

The details of samples can be found in table1 from serial number 1-6. The LIBS spectra of the samples are shown in [figure 15](#). As observed in the table each compound contain peaks majorly related to C, H, N, O and other inorganic compounds like Na, Mg, Mn and Ni.

2.3.3.1 PRINCIPAL COMPONENT ANALYSIS

The details of PCA can be obtained in the section 1.6. PCA is first performed on the entire 85 sample dataset of pharmaceutical tablets. Figure 24 shows the first three principal components, which reveal the dimensions that explain most of the variance present in the dataset. These components, while abstract in form as they are obtained through a mathematical change of basis (*e.g.* via singular value decomposition), are useful in indicating the informative spectral features associated with the different

samples. Here, it is observed that the first PC appears strikingly similar to the glucosamine spectra (compared to Figure 16) and the corresponding magnitudes of the scores for the glucosamine samples are larger than

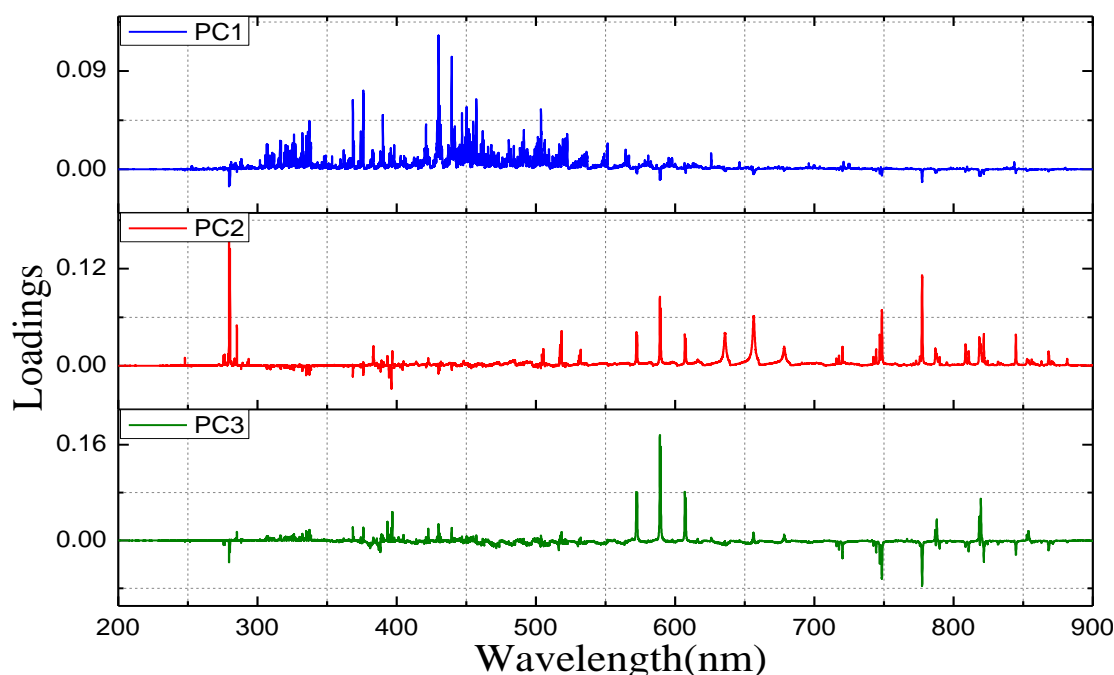


Figure 24: Loadings plot of LIBS data of pharmaceutical compounds. The first three principal components corresponding to the entire spectral dataset acquired from the pharmaceutical samples. These three principal components, combined, explain 96.29% of the net variance in the dataset.

that for the other samples. Specifically, the distinctive calcium lines appear solely in PC1. On the other hand, PC2 shows a substantive influence of manganese alongside the contributions of the API elements such as carbon, hydrogen, nitrogen and oxygen. This is reflected in the higher PC scores for the paracetamol and vitamin C samples. The first two PCs explain 90.67% of the total variance in the dataset (as computed from the cumulative contribution of their Eigen values). After incorporating the third PC, this metric rises to 96.29%. The relatively less significance of the third PC can also be visualized from Figure 24, as some of the spectral features are repetitions from the previous PCs. The subsequent PCs (*i.e.* fourth, fifth etc.) are fairly noisy and their incorporation deteriorates the quality of the model. As a consequence, it is employed only the first three PCs for understanding the cluster behavior of the samples. Figure 25 shows the corresponding scores along the first three PCs. Clearly, the samples of each class tend to cluster together and in almost all cases are fairly well

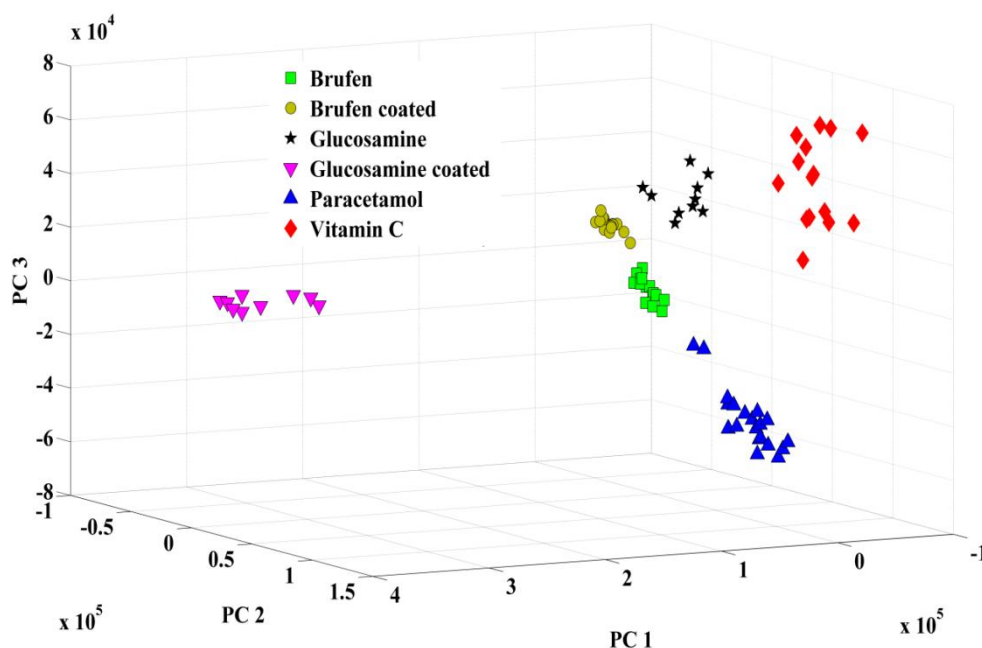


Figure 25: PCA scores plot of the first three principal components of the spectral dataset acquired from the six classes of pharmaceutical samples.

separated from the other classes. Of these, vitamin C and paracetamol appear to be the easiest to distinguish based on their distance from the other classes. It is also interesting to probe the dispersion of the different classes along the PC directions. For example, vitamin C and glucosamine tablets show a considerably larger dispersion as compared to both the coated and uncoated brufen samples which exhibit a more uniform pattern. This provides a novel insight into the tablet-to-tablet intra-class variations (*e.g.* arising from the heterogeneity of each tablet composition). Additionally, it can be observed outliers based on the spectral data, notably one coated glucosamine sample and two paracetamol samples as seen in Figure 25.

Based on these results it can be reasonably infer that: (a) the LIBS spectra provides vital information which can be used in routine sample monitoring for pharmaceutical tablets (or at least the ones used in this study); and (b) the PCA classification scheme is able to identify the primary elements which help in distinguishing the various classes, which corresponds to the existing knowledge of the sample composition. The latter proves that there is a direct causality between PCA classification and chemical basis of the samples, rather than an arbitrary (potentially spurious) correlation which cannot be successfully reproduced in prospective application. Nevertheless, while PCA is a valuable tool for recognizing similarities between sample

types, it does not automatically provide class memberships due to its unsupervised nature. To assign class membership to the tested tablets, SIMCA is employed.

2.3.3.2 SIMCA

The SIMCA approach has been widely adopted for a wide range of applications as it offers a multiple advantages over other classification techniques. As discussed in earlier section. First, SIMCA inherits the strengths of PCA-based dimensionality reduction, which as mentioned before reduces the chances of spurious correlations and noise corruption. Second, the prospective sample is only assigned to a class with a sufficiently high probability (above a pre-assigned threshold value) rather than to "the nearest class". This reduces the chances for misclassification as for any sample there always exists a nearest class, even if the sample is an outlier or belongs to a class which is not represented in the training set. Third, SIMCA provides a robust classification scheme in the sense that it does not require a large training dataset nor does it constrain the number of measurement variables. This represents a significant improvement over standard discrimination techniques which do not work under such scenarios due to well-characterized problems of collinearity and chance classification. This feature makes SIMCA a particularly attractive classification tool for spectroscopic characterization, where a large number of measurement variables is almost always encountered.

SIMCA was subsequently performed on the spectral dataset in conjunction with the class membership information. In this investigation, 30 test samples (5 samples per each of the 6 classes of tablets) were randomly chosen and kept aside for prospective application. The construction of an independent test set is a standard chemometrics approach employed to diminish and/or examine for the presence of spurious correlations. Subsequently, the 55 training samples were used to develop the SIMCA models using a modified version of the LIBRA toolbox for MATLAB originally developed by Verboven *et al.*[14]. To ensure the reproducibility of the classification results, 100 iterations are performed to obtain an average classification accuracy.

A key point in SIMCA model development is to decide how many of the principal components should be retained in the subsequent analysis. In an ideal situation, this number should equal the number of sample constituents. However, in real life applications (such as in this study), the number is rarely known *a priori*; moreover,

correlations between sample constituents, system drift and noise also play a key role in the final number of principal components retained in real world situations. Here, we employed a standard leave-one-out cross-validation procedure to determine the number of PCs for each PCA model. The optimal number of PCs was observed to be in the range of 3-5 for each of the classes under investigation.

Finally, equally weighted scaled orthogonal and score distances for assignment of class membership are employed. In addition, an unclassification criterion was defined to prevent the misclassification of potential samples that were not close to the center of any of the PCA models (*e.g.* spectral outliers). Similar to the scheme outlined in the study performed by Sirven and co-workers [15], we assumed that the distances of the training sample to the center of the corresponding class followed a normal distribution. This normal distribution was then employed to compute the probability of class membership of any test spectrum, given its distance to the center of the different classes. In the event that the membership probability for every class was observed to be less than 5%, we assigned it as an "unclassified" sample and removed it from further classification analysis. This protocol helps reduce the number of misclassifications and the decision threshold (*i.e.* statistical significance of 5%) can be tuned depending on the requirements for the specific monitoring technology. It should be noted that due to the limited number of training samples per class (approximately 10), we used a single 'universal' distribution for our unclassification assessment - although a more accurate procedure in real life applications could potentially utilize separate class-specific distributions.

As mentioned above, SIMCA computes a PCA model for each of the six classes of pharmaceutical samples and identifies the prospective samples based on their distance to the respective models. Based on this key concept and the unclassification criterion detailed above, class memberships are assigned to 30 prediction samples for each iteration (where for every iteration re-splitting of the dataset into training (55) and test (30) samples is performed). Figure 26 (below) shows a bar plot visualization of SIMCA classifications for a representative set of 30 test samples. Here, none of the samples are unclassified (unclassified samples are assigned a class membership of 0) and three of them are misclassified.

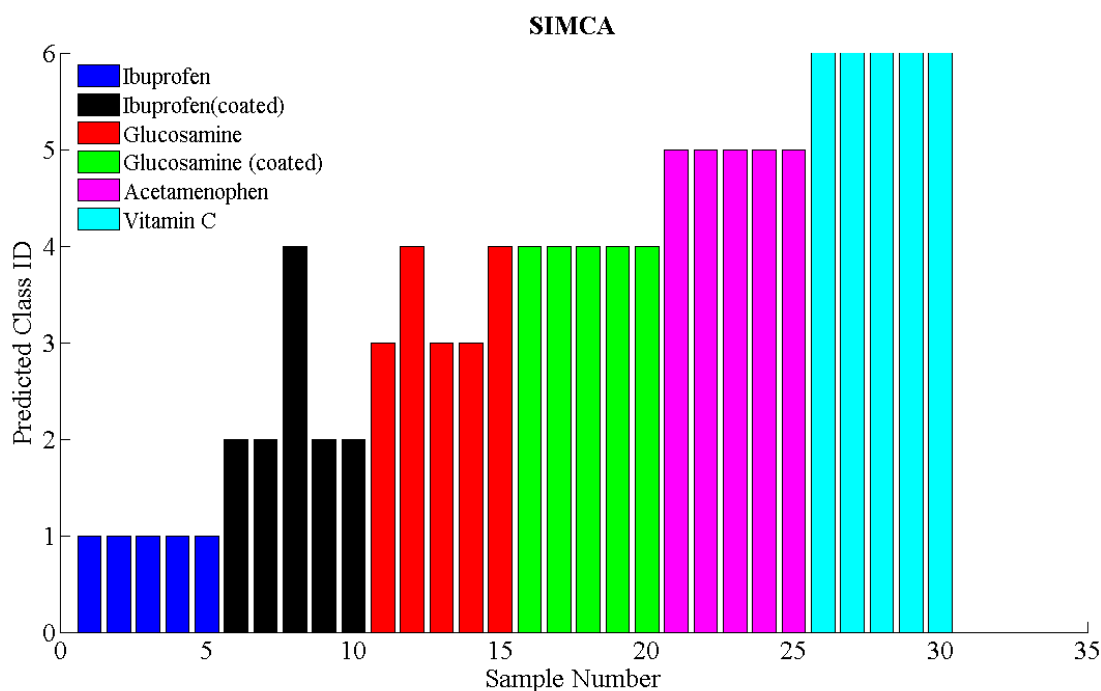


Figure 26: Bar plot of SIMCA classifications for a representative set of 30 test samples.

Bar plot of SIMCA classifications for a representative set of 30 test samples. The test samples are ordered such that each successive subset of 5 samples belong to a different class, as shown by the different bar coloring. Predicted Class ID are as follows: 0: unclassified; 1: brufen; 2: brufen (coated); 3: glucosamine; 4: glucosamine (coated); 5: paracetamol; and 6: vitamin C. Here, three misclassifications are observed, one corresponding to coated brufen and two corresponding to glucosamine.

To obtain a more comprehensive sense of the results, the average rate of unclassification, misclassification and correct classification over 100 iterations (Table 6) are computed. It is observed that on average acceptable rate of correct classification for all the classes of pharmaceutical samples (>90%). In addition, rate of unclassification is fairly low indicating that the acquired data consisted of few outliers and the training model is sensitive to all remaining samples (*i.e.* all but the outliers). Unclassification is relatively higher for the paracetamol (5.8%) and coated glucosamine (3.8%) samples. This is not surprising based on the PCA scores plot (Figure 26 above), where one could clearly observe one coated glucosamine and two paracetamol spectral outliers.

Furthermore, it can be find that the paracetamol, vitamin C and coated glucosamine samples are almost perfectly classified (*i.e.* there is no misclassification), which is again consistent with the clear separation of these samples observed in the PCA scores plot (Figure 25). The others, namely the brufen (both coated and uncoated) and

uncoated glucosamine sample, are significantly more difficult to classify resulting in higher errors (7.6%, 6.2% and 8.8%, respectively).

Average rate of ...	Correct classification	Wrong classification	Unclassification
Brufen	0.908	0.076	0.016
Brufen-coated	0.934	0.062	0.004
Glucosamine	0.904	0.088	0.008
Glucosamine-coated	0.940	0.002	0.058
Paracetamol	0.968	0	0.032
Vitamin C	0.988	0	0.012
Average	0.9403	0.038	0.0217

Table 7: SIMCA classification results obtained from 30 test samples over 100 iterations

An interesting point in this regard is the striking difference in classification accuracy for the coated and uncoated glucosamine samples. The significantly better classification accuracy for the coated samples can be primarily attributed to the distinctive coating composition of these tablets. Coating is primarily made of titanium oxide, which is evidently absent in the tablets as is shown in Figure 15. Evidently for the brufen (coated and uncoated) and the uncoated glucosamine tablets, the variance between the classes is of the scale of the variance within the respective classes, thereby impeding prospective SIMCA analysis. This problem can be potentially solved by employing an alternate classification scheme such as PLS-DA (partial least squares based discriminant analysis) in the future [34]. PLS-DA seeks to establish the maximum separation between classes as opposed to PCA which does not discriminate between class-to-class and intra-class variability in explaining the total variance in the dataset. Consequently, PLS-DA may present a better alternative when these two variability numbers are of the same order, as observed above for brufen and uncoated glucosamine samples.

2.3.3.3 PARTIAL LEAST SQUARE DISCRIMINANT ANALYSIS

The additional information about PLS-DA can be found in the section .. chapter PLS-DA is a similar supervised classification technique, which incorporates PLS analysis for

noise reduction and variable selection in place of PCA [16]. Notably, PLS-DA seeks to determine the maximal separation between each class by fitting one global model to the entire dataset. This key feature enhances its ability to successfully discriminate between samples even when the variability within a class is similar to the inter-class variability. Previous researchers have reported on the viability of PLS-DA.

In this present context, PLS-DA was applied to the dataset of pharmaceutical compounds. The total class memberships are stored in the form of a 85x6 matrix of dummy variables, where the columns are equal to the number of classes and rows are equal to the number of samples. Each column in this matrix consists of '1's for the corresponding class; otherwise it consists of '0's. To test the model for the unknown spectrum, it yields a 1x6 vector containing predicted membership probability values. It is worth emphasizing that for PLS-DA, only a single "global" model is developed for classification purposes, in contrast to the independent PCA sub models created for the aforementioned SIMCA analysis. Similar to the SIMCA analysis, probability of class membership for proper assignment has been used. For PLS-DA, the same 0.3% decision threshold to unclassify samples is used.

Average rate of ...	Correct classification	Wrong classification	Unclassification
Brufen	0.980	0	0.020
Brufen-coated	0.916	0.03	0.054
Glucosamine	0.978	0	0.022
Glucosamine-coated	0.888	0.004	0.108
Paracetamol	0.980	0.008	0.012
Vitamin C	0.998	0	0.002
Average	0.9567	0.007	0.0363

Table 8: Pharmaceutical compounds predictions for each sample using PLS-DA

The non coated samples of Brufen and Glucosamine are predicted greater than the prediction accuracy of SIMCA and in the case of with other common tablets like paracetamol and vitamin C. It is interesting to note that the wrong classifications of all tablets are found to be less than 0.03. Unlike SIMCA, the average wrong classification is decreased and enhanced in average correct classification. Vitamin C, the non coated

samples of brufen and Glucosamine wrong classification accuracies are zero. Comparing both table 4 and table5 the overall correct classification is increased with decreasing in unclassification

2.4 Conclusions

A wide variety of organic compounds, total 28 samples were chosen for LIBS study. Three methodologies were used . Out of which, first two methods failed to provide an acceptable classification. First, the relative intensities of elements are not proportional with the stoichiometry of elements, towards identification of samples. Second, The ratios O/N, C/O, C/N,H/O and H/N never find correlation with the empirical ratios obtained from chemical formulae. Based on these two approach it is impossible to identify real world samples. To tackle this situation real world samples i.e., over the counter pharmaceutical tablets are used for multivariate statistical analysis. PCA, SIMCA and PLSDA have been utilized for discrimination. Visual inspection of score plot of PCA clearly shows that compounds are six types and are grouped far apart. SIMCA and PLSDA estimated corrected classification more than 0.94. Hence using chemometrics it is possible to identify compounds as a universal marker. Multivariate methods combined with LIBS can be ideal solution for the identification of samples.

2.5 References

- [1] Myakalwar, A.K., Sreedhar, S., Barman, I., Dingari, N.C., Rao, S.V., Kiran, P.P., et al. Laser-induced breakdown spectroscopy-based investigation and classification of pharmaceutical tablets using multivariate chemometric analysis. *Talanta*. 2011, 87, 53-9.
- [2] Pellegrino, P.M., Holthoff, E.L., Farrell, M.E. *Laser-Based Optical Detection of Explosives*, CRC Press, 2015.
- [3] Lasheras, R., Bello-Gálvez, C., Anzano, J. Identification of polymers by libs using methods of correlation and normalized coordinates. *Polymer Testing*. 2010, 29, 1057-64.
- [4] Fortes, F., Laserna, J. The development of fieldable laser-induced breakdown spectrometer: No limits on the horizon. *Spectrochimica Acta Part B: Atomic Spectroscopy*. 2010, 65, 975-90.
- [5] Mowery, M.D., Sing, R., Kirsch, J., Razaghi, A., Béchard, S., Reed, R.A. Rapid at-line analysis of coating thickness and uniformity on tablets using laser induced breakdown spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis*. 2002, 28, 935-43.

- [6] Archambault, J.-F., Vintiloù, A., Kwong, E. The effects of physical parameters on laser-induced breakdown spectroscopy analysis of intact tablets. *AAPS PharmSciTech*. 2005, 6, E253-E61.
- [7] Doucet, F.R., Faustino, P.J., Sabsabi, M., Lyon, R.C. Quantitative molecular analysis with molecular bands emission using laser-induced breakdown spectroscopy and chemometrics. *Journal of Analytical Atomic Spectrometry*. 2008, 23, 694-701.
- [8] Blomquist, G., Johansson, E., Söderström, B., Wold, S. Classification of fungi by means of pyrolysis-gas chromatography-pattern recognition. *Journal of Chromatography A*. 1979, 173, 19-32.
- [9] Lal, B., Zheng, H., Yueh, F.-Y., Singh, J.P. Parametric study of pellets for elemental analysis with laser-induced breakdown spectroscopy. *Applied optics*. 2004, 43, 2792-7.
- [10] Missaghi, S., Fassihi, R. A novel approach in the assessment of polymeric film formation and film adhesion on different pharmaceutical solid substrates. *AAPS PharmSciTech*. 2004, 5, 32-9.
- [11] Rai, S., Rai, A.K., Thakur, S.N. Identification of nitro-compounds with LIBS. *Applied Physics B: Lasers and Optics*. 2008, 91, 645-50.
- [12] Miziolek, A.W., Palleschi, V., Schechter, I. *Laser Induced Breakdown Spectroscopy*, Cambridge University Press, 2006.
- [13] http://physics.nist.gov/PhysRefData/ASD/lines_form.html.
- [14] Verboven, S., Hubert, M. LIBRA: A MATLAB library for robust analysis. *Chemometrics and Intelligent Laboratory Systems*. 2005, 75, 127-36.
- [15] Sirven, J.B., Sallé, B., Mauchien, P., Lacour, J.L., Maurice, S., Manhès, G. Feasibility study of rock identification at the surface of Mars by remote laser-induced breakdown spectroscopy and three chemometric methods. *Journal of Analytical Atomic Spectrometry*. 2007, 22, 1471-80.
- [16] Berrueta, L.A., Alonso-Salces, R.M., Héberger, K. Supervised pattern recognition in food analysis. *Journal of Chromatography A*. 2007, 1158, 196-214.

3. EXPLOSIVE DETECTION USING LIBS

In this chapter, two main objectives are involved: one is to identify explosive and other is to identify a specific explosive. LIBS data was collected from a set of secondary explosives and several non-explosives using a gated spectrometer. Neural network analysis has been introduced as a classifier and results are compared with that of PLSDA. Judicious feature selection approaches - specific peaks, sub-spectral regions have been investigated. With specific peak of hydrogen as feature input, a highest rate of classification rate of ~ 97% was obtained among explosive and non-explosives. All the four sub-spectral regions and the full spectrum as feature input lead to almost similar classification rates ranging between 96-98%. For classification among the explosives, the full spectral input resulted in classification rate of ca.92%, and sub-spectra from carbon to hydrogen-specific features (~ 60% of the total intensity values) resulted in near identical performance.

3.1 Introduction

“In the simplest terms, an explosive is defined as a substance, which on initiation by friction, impact, shock, spark, flame, heating, or any simple application of an energy pulse, undergoes a rapid chemical reaction evolving a large amount of heat and so exerting a high pressure on its surroundings. The vast majority of explosives release gaseous products on explosion but this is not an essential requirement as in the case of some metal acetylides [1]”. Explosives in the broader sense sometimes termed High Energy Material (HEM). HEMs can be defined as:

“...In order to camouflage research on explosives, propellants and pyrotechnics, a new term ‘high energy materials’ (HEMs) was coined by the explosives community for them. Thus all explosives, propellants and pyrotechnics can be referred to as high energy materials (HEMs) or energetic materials (EMs). In other words, the other name of HEMs/EMs is explosives, propellants and pyrotechnics depending on their formulations and intended use. Nowadays, the term HEMs/EMs is generally used for any material that can attain a highly energetic state mostly by chemical reactions[2].”- Jai Prakash Agrawal

HEMs can be divided into two main groups on the basis of its expansion after explosion: low explosives and high explosives. Low explosives includes black powder, smokeless gunpowder and low vulnerability ammunition. High explosives are divided into three subgroups: primary, secondary and tertiary explosives. Primary explosives include compounds like Lead Azide and Hexa Methylene Triperoxide Diamine (HMTD), secondary explosives comprise Tri Acetone Tri Peroxide (TATP), Royal Demolition Explosive (RDX), C4, High Melting Explosive (HMX), dynamite and Trinitrotoluene (TNT). Tertiary explosives include ANAL and ANFO. Figure 27 shows categorical classification of explosives based on their expansion speed, application and analytical relevance [3].

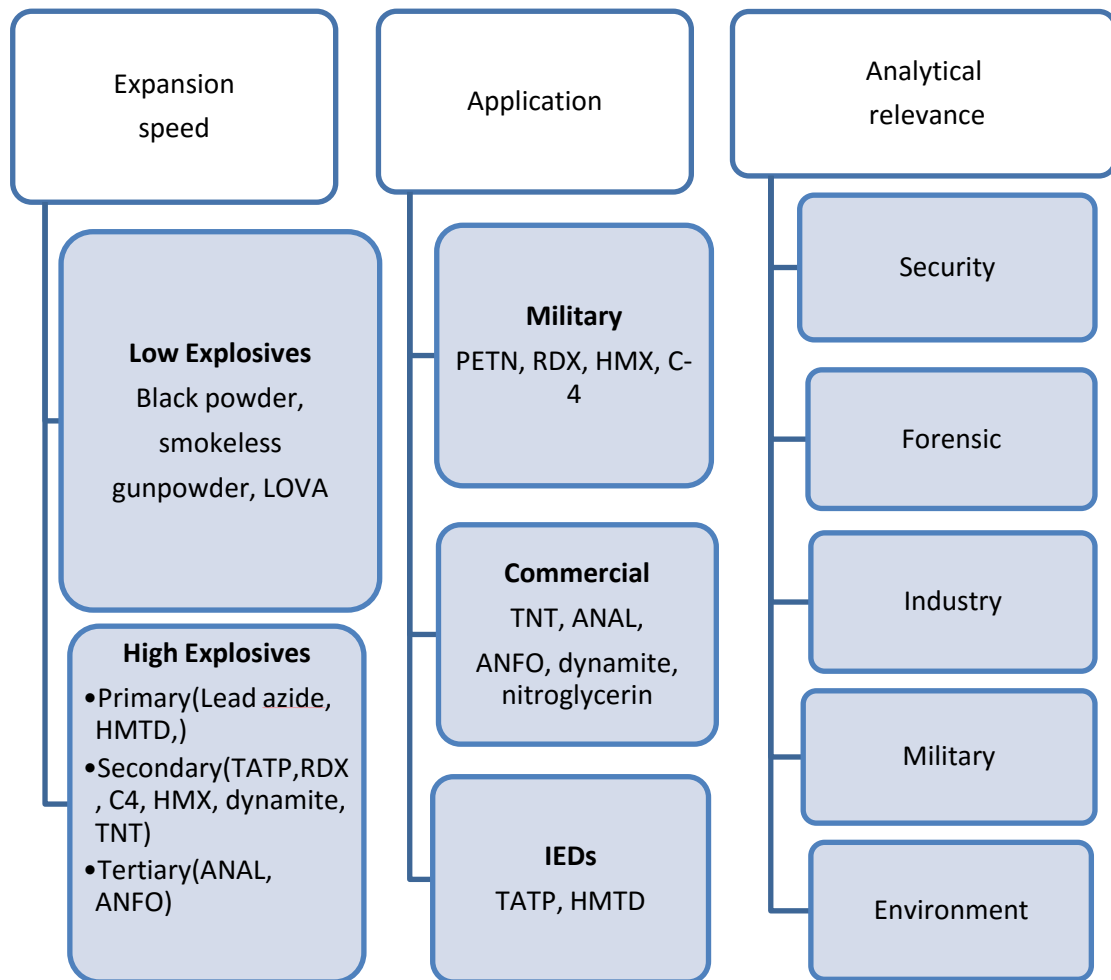


Figure 27: Classification of HEMs based on their performance, application and analytical importance

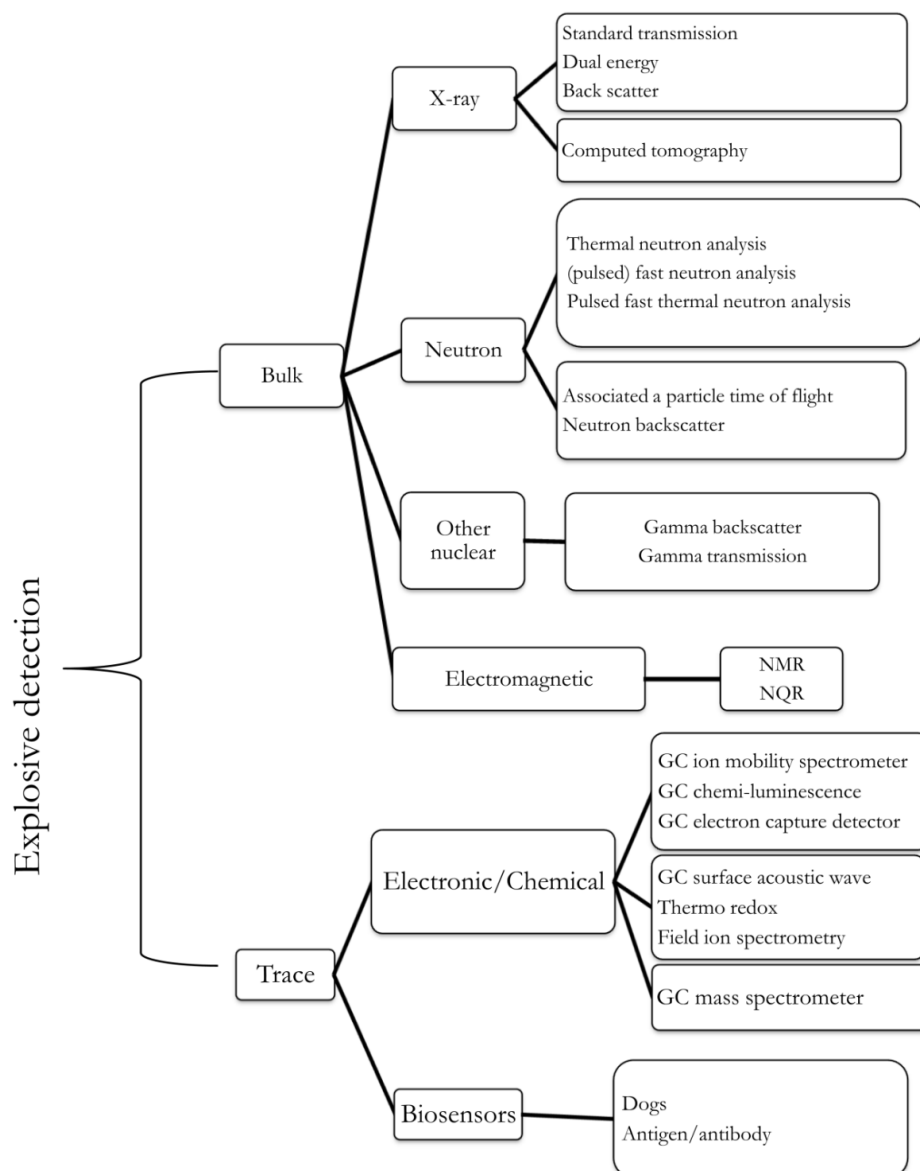


Figure 28. Organization of current detection technologies (Adopted from Explosives Detection Using Magnetic and Nuclear Resonance Techniques- NATO Science for Peace and Security Series)

Explosive detection has been a subject of research domestically and internationally due to homeland security and anti-terrorism applications. The explosive sensing is generally broken down into trace and bulk detection. Detecting the trace amount of HEM, the general outlook is to sense vapors from the explosive or explosive particles adhering to surfaces, with addition, in some cases to hunt for a specific chemical from a library of target compounds, while bulk detection looks for a definite property, which is an indicative of HEM (nitro rich or oxy rich compound) i.e., either by imaging or by

detection of the explosive itself. figure 28 summarizes a broad class of detection techniques based on this perspective of detection schemes.

Several technologies were discussed in the SPIE conferences on detection. Large number of companies have been established and exploited path breaking technologies to solve the problem of explosive detection. All of these actions are a direct consequence of the necessity of the detection problem.

The proposals made by NATO committee for standoff detection mostly can be achieved through laser-based techniques, which have become a basic tool for remote and stand-off spectroscopy and sensing for environmental and security applications. In particular, notably Raman spectroscopy [5,6], laser induced fluorescence (LIF) [7], infrared (IR) absorption [8] and laser induced breakdown spectroscopy (LIBS) [9] has received considerable attention for HEMs detection because of the lack of sample preparation requirements and ability to provide real-time assessment in stand-off mode. While the vibrational spectroscopic approaches provide a wealth of molecular information, LIBS furnishes complementary elemental information and offers higher signal sensitivity due to the intense emission lines. Researchers have exploited the single-shot and facile elemental analysis capability of LIBS in diverse fields ranging from planetary exploration [10,11], archaeology [12] pharmaceutical formulation identification [13], HEM screening [14,15] as well as in other biomedical [16-18] and polymer applications [19]. The incorporation of statistical models in the analysis of LIBS coupled with the seminal work by Miziolek and co-workers [20]

In 2004–2005 the NATO-Russian Explosive Detection Working Group (EDWG) of the NATO Advisory Panel on Security-Related Civil Science and Technology made a proposal to detect explosive in standoff mode. The EDWG has prepared requirements for technology [4]:

- *Detection should be harmless and unnoticeable to person under inspection;*
- *Detection device may be portable or stationary;*
- *The sensor (or sensors) may be separate from the registration unit;*
- *The detection distance should be between 1 and 10 m;*
- *The HEX carrier person will be moving, so detection must be done in “realtime” — preferably in times less than one second-*

and stand-off operations at Los Alamos National Laboratory [21] has further enhanced its attractiveness for evaluation of complex materials.

Specific to HEM detection, there are two principal research goals: first, to identify a specimen as a HEM/non HEM and second, to classify the specimen as a specific HEM. Yet, HEMs typically are nitro-rich molecules and exhibit grossly similar spectral profiles with prominent emission lines corresponding to nitrogen, carbon, hydrogen and oxygen. Furthermore, the spectral interference emanating from trace quantities of grease, oils [22] and biomaterials [23] as well as the presence of oxygen and nitrogen in the ambient air, and hydrogen and oxygen in the atmospheric moisture impedes the reliable detection of the materials in the field. Consequently, the basic spectroscopic approach of employing atomic (and in cases molecular) intensity ratios for these kinds of materials yield very limited results. Also, the substrate effects for HEM residue detection, which have been detailed by Gottfried *et al* [24], exacerbates the segmentation challenges due to the potential alterations to the light emission from the micro plasma. Taken together, it is not surprising that non- analyte specific peaks appear in the spectral profile due to the entrainment of air and the (variable) substrate information [25]. Approaches that partially alleviate these issues include noble gas purging and double pulse method. However, purging is not a viable option for on-field application and while the double pulse method shows some improvement in the signal enhancement [26], purely ambient air peaks (*e.g.* oxygen and nitrogen) cannot be avoided.

In this milieu, the application of multivariate methods to LIBS datasets has enabled the extraction of information both amenable to and hidden from human examination and the resultant combination has been successfully utilized for materials identification [27,28]. The underlying hypothesis is that while assessing the spectral signatures of several specimen shows significant complexity, comparing hundreds of such spectra reveal subtle, yet reproducible, patterns that are valuable for objective sample identification. Various multivariate segmentation methods including principal component analysis (PCA) [20,29], soft independent modeling of class analogy (SIMCA) [25], partial least squares-discriminant analysis (PLS-DA) [20,25,27] , artificial neural network (ANN) [30] and support vector machine (SVM) [31] have been exploited to analyze the rich LIBS datasets. Concomitant advances in instrumentation, especially the use of

echelle spectrographs, have further enabled measurements of high resolution thereby providing larger number of spectral features for classifier development.

However, the massive amount of data collected puts forth a different challenge as far as the development of a robust classifier is concerned. Given the high dimensionality of the LIBS dataset and the relative sample sparsity (stemming from the unavailability of large quantity of HEM samples), the quest for the ‘perfect’ classifier can often produce unwarranted conclusions and lead to apparently functional algorithms that cannot be used prospectively. Boosting the sample per feature ratio by appropriate selection of spectral regions can not only produce classifiers that are more insensitive to outliers and possess higher generalization power but also aid in interpretability by identification of the smallest possible subset of maximally discriminatory features. Recently, considerable effort has been directed towards developing and evaluating different procedures that objectively identify wavelength bands that contribute useful information and/or eliminate regions that contain mostly noise or spurious information [5,25,32]. In principle, judicious wavelength selection for spectroscopic datasets focusing on marker-specific features could even result in models having a greater predictive ability. Despite these advantages and the suitability of LIBS for feature selection due to its inherently narrow emission peaks, wavelength selection for building LIBS-based segmentation models has received scant attention in the literature. This chapter highlights on feature selection primarily on region selection followed by calculating detection accuracy using PLSDA and additionally the explosive detection algorithm using neural network which potentially point towards a reduced spectral window rather than entire UV-VIS-NIR region from 200-900nm.

3.2 Materials

A set of HEM samples shown in Table 9 namely octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine (also known as high-molecular-weight RDX (HMX)), nitrotriazalone (NTO), pentaerythritol tetranitrate (PETN), 1,3,5-trinitro-1,3,5-triazacyclohexane (RDX), 2,4,6-trinitrotoluene (TNT) were acquired from High Energy Materials Research Laboratory (HEMRL), Pune, India. The fine powder material was pressed into 1 cm diameter pellets by a die-hydraulic pressing machine by applying 3-4 tons of pressure. The pelletization operation improves signal reproducibility, as the position of the focal

spot is almost unchanged for all the laser pulses arriving at the pellet surface in sharp contrast to a powder specimen.

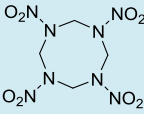
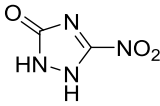
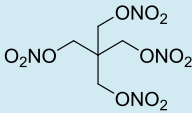
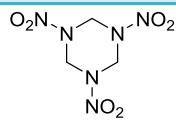
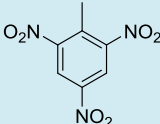
S. No	HEMs	Formulae	Structure	No. of spectra
1	HMX	$C_4H_8N_8O_8$		133
2	NTO	$C_2H_2N_4O_3$		75
3	PETN	$C_5H_8N_4O_{12}$		60
4	RDX	$C_3H_6N_6O_6$		61
5	TNT	$C_7H_5N_3O_6$		143

Table 9. Details of HEMS: chemical formulae, chemical structure and usable number of spectra

The samples were placed on a manual X-Y-Z translation stage such that spectra could be collected from different spatial locations on the pellet to estimate measurement variability. For the present investigation, 472 spectra were acquired from the HEM samples. Specifically, 133 spectra for HMX, 75 for NTO, 60 for PETN, 61 for RDX and 143 for TNT were included for the ensuing classification analysis. During the entire process of grinding, pelletization or laser irradiation, no explosions or burning of the samples was observed. As observed from [Figure 29](#), the acquired spectra show the presence of emission peaks associated with C (247.8 nm), Mg (279.5 & 280.3 nm), Ca (393.3, 396.8, 422.7 nm), H (656.3 nm), N (742.4, 744.3, 746.9, 818.4, 818.8, 821.6, 824.2 nm), O (777.2, 777.4, 794.8, 822.2, 822.7, 844.6, 868.1 nm), Na (589, 589.6 nm) and C_2 (467.11, 558.86 nm). Evidently, the oxygen peak at 777.3 nm has the highest intensity followed by CN at 388.3 nm, which along with the C_2 peaks are representative of organic molecules. In addition, the CN violet bands are observed corresponding to

$B^2\Sigma^+ \rightarrow X^2\Sigma^+$ transitions at three regions with $\Delta v = -1, 0$ and $+1$ at 357-360 nm, 384 – 389 nm and 414 – 423 nm ranges, respectively. C_2 swan bands

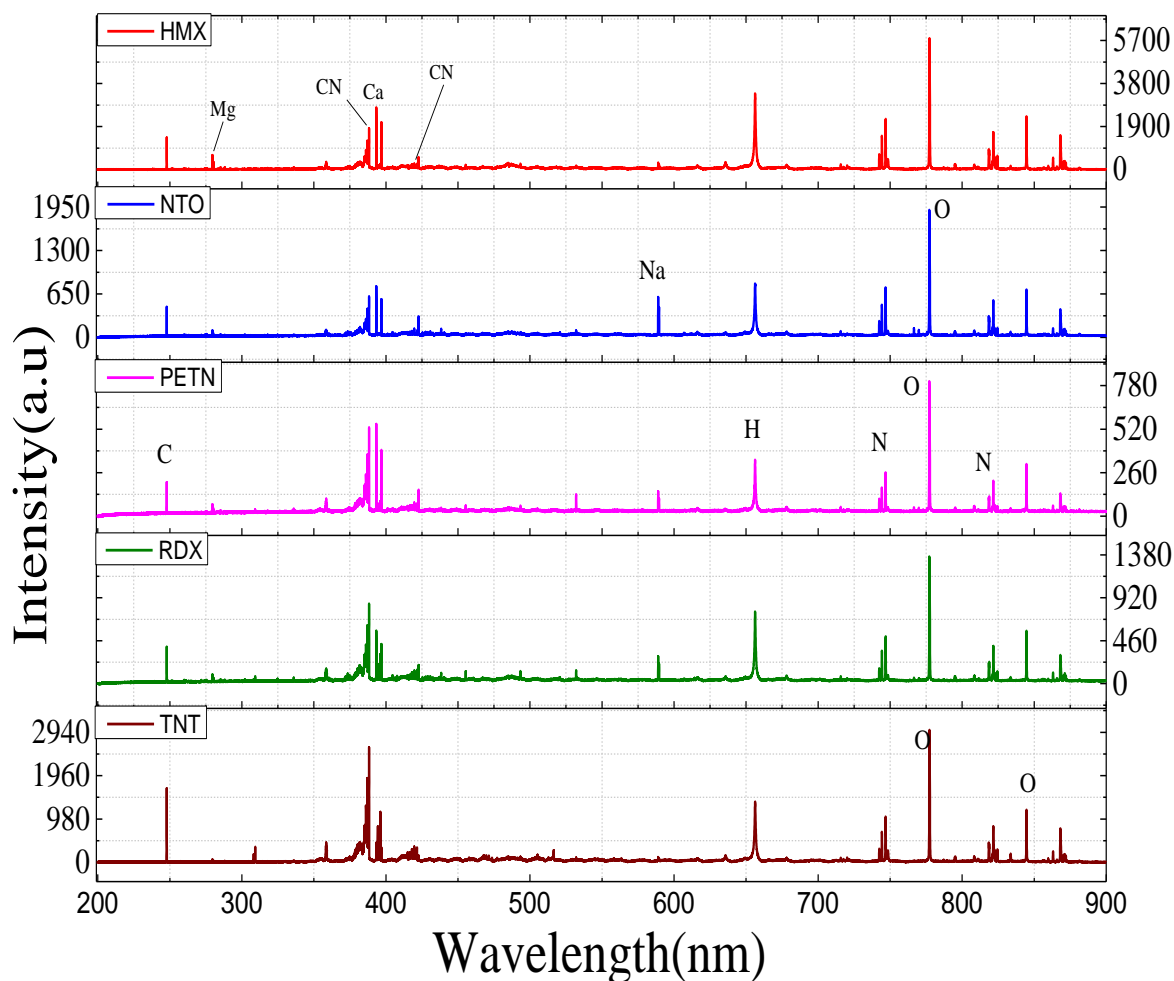


Figure 29. Representative spectra of HEMs used in the LIBS study

corresponding to $D^3\Pi_g \rightarrow a^3\Pi_u$ transitions are also seen in two regions with $\Delta v = -1$ and $+1$ at 460 – 475 nm and 550 – 565 nm ranges, respectively. Different possibilities for the molecular formation of CN and C_2 in the plasma have been proposed in the literature - (a) The sample fragmentation leads to direct release of radicals (b) recombination of atomic constituents from surroundings with the constituents from the sample leads to formation of CN (c) carbon and nitrogen atoms from the compound combine to form CN. Intermediary diatomic molecule C_2 form during plasma evolution The kinetic model based on RDX sample also shows that numerous reaction possibilities may exist in the plasma [33]. Looking at the spectra (Figure 29), the HEM specimen show expectedly near-identical peaks, however careful inspection reveals the subtle, but reproducible, peak intensity variations across the samples. A set of known samples have been chosen as non-

explosives. Table 10 shows the list of the samples considered in this class. The organic nature and ease/readily availability was

Compounds	Formulae	No of spectra used
Dextrose	$C_6H_{12}O_6$	23
Dfructose	$C_6H_{12}O_6$	30
Ltryptophan	$C_{11}H_{12}N_2O_2$	61
Cytosine	$C_4H_5N_3O$	34
SLS(Sodium lauryl sulphate)	$NaC_{12}H_{25}SO_4$	46
Glycene	$C_2H_5NO_2$	50
Hypoxanthine	$C_5H_4N_4O$	31
Methionine	$C_5H_{11}NO_2S$	43
Neomycin	$C_{23}H_{46}N_6O_{13}$	37
Pyridine	C_5H_5N	35
Thymine	$C_5H_6N_2O_2$	32
Protein	-	54

Table 10. Details of organic compounds designated as non explosive class

the criterion for the selection of these samples. A total of 476 spectra were recorded under identical excitation and collection conditions. Each spectral profile represents the signal averaged over 3 consecutive pulses. In Figure 4 spectra of HEMs and non-HEMs are plotted together to observe similarities and differences. Five different compounds viz., dextrose, HMX, RDX, cytosine and glycine spectra are plotted together. It can be observed from the spectra that though inorganic peaks such as Na, Ca, K and Mg have minor differences, while the major peaks related to C, H, N and O exhibit high degree of similarity irrespective of their category. The details of elements tabulated in table 2.

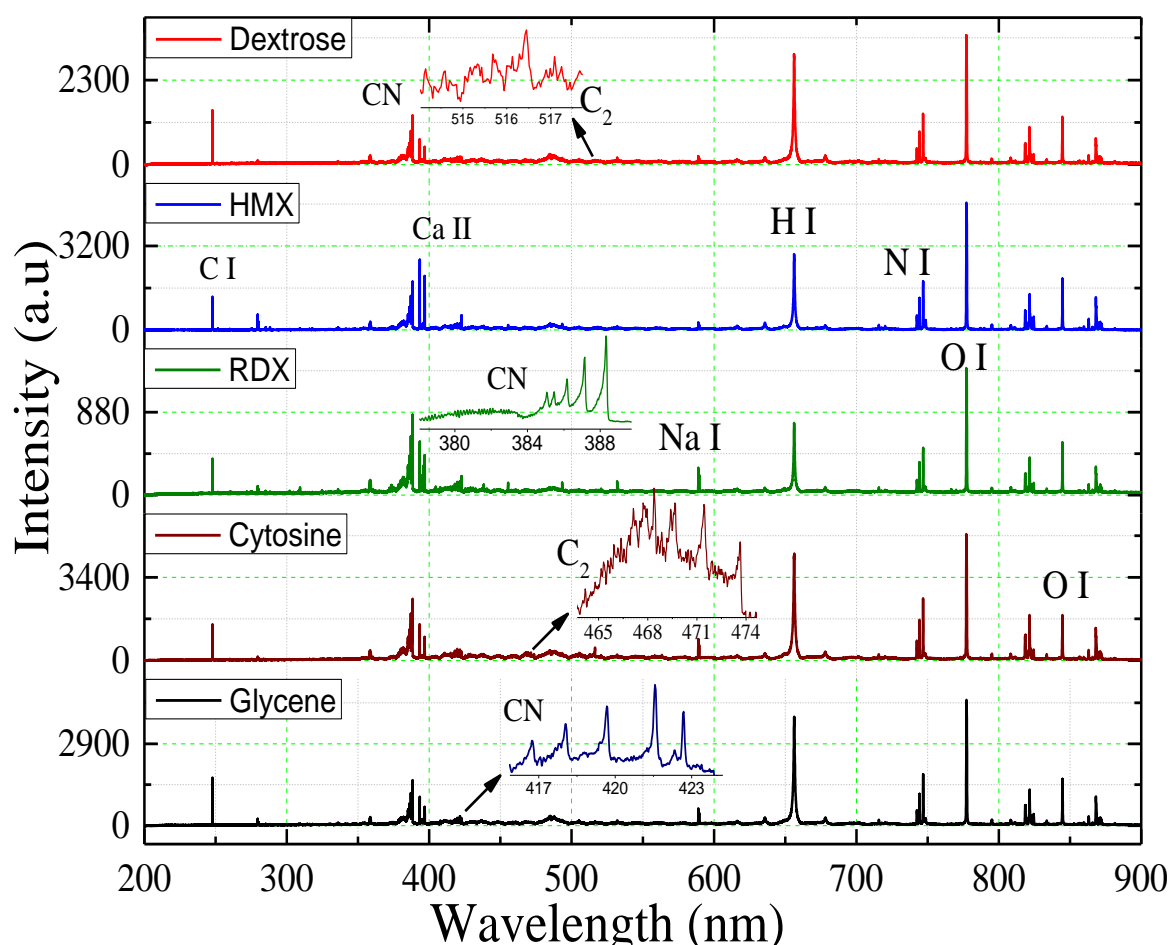


Figure 30. Representative spectra of Dextrose, HMX, RDX, Cytosine and Glycine.

The atomic and molecular peaks decay with time. Figure 5 shows the time evolution of the LIBS spectra of TNT compound. An initial delay of $1\mu\text{s}$ and width of $1\mu\text{s}$ was used. The delay was stepped $1\mu\text{s}$ for recording the time evolution spectrum. The Na peak is not observed explicitly in the TNT spectra. Other spectral features related to CN, C, H, N and O were found to decay with different rates. CN molecular peak stays longer, while the other peaks are short lived. [figure 31](#) shows the decay of LIBS signal of TNT compound with a function of time delay. The time evolution is modeled as single exponential decay and the decay constant is $1.3\mu\text{s}$. Carbon peak dies down faster when compared to other atomic peaks.. It can be observed that C and N are the shortest lived peaks and CN lasts for $> 7\mu\text{s}$. The results of the decay time constants are summarized in table provided in [Figure 32](#). While the other elemental peaks have intermediate decay times.

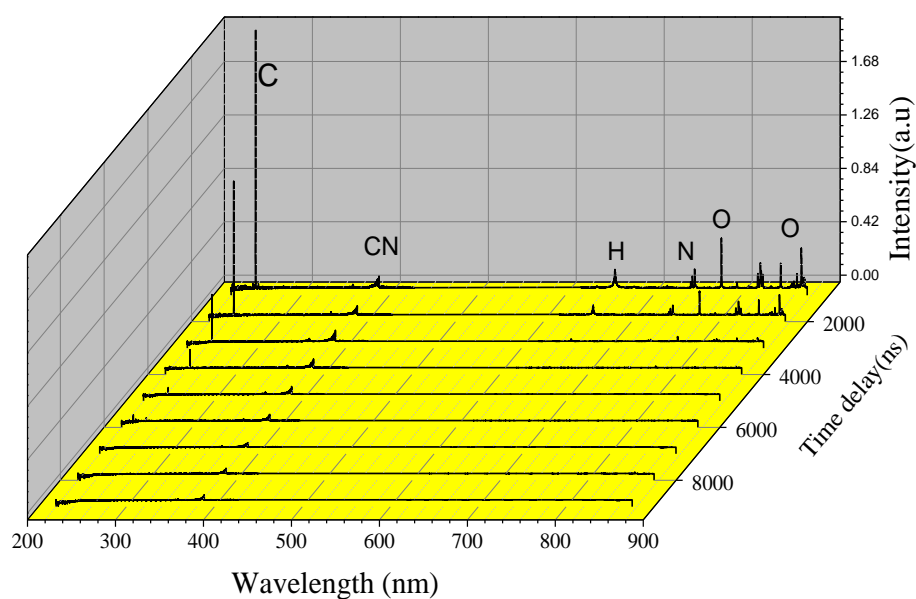


Figure 31. The Decay of LIBS signal of TNT compound with a function of time delay

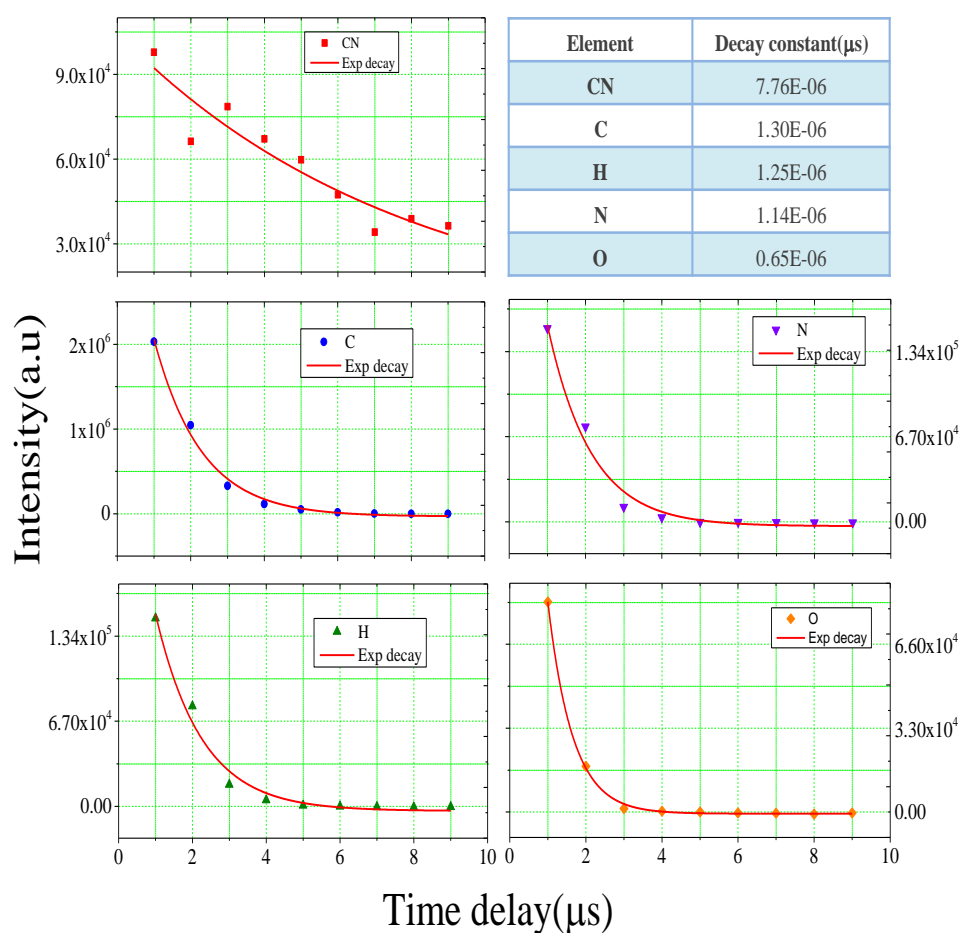


Figure 32. The decay of elemental peaks C, H, N, O and molecular peak CN as a function of time delay. Upper quarter table shows the decay constants of the above peaks.

Wave Length	Ele Ment	Dex Trose	Dfru Ctose	HMX	L Trypto	NTO	PETN	RDX	TNT	Cyto	Gly	Hypo	Meth	Neo	Prot	Pyri	Sls	Thy
247.89	C I	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
384.82	CN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
385.2	CN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
385.86	CN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
386.85	CN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
388.05	CN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
393.19	Ca II	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
415.02	CN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
415.59	CN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
416.61	CN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
417.89	CN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
419.51	CN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
421.42	CN	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
467.752	C ₂	*	*	N	*	N	N	N	N	*	N	*	N	N	*	*	*	*
468.71	C ₂	N	N	*	*	N	N	*	*	*	N	*	*	N	*	*	*	*
469.71	C ₂	*	*	N	*	N	N	N	*	*	N	*	*	N	*	*	*	*
47140	C ₂	N	N	N	N	N	N	N	*	*	N	N	*	*	*	*	*	*
473.6	C ₂	N	N	N	*	N	N	N	*	*	N	*	*	N	*	*	*	*
512.94	C ₂	*	*	*	*	*	*	*	*	*	*	*	*	N	*	*	*	*
516.35	C ₂	*	*	*	*	*	*	*	*	*	*	*	*	N	*	*	*	*
558.41	C ₂	N	N	N	*	N	N	N	N	*	N	*	N	N	N	*	*	*
563.46	C ₂	N	N	N	*	N	N	N	N	*	N	*	N	*	N	*	*	*
656.45	H I	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
742.46	N I	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
744.36	N I	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
746.89	N I	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
777.52	O I	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
766.51	K I	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
769.96	K I	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
589.041	Na I	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
589.71	Na I	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

Table 11. Emission lines from the organic compounds spectra (* indicates presence of element, N indicates the absence of the element)

However, similar data for the other explosives could not be recorded because of the non availability of the adequate quantity of sample. More amount of sample is needed to record the time evolution spectra as multiple shot averaging at each time step is involved here. The time evolution information can be crucial in selecting the sub- spectra for the classification of the samples. When dealing with large database of several compounds, it is important to identify the important variables (features) which account for discrimination. The underlying basis of feature selection is to exploit these intensity

variations not only at the notable peaks (which often do not provide the highest diagnostic power) but also across the spectrum to obtain an accurate classifier and, ideally, identify the smallest spectral subset of discriminatory features.

3.3 Plasma parameters

Two important properties, Temperature and electron density, of the laser induced plasma can be estimated from the LIBS spectrum. The details of the calculations were demonstrated in section 1.3 & 1.4. From Equation 7, it is evident that the emission intensity is dependent on the temperature of the plasma and the emission is the direct consequence of the plasma cooling down. Assuming local thermodynamic equilibrium, the plasma temperature is estimated using Boltzmann equation 1.3.1. Oxygen lines at 777.5 nm, 794.5 nm and 844.4 nm are used. The intensity is considered as an area under these peaks. The mean temperature for laser induced plasma of explosives is found to be of the order of 11000 K. The statistics of the temperature for all explosive compounds is shown as a box plot in figure 33(a). Electron density is estimated using stark broadening of oxygen peak at 844.4 nm. The peak is fitted using lorentzian line shape and the FWHM is related to the number density of electrons present in the plasma. The mean electron densities are found to be of the order of $\sim 3.5 \times 10^{15} \text{ cm}^{-3}$. Figure 33(b) depicts the statistics of the electron densities for all explosive compounds. Considering the error bars, which are obtained from estimating the parameters from multiple LIBS spectra and the identical

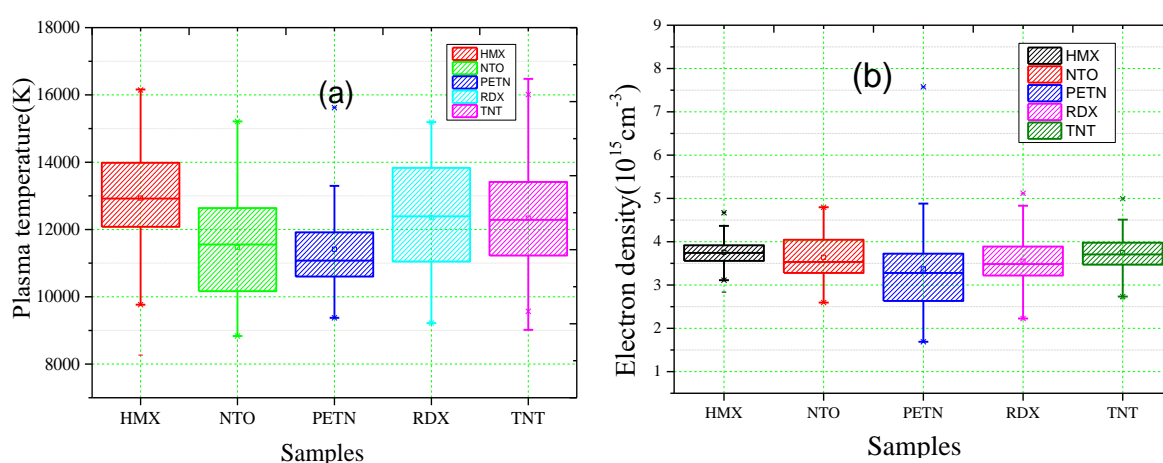


Figure 33. Box plot illustrating the statistics of (a) plasma temperature and (b) electron density of different explosive compounds.

excitation and collection conditions, the estimated temperatures and electron densities for all the explosives appear to be of similar magnitude. Similarly the laser induced plasmas of non-explosives are characterized for their temperature and electron densities. Figure 34(a) and Figure 34(b) shows the bar plot for the estimated temperature and electron densities. The mean temperature for laser induced plasma of non-explosives is

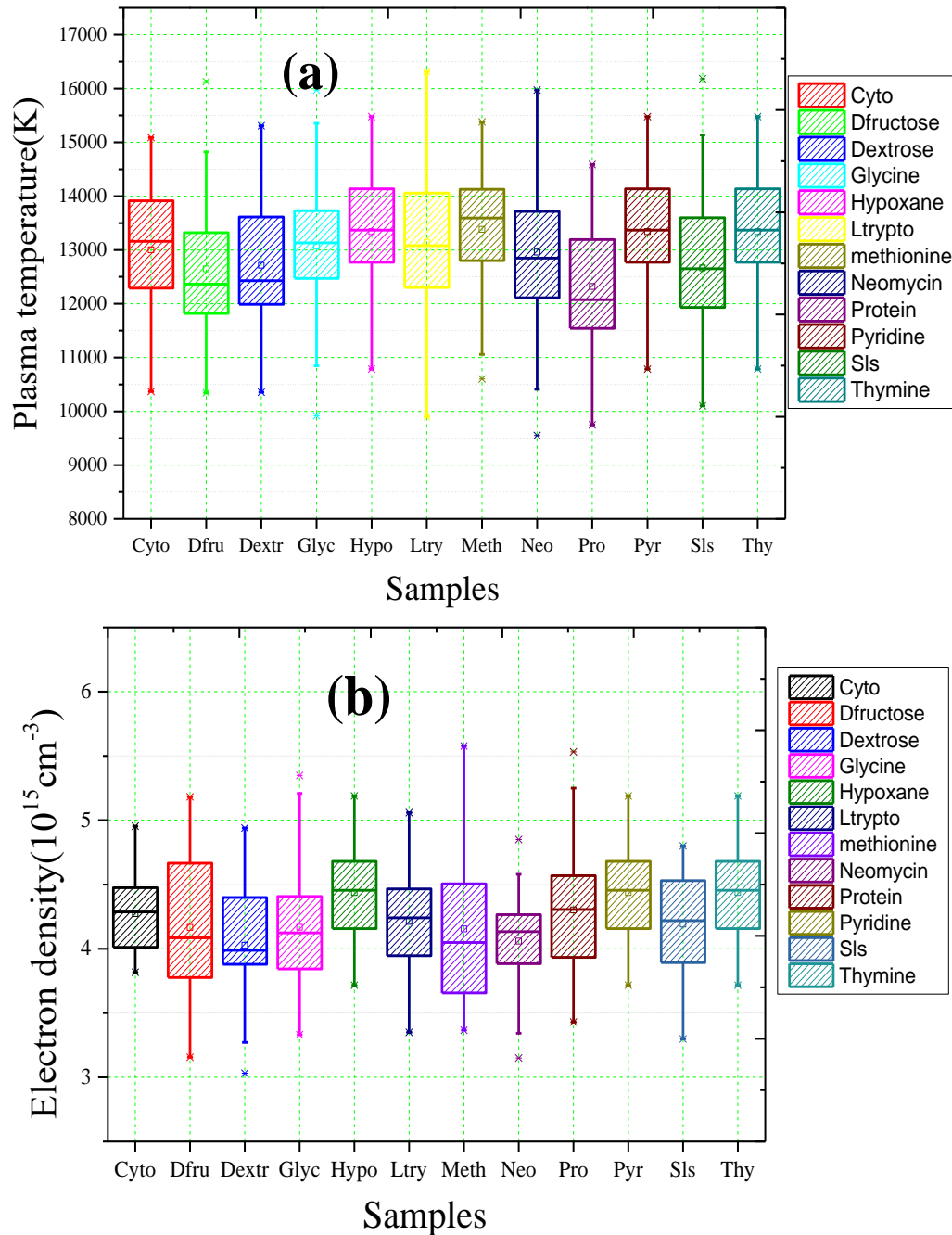


Figure 34. The statistics of (a) plasma temperature, (b) electron density using non explosives.

of the order of 12500 K and the mean electron densities are found to be of the order of $\sim 4.2 \times 10^{15} \text{ cm}^{-3}$. Though both the parameters are slightly higher for the non-explosives when compared to the explosives, considering the error bars for both the classes this difference may be statistically insignificant.

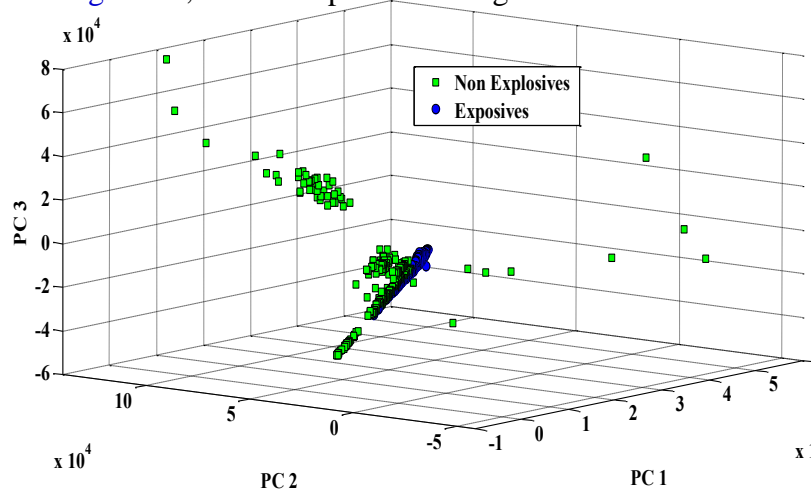
3.4 Results and Discussion

3.4.1 EXPLOSIVE AND NON EXPLOSIVE IDENTIFICATION

In the area of explosive detection, most of the LIBS community's interest is to identify small traces, while deriving the research, the implications of atmosphere contribution to the signal and feature selection is focused. All the spectra used in this chapter were acquired using gated spectrometer. The spectra of compounds, listed in [table 9](#) are utilized for this study. The total input data consisted of 472 spectra for explosives and 476 for non- explosives.

The recorded spectra consist of intensities at 25,699 points, spanning a wide spectral region from the ultraviolet to near-IR. The large number of spectral points in most datasets in LIBS multivariate analysis often renders the prediction of a compound unreliable which is sometimes referred to as '*curse of dimensionality*'. While the LIBS spectrum consists of several sharp peaks attributed to the elemental (and some molecular) emission lines, a substantial portion is composed of minimal signal beyond the noise floor ([Figure 29](#)). However, very rarely is there a concerted effort in choosing the spectral variables that is rooted in the context of the application (for example, chemical composition of the samples) – rather for reasons of expediency the full LIBS spectrum is routinely employed in classification. Though inclusion of noise [34] and continuum [35] may not significantly hamper the prediction performance in specific cases, feature selection in general is critical to eliminate spurious correlations especially in applications where the inter-class differences are subtle and also for developing classification models that are faster, which is a crucial requirement for a on -field device.

Full spectra of compounds were selected as input to the unsupervised classification method i.e., PCA. Figure 35, is a score plot involving the data of both the classes to



enable

Figure 35: Score plot of Non explosives and explosives

exploratory data analysis. No preprocessing was applied prior to the analysis. Discrete localized clusters were observed in the score plot. The points representing non-explosives in the score plot were found to be more diverged and have distinct clusters, which also intersected with the explosives cluster.

The first three PCs account for 66.57%, 15.43% and 12.43% of the variance of the input data. The first five PCs cumulative account for 96% of the variance in the data. The total of all subsequent PC combined account for the remaining 4 % of the variance of the data. While, score plot indicates that there is chemical/spectral information in the data which leads to separate clusters for the explosives and non-explosives, it cannot give any information of the features responsible for the classification. Figure 36 shows the loadings plot for the first three PCs. Most of the coefficients of first principal component have strong positive correlation with the

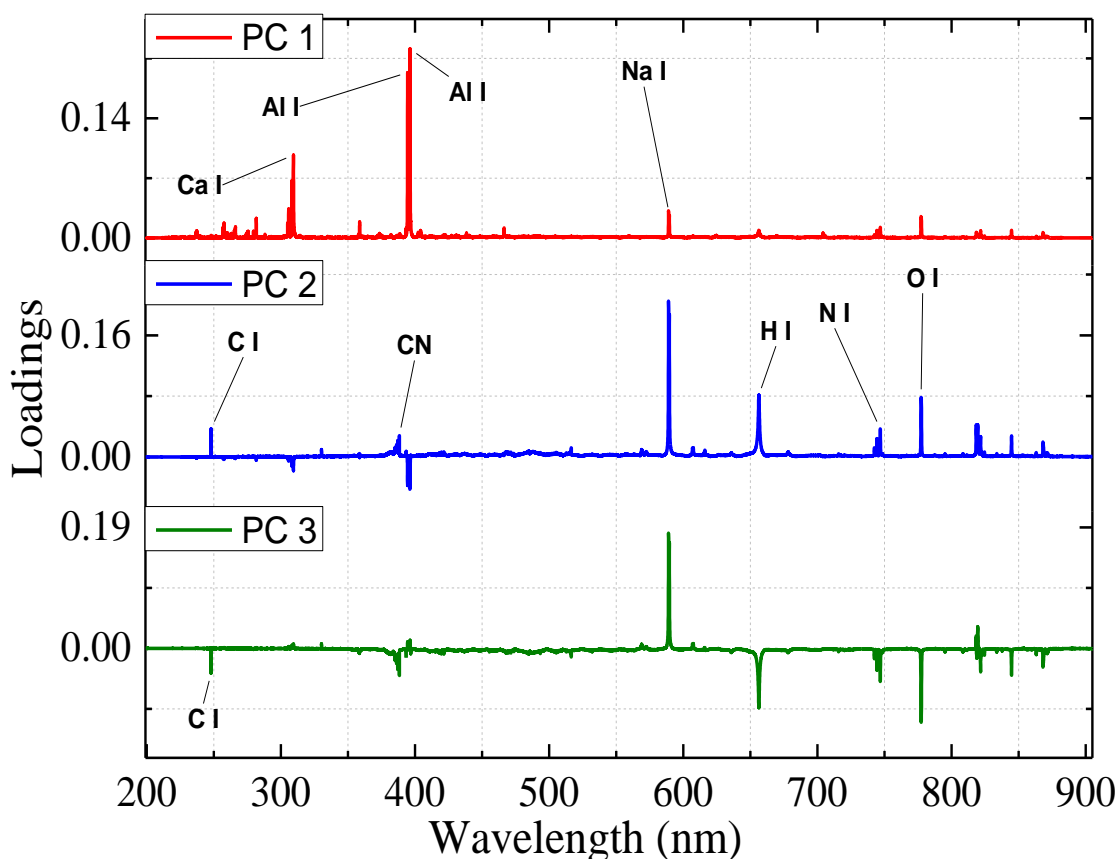


Figure 36: Loadings plot of explosives and non explosives

inorganic peaks calcium, aluminum and sodium with very weak correlation with the other peaks of carbon, hydrogen, nitrogen or oxygen. While the second PC has a positive correlation with the carbon, CN peak, sodium, hydrogen, nitrogen and oxygen; negative correlation with aluminum and calcium peaks. The third PC has positive correlation with sodium, while negative correlation with carbon, hydrogen, nitrogen and oxygen. Taken together, all the results from [figure 35-36](#), indicate that that the separation among explosive and non-explosives predominantly relies on inorganic peaks, while the other peaks contributing to a lesser extent. It is well known that inorganic peaks are present in these sample as minor constituents or impurities. IN the context of an actual application, these entities are not in the control of a user and hence may lead to spurious results.

Further, to investigate the problem of inorganic elemental features serving as PCs for classification; every spectrum was normalized using area under the spectrum i.e., sum of all intensity spanning the entire full spectrum, which was successfully demonstrated by sirven et al [27]. The main aim of the task is not only to reduce the fluctuation caused in the score but also to see if better discrimination can be achieved by having more localized

clusters. The score plot obtained using the normalized data is shown [Figure 37](#). By visualizing this three dimensional plot and comparing it with the former case as shown in [figure 35](#), it is apparent that scattering in the clusters is reduced and the separation also increased between clusters. On the contrary the cumulative variance was decreased. The cumulative variance using five PCs in this case is 82. 51 %; while, in the earlier case it was 96.36 %. Loadings plot is shown in [Figure 38](#). PC1 contains the very dominant positive correlation with Na peak and negative correlation with CN peak. PC2 contains positive correlation with C, CN, Na, H, N and O peaks. However it should be observed that in both the above cases, the separation of clusters of hem and non-hems seems to be emanating primarily from Na peak and the other inorganic elemental peaks.

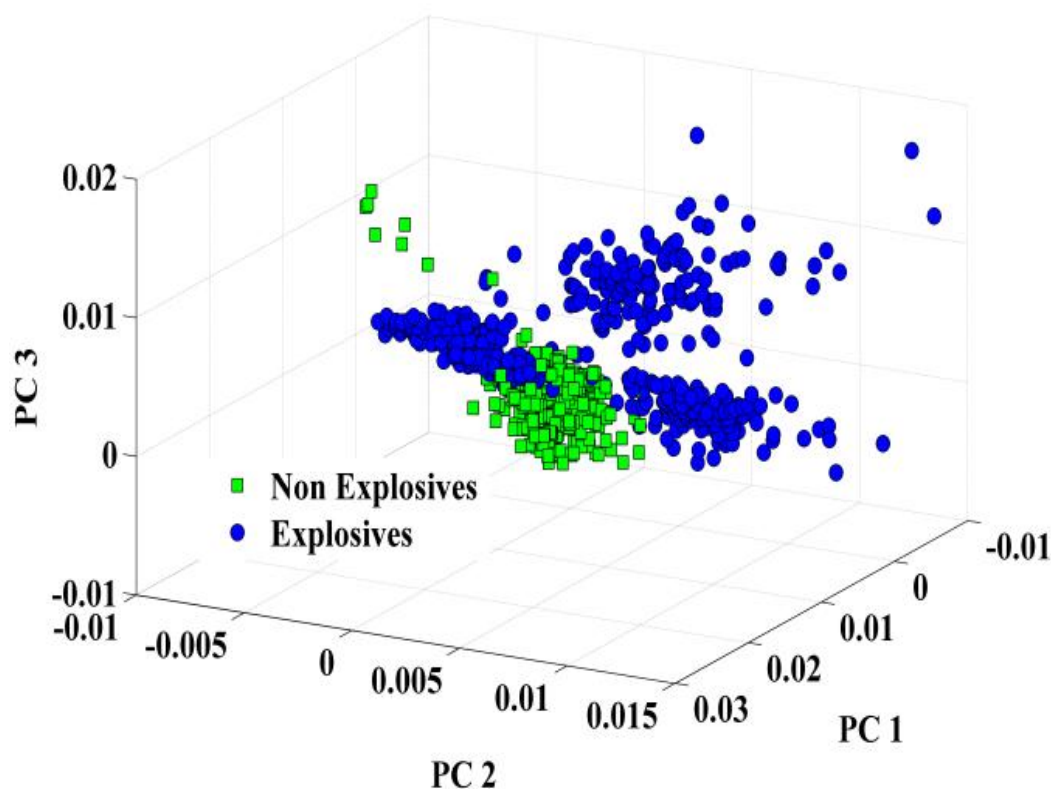


Figure 37 Score plot of non-explosives and explosives with data normalized to area under the curve.

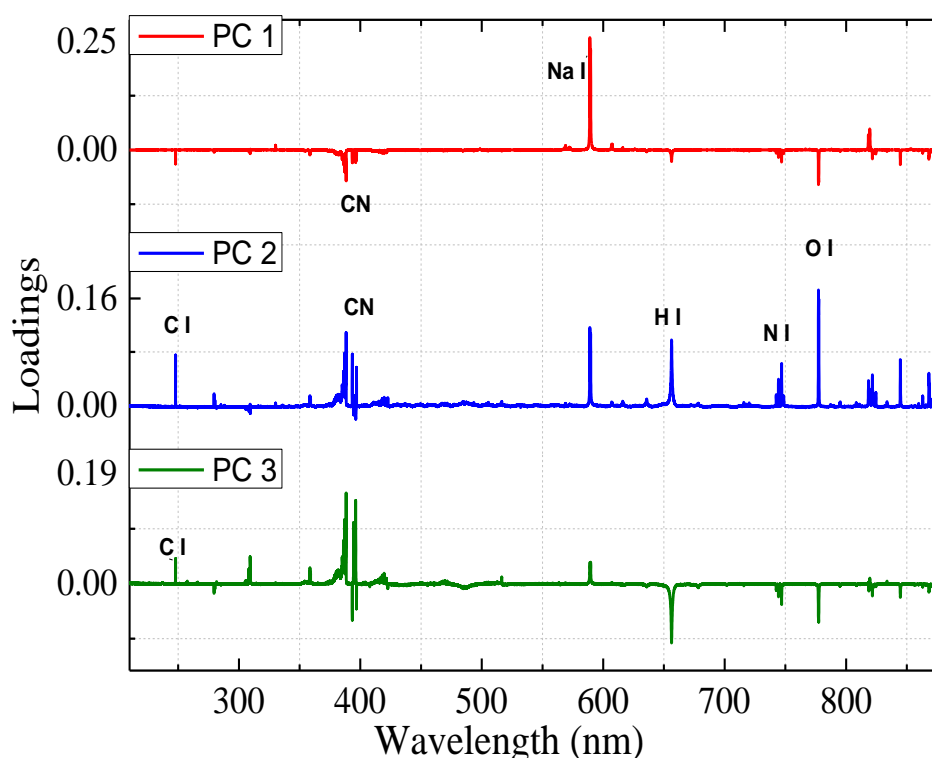


Figure 38. Loadings plot using normalised data of explosives and non-explosives

In order to overcome this problem of classification depending on the inorganic peaks, it is necessary to adopt variable selection for improved classification. Large datasets may sometimes make outcome unreliable and hinder the results. In the next section, systematic study has been undertaken to find and understand the potential variables which can contribute useful information towards classification and preferably avoiding the inorganic peaks. The variable selection was performed with the prior knowledge of elemental constituents of the material. Two different spectral variables were extracted for classification using multivariate analysis -

- (i) Sub - spectrum : Here a region from the full spectrum is taken as sub- spectrum depending emission region of interest.
- (ii) Specific peaks : Here, individual peak has been considered as an input. The entire spectral region of the peak has been considered.

(I) SUB SPECTRUM:

Based on the prominent spectral features that are representative of elemental constituents of molecule ,CN, C₂, C, H, O and N, four different regions R1, R2, R3 and R4 detailed

in Table 12 and Figure 39 have been selected as sub-spectra. It is important to note that, now the size of the input matrix becomes smaller depending on the region selected. Sub-spectrum, R1, with ~ 3% of the spectral points, consists of carbon and hydrogen-specific peaks. The reason for selection of this region is that these are dominant peaks for HEMs and are less influenced by the surrounding ambient atmosphere in comparison with oxygen and nitrogen emissions. Region R2 is based on the peaks representative of C and CN. The formation of CN may depend on the content of carbon and nitrogen in the plasma, with the possible nitrogen contribution from the atmosphere as well. Region R3 provides the cumulative information of R1, R2 and C₂ which is expected to indicate the presence of aromatic double bonds (C=C) [36]. Finally, region R4 consists of the higher wavelength spectral region following the hydrogen peaks and can have contributions from both the HEM specimen and the atmosphere. It was evident in the previous section that the application of PCA on full spectrum, picks information mostly inorganic peaks. In contrast, in this section PCA is applied on four different regions. The data has been normalised prior to application of PCA. Further it should be noted that the H peak is possibly the least affected by the atmosphere.

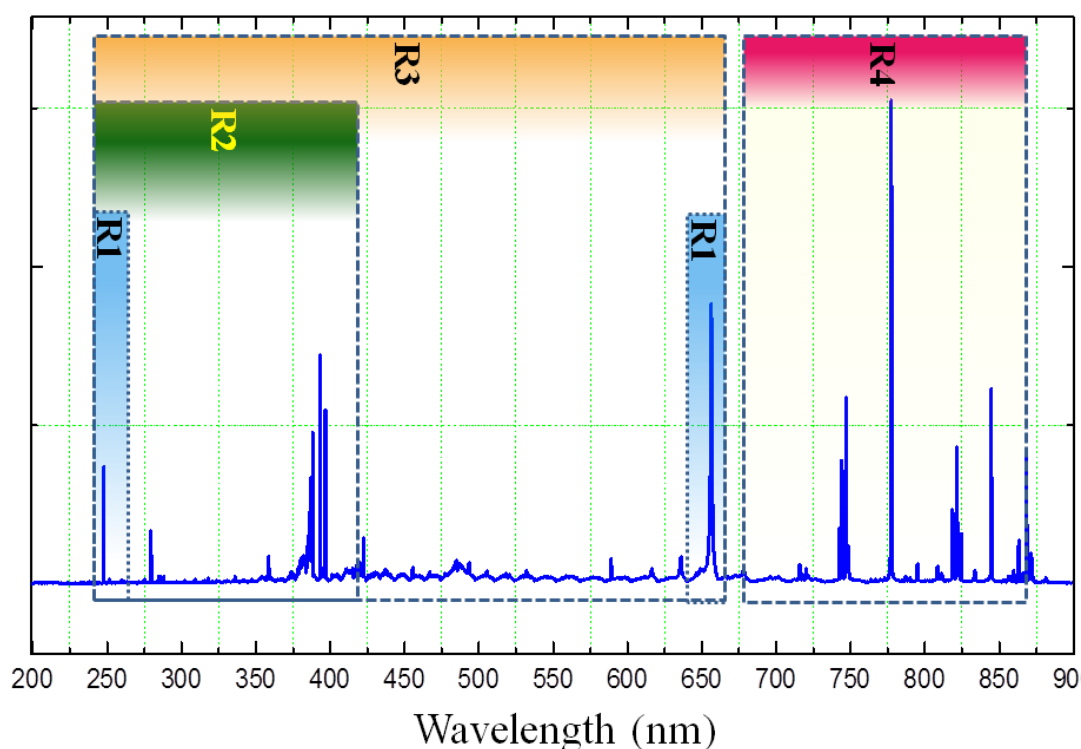


Figure 39. Full spectrum division into four sub spectral regions.

Sl. No.	Region Name	Description	Wavelengths included (nm)	% of full spectrum
1	R1	C and H peaks	246.98-248.05; 640-670	3.16
2	R2	C to CN region	246.98-424.04	34.05
3	R3	C to H region	246.98-670.04	62.85
4	R4	Post-H region	670-875.99	17.06
5	Full	Entire spectrum	199-981	100

Table 12: Details of selected spectral data regions based on the sample compositional information

PCA is applied on R1 and the corresponding score plot is shown in [figure 40](#). It can be observed that both the classes are classified clearly with less scattering as compared with two score plots of full spectrum in the previous sections. The loadings have positive correlation with carbon and hydrogen peaks and inevitably the separation between classes in score plot is due to these peaks. The variance is moderately distributed across all PCs. First three PCs explain the variance of about 32.12%, 20.45% and 3.84% respectively. The cumulative variance explained using first five PCs is about 60.84 %. The clear separation between the clusters implies towards the use of this region for explosive identification using supervised statistical methods.

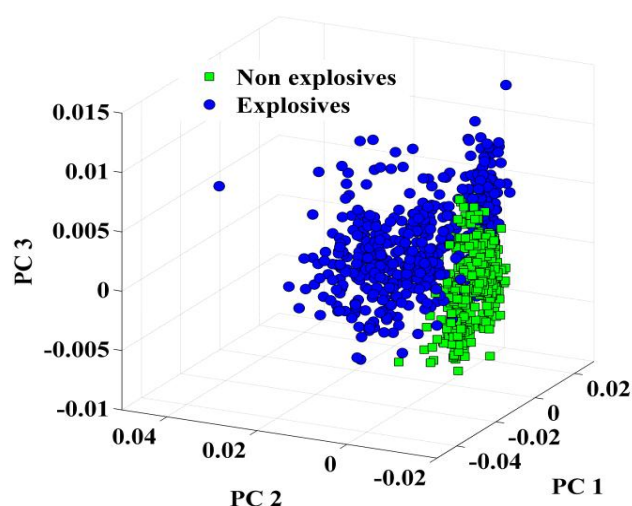


Figure 40. Score plot using sub spectrum R1

The sub spectrum R2 contains the region spanning from C to CN. The score plot using R2 is shown in figure 41. Two localized clusters of both classes can be seen in the

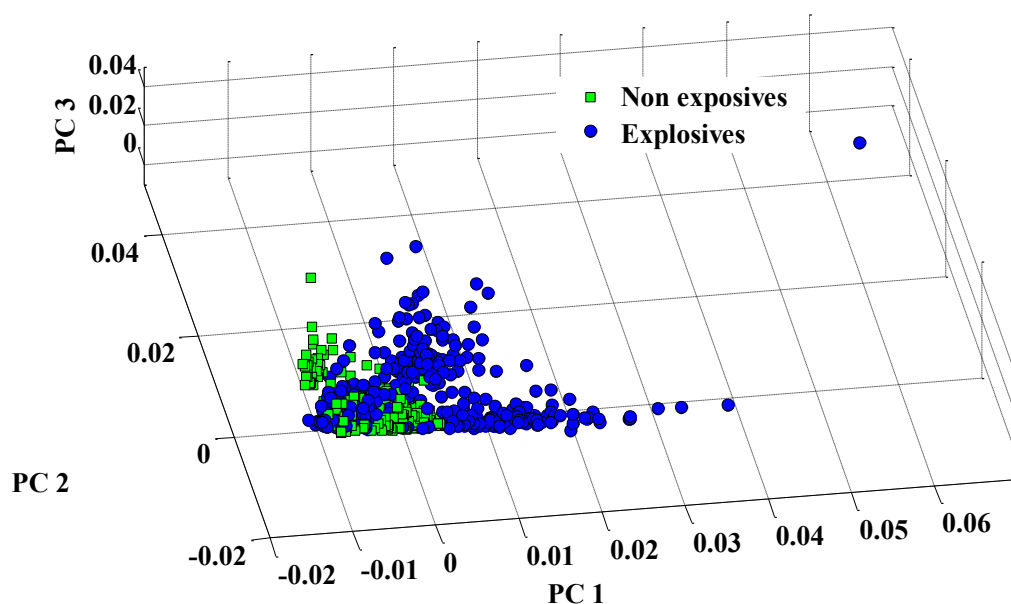


Figure 41. Score plot using R2 sub spectrum.

score plot and within the non-explosives, two more sub clusters are observed. These sub clusters may indicate towards different classes of samples within the non-explosives.

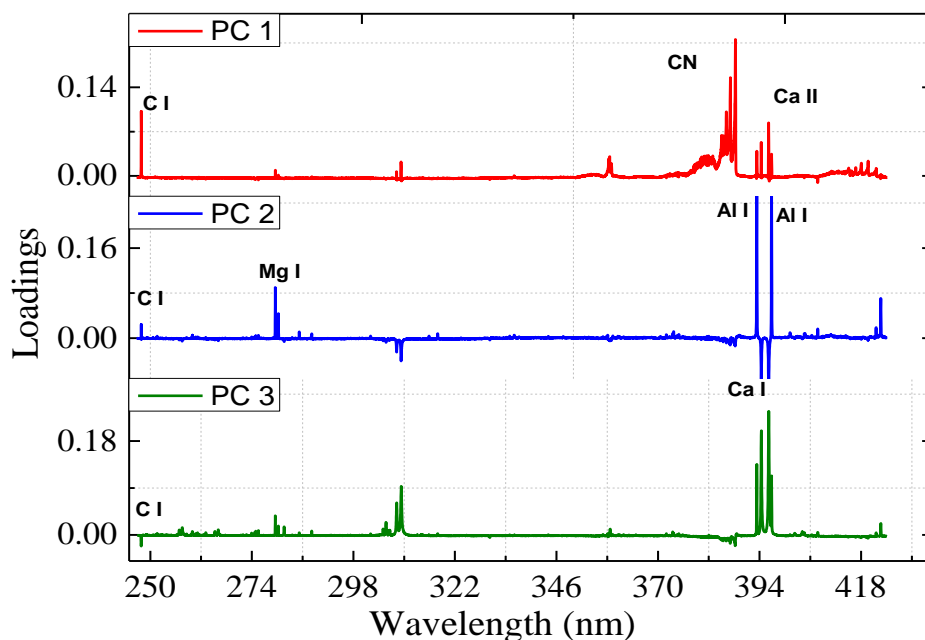


Figure 42. Loadings plot using R2 sub spectrum.

From the loadings plot shown in [figure 42](#), it is evident that carbon, CN peak and singly ionized calcium have positive correlation with PC1. PC2 has positive correlation with carbon, magnesium and aluminum, where as PC3 has with carbon, magnesium, aluminum and calcium. The first three PCs explained variance of the sub spectrum is about 25.19%, 21.14% and 7.03% respectively. Cumulatively variance accounted with the first five PCs is 61.09%.

The sub spectrum R3 is an extension of R2 with additional peaks such as C_2 , Na and H. It is already observed that Na is contributing more for classification in the case of full spectrum. In this section, an attempt is made to look Na involvement towards clustering of classes in score plot. Score plot is shown in [figure 43](#) and the corresponding loadings plot in [figure 44](#). The coefficients are similar as that of full spectrum with Na having positive correlation with PC1 and negative correlation with carbon and CN peaks.

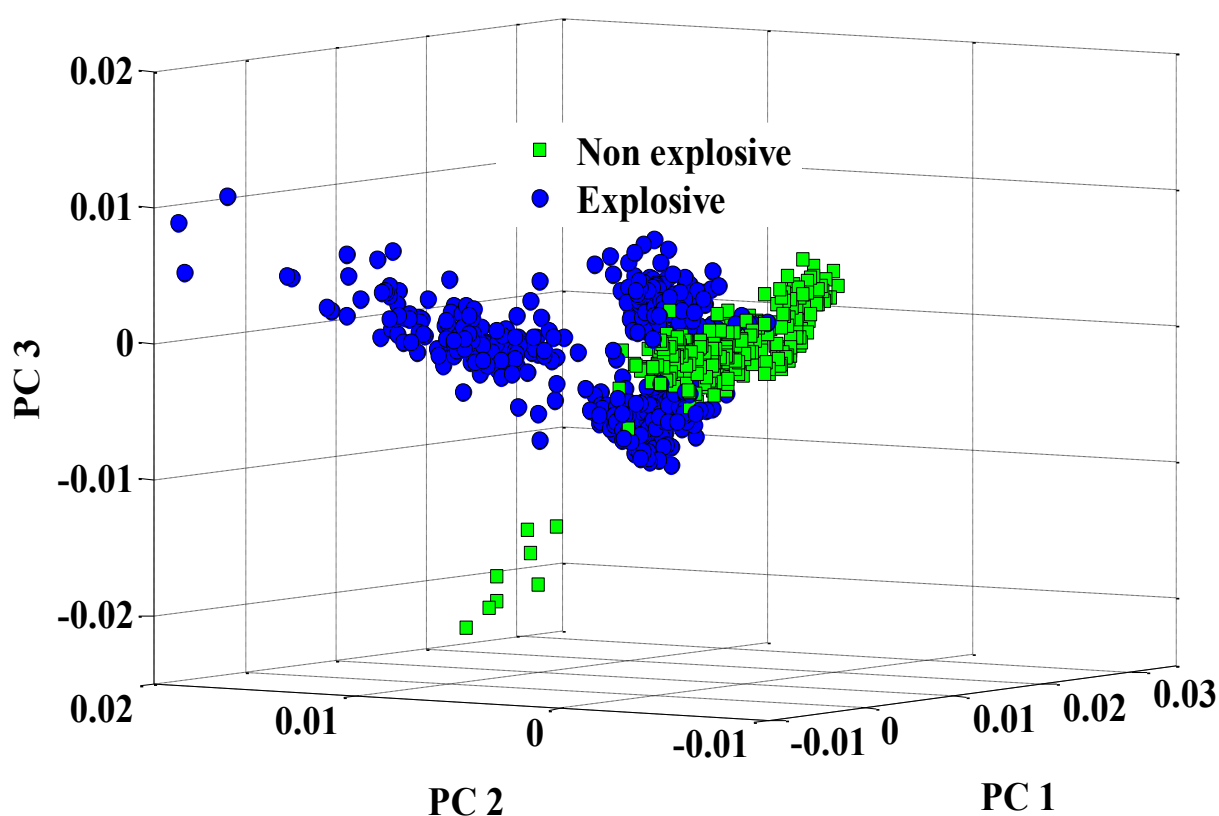


Figure 43. Score plot using sub spectrum R3

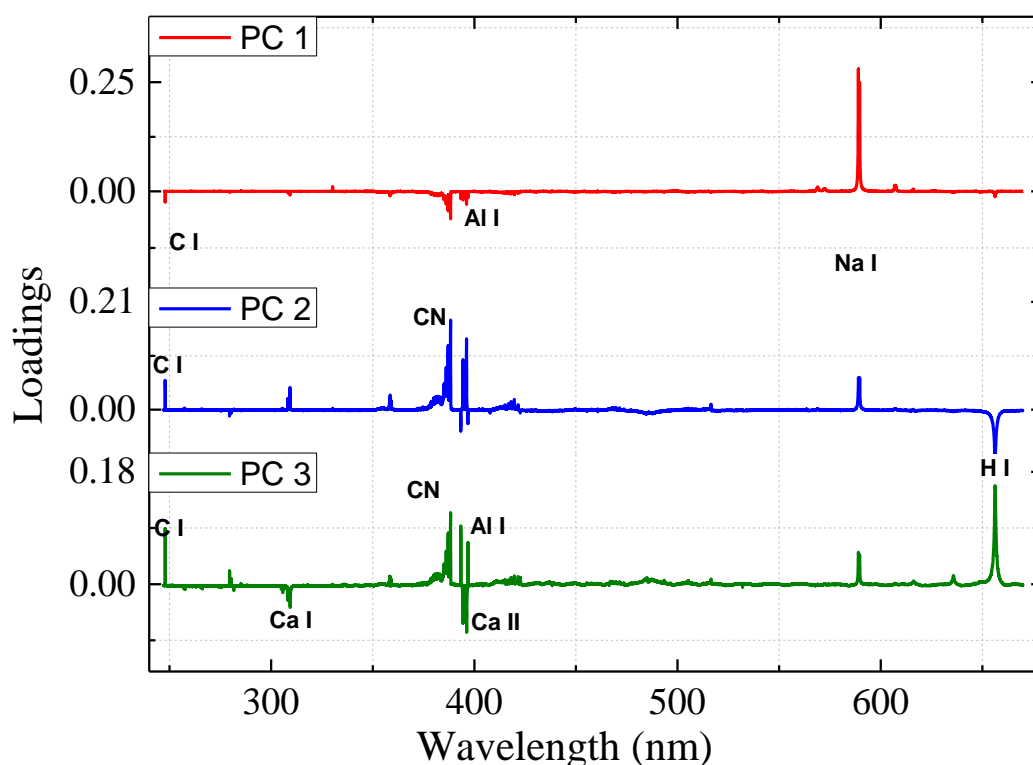


Figure 44. Loadings plot using subspectrum R3.

The variance explained by the first three PCs is 25.33%, 20.42% and 16.73% respectively. The cumulative explained variance of first five PCs together accounted 71.71% of the R3 sub spectrum. The two classes of explosives and non-explosives are very well separated.

The region R4 is dominated by peaks corresponding to oxygen and nitrogen, which are also the constituents of surrounding atmosphere. This implies that the oxygen/nitrogen emission intensity in the spectrum can have contributions from the atmosphere as well. The score plot and corresponding loadings plot have been shown [figure 45](#) and [figure 46](#). Separation has been observed between the classes in the score plot. PC1 is similar to the sub-spectrum consisting of all the oxygen and nitrogen emission lines. The first three PCs explain the variance of 38.47%, 17.60% and 7.38% respectively. Cumulatively the variance of first five PCs together accounted 67.47 of the sub spectrum.

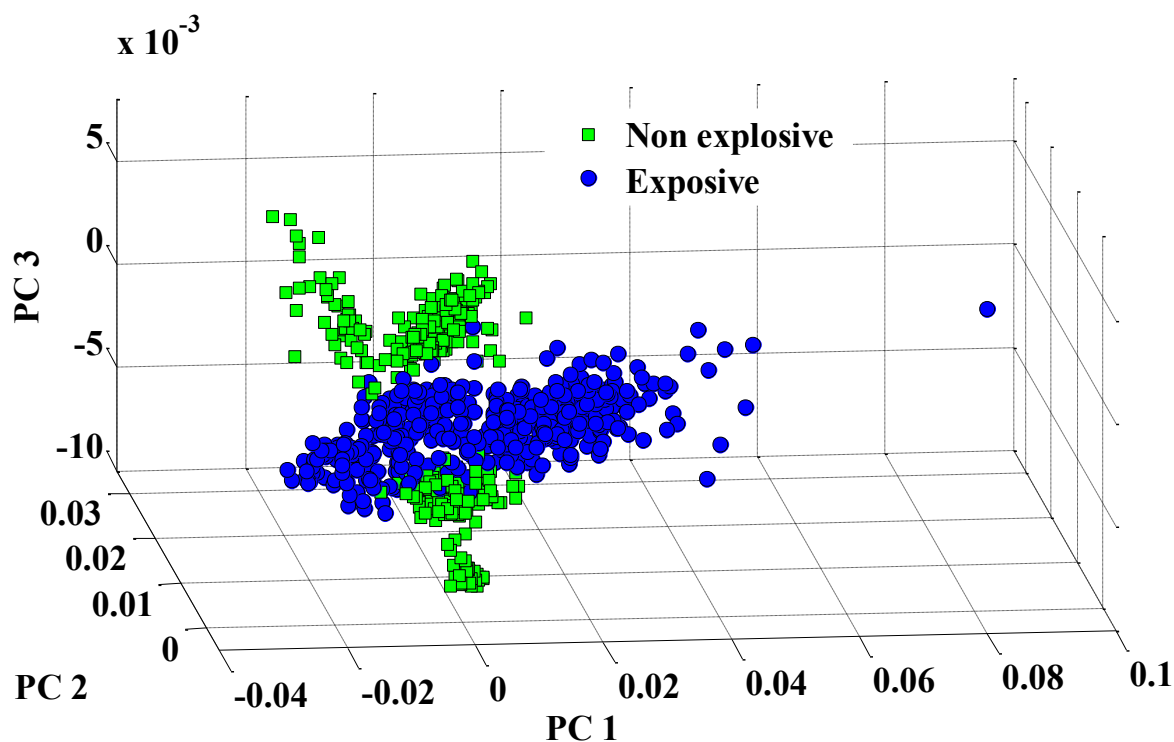


Figure 45. Score plot using R4 sub spectrum

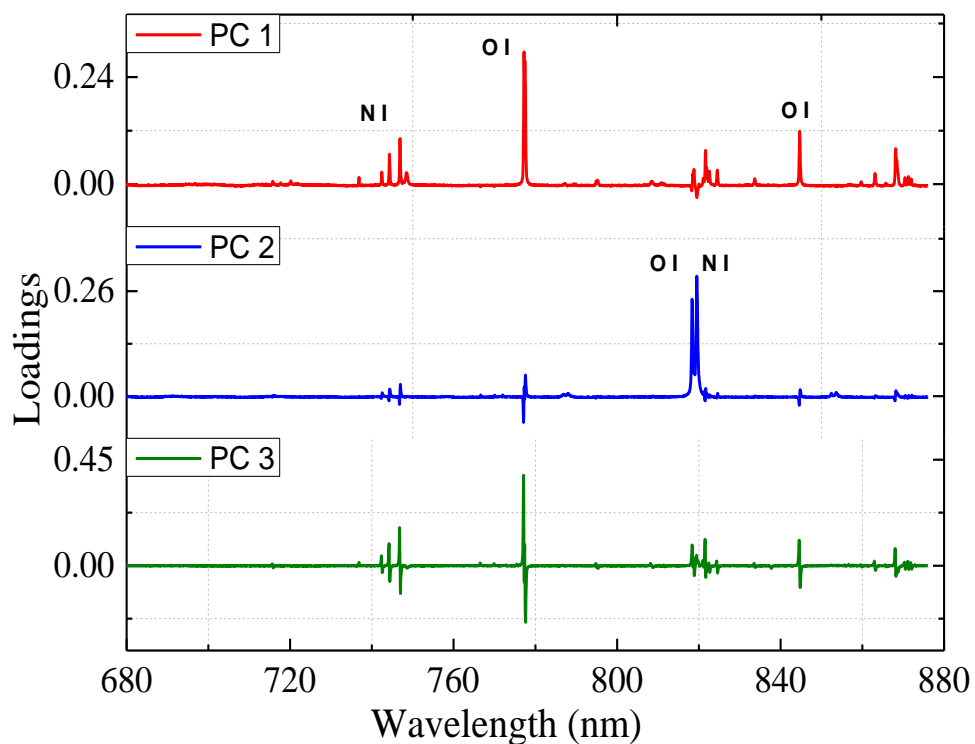


Figure 46. Loadings plot using sub spectrum R4.

All the four sub-spectra give a good binary clustering and sometimes the non-hem cluster showing sub clusters. However the two clusters are well separated. The results suggest that even in the absence of the inorganic elemental peaks, such as in R1, there is enough spectral information in other peaks that can lead to separation of HEMs and non-HEMs.

To assess the classification efficiency of using these regions, supervised classification technique ANN have been implemented for full spectrum and the sub-spectra. Bohling and his coworkers were successfully demonstrated mine detection using neural network [30]. In the similar fashion, in this chapter ANN was used for classification problem of HEMs and non HEMs. ANN belongs to a class of nonlinear, non-parametric classification and predictive methods, which have lately received attention in LIBS spectral analysis. For example, it has been used to predict chromium element concentration in soils [37], concentration of vanadium and nickel in complex crude oils[38], and for material identification in plant materials [39].

For ANN analysis, a two-layer feed-forward network model was used for classification consisting of one input layer, one hidden layer and one output layer. The input layer comprises of raw data intensity values at various wavelengths. Sigmoid functions were used as neurons for the input and hidden layers. The final output consists of binary encoding used as target, where 1 is for correct assignment and 0 otherwise. The ANN was trained with a classical scaled conjugate gradient back-propagation algorithm. The data division control was in-built in the matlab toolbox. Out of the total input spectra, 70% were presented to the network during training, and the network was adjusted according to its error. 15% of the spectra was used to measure network generalization and to halt training indicating an increase in the mean square error of the validation samples. The remaining 15% spectra were held out of the training and validation process and therefore provide an independent measure of network performance.

In this binary decision problem, the prediction made by the classifier can be represented by a structure known as contingency table/confusion matrix. The resulted confusion matrix after performing ANN is shown in [table 13](#). The diagonal entries in this table are percent correctly classified (PCC) classes; while off diagonal elements are percent misclassified (PMC) classes.

Targeted Labels	Predicted Labels	
	Explosive	Non explosive
	Explosive	Non explosive
Explosive	97.01	2.99
Non explosive	1.88	98.12

Table 13. Confusion matrix computed using neural network using explosive and non explosive data. diagonal elements in the matrix represent correctly identified, while off diagonal elements are called misclassified.

The input to the ANN classifier for full spectrum is 948 x 25699, where 476 belong to the class of non explosives and 472 explosives. The classifier was able to predict 97% as explosives where misclassification was about 3%. In the case of non explosives 98% was correct classified was and misclassification is 2%.

In the similar fashion, ANN is also applied on sub spectra to understand the impact of variables on the classification model. The correct classification performance given by all sub spectra is > 96% as shown in [table 14](#). In the case sub spectrum R1, whose variables are carbon and hydrogen, the average correct classification is 97.49 %.

Sub Spectrum	Correct Classification (%)	Mis-Classification (%)
R1	97.49	2.51
R2	95.99	4.01
R3	97.71	2.29
R4	98.15	1.85

Table 14: Average classification performance of neural network model on subspectrum data of HEM and non HEMS over 100 iterations

But with the sub spectrum R2, the mis-classification shows a slight increases to 4.01. The correct classification performance of other region R3 & R4 is 97.71 and 98.15% respectively. The results suggest that irrespective of input viz., full spectrum or sub-spectrum, the neural network model can perform the task of classification very effectively. However, the number of spectral points used differs for each sub spectrum

region. The result gives multiple options for the choice of the sub-spectrum. Except R1 and R4, the other two sub-spectra have peaks corresponding to inorganic peaks. Further, R4 can have major influence from the ambient atmosphere. R1 which consists of only ~3% of the spectral data is sufficient to get the performance as high as the full spectrum and is minimally affected by ambient and has no inorganic peaks. In the next section, the study has been extended to individual peaks alone to explore if it is sufficient to achieve classification.

(II) SPECIFIC PEAKS

In this section, individual elemental/molecular emission profiles viz., carbon, hydrogen, CN, nitrogen and oxygen are explored for their power to distinguish between HEM/non-HEMS. Two methods analogous to previous section, PCA and ANN are applied on these six spectral features as input. Carbon peak lies in the ultraviolet region at 247 nm, hydrogen lies in the centre of visible region at 656.6 nm, oxygen and nitrogen lie towards the end of the visible region. When carbon peak profile is taken spectral input, the score plot as shown in [figure 47](#) shows two clusters however the inter-cluster distance between them is very low and their surfaces merge.

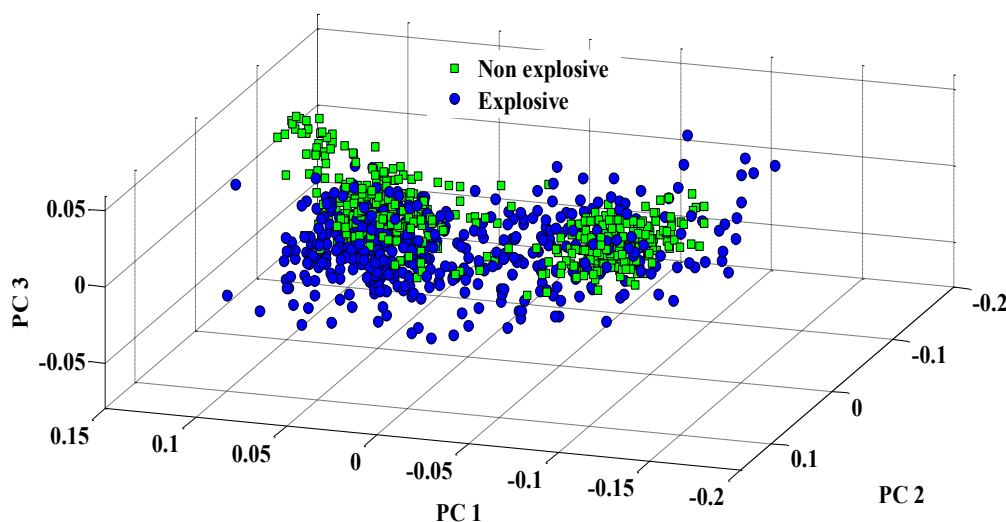


Figure 47. Score plot using carbon peak profile centered at 247.8nm.

The first three PCs explain the variance of 64.9%, 21.75%, 5% respectively. The cumulative variance explained by the first five PCs is 94.95%.

When carbon reacts with atmospheric nitrogen or compound's nitrogen in plasma, it creates CN molecule. The CN spectral profile at 370 - 390nm was chosen as input to

PCA The score plot is shown in [figure 48](#). Looking at the clusters it is evident that there is a wide spread of the points in both the clusters and there is an overlap at their centers. The reason

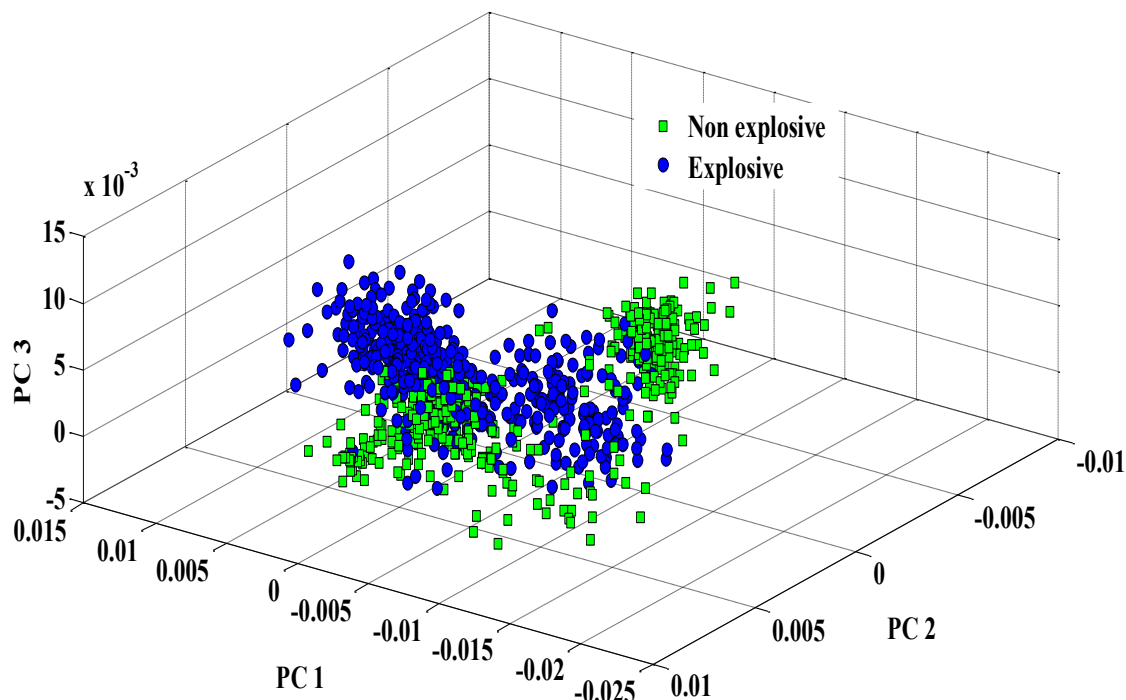


Figure 48. Score plot using CN peak profile .

may be the stochastic nature of CN formation and it cannot be explained without accounting plasma complex nature, which is random in different compounds. The first three PCs explain the variance of 24.68%, 8.57%, 2.4% respectively. The cumulative variance explained by the first five PCs is 39.31%.

Hydrogen peak at 656.6nm is the least affected by surrounding atmosphere and also it lies at the centre of CCD spectral sensitivity. PCA was applied on the hydrogen peak profile of the spectra and corresponding score plot is shown in [figure 49](#). The classes are very clearly well separated. Though the variance explained is less in number. The first three PCs explain the variance of 34.12%, 7.63% and 4.25% respectively. The cumulative variance explained by the first five PCs is 51.78%.

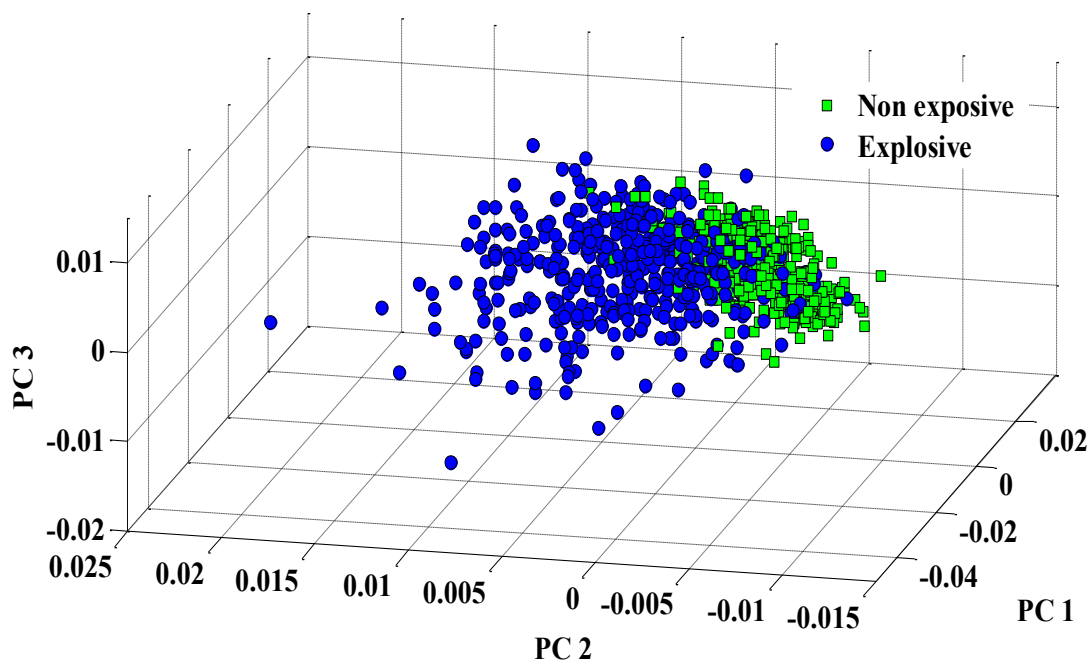


Figure 49. Score plot using hydrogen peak of non explosives and explosives at 656.6 nm

PCA was also applied on nitrogen peak profile. The score plot is shown in [figure 50](#). There were Local clusters of non-explosives have been observed with scattering points, though the explosive cluster is spread but it was localized.

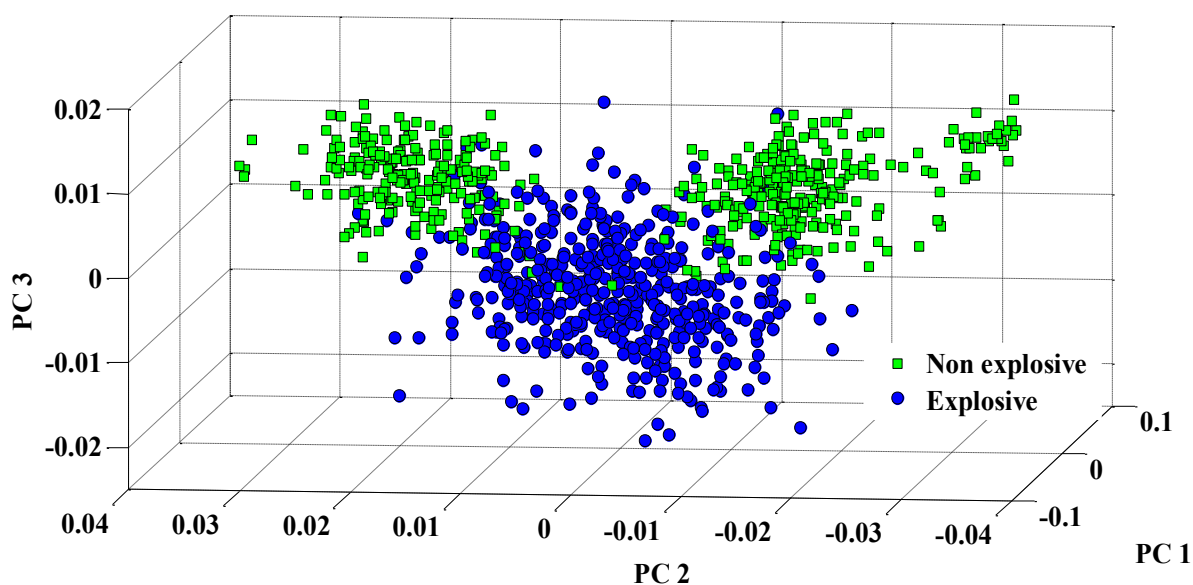


Figure 50. Score plot using nitrogen peaks of non explosives and explosives around 742, 744, 746 nm.

Both the classes are not well separated, but explosives are clustered well. The first three PCs explain the variance of 32.66%, 20.87%, 6.69% respectively. The cumulative variance explained by the first five PCs is 67.17%.

In the case of oxygen line, the score, loadings and scree plot is shown in [figure 51](#).

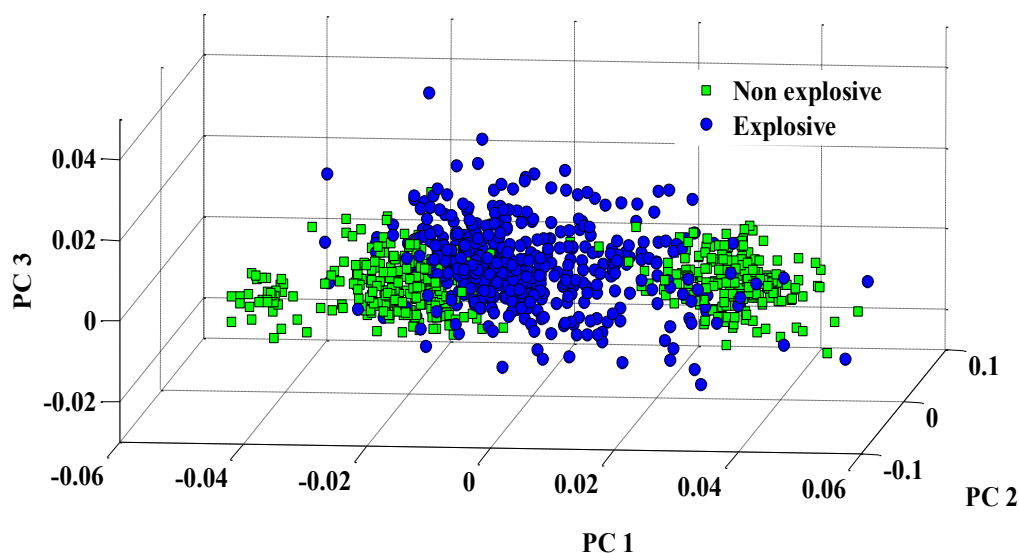


Figure 51. Score plot using oxygen peak of non explosives and explosives centred at 777.5 nm.

Oxygen line profile is also following the same trend as nitrogen peaks profile. Both classes scores are not well separated, but explosives are closely grouped. Non-explosives are grouped as two localized clusters with scattering points in between them. It is a well known fact that earth atmosphere contains large amount of oxygen and nitrogen. When experiments are performed in the ambient air, both of the elements contribution from the air removal is a challenging task. Though the experiments performed with double pulse LIBS, argon as buffer gas may be a good choice, however these methods demand additional experimental requirements, increasing the cost and complexity of the system. These attributes (selection of O and N features) are particularly not desirable for field application. On account of interference from ambient air. The first three PCs explain the variance of 50.88%, 15.90%, 6.65% respectively. The cumulative variance explained by the first five PCs is 84.12%.

PCA analysis of the individual elemental peaks shows that the clusters are formed for all the peaks. However, similar to the sub-spectrum results, there are sub-clusters within the

non-explosive clustersexcept for the hydrogen. Out of the five peaks considered, except for hydrogen all the other peaks either can have contributions from the ambient atmosphere or participate in the formation CN. In some cases, particularly on field applications, hydrogen signal may sometimes have influence from the ambient hydrogen contribution from H₂O. Cumulatively, hydrogen appears to be the most suitable candidate for the individual peak as spectral input for a better classification.

ANN model is applied on the specific peaks data. The model is iterated over 100 trials and the average results are listed in [table 15](#). The correct classification performance of C, CN, H, N and O are 87.78%, 92.57%, 96.98%, 94.59% and 93.62 % respectively. The reasons for carbon line exhibiting lower performance could be - (a) The number of points in the peak are very less compared to other peaks (b) The peak lies in ultraviolet region, which is at the edge of spectrometer spectral sensitivity where the quantum efficiency is moderate (c) A part of carbon atoms in the plasma may react with nitrogen of surrounding atmosphere or compound itself and forms CN molecule. The classification performance related to CN, N and O sounds to be appealing which is more than 92.5%. However, as discussed earlier, it is well known fact that these may be influenced by the surrounding atmosphere and may not be good choice for field applications. Hydrogen peak profile which is in the center of CCD spectral range and the quantum efficiency of the detector is also flat, the classification performance is found to higher compared with features and equal to full spectrum performance. To investigate further the average spectra of compounds were calculated and elemental profiles were selected to calculate euclidean distances.

Elemental profile	Correct Classification (%)	Mis-Classification (%)
C	87.79	12.22
CN	92.57	7.43
H	96.99	3.02
N	94.59	5.41
O	93.62	6.38

Table 15: ANN model average classification performance over 100 iterations using specific peaks C, CN, H, N and O

The Euclidean distance metrics of hydrogen peak dendrogram is shown in figure 52. It is evident from the dendrogram that the explosives and non-explosives are well separated. This further indicates towards the potential of hydrogen as a key feature for explosive discrimination. The dendrograms constructed from other peaks, such a clear separation was not observed. Even within HEMs, HMX is a nitroamine high explosive, which is chemically similar to RDX. The euclidean distance metrics correlates the same similarity index and connected both with '∩' shaped line between RDX and HMX where as PETN, NTO and TNT are separately grouped.

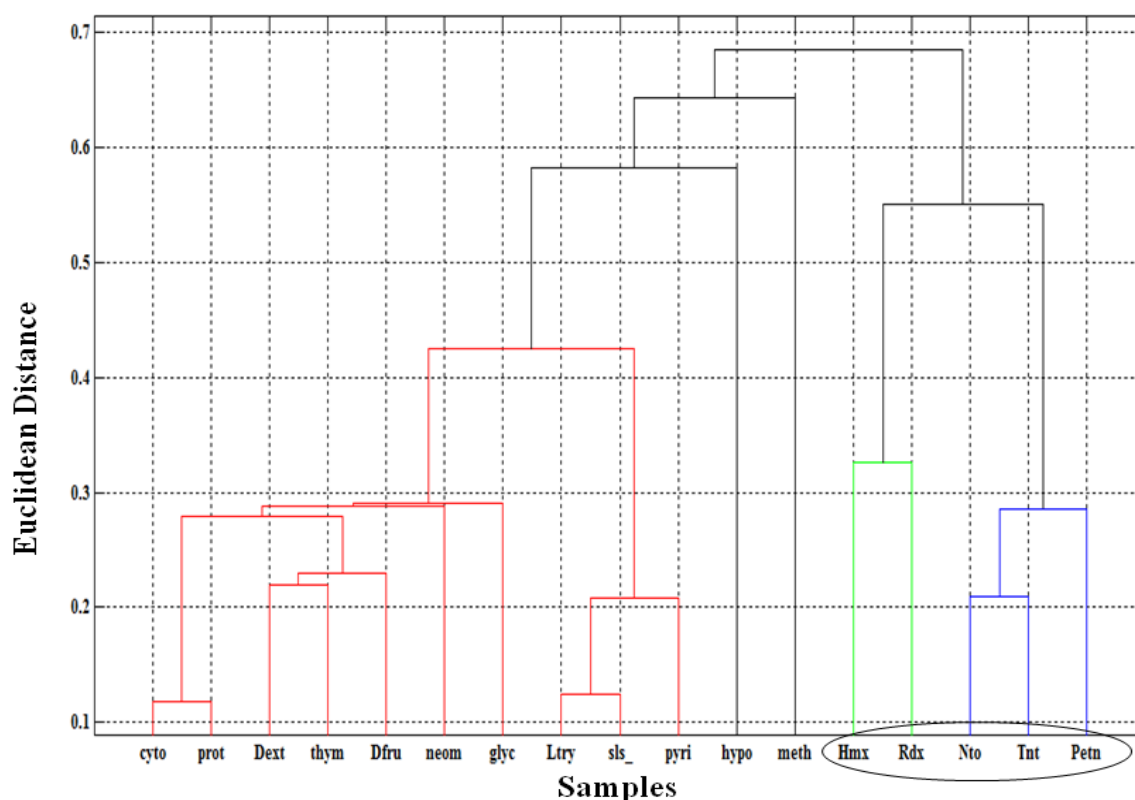


Figure 52: Dendrogram constructed between explosives and non explosives using euclidean distance as metrics.

3.4.2 SPECIFIC EXPLOSIVE IDENTIFICATION

LIBS emission spectra is used in the previous section for explosive/non-explosive HEMs identification, in this section the classification is extended towards identification of compounds within HEMS. The efficacy of data selection and strategies of statistical methodologies is evaluated. The comparison of PLSDA and ANN model is carried out. PCA has been applied on the full spectrum of explosives. The score plot of explosives is

shown in figure 53(a). Five distinct clusters are observed with a good separation with scattering point of HMX. The coefficients associated with the PCs are shown in loadings plot (figure 53(b)). The first PC contain maximum variance and the variables are similar to the full spectrum. the variance information is explained using scree plot as shown in figure 57(c). Cumulatively, the first four PCs explain 82.25 % of variance related to data.

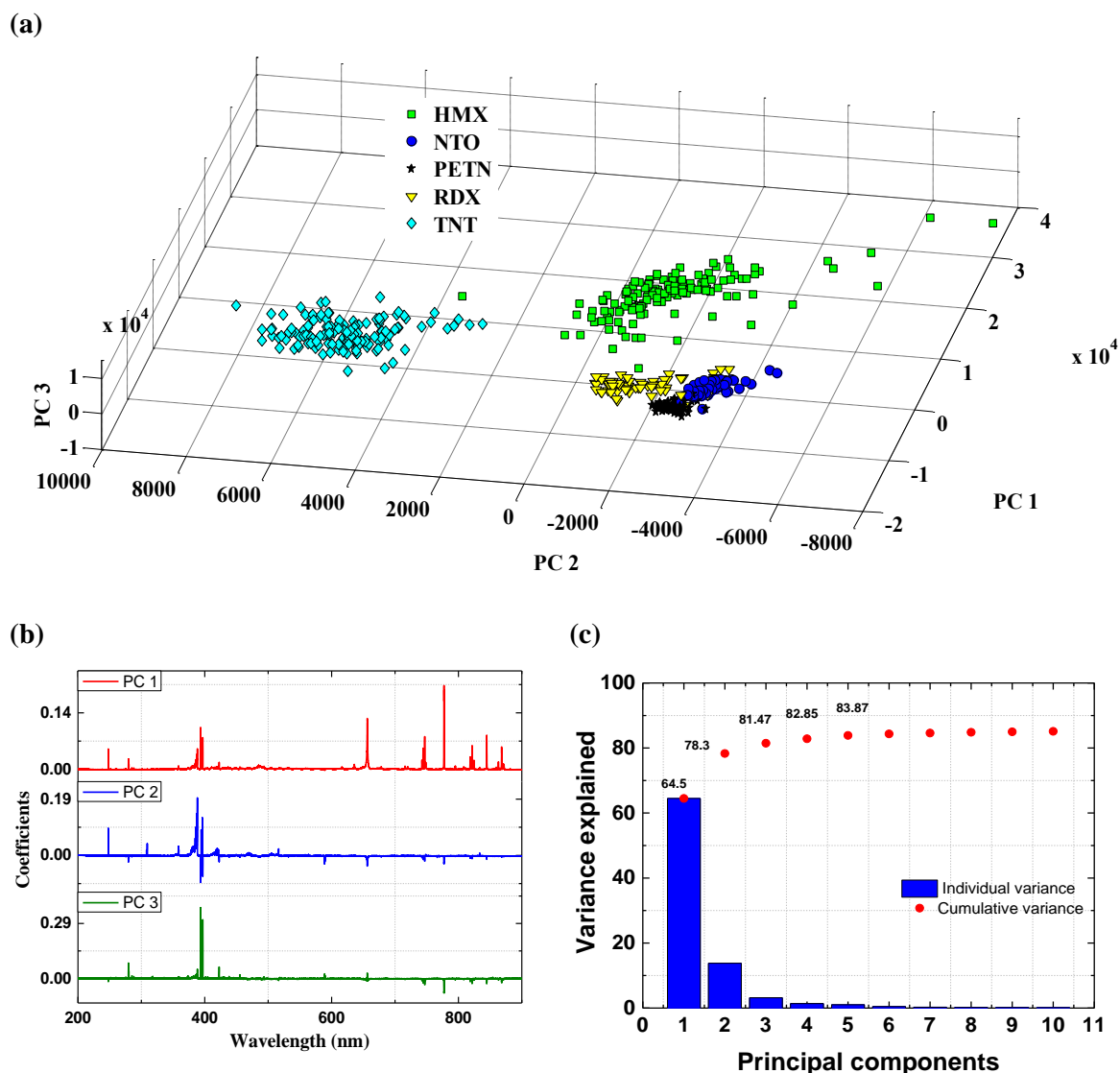


Figure 53(a) Score plot, (b)loadings plot, (c)Scree plot using full spectrum of explosives.

The associated spectral variable, which maximizes the discriminatory feature using full spectrum can be understood by the loadings. To understand even specific variables which contribute the maximum information for classifiers, it is necessary to revise the sub-spectra feature selection concept to this application. The score plot using four regions are

shown in figure 54. The scores corresponding to R1 have three different clusters, in which one cluster is a group of NTO, PETN and RDX. Sub-spectra R1 is a union peak containing carbon and hydrogen. The region R2 have the range between carbon to CN peak. The score plot is similar to that of R1, except dense cluster. R3 sub-spectrum contain additional peaks over R2

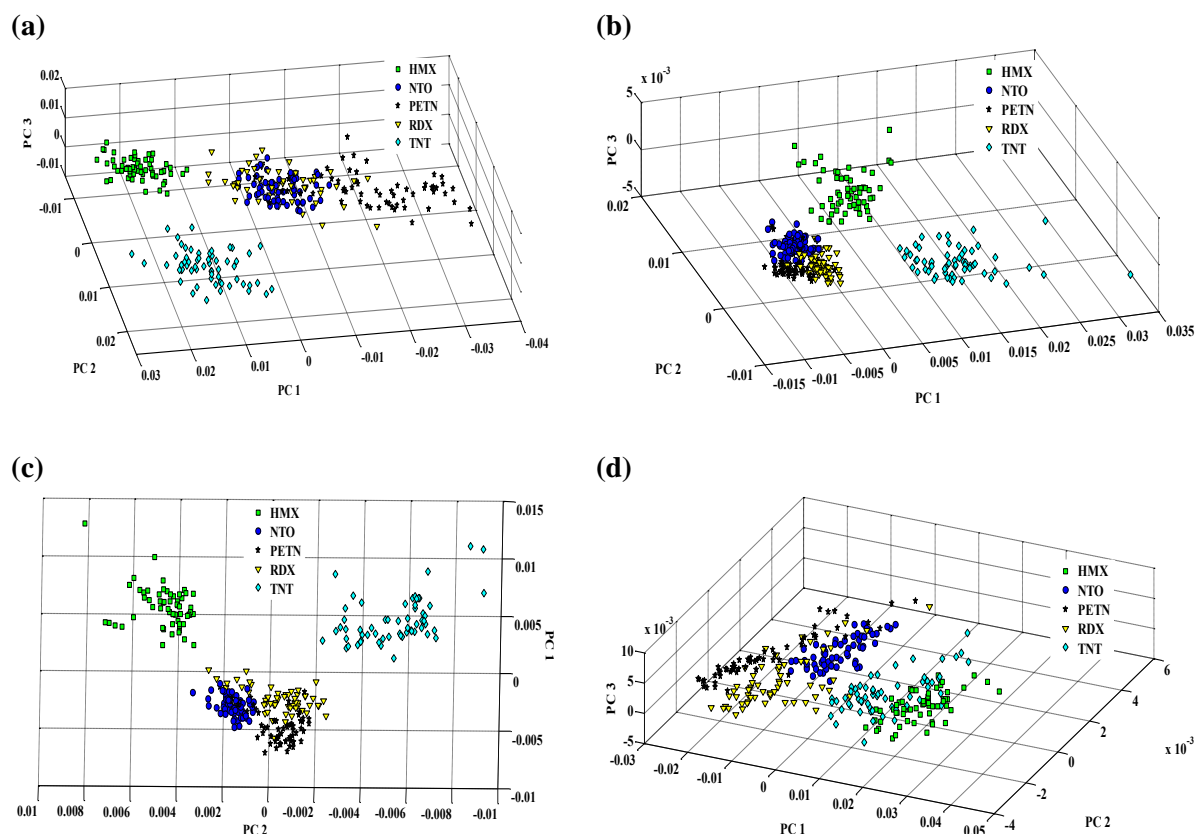


Figure 54. Score plots using (a) sub spectrum R1, (b) sub spectrum R2, (c) sub spectrum R3 and (d) sub spectrum R4.

and from the figure 54 (b), it is explicitly clear that five distinct clusters are formed. But in the case of region R4, which may be hampered by atmospheric contribution have also separate clusters. R4 sub-spectra score plot also point towards use as variables to differ compounds within explosives category. Cumulative variance versus principal components plots are shown in figure 55. Out of the all spectral regions, sub-spectrum R4 has the sources of maximum variance associated with the starting four PCs. Other regions R1, R2 and R3 have cumulative variance with combining five starting PCs is about 69.24%, 70.55% and 84.53% respectively. Out of the three sub spectra, R3 can be considered as meaningful where five compounds are separated clearly. Also add a general discussion on the clustering behaviour of all the regions together (adopt from earlier sections)

To understand further, the discriminatory features of these subspectra; supervised techniques PLSDA and ANN is applied. Using PLS-DA derived classifiers classification (write similar to what has been written in earlier sections) performance was determined for each of the spectral regions. Although it was originally developed as a multivariate regression method, PLS [40] has been extensively used to solve classification problems by encoding the class membership of the measured samples in the target

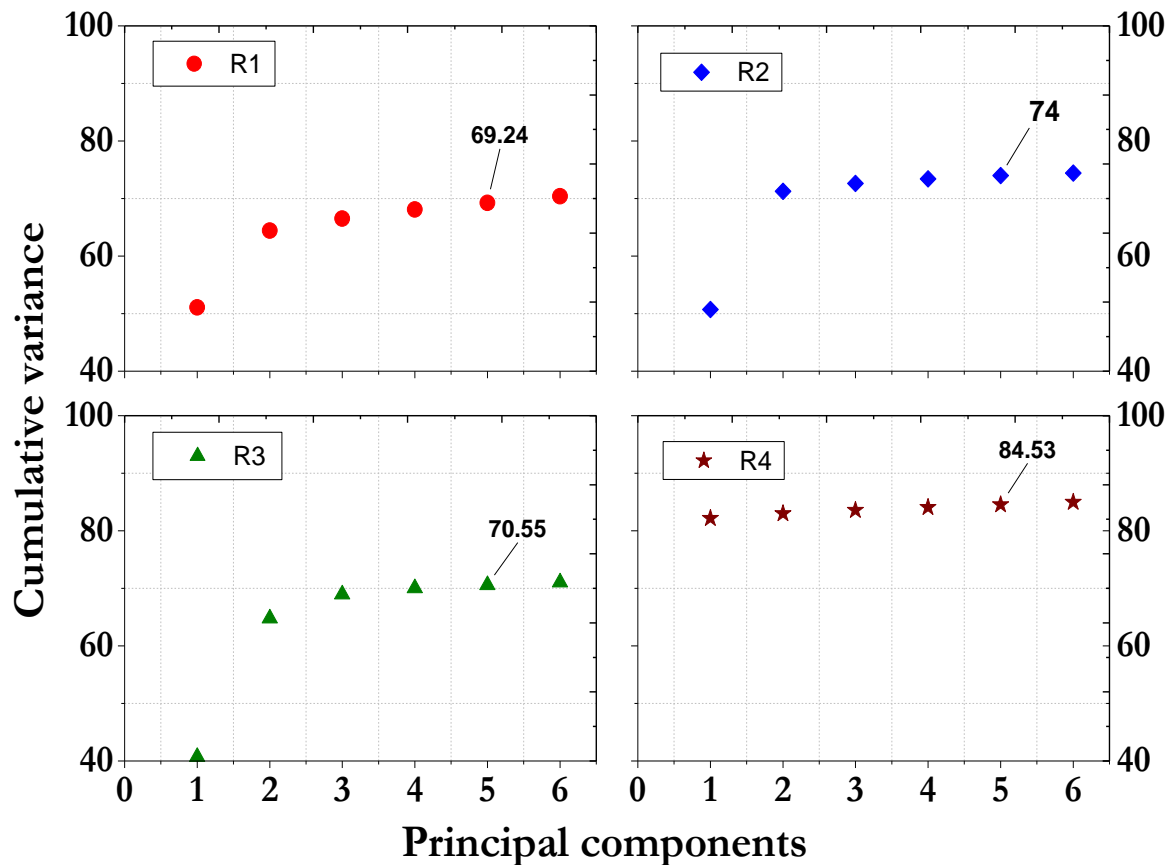


Figure 55. Cumulative variances of PCs of four spectral regions R1, R2, R3 and R4.

matrix[41,42]. An important feature of this supervised technique is that it is specifically suited to deal with problems in which the number of variables is large (compared to the number of observations) and collinear, two major challenges encountered when LIBS spectra are used. Furthermore, PLS-DA provides a powerful tool for discrimination even when the variability within a class is similar to the inter-class variability by virtue of establishing the maximal separation between each class via fitting one global model to the entire dataset. In PLS-DA, a PLS regression model is calculated that relates the independent variables (spectral data) to a binary response encoded as: [1,0,0,0,0] for sample belonging to class 1, [0,1,0,0,0] for sample belonging to class 2, and so on until

class 5. Typically, a threshold value based on the Bayesian method[43], is defined between 0 and 1, and calculated on the basis of the values predicted during the classification process, so an object is assigned to a particular class if its prediction is larger than the threshold value for this class. As in the PLS regression model, the optimal number of latent variables (LVs) retained are chosen before the modeling process and this is determined by using the fractional misclassification error rate in cross validation[44].

Table 16 enumerates the results of correct classification,

Region Selected	Correct Classification (%)	Misclassification (%)	Unclassification (%)
R1	69.65	28.32	2.02
R2	87.48	10.24	2.27
R3	90.42	6.78	2.79
R4	85.42	12.34	2.23
Full Spectrum	92.61	4.65	2.73

Table 16: Classification performance of PLS-DA models featuring selected spectral regions misclassification and unclassification rates. Here the maximum correct classification rate (92.61%) was obtained when the full spectral profile was considered as the input to the classifier. The misclassification of 4.65% was mainly due to the incorrect recognition of RDX samples that were misclassified as HMX. This misclassification may be attributed to the similar chemical formulation of HMX and RDX and, notably, to the significantly larger number of HMX samples in the dataset that likely ‘skew’ the decision algorithm. The unclassification rate was almost the same for all classes and, as noted from Table 16, for the four spectral regions and the full spectrum. On the other hand, the misclassification rates varied for the selected spectral regions from *ca.* 6.8% for R3 to 28.3% for R1. It is noteworthy that despite the presence of the prominent carbon and hydrogen peaks in R1, its diagnostic power is significantly lower due to the lack of significant variations in this region between the 5 HEM sample classes. Perhaps, more importantly, we observe that there is barely a 2% reduction in the classification performance between R3 and the full spectrum, even though R3 constitutes only ~63% of the full spectrum intensity values.

Similarly, ANN model is also used to the four spectral regions and full spectrum. The results are as similar to that of PLSDA as shown in table 17 below. The classification

performance of region R3 is comparable to full spectrum classification performance. Evidently, decreasing the size of sampled points not only reduces the redundant and uninformative regions but also raises the possibility of using a spectrometer that is focused only on the lower wavelength region for HEM classification with equal efficacy.

Region Selected	Correct Classification (%)	Misclassification (%)
R1	79.43	20.57
R2	87.78	12.22
R3	90.33	9.67
R4	83.17	16.83
Full Spectrum	91.6	8.4

Table 17. The classificatin performance of ANN model using full spectrm and four spectra regions.(In the analysis of ANN modelling, no unclassification criterion is established. hence the table doesn't contain the unclassification column.)

3.5 Conclusions

This chapter explored the application of LIBS for explosive classification. A set of explosives and non-explosives were considered for the study. Being organic in nature, all the samples showed similar spectra. The plasma is characterized by estimating the temperatures and electron densities. Within the error bars the plasma parameters showed no statistically significant variations. The problem of classification with different spectral variables as features inputs has been studied in detial. The issues of explosive/non-explosive and specific explosive classification have been studied.

Apart from full spectrum, feature selection was performed in two different ways. With sub-spectrum as input, the classification performance is almost similar (~ 96 %) , however R3 with only ~ 3% of the spectral points may be an ideal choice as it can reduce the requirement of data storage and enhance the speed of identification process, both of which are very crucial for an on field application PCA analysis was performed for exploratory analysis and neural network/PLSDA classifiers were selected to classification analysis. For explosive/non-explosive, the clustering was mainly aided by the emission peaks of inorganic entities. With sub-spectrum as input, all the four regions resulted in a

good binary clustering and sometimes the non-hem cluster showing sub clusters. However the two clusters are well separated. The results suggest that even in the absence of the inorganic elemental peaks, such as in R1, there is enough spectral information in other peaks that can lead to separation of HEMs and non-Hems. PCA analysis of the individual elemental peaks shows that the clusters are formed for all the peaks. However, similar to the sub-spectrum results, there are sub-clusters are all peaks except the hydrogen. Out of the five peaks considered, except for hydrogen all the other peaks either can have contributions from the ambient atmosphere or participate in the formation of CN. In some cases, particularly on field applications, hydrogen signal may sometimes have influence from the ambient hydrogen contribution from H₂O. Cumulatively, hydrogen appears to be the most promising candidate for the individual peak as spectral input for a better classification. As many of the detectors have a higher QE at this wavelength and this emission line is least perturbed by the ambient atmosphere.

Classification among the explosives has been explored on the similar lines. The clustering behavior showed well separated clusters with any of the inputs. The full spectrum model resulted in classification accuracies of 92.61 % and 91.60 % with PLSDA and ANN respectively. R3 showed highest classification with both the classifiers (90.42 % and 90.33 %). ANN shows an almost equal classification performance with the well reported classifier of PLSDA for explosives and can be suggested as alternative.

Though the results presented in this chapter used an echelle spectrograph combined with an ICCD, one can envision architecting a compact, inexpensive diagnostic platform that can operate with a compact handheld CCD spectrometer. This combined with optimization of the laser excitation wavelength can enable higher pulse energies and improved discrimination. Through all computational work on feature selection in the sparse LIBS dataset, results put together, a critical consideration in this regard is the possibility of acquiring lower resolution data and its influence on the quality of the classification model. This work greatly expands the feasibility of field applications while highlighting the untapped opportunities in exploiting computational approaches to make instrumentation innovations and advances. With minimal alterations, the approach described here can be equally exploited for other applications involving pharmaceutical samples, forensic specimen as well as in other recalcitrant process monitoring applications.

3.6 References

1. Agrawal JP, Hodgson R (2007) Organic Chemistry of Explosives: Wiley.
2. Agrawal JP (2010) High Energy Materials: Propellants, Explosives and Pyrotechnics: Wiley.
3. López-López M, García-Ruiz C (2014) Infrared and Raman spectroscopy techniques applied to identification of explosives. *TrAC Trends in Analytical Chemistry* 54: 36-44.
4. Schubert H, Rimski-Korsakov A (2006) Stand-off Detection of Suicide Bombers and Mobile Subjects: Springer.
5. Moore DS, Scharff RJ (2009) Portable Raman explosives detection. *Analytical and bioanalytical chemistry* 393: 1571-1578.
6. Carter JC, Angel SM, Lawrence-Snyder M, Scaffidi J, Whipple RE, et al. (2005) Standoff detection of high explosive materials at 50 meters in ambient light conditions using a small Raman instrument. *Applied Spectroscopy* 59: 769-775.
7. Wynn C, Palmacci S, Kunz R, Rothschild M (2010) Noncontact detection of homemade explosive constituents via photodissociation followed by laser-induced fluorescence. *Optics express* 18: 5399-5406.
8. de la Ossa MÁF, Amigo JM, García-Ruiz C (2014) Detection of residues from explosive manipulation by near infrared hyperspectral imaging: A promising forensic tool. *Forensic science international* 242: 228-235.
9. De Lucia FC, Harmon RS, McNesby KL, Winkel RJ, Miziolek AW (2003) Laser-induced breakdown spectroscopy analysis of energetic materials. *Applied optics* 42: 6148-6152.
10. Colao F, Fantoni R, Lazic V, Paolini A, Fabbri F, et al. (2004) Investigation of LIBS feasibility for in situ planetary exploration: an analysis on Martian rock analogues. *Planetary and Space Science* 52: 117-123.
11. Laxmiprasad A, Raja VS, Menon S, Goswami A, Rao M, et al. (2013) An in situ laser induced breakdown spectroscopy (LIBS) for Chandrayaan-2 rover: Ablation kinetics and emissivity estimations. *Advances in Space Research* 52: 332-341.
12. Gaudiuso R, Dell'Aglio M, De Pascale O, Loperfido S, Mangone A, et al. (2014) Laser-induced breakdown spectroscopy of archaeological findings with calibration-free inverse method: Comparison with classical laser-induced breakdown spectroscopy and conventional techniques. *Analytica chimica acta* 813: 15-24.

13. Myakalwar AK, Sreedhar S, Barman I, Dingari NC, Rao SV, et al. (2011) Laser-induced breakdown spectroscopy-based investigation and classification of pharmaceutical tablets using multivariate chemometric analysis. *Talanta* 87: 53-59.
14. DeLucia Jr FC, Samuels AC, Harmon RS, Walters R, McNesby KL, et al. (2005) Laser-induced breakdown spectroscopy (LIBS): a promising versatile chemical sensor technology for hazardous material detection. *Sensors Journal, IEEE* 5: 681-689.
15. Munson CA, De Lucia FC, Piehler T, McNesby KL, Miziolek AW (2005) Investigation of statistics strategies for improving the discriminating power of laser-induced breakdown spectroscopy for chemical and biological warfare agent simulants. *Spectrochimica Acta Part B: Atomic Spectroscopy* 60: 1217-1224.
16. Singh VK, Rai AK (2011) Prospects for laser-induced breakdown spectroscopy for biomedical applications: a review. *Lasers in medical science* 26: 673-687.
17. Rehse SJ, Mohaidat QI, Palchaudhuri S (2010) Towards the clinical application of laser-induced breakdown spectroscopy for rapid pathogen diagnosis: the effect of mixed cultures and sample dilution on bacterial identification. *Applied Optics* 49: C27-C35.
18. Kanawade R, Mahari F, Klämpfl F, Rohde M, Knipfer C, et al. (2015) Qualitative tissue differentiation by analysing the intensity ratios of atomic emission lines using laser induced breakdown spectroscopy (LIBS): prospects for a feedback mechanism for surgical laser systems. *Journal of biophotonics* 8: 153-161.
19. Unnikrishnan V, Choudhari K, Kulkarni SD, Nayak R, Kartha V, et al. (2013) Analytical predictive capabilities of laser induced breakdown spectroscopy (LIBS) with principal component analysis (PCA) for plastic classification. *RSC Advances* 3: 25872-25880.
20. De Lucia FC, Gottfried JL, Munson CA, Miziolek AW (2008) Multivariate analysis of standoff laser-induced breakdown spectroscopy spectra for classification of explosive-containing residues. *Applied optics* 47: G112-G121.
21. Sallé B, Lacour J-L, Vors E, Fichet P, Maurice S, et al. (2004) Laser-induced breakdown spectroscopy for Mars surface analysis: capabilities at stand-off distances and detection of chlorine and sulfur elements. *Spectrochimica Acta Part B: Atomic Spectroscopy* 59: 1413-1422.
22. Lopez-Moreno C, Palanco S, Laserna JJ, DeLucia Jr F, Miziolek AW, et al. (2006) Test of a stand-off laser-induced breakdown spectroscopy sensor for the detection of explosive residues on solid surfaces. *Journal of Analytical Atomic Spectrometry* 21: 55-60.
23. Gottfried JL, De Lucia FC, Munson CA, Miziolek AW (2007) Double-pulse standoff laser-induced breakdown spectroscopy for versatile hazardous materials detection. *Spectrochimica Acta Part B: Atomic Spectroscopy* 62: 1405-1411.

24. Gottfried JL, De Lucia Jr FC, Miziolek AW (2009) Discrimination of explosive residues on organic and inorganic substrates using laser-induced breakdown spectroscopy. *Journal of Analytical Atomic Spectrometry* 24: 288-296.
25. Gottfried JL, De Lucia Jr FC, Munson CA, Miziolek AW (2008) Strategies for residue explosives detection using laser-induced breakdown spectroscopy. *Journal of Analytical Atomic Spectrometry* 23: 205-216.
26. Gottfried J, De Lucia Jr F, Munson C, Miziolek A (2009) 'Laser-Induced Breakdown Spectroscopy for Explosive Residue Detection: A Review of the Challenges, Recent Advances, and Future Prospects. *Anal Bioanal Chem* 395: 283-300.
27. Sirven J-B, Salle B, Mauchien P, Lacour J-L, Maurice S, et al. (2007) Feasibility study of rock identification at the surface of Mars by remote laser-induced breakdown spectroscopy and three chemometric methods. *Journal of Analytical Atomic Spectrometry* 22: 1471-1480.
28. Yueh F-Y, Zheng H, Singh JP, Burgess S (2009) Preliminary evaluation of laser-induced breakdown spectroscopy for tissue classification. *Spectrochimica Acta Part B: Atomic Spectroscopy* 64: 1059-1067.
29. Brown C, Baudelet M, Bridge C, Fisher M, Sigman M, et al. Atmosphere issues in detection of explosives and organic residues; 2009. International Society for Optics and Photonics. pp. 73041D-73041D-73012.
30. Bohling C, Scheel D, Hohmann K, Schade W, Reuter M, et al. (2006) Fiber-optic laser sensor for mine detection and verification. *Applied optics* 45: 3817-3825.
31. Dingari NC, Barman I, Myakalwar AK, Tewari SP, Kumar Gundawar M (2012) Incorporation of support vector machines in the LIBS toolbox for sensitive and robust classification amidst unexpected sample and system variability. *Analytical chemistry* 84: 2686-2694.
32. Dingari NC, Barman I, Kang JW, Kong C-R, Dasari RR, et al. (2011) Wavelength selection-based nonlinear calibration for transcutaneous blood glucose sensing using Raman spectroscopy. *Journal of biomedical optics* 16: 087009-087009-087010.
33. Babushok VI, DeLucia Jr FC, Dagdigian PJ, Gottfried JL, Munson CA, et al. (2007) Kinetic modeling study of the laser-induced plasma plume of cyclotrimethylenetrinitramine (RDX). *Spectrochimica Acta Part B: Atomic Spectroscopy* 62: 1321-1328.
34. Sirven J-B, Bousquet B, Canioni L, Sarger L (2006) Laser-induced breakdown spectroscopy of composite samples: comparison of advanced chemometrics methods. *Analytical chemistry* 78: 1462-1469.
35. Myakalwar AK, Dingari NC, Dasari RR, Barman I, Gundawar MK (2014) Non-Gated Laser Induced Breakdown Spectroscopy Provides a Powerful Segmentation Tool on Concomitant Treatment of Characteristic and Continuum Emission.

36. Lucena P, Dona A, Tobaría L, Laserna J (2011) New challenges and insights in the detection and spectral identification of organic explosives by laser induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* 66: 12-20.
37. Sirven J-B, Bousquet B, Canioni L, Sarger L, Tellier S, et al. (2006) Qualitative and quantitative investigation of chromium-polluted soils by laser-induced breakdown spectroscopy combined with neural networks analysis. *Analytical and bioanalytical chemistry* 385: 256-262.
38. Tarazona JL, Guerrero J, Cabanzo R, Mejía-Ospino E (2012) Construction of a predictive model for concentration of nickel and vanadium in vacuum residues of crude oils using artificial neural networks and LIBS. *Applied optics* 51: B108-B114.
39. Nunes LC, da Silva GA, Trevizan LC, Júnior DS, Poppi RJ, et al. (2009) Simultaneous optimization by neuro-genetic approach for analysis of plant materials by laser induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* 64: 565-572.
40. Geladi P, Kowalski BR (1986) Partial least-squares regression: a tutorial. *Analytica chimica acta* 185: 1-17.
41. Barker M, Rayens W (2003) Partial least squares for discrimination. *Journal of chemometrics* 17: 166-173.
42. Ståhle L, Wold S (1987) Partial least squares analysis with cross-validation for the two-class problem: A Monte Carlo study. *Journal of Chemometrics* 1: 185-196.
43. Duda RO, Hart PE, Stork DG (2012) *Pattern Classification*: Wiley.
44. Kryger L (1981) Interpretation of analytical chemical information by pattern recognition methods—a survey. *Talanta* 28: 871-887.

This page intentionally left blank

4. NONGATED LIBS TOWARDS FIELD APPLICATIONS

In this chapter application of non-gated LIBS for classification of organic materials viz., pharmaceutical tablets and High energy materials (HEMs) has been demonstrated. While use of such a system introduces substantive continuum background in the spectrum, appropriate treatment of the continuum and characteristic emission results in accurate discrimination of organic compounds of similar stoichiometry. Specifically, results suggest that near-perfect classification can be obtained by employing suitable multivariate analysis methods on the acquired spectra, without prior removal of the continuum background. Indeed, this chapter even predicts that pre-processing in the form of background removal can reduce the average correct classification performance from 92% to 88%. Explosives and diverse classes of nonexplosives are classified with high degree accuracies. This has an implication towards application of a portable, non-gated LIBS system as a process analytical tool, for online, insitu and standoff monitoring.

4.1 Introduction

Gating the collection time window is primarily choice for LIBS experiments. First, it controls the time delay between laser pulses to signal collection to avoid continuum emission in the early phase of plasma evolution. Secondly it, controls integration time to enhance signal collection with good signal to noise ratio. Gated detectors are beneficial for the diagnostics of plasma, calibration free LIBS, calibration LIBS and determining elemental lifetimes in plasma. Particularly this arrangement is mostly useful for those applications where atomic/molecular peak profiles are critical. Background estimation and subtraction are important because of the strong interferences between isolated lines and continuum. However, continuum is generally treated as not important though it is emitted because of free-free transitions and free-bound transitions. It can create hurdles when analysis being carried out with line emissions. As most of the researchers' general interest have been towards atomic and molecular emission studies. It is better to review the methods, which are used for continuum removal. Continuum removal/reduction can

be made with the choice of lasers, detectors, externally controlled electronics and with data analysis by adopting suitable background correction. Diode pump solid state (DPSS) lasers create a smaller lifetime of continuum as compared with continuum generated by flash lamp pumped solid state laser (FLPSS). CCD based detector conjunction with DPSS laser may serve as better solution to deal with continuum [1]; Even the femto second lasers also relatively produce less continuum as compared to nanosecond lasers[2,3]. Now a days intensified charge coupled device (ICCD) coupled to echelle type grating is a most common detector and it offers several advantages: (1) low light imaging (2) fast gating in the order of nanoseconds (3) high signal to noise ratio.

However, it has some drawbacks: (1) the intensifier involves a pulsed high-voltage supply, (2) the spectral sensitivity of ICCD system can be restricted by the intensifier, which may be problematic for simultaneous measurement of many elements, (3) cooling is required (4) acquisition time is more (5) ICCD with spectrometer delivers spectral data is large and finally,(6) ICCD is more expensive, bulky with electronics and need skilled manpower to operation. Keeping view of the above disadvantages, researchers at Kyoto University, Japan proposed utilization of acousto-optic modulator as a gating element inserted between collection optics and a normal spectrometer [4]. The similar arrangement later shown by pawel et al. with high repetition rate lasers, making detector portable and cost reduced [5]. Wang and his coworkers proposed utilization of a chopper wheel as a fast shutter placed in infront of collection optics [6]. Some companies also manufacturing CCD detectors with gated electronics. But they also need peltier cooling [6]. Generally, fiber-coupled spectrometers equipped with CCD detectors can also be used for time-resolved spectroscopy by applying an external trigger and with a short exposure time.

However, all the above solutions are pertaining to removal/avoiding of continuum. Application of gated detection is based on the prevailing view that discrimination against the early “uncharacteristic” continuum signal (from radiative recombination and Bremsstrahlung emission) is critical for quantitative analysis [7]. As a consequence, most of the reports have focused on a suitable time window of acquisition where the condition of local thermodynamic equilibrium (LTE) is satisfied [8]. While this perspective is largely justified for trace element analysis as well as suppression of matrix effects in certain specimen[9], such gating may not be necessary for classification applications even

when dealing with samples of similar chemical composition. It is possible to obtain similar levels of performance without gated detection by appropriately utilizing the features across the entire spectral window of collection. In fact, the distinctiveness of the plasma produced by different samples may result in subtle differences in the broad continuum signals, which in turn could positively aid performance of the multivariate analysis models. Classical ratiometric analysis based on a study of a few channel (wavelength) traces has limited capability of dealing with segmenting such spectral datasets due to the overlap of the continuum and characteristic emission signals and to the impossibility of detecting interfering species in the measured signal. Multivariate classification of non-gated LIBS spectra can be successfully implemented despite the presence of the continuum background signals as it was observed earlier in the case of Raman spectra [10].

According to the U.S. Food and Drug Administration, the main objectives of process analytical technologies, is the development of novel sensors that can be incorporated in the manufacturing process loop to enable in-process material characterization. Such a method can assist in better monitoring each step of the formulation development and manufacturing process and, therefore, in real-time control of the process itself. However, there is a lack of analytical tools that can perform rapid on-line determination of the consistency of the drug constituents (especially the active pharmaceutical ingredient, API) in order to ensure the potency, purity and bioavailability of the final product. Laser-induced breakdown spectroscopy (LIBS) is an emerging instrument in the analytical toolkit, as it can provide real-time analysis with minimal or no sample preparation. Because of its real-time diagnostic capability, LIBS can be potentially used for testing a larger number of samples in comparison to existing analytical tools (*e.g.* high performance liquid chromatography (HPLC)), with the additional possibility of high-resolution surface mapping and depth profiling. However, despite these intrinsic advantages, LIBS systems have hitherto not been employed for online process monitoring [11,12]. The primary bottlenecks towards more extensive usage, particularly for screening applications, is the lack of robustness and resource-intensive, unwieldy nature of the conventional LIBS spectrometers that use gated intensified detectors (primarily intensified charge-transfer devices (ICCD)) for spectral recording. This is further compounded by the large spatial footprint and weight of these systems as well as the

considerable maintenance and technical expertise required for its routine use. In the history of LIBS, the journey of portable LIBS instrument started with Cremers and his coworkers, who made backpack LIBS system to monitor pb in paints [13]. After 9/11 attacks, global security is the major concern of all nations. The pioneer work in the detection of explosive on field was led by US army lab and researchers at university of Malaga, Spain. With the association of Ocean Optics, US army research lab made backpack LIBS system to monitor hazardous materials and mines detection [14]. Most of the portable instruments are CCD based.

Unfortunately, most of the spectrometers have minimum exposure time in the order of a milli-second; preventing time-resolved measurements on shorter time-scales. The CCD-based spectrometer used for experiments in this thesis, has exposure time ~ 6 ms. Hence the continuum is prevailed in measurements and efforts has been putforth to use these features as an important variables in the anaysis and differentiation of pharmaceutical samples and high energy materials has been demonstrated, by concomitant treatment of continuum and characteristic emission. This work extends the recent efforts in understanding the role of continuum radiation in LIBS [15] and its application in metal alloy identification [16] to analysis of organics (*e.g.* API) of similar composition. In this chapter, the classification with the aid of multivariate analysis on pharmaceutical formulations and explosives in solid dosage forms, without employing gated detectors or echelle spectrographs have been reported. These non-gated LIBS measurements reveal that despite the presence of substantive continuum emission, the acquired signals exhibit subtle, but consistent, changes in spectral features. By correlating the spectra with the corresponding class of samples, we have developed predictive models based on soft independent modeling of class analogy (SIMCA), artificial neural networks (ANN) and partial least squares discriminant analysis (PLSDA). Multivariate analysis shows very high diagnostic power with correct classification accuracy levels in excess of 90% with SIMCA and 100% with ANN and PLSDA, with measurements performed under ambient air conditions. Taken together, these results provide a powerful toolkit in the minimally perturbative process monitoring (on-line and at-line) and quality control domains in the pharmaceutical industry; and as well as in on field explosive detection.

4.2 Materials and methods

Two classes of samples, pharmaceutical and High energy materials are chosen for this study. Pharmaceutical samples viz., Cetirizine dihydrochloride, Metformin hydrochloride, Cipro pure and Ciprofloxacin hydrochloride were acquired from the local drug manufacturer Yegna Manojavam Drugs and Chemicals Ltd., Nalgonda, India in powder form as shown in table 18, while high energy materials (HEM) NTO, PETN, TATB and TNT were acquired from the High Energy Materials Research Laboratory (HEMRL), Pune, India as shown in table 19. The fine powder material was pressed into ca. 1 cm diameter pellets by a die-hydraulic press combination. The details of lasers and experimental arrangement were as discussed in chapter 1.

S. No	Name of the sample	Formula	Spectra used in the analysis
1	Cetirizine Di hydrochloride IP	$C_{21}H_{27}O_3N_2Cl_3$	110
2	Cipro pure	$C_{17}H_{19}ClF-N_3O_3.H_2O$	110
3	Metformin hydrochloride IP	$C_4H_{12}ClN_5$	110
4	Ciprofloxacin hydrochloride USP	$C_{17}H_{18}FN_3O_3.HCl$	110

Table 18. Details of the pharmaceutical samples: chemical formula and spectra used.

S. No	Name of the sample	Formula	Spectra used in the analysis
1	Nitrotriazalone (NTO)	$C_2H_2N_4O_3$	80
2	Pentaerythritol tetranitrate (PETN)	$C_5H_8N_4O_{12}$	80
3	Triaminotrinitrobenzene (TATB)	$C_6(NO_2)_3(NH_2)_3$	80
4	Trinitrotoluene (TNT)	$C_7H_5N_3O_6$	80

Table 19. Details of the HEM samples: chemical formula and spectra used.

In short, the LIBS system used a frequency-doubled Nd:YAG laser ($\lambda_{ex} = 532$ nm, 7 ns pulse width, Spit light 1200, InnoLas Laser GmbH, Germany) for excitation and the emission signal was recorded using a non-gated spectrometer (Ocean Optics, MAYA 2000). Notably, the detector unit comprised of a conventional Czerny-Turner spectrograph

for dispersion instead of the higher resolution echelle spectrograph allowing us to assess a lower bound for the classification performance. The pharmaceutical samples were subjected to ca. 6 mJ of laser energy, while HEMs subjected to ca. 110 mJ. All of the LIBS spectra were collected under ambient conditions. A motorized XYZ stage was employed to enable a fresh portion of the sample to be interrogated after acquiring multiple spectra from a single location. Here it should be observed that pharmaceutical nongated LIBS was performed two years earlier than HEMs nongated LIBS. The experiments pertaining to pharmaceuticals have been performed manually because laboratory settings were at early stage of development. After gaining of good knowledge in data analysis and instrumentation, automation and interface of instruments have been performed which was used for the recording the data HEMs. The following section describes the potentiality and need of LABVIEW based instrument interface for online monitoring process.

4.3 Instrument interface by LABVIEW

The automation of experiments allows the fast and easy collection and analysis of data, facilitating easiness of measurements. In this section details of LABVIEW virtual instruments (VI) are presented.

4.3.1 SPECTROMETER

Prior to automation of spectrometer, in the early stage of development of laboratory, it was very hard to acquire even single LIBS spectrum per minute and Nd YAG Laser repetition raterange is 1-10 Hz. Acquiring the spectrum with save option provided in the software provided by the manufacturer is lengthier, tedious and time consuming. As discussed in the introduction section of this chapter, non gated spectrometers at our laboratory have a limitation of 6 msintegration time. To capture every pulse response with 10Hz

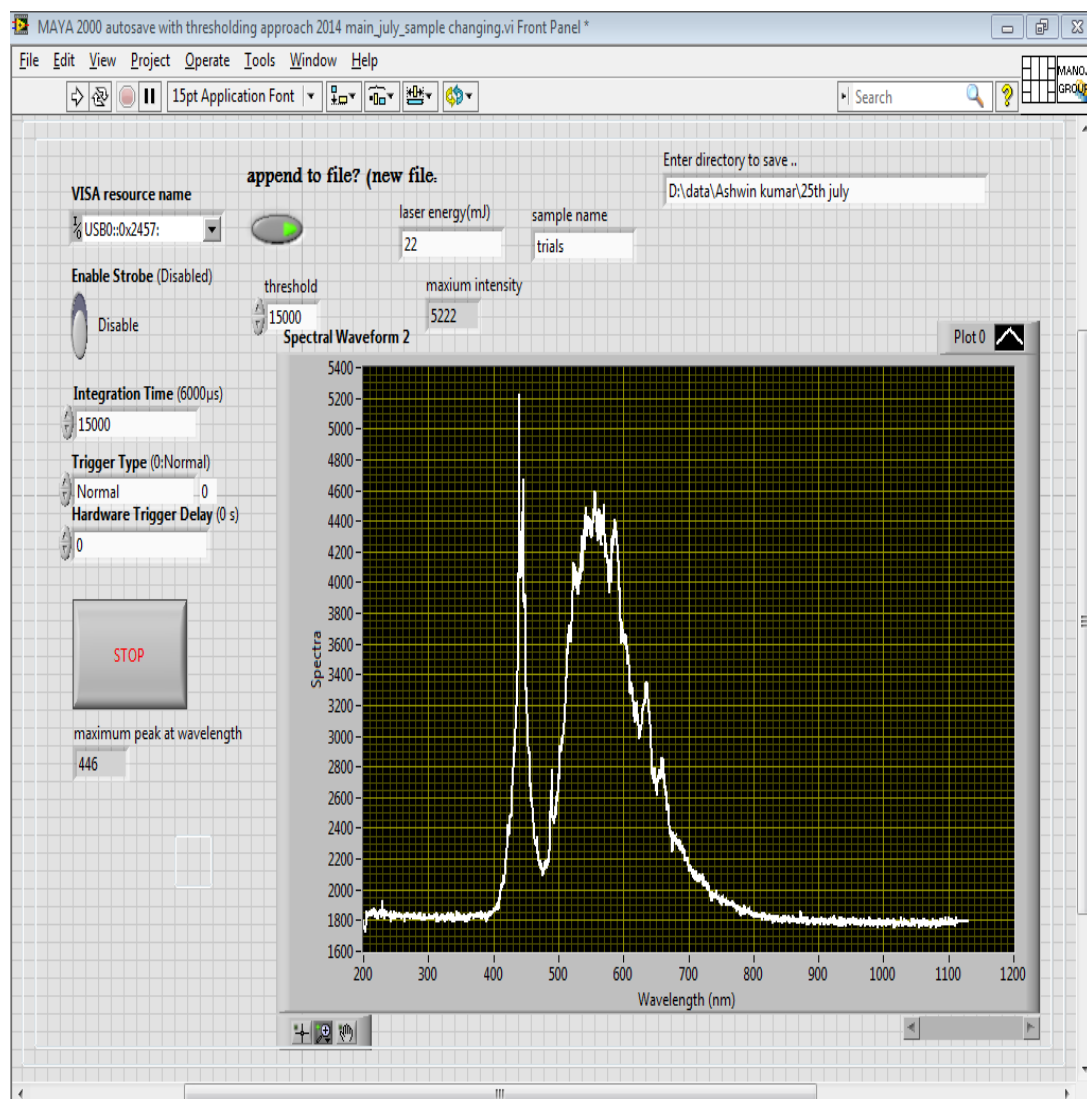


Figure 56. Screenshot of the labview VI developed for non gated spectrometer(MAYA 2000)

repetition rate manually is therefore an impossible task. To eradicate the time and sample consumption, VI was constructed for two nongated CCD based spectrometers viz., MAYA 2000. With the help of this VI every single pulse response signal can be recorded and automatically stored. The screen shot of Labview controlled spectrometer has been shown in Figure 56. The Figure 56 contains a typical LED light spectrum shown in front panel of LABVIEW. The instrument interfaced through USB connection with the computer and each parameter can be changed through this VI.

4.3.2 POWER METER:

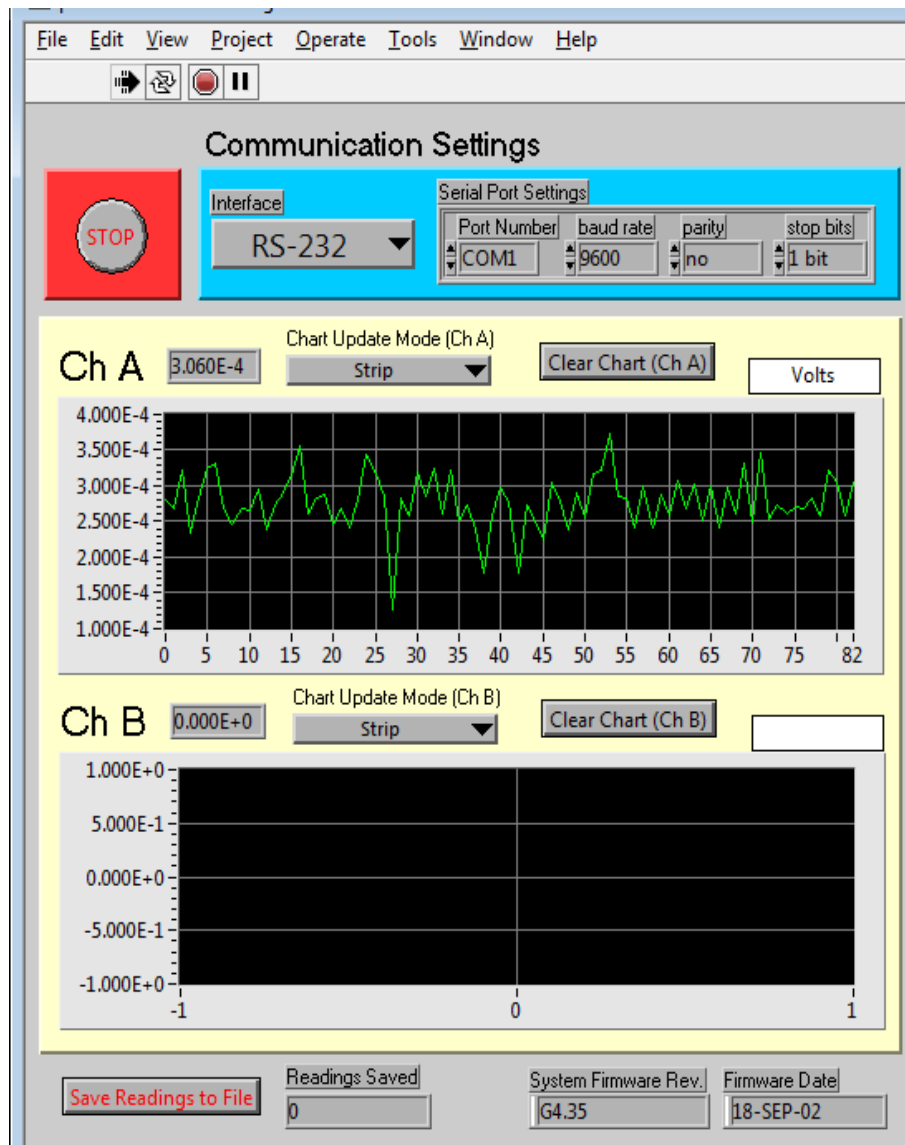


Figure 57. Screenshot of the labview VI developed for measurement of laser power using power meter

The discriminatory powers of classifiers may alter with stochastic nature of plasma signal emissions, which in turn produces peculiarities in the data of LIBS. To account the fluctuations in the plasma emission, it is necessary to monitor laser pulse energy variations by measuring energy with energy/power meter. Statistics of laser pulse energies (mean, standard deviation and relative standard deviation) are determined with LABVIEW by monitoring some part of laser radiation scattered from mirror. Figure 57

show screen shots of an independent control of epm2000, coherent, Molectron Inc, powermeter

4.3.3 STAGE CONTROLLER:

Moving a stage using controller with the frequency of laser repetition and obtaining each LIBS signal arising from fresh portion of the sample is an important step in the experiment. Figure 59 shows the screen shot of a typical VI used for stage controller taking consideration of laser repetition rate and sample dimensions. This VI used for the simultaneous measurement of laser power, acquisition of spectral data and movement of stage.

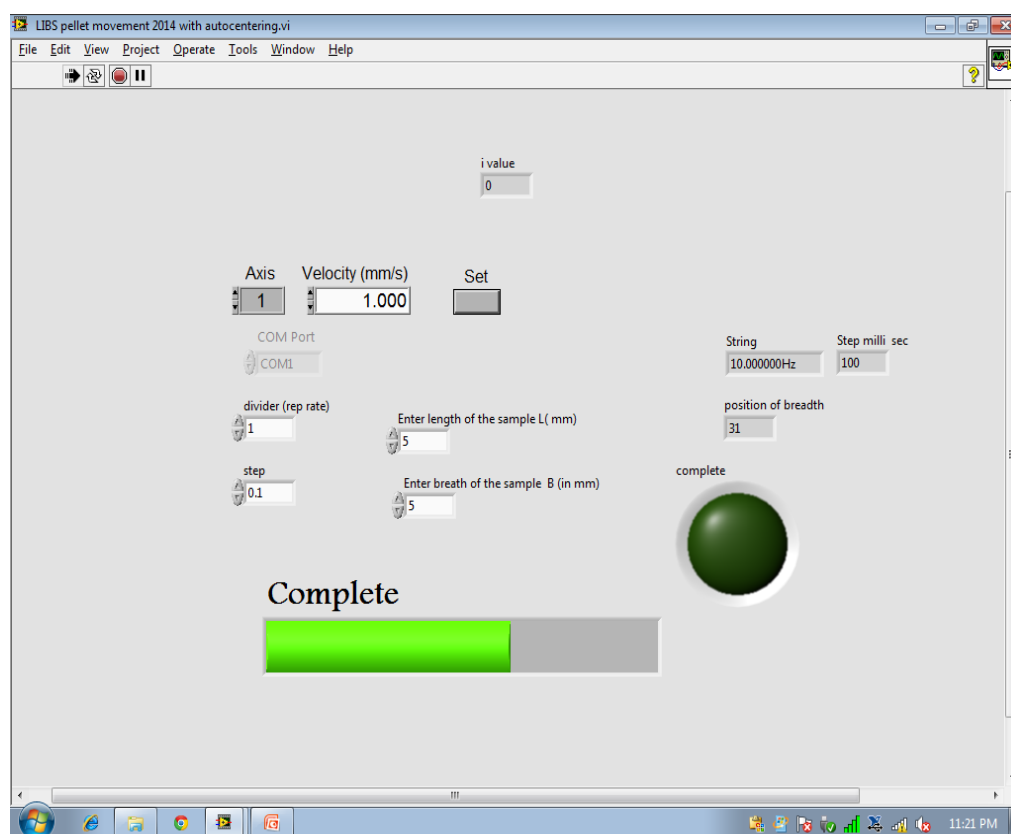


Figure 58. Screenshot of the labview VI developed for stage controller. Pulse operating frequency and the size of the sample also considered for movement of the axis.

Subsequent to acquisition and pre-processing of the spectral data, principal component analysis (PCA) was employed as a data exploration and dimensionality reduction step. For quantifying the classification ability of the LIBS dataset, we selected SIMCA, PLS-DA and ANN as representative methods. Multivariate analysis calculations, including dendrogram analysis-based outlier detection and subsequent classification, were

conducted using Matlab R2014 (Math Works, India). Following removal of outliers, a total of 403 LIBS spectra were used in the classification analysis, with more than 90 spectra from each of the four pharmaceutical formulations. In the case of explosive data, using automated setup total 2857 spectra were collected. However, using correlation analysis each of 80 spectra were selected for further analysis. The selection of equal numbers will help in distribution of data in training, test and validation of classifiers. Total 320 spectra were used for classification within the explosives and also used for the classification of specific explosive.

4.4 Spectral description

Figure 60 shows representative spectra acquired from the Cetirizine dihydrochloride,

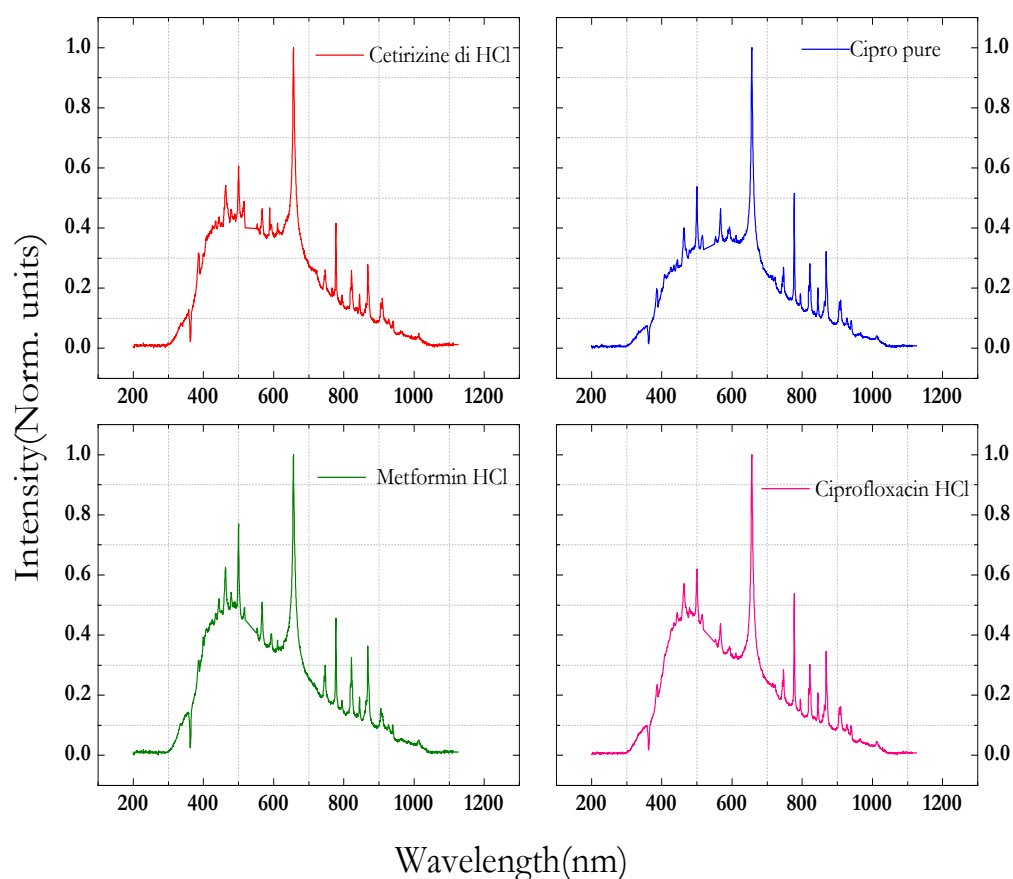


Figure 59. Representative LIBS spectra acquired from the pharmaceutical formulation (a) Cetirizine dihydrochloride; (b) Cipro pure; (c) Metformin hydrochloride; (d) Ciprofloxacin hydrochloride. Intensity on the y-axis is normalized with respect to the characteristic hydrogen emission peak at 656 nm.

Cipro pure, Metformin hydrochloride, and Ciprofloxacin hydrochloride pellets. Similarly, Figure 61 shows the representative spectra acquired from the NTO, RDX, TATB and TNT explosive compounds.

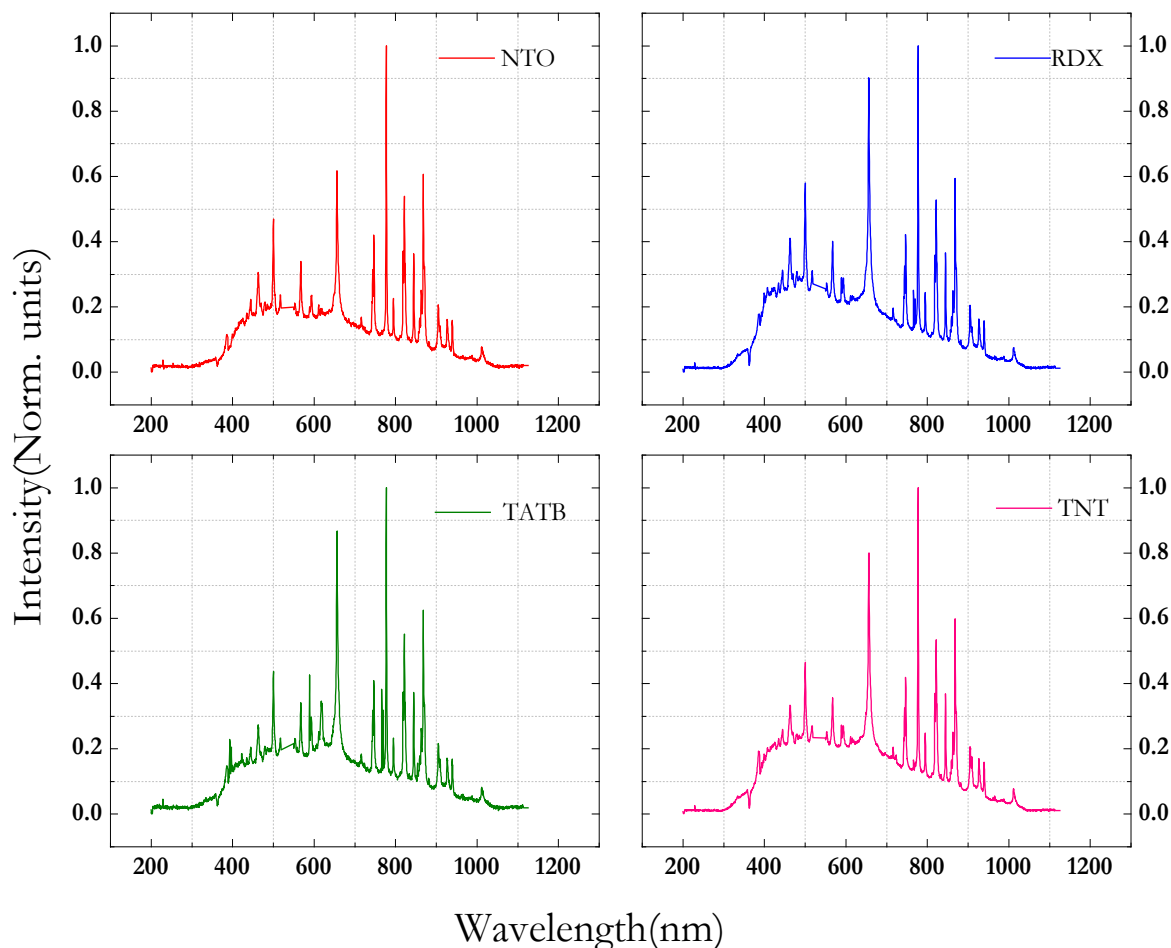


Figure 60. Representative LIBS spectra acquired from high energy material pellets (a) NTO; (b) RDX; (c) TATB; (d) TNT. Intensity on the y-axis is normalized with respect to the characteristic oxygen emission peak at 777 nm.

The discontinuity in the wavelength axis is due to the omission of a 30 nm wide band surrounding the 532 nm laser line, which otherwise causes significant interference in the non-gated signal of both the class of samples. The corresponding intensity values are disregarded from the ensuing analysis. From Figure 5 and Figure 6, it is notable how similar the LIBS spectra from each of the pharmaceutical and HEM samples within the classes. Further, any differences that may exist (for example, the absence of oxygen in metformin HCl) are obviated by the presence of the corresponding element(s) in air. A complete listing of the prominent peaks in both the above figures are listed in Table 20

Wavelength(nm)	Element	pharmaceutical	HEMs
384-389	CN	√	√
479.5	Cl II	√	X
500.5	N II	√	√
589.0	Na I	√	√
589.6	Na I	√	√
657.0	H I	√	√
742.3	N I	√	√
744.6	N I	√	√
747.2	N I	√	√
766.0	K I	X	√
777.7	O I	√	√
833.8	Cl I	√	X
838.2	Cl I	√	X
845.2	O I	√	√
868.7	O I	√	√
940.3	C I	X	√

Table 20. Elemental assignments of the major emission lines observed in the LIBS spectra acquired from the pharmaceutical and HEM samples used in this study

Evidently, the early continuum emission (present as broad featureless background) constitutes a substantive component of the acquired spectra as shown in [figure 61](#) for multiple trails. Despite the apparent featureless nature, however, it is possible that the subtle differences in the continuum emission signals may provide discriminatory power – particularly when viewed in light of the similarity of the characteristic emission lines in [figure 59](#) and [figure 60](#).

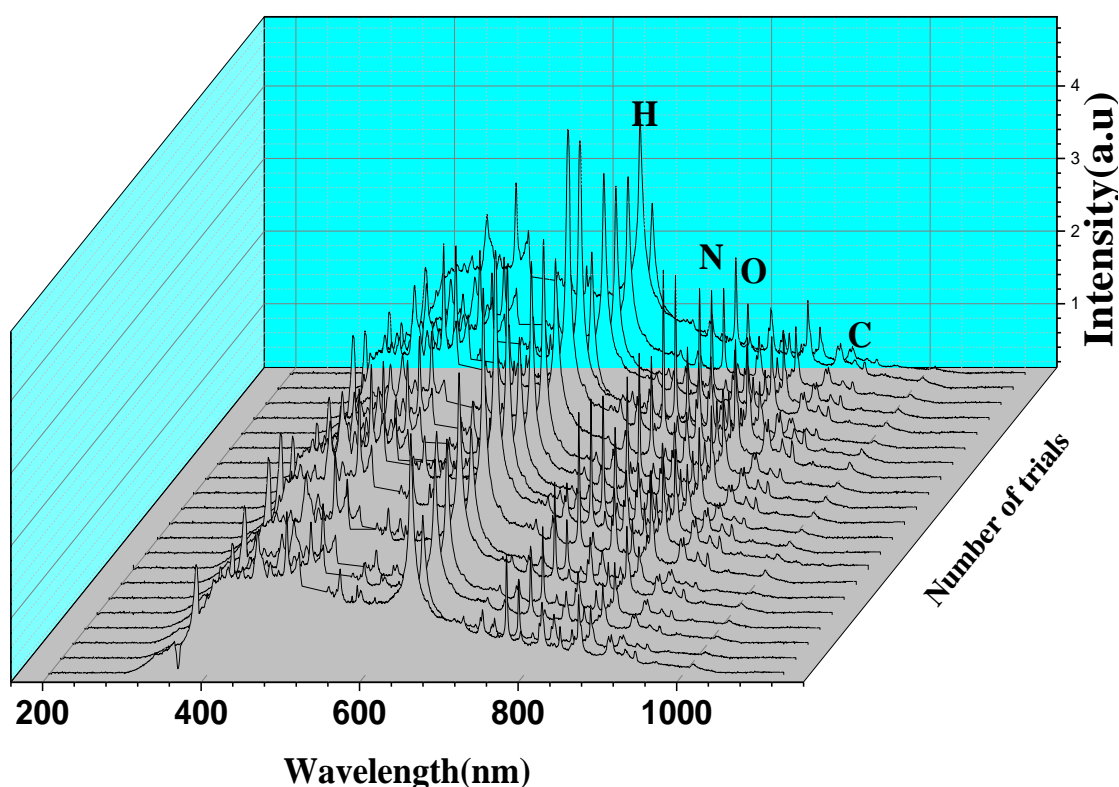


Figure 61. Nongated LIBS signal dominated with continuum . The atomic lines related to C,H,N and O are sitting on broad emission of continuum

4.5 Results

4.5.1 COMPOUND SPECIFIC CLASSIFICATION

To systematically investigate the presence of subtle differences between the LIBS spectra of each type of pharmaceutical sample, principal component analysis (PCA) was used. The principal components are linear combinations of the acquired signals and capture the spectral variance in a reduced dimensional space. [Figure 62](#) shows the first three principal components, which together account for 96.5% of the net variance in the spectral dataset. In particular,

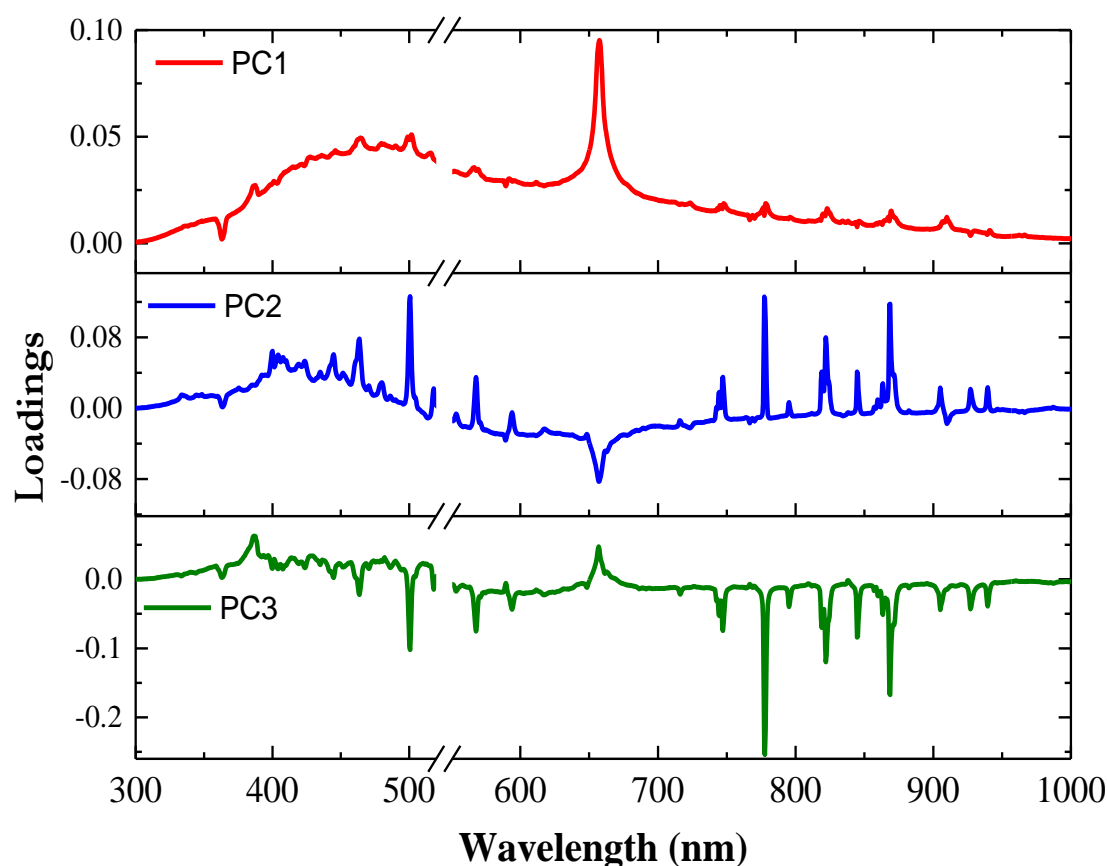


Figure 62. Plot of the first three principal components corresponding to the entire spectral dataset acquired for all pharmaceutical samples. These three principal components, combined, explain 96.5% of the net variance in the dataset.

the first two PCs explain 93% of the variance present in the dataset, with the first one contributing 84.3%. Expectedly, the subsequent PCs after the first three are mostly dominated by noise. PC1 shows only a single dominant hydrogen emission line at *ca.* 657 nm on a broad background that can be ascribed to the early continuum emission. PC2 and PC3 exhibit mostly characteristic emission lines including those at *ca.* 478 nm (chlorine), 500 nm (nitrogen), 567 nm (nitrogen), 747 nm (nitrogen), 777 nm (oxygen) and 868 nm (oxygen). The primary differences between PC2 and PC3 are in the relative intensities of these lines and also in the presence of a small background for PC2, especially in the lower wavelength region. The corresponding scores plot for PC 1, 2 and 3 is provided in [Figure 63](#). From the figure, it is evident that the samples of each class tend to form a cluster and in general appear

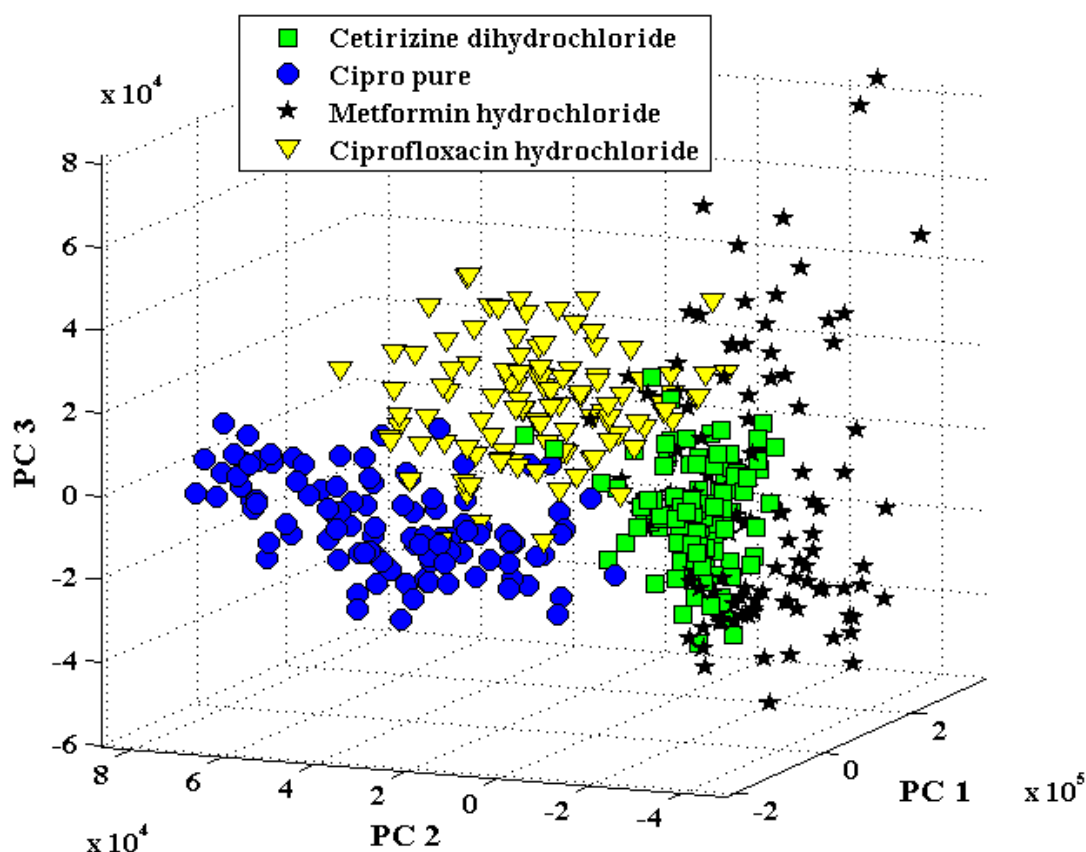


Figure 63. Discrimination of pharmaceutical samples based on their LIBS spectra: Scores plot corresponding to principal components 1, 2 and 3 for the spectral dataset acquired from the four samples. The data points corresponding to Cetirizine dihydrochloride, Cipro pure, Metformin hydrochloride and Ciprofloxacin hydrochloride are indicated by green squares, blue circles, black asterisks and yellow inverted triangles, respectively.

to show good separation from the other classes. In particular, Cetirizine dihydrochloride, Cipro pure and Ciprofloxacin hydrochloride are clearly distinguishable based on the PC scores obtained from the acquired LIBS spectra. However, the scatter in the PC scores for the Metformin hydrochloride spectra makes its separation more challenging, particularly from the Cetirizine dihydrochloride cluster. The scatter in the PC scores arises primarily from the inhomogeneity of the fabricated pellets as well as the potential non-sample specific variance introduced by non-gated detection. Additionally, the overlap in the Metformin hydrochloride and the Cetirizine dihydrochloride clusters can also be attributed to the absence of any distinguishing element between the two formulations, *i.e.* both formulations in ambient air conditions exhibit emission lines of carbon, hydrogen, oxygen, nitrogen and chlorine. This is in contrast to the Cipro pure and Ciprofloxacin hydrochloride samples that contain fluorine in their corresponding API. Nevertheless, in

totality, this exploratory data analysis of the non-gated LIBS data reveals that the chemical basis in the form of PCs gives rise to a substantive degree of separation of the data points corresponding to a particular pharmaceutical formulation – which is promising for the development of models for classifying and screening these and similar pharmaceutical tablets.

In a similar fashion to understand the similarity and classification PCA is applied on HEMs data. The cumulative variance of first three PCs accounted for 99.5% of total data. The first PC explain 95.06% and cumulatively with second PC explain 98.46% of the variance associated with data as shown in [figure 65](#). The first PC has the largest variance

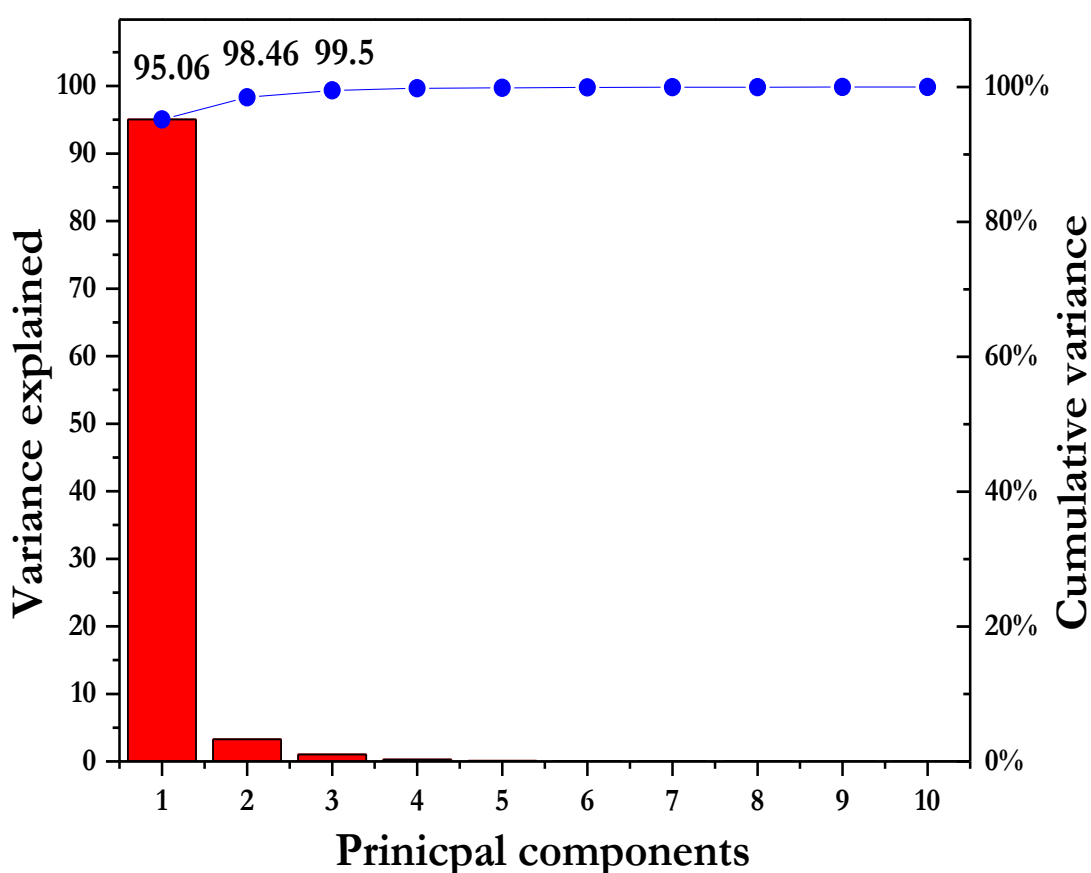


Figure 64. Scree plot of HEMs used in the study

dominated by hydrogen, oxygen and nitrogen peaks as shown in [figure 65](#). Contrary to pharmaceutical samples, the first PC mostly has the peaks of hydrogen, oxygen and nitrogen.

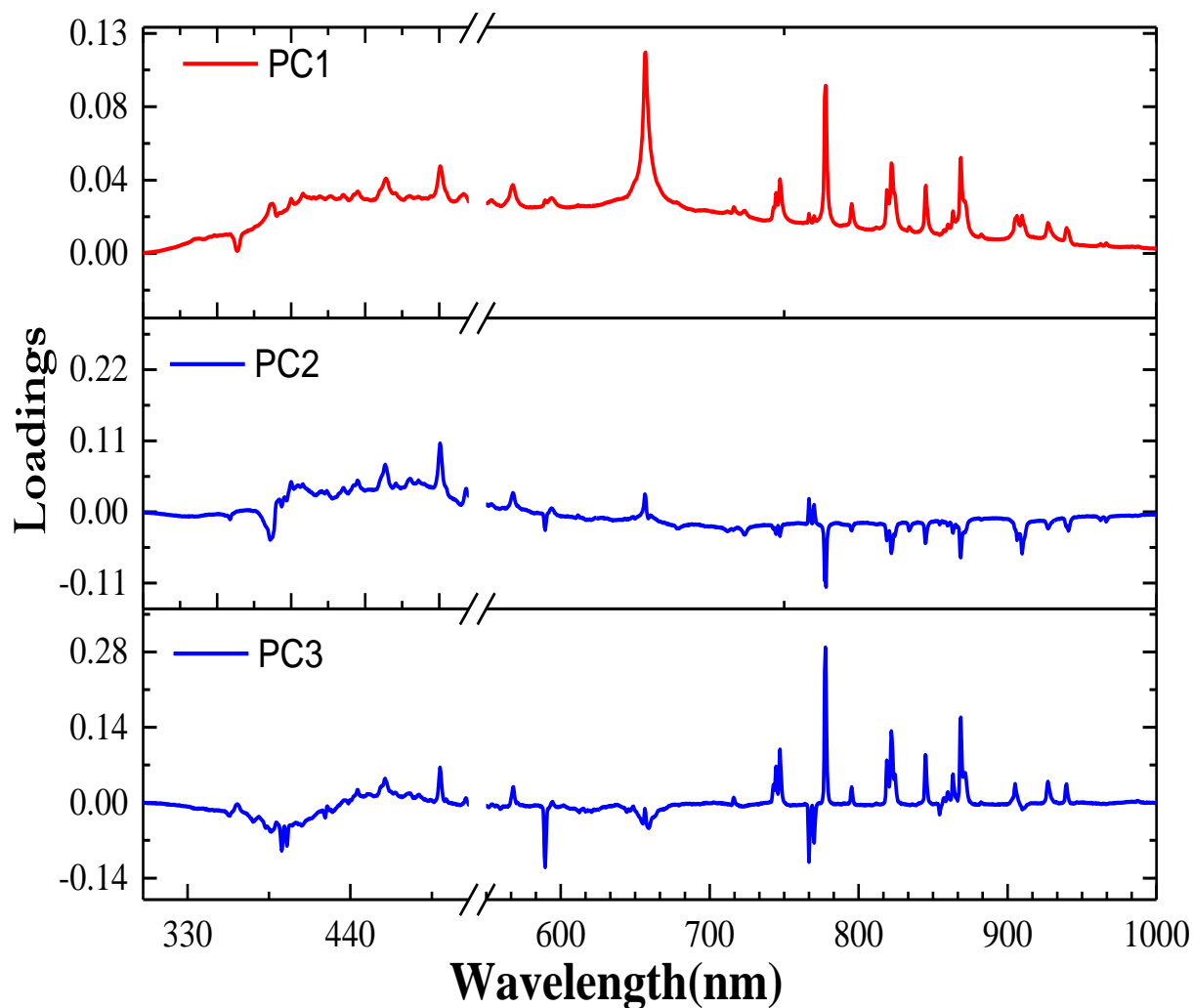


Figure 65. Plot of the first three principal components corresponding to the entire spectral dataset acquired for all samples of HEMs. These three principal components, combined, explain 99.5% of the net variance in the dataset.

Score plot which consists of maximum variance has been plotted and is shown in [figure 66](#). The data points corresponding to each compound are clustered together and well separated with other compound data points. The datapoint forming RDX cluster is scattered compared to other data points, whereas TNT appeared to be close.

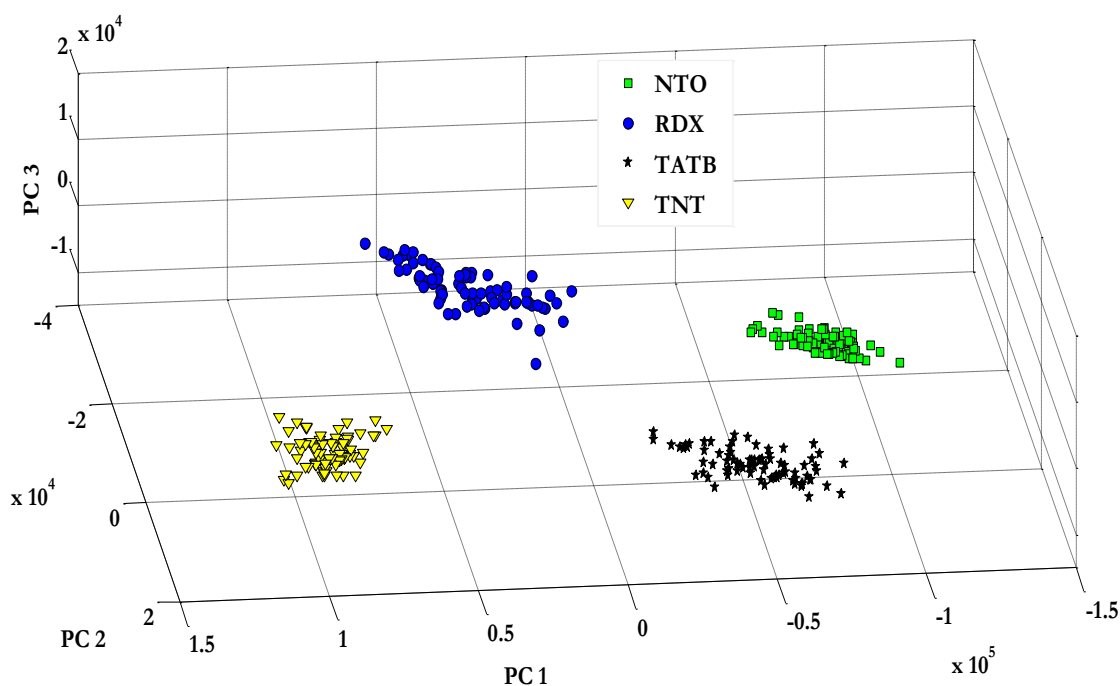


Figure 66. Discrimination of HEMs based on their LIBS spectra: Scores plot corresponding to principal components 1, 2 and 3 for the spectral dataset acquired from the four samples.

To ensure suitable quality of the data for the development of classification algorithms, hierarchical clustering by means of dendrogram analysis was first pursued. In this case, the grouping of the objects in a dendrogram to detect outliers based on the presence of isolated branches is exploited. Moreover, dendrogram analysis was also used to obtain class similarity between the different pharmaceutical formulations and HEMs. In this case, the mean spectrum for each formulation was computed from all the spectra acquired from that formulation barring the detected outlier spectra. [Figure 67](#) and [figure 68](#) shows the resultant dendrogram obtained by analyzing the mean spectrum of each four formulations. This corroborates the PCA findings as one clearly observe that the Metformin hydrochloride exhibits spectral similarity to Cetirizine dihydrochloride in comparison to the other formulations, but in the case of HEMs clear separation has been observed

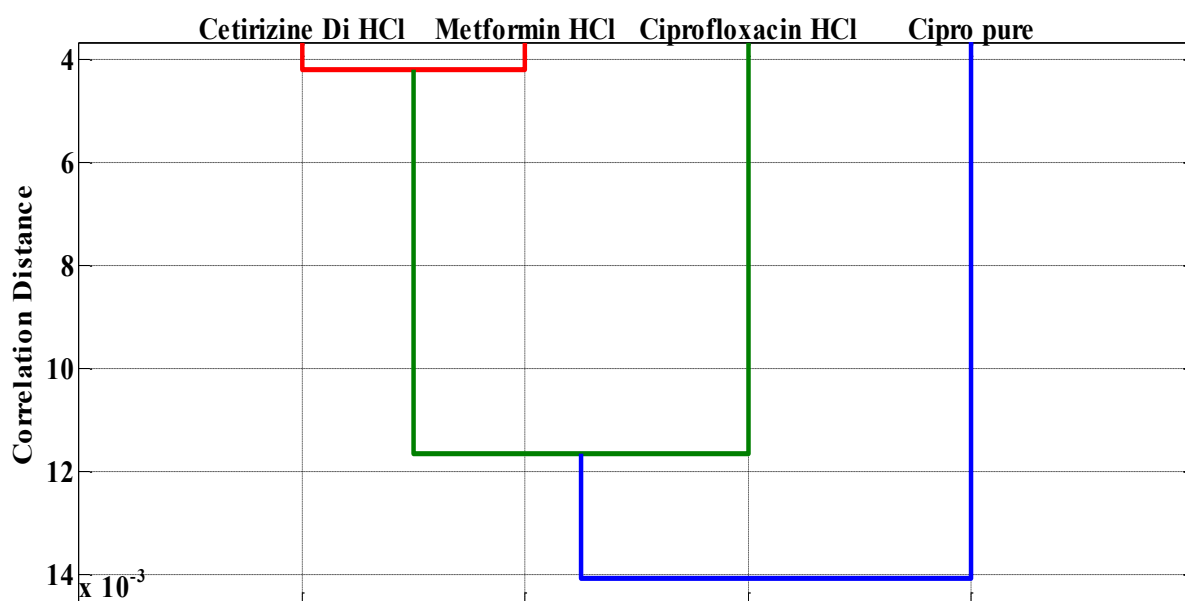


Figure 67. Hierarchical clustering using the dendrogram representation for LIBS spectra acquired from the 4 sets of pharmaceutical samples.

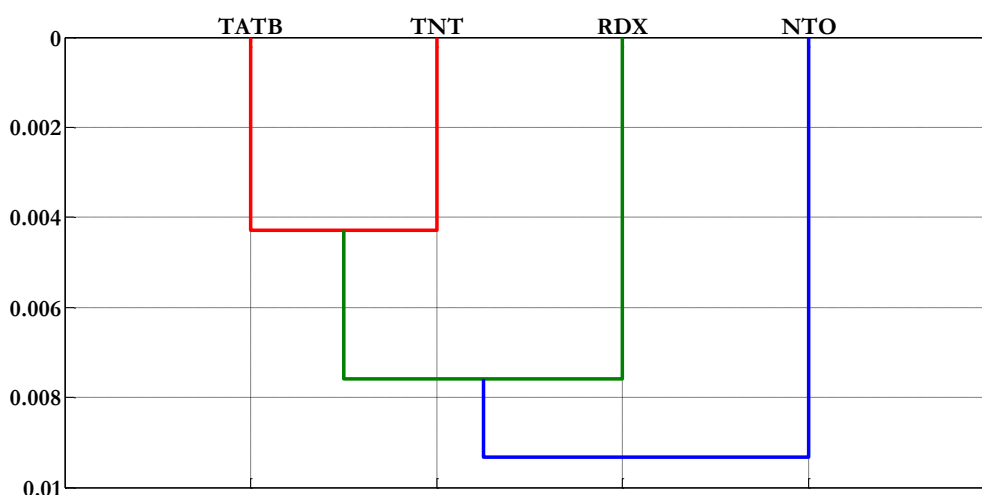


Figure 68. Hierarchical clustering using the dendrogram representation for LIBS spectra acquired from the four sets of HEM samples.

In order to evaluate the suitability of the non-gated LIBS data for classification purposes, first SIMCA was chosen to develop the discrimination algorithm. Specific to the SIMCA model development, 70% of the samples were designated as training data while the rest of the samples (30%) were held out of the model building and served as the test set. It was ensured that the spectra corresponding to each sample were split as per these representative percentage values. In order to obtain a representative estimate of the rates of correct classification, misclassification and unclassification, 100 independent

iterations were performed by re-splitting the entire data into training and test sets. Here, the model is judged to be sensitive if the correct classification rate is high and the unclassification rate is low. This analysis was separately conducted for Pharmaceuticals and HEMs

Table 21 gives the results of the classification analysis for the SIMCA-derived discrimination model for the four pharmaceutical formulations, whereas Table 22 for HEMs, based on the 3σ unclassification threshold. The misclassification rate is observed to be 0% in both cases, irrespective of the pharmaceutical or HEMs in question. The correct classification accuracy

Average rate of...	Correct	Mis	Unclassification
Cetirizine dihydrochloride	100	0	0
Cipro pure	86	0	14
Metformin hydrochloride	82	0	18
Ciprofloxacin hydrochloride	96	0	4
Average	91	0	9

Table 21 SIMCA classification results obtained from the pharmaceutical test samples over 100 iterations

varies from 82% (Metformin hydrochloride) to 100% (Cetirizine dihydrochloride) in the case of pharmaceuticals, in case of HEMs varies from 89% (RDX) to 94% (NTO). Based on the PCA and dendrogram analysis, the relatively lower classification accuracy (and correspondingly,

Average rate of...	Correct	Mis	Unclassification
NTO	94	0	6
RDX	89	0	11
TATB	93	0	7
TNT	93	0	7
Average	92	0	8

Table 22 : SIMCA classification results obtained from the HEM test samples over 100 iterations

the higher unclassification rate) of the Metformin hydrochloride sample and RDX(cluster spread is more in PCA) is not surprising. The inherent scatter of the sample data seems to be the predominant factor in the relatively inferior performance of the SIMCA model for Metformin hydrochloride and Cipro pure classification with unclassification rates observed to be 18% (Metformin hydrochloride) and 14% (Cipro pure), respectively. On average, the SIMCA-derived model provides a correct classification rates of 91% and ~92% for pharmaceutical and HEMs respectively, which suggests that the non-gated LIBS spectral measurements, even under ambient air conditions, provides sufficient information for sensitive discrimination of the studied formulations. It is worth mentioning that these results are comparable to results obtained from gated LIBS spectra of similar pharmaceutical samples [17,18](and also the same in the chapter 3) indicating that the presence of the continuum emission signals does not hinder the statistical performance, especially on application of suitable multivariate algorithms. To the best of knowledge, this provides the first experimental demonstration of the suitability of non-gated data for classification of complex organic samples having similar composition.

Finally, to more comprehensively examine the classification capability of the acquired non-gated data, PLS-DA and ANN models were also constructed. For PLS-DA, a single “global” model is obtained for classification, in contrast to the independent PCA submodels developed for SIMCA analysis. Using this single global PLS-DA model, for Pharma and HEMs separately we were able to obtain 100% correct classification accuracy for all formulations in both cases(Table 23 for pharmaceuticals and Table 24 for HEMs). While the underlying principle of PLS-DA to obtain maximum separation

Average rate of...	Correct classification	Misclassification	Unclassification
Cetirizine dihydrochloride	100	0	0
Cipro pure	100	0	0
Metformin hydrochloride	100	0	0
Ciprofloxacin hydrochloride	100	0	0
Average	100	0	0

Table 23: PLSDA classification results using pharmaceutical formulations obtained from the test samples over 100 iterations

between classes (as opposed to modeling the maximum variance in each individual class in SIMCA) may be partially responsible for the improvement, the presence of dataset-specific factors cannot be neglected. For example, previously observed for the corresponding gated data that SIMCA was marginally more sensitive, even though PLS-DA was significantly more robust in classifying unknown samples please refer chapter 3 [14].

Average rate of...	Correct classification	Misclassification	Unclassification
NTO	100	0	0
RDX	100	0	0
TATB	100	0	0
TNT	100	0	0
Average	100	0	0

Table 24: PLSDA classification results obtained from the test samples over 100 iterations

Furthermore, application of ANN also yields no misclassification or unclassification (Table 25 and Table 26). Given the nonlinear nature of the ANN algorithm and the potential interferent sources in the non-gated data, the enhanced performance in this case, with respect to SIMCA, is not wholly unexpected. Variability in the continuum emission, plasma self-absorption and matrix effects could contribute towards such nonlinear changes in intensity ratios in the features of interest.

Average rate of...	Correct classification	Misclassification	Unclassification
Cetirizine dihydrochloride	100	0	0
Cipro pure	100	0	0
Metformin hydrochloride	100	0	0
Ciprofloxacin hydrochloride	100	0	0
Average	100	0	0

Table 25: ANN classification results obtained from the HEMs test samples over 100 iterations

Average rate of...	Correct classification	Misclassification	Unclassification
NTO	100	0	0
RDX	100	0	0
TATB	100	0	0
TNT	100	0	0
Average	100	0	0

Table 26: ANN classification results obtained from the HEMS test samples over 100 iterations

4.5.2 EXPLOSIVE/NON EXPLOSIVE CLASSIFICATION

Detection of explosive on site is a challenging assignment to researchers particularly when using vast amount of spectral data. Anticipating the results from linear models with larger number of compounds in LIBS database may lead to false positive. In this section nonlinear classifier ANN is used for identifying explosive from huge data bank of LIBS. It is more beneficial for homeland security, environmental monitoring and also for demining. In a real time application first it is critical to identify if a given material is explosive or not before proceeding further

To cope the real world complexity, extensive nongated LIBS of variety of organic and inorganic compounds are used for modelling. Total 3567 spectra of various explosives viz., NTO (719+70), HMX 72, PETN 11, RDX (49+266), TATB 745, TNT (101+1127), pyrazole 71, 1 Nitro pyrazole (25+74), 3 Nitro pyrazole (54+59), 4 Nitro pyrazole (44+80); In the case of Non explosive sample 3794 viz., Aluminium(90 grade) 267, CD 257, Dextrose 20, D fructose 30, Ltrito 43, PVC 490, sodium lauryl sulphate 133, Zirconium strantium titanate 94, Anthodiasta Nobilis 98, brass 388, cytosine 21, edta 131, Elasis guniosis 80, Entando dphargm 53, Glass 294, glycine 33, hypoxanthanine 19, methionine 30, mica 275, naphthalene 219, neomycine sulphate 26, polythalia longiphalia 130, protein 36, pyridine 30, silicon 101, Stainless steel 152, teflon 342, Terminalia evorinsis 162, thymine 24, triethonalamine 16 were acquired. The numbers

adjacent to sample names are number of spectra, while the numbers in brackets are number of recorded data at different instances.

Here, a two-layer feed-forward network model was used for classification consisting of one input layer, one hidden layer and one output layer. The input layer comprises of raw data intensity values at various wavelengths. Sigmoid functions were used as neurons for the input and hidden layers. The final output consists of binary encoding used as target, where 1 is for correct assignment and 0 for absent. The ANN was trained with a classical scaled conjugate gradient back-propagation algorithm. The data division control was in-built in the toolbox. Out of the 7361 total spectral data, 70% were presented to the network during training, and the network was adjusted according to its error. 15% of the data was used to measure network generalization and to halt training indicating an increase in the mean square error of the validation samples. The remaining 15% were held out of the training and validation process and therefore provide an independent measure of network performance. The performance of this supervised learning algorithm can be visualized by confusion matrix as shown in table 27.

		Predicted Labels	
Actual labels		Explosive	Non explosive
	Explosive	99.97	0.03
	Non explosive	0.77	99.23

Table 27: Confusion matrix showing classification rate in percentage. The diagonal elements of this matrix is correctly identified; while off diagonal elements are misclassified.

It shows the near perfect classification ~100% is achieved. No preprocessing has been performed on this data except 532 laser affected 40 nm bandwidth points are removed;. The following matrix is the resultant over 1000 iterations.

4.6 Discussion

In addition to quantifying the classifier performance based on non-gated LIBS spectra, it is imperative to precisely understand the impact of the continuum emission background. As alluded to above, the prevailing thought in the LIBS community has centered around the uncharacteristic nature of the continuum emission, as it emanates from radiative

recombination and Bremsstrahlung emission that do not depend on the identity of the element/ion. Considerable contemporary attention has been focused on the choice of a proper delay time when the ratio of line emission to background (continuum) emission is very high, since “only the lines emission from the plume is important for the compositional analysis of the sample target”.

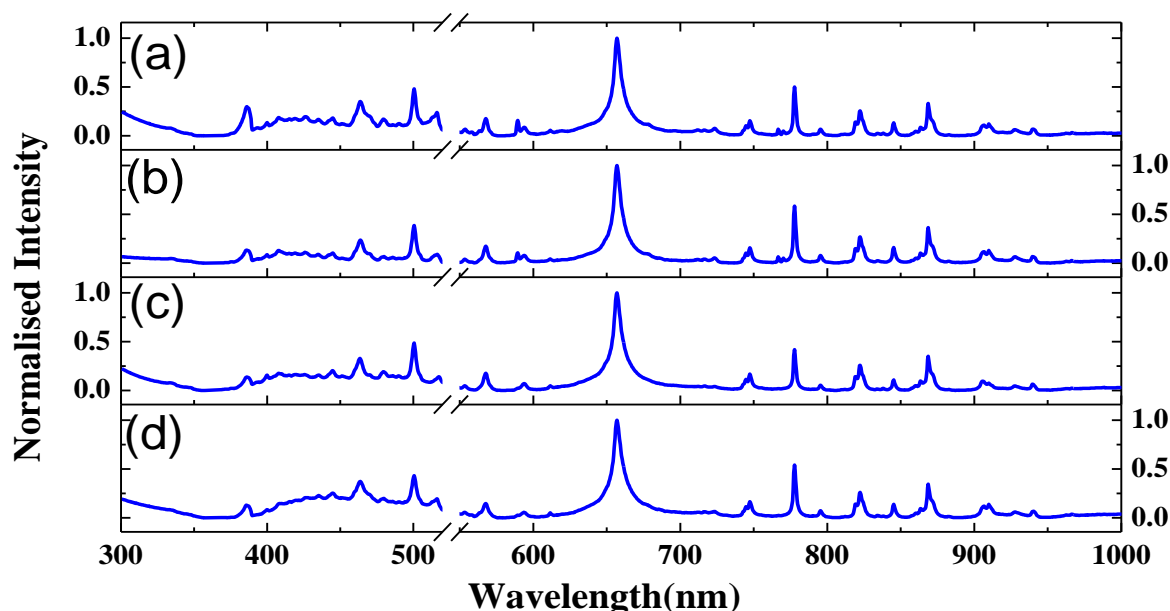


Figure 69: Background corrected LIBS spectra acquired from the pharmaceutical formulations: (a) Cetirizine dihydrochloride; (b) Cipro pure; (c) Metformin hydrochloride; (d) Ciprofloxacin hydrochloride. Intensity on the y-axis is normalized with respect to the characteristic hydrogen emission peak at 656 nm. The continuum background is approximated and removed using a least squares-based polynomial curve-fitting algorithm.

To test this hypothesis, compared the aforementioned results using non-gated LIBS spectra to those where the continuum background is removed using a lower order polynomial. Since the continuum background is broad and featureless, previous investigators have employed numerical post-processing schemes to approximate and remove it in non-gated LIBS data. Here, the background of the LIBS spectra was removed by application of an iterative least squares-based curve-fitting algorithm that uses a polynomial (6th order for our dataset) with non-negativity constraints. This algorithm (and its variants) have been extensively used for addressing broad backgrounds in similar spectroscopic data featuring small characteristic peaks and large backgrounds. Representative background-removed LIBS spectra from each pharmaceutical formulation are shown in Figure 55. Subsequent to background removal, all the spectra were

subjected to the same protocol as described above, namely dendrogram analysis for outlier removal and SIMCA for development of the discrimination algorithm. Using 100 iterations on the partitioned test sets (30%), we observe that, in this case, the SIMCA-derived models provide an average correct classification rate of 88% with a corresponding misclassification and unclassification rates of 0% and 12%, respectively. Based on the average correct classification rates and the corresponding standard deviations for the non-gated LIBS data with and without background removal, the two-tailed p -value is computed to be less than 0.0001. By conventional criteria (*i.e.* rejection of null hypothesis at p -value less than 0.01), this difference can be considered to be extremely statistically significant. It is also observed that the reduction in correct classification rate relative to Table 21 is consistent for all the pharmaceutical formulations. The difference in correct classification rate on application of SIMCA on the two datasets could be attributed to: (A) the presence of diagnostic information in the continuum background; and/or (B) introduction of artifacts due to the background removal procedure that results in deterioration of model performance. Further experiments with a range of time delays (currently underway in our laboratories) are necessary to elucidate whether the continuum background may indeed aid specific classification analyses.

Finally, it is worth analyzing the influence of the additional noise component incorporated in the non-gated LIBS data due to the continuum emission background. Overall, the relative root-mean-square noise in a spectrum has two contributions, namely the constant fixed-pattern noise characterizing the non-uniform response of the CCD pixels and the shot noise. Given the high signal levels of LIBS data, the latter has lesser significance and can always be reduced relative to the signal level by acquiring for longer time periods. The better signal-to-noise characteristics of the CCD detectors employed for non-gated detection in relation to the ICCD used for gated detection also alleviates this problem. Additionally, employing a suitable flat-field correction scheme can eliminate the more relevant fixed-pattern noise for high intensity LIBS data. Recently, researchers have detailed a promising fixed-pattern noise removal approach that focuses on subtraction of spectra acquired before and after shifting of spectrometer grating, which can be suitably employed in future non-gated LIBS studies to enhance classifier performance.

4.7 Conclusion

In summary, the potential of non-gated LIBS for identification and classification of pharmaceutical formulations and HEMs with similar elemental compositions is proposed and demonstrated. It is observed that the non-gated spectra shows high efficacy in discrimination with an average correct classification rate of 91% and 92% for the classification using SIMCA. An excellent classification has been achieved using PLSDA and ANN, which is found to be 100% average over 100 iterations. Importantly, from a scientific standpoint, the presence of the continuum background in the non-gated LIBS dataset does not impede the classification performance and also envision that the significant advantages of this detection method in terms of cost, maintenance and system portability, coupled with its sensitivity, will ideally complement not only the existing analytical technologies for determination of surface and internal distribution of API and excipients in addition to verification of the drug content in the finished dosage form but also for detection of explosives on field with standoff arrangement. The sensitivity of this LIBS approach can be further enhanced by implementing hybrid classification schemes, which feature a range of (linear and non-linear) multivariate analysis strategies. With further refinements in the classification methodologies as well as development of a hand-held LIBS monitor, the approach presented here can be extended for sensitive identification in other critical applications including standoff detection of high energy and hazardous materials and inline monitoring of pharmaceuticals.

4.8 References

1. Hoehse M, Gornushkin I, Merk S, Panne U (2011) Assessment of suitability of diode pumped solid state lasers for laser induced breakdown and Raman spectroscopy. *Journal of Analytical Atomic Spectrometry* 26: 414-424.
2. Sirven JB, Bousquet B, Canioni L, Sarger L (2004) Time-resolved and time-integrated single-shot laser-induced plasma experiments using nanosecond and femtosecond laser pulses. *Spectrochimica Acta Part B: Atomic Spectroscopy* 59: 1033-1039.
3. Eland KL, Stratis DN, Gold DM, Goode SR, Angel SM (2001) Energy Dependence of Emission Intensity and Temperature in a LIBS Plasma Using Femtosecond Excitation. *Applied Spectroscopy* 55: 286-291.
4. Sakka T, Irie K, Fukami K, Ogata YH (2011) Emission spectroscopy of laser ablation plasma with time gating by acousto-optic modulator. *Review of Scientific Instruments* 82: 023112.

5. Pořízka P, Klessen B, Kaiser J, Gornushkin I, Panne U, et al. (2014) High repetition rate laser-induced breakdown spectroscopy using acousto-optically gated detection. *Review of Scientific Instruments* 85: 073104.
6. Wang S, Wendt AE, Boffard JB, Lin CC (2015) Time-resolved spectroscopy using a chopper wheel as a fast shutter. *Review of Scientific Instruments* 86: 013111.
7. Mueller M, Gornushkin IB, Florek S, Mory D, Panne U (2007) Approach to Detection in Laser-Induced Breakdown Spectroscopy. *Analytical Chemistry* 79: 4419-4426.
8. Tognoni E, Cristoforetti G, Legnaioli S, Palleschi V (2010) Calibration-Free Laser-Induced Breakdown Spectroscopy: State of the art. *Spectrochimica Acta Part B: Atomic Spectroscopy* 65: 1-14.
9. Diwakar PK, Jackson PB, Hahn DW (2007) The effect of multi-component aerosol particles on quantitative laser-induced breakdown spectroscopy: Consideration of localized matrix effects. *Spectrochimica Acta Part B: Atomic Spectroscopy* 62: 1466-1474.
10. Barman I, Kong C-R, Singh GP, Dasari RR (2011) Effect of photobleaching on calibration model development in biological Raman spectroscopy. *Journal of Biomedical Optics* 16: 011004-011004-011010.
11. Harmon RS, DeLucia FC, McManus CE, McMillan NJ, Jenkins TF, et al. (2006) Laser-induced breakdown spectroscopy – An emerging chemical sensor technology for real-time field-portable, geochemical, mineralogical, and environmental applications. *Applied Geochemistry* 21: 730-747.
12. Singh JP, Thakur SN (2007) *Laser-Induced Breakdown Spectroscopy*: Elsevier Science.
13. Yamamoto KY, Cremers DA, Ferris MJ, Foster LE (1996) Detection of Metals in the Environment Using a Portable Laser-Induced Breakdown Spectroscopy Instrument. *Applied Spectroscopy* 50: 222-233.
14. DeLucia FC, Jr., Samuels AC, Harmon RS, Walters RA, McNesby KL, et al. (2005) Laser-induced breakdown spectroscopy (LIBS): a promising versatile chemical sensor technology for hazardous material detection. *Sensors Journal, IEEE* 5: 681-689.
15. De Giacomo A, Gaudiuso R, Dell'Aglio M, Santagata A (2010) The role of continuum radiation in laser induced plasma spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* 65: 385-394.
16. Lopez-Moreno C, Amponsah-Manager K, Smith BW, Gornushkin IB, Omenetto N, et al. (2005) Quantitative analysis of low-alloy steel by microchip laser induced breakdown spectroscopy. *Journal of Analytical Atomic Spectrometry* 20: 552-556.
17. Dingari NC, Barman I, Myakalwar AK, Tewari SP, Kumar Gundawar M (2012) Incorporation of Support Vector Machines in the LIBS Toolbox for Sensitive and Robust Classification Amidst Unexpected Sample and System Variability. *Analytical Chemistry* 84: 2686-2694.

18. Myakalwar AK, Sreedhar S, Barman I, Dingari NC, Venugopal Rao S, et al. (2011) Laser-induced breakdown spectroscopy-based investigation and classification of pharmaceutical tablets using multivariate chemometric analysis. *Talanta* 87: 53-59.
19. Hahn DW, Omenetto N (2010) Laser-Induced Breakdown Spectroscopy (LIBS), Part I: Review of Basic Diagnostics and Plasma-Particle Interactions: Still-Challenging Issues Within the Analytical Plasma Community. *Applied Spectroscopy* 64: 335A-366A.
20. Beier BD, Berger AJ (2009) Method for automated background subtraction from Raman spectra containing known contaminants. *Analyst* 134: 1198-1202.
21. Kwiatkowski A, Gnyba M, Smulko J, Wierzba P (2010) Algorithms of Chemicals Detection Using Raman Spectra. *Metrology and Measurement Systems*. pp. 549.
22. Lieber CA, Mahadevan-Jansen A (2003) Automated Method for Subtraction of Fluorescence from Biological Raman Spectra. *Applied Spectroscopy* 57: 1363-1367.
23. Schulmerich MV, Reddy R, Kodali AK, Elgass LJ, Tangella K, et al. (2010) Dark Field Raman Microscopy. *Analytical Chemistry* 82: 6273-6280.

This page intentionally left blank

5. DISCRIMINATION OF ORGANIC ISOMER COMPOUNDS

LIBS has surfaced as an attractive alternative to mass spectrometry and wet chemistry methods for chemical identification, driven by its real-time, label-free nature. Rapid analysis needs, especially in high-energy materials and pharmaceutical compounds, have further fueled an increasing number of refinements in LIBS. Yet, isomers, for example of pyrazoles, are seldom identifiable by LIBS as they generate nearly identical spectra. A set of three homemade isomeric pyrazole compounds are studied and combined with a suite of data analysis approaches to exploit the subtle, but reproducible, differences in LIBS spectra to develop a sensitive and reliable segmentation method. PCA and ANN are used on the series of data obtained using gated and nongated spectrographs. The key mechanistic principles (causation) are investigated behind spectral variations and confirmed their statistically significant nature that empowers the excellent classification performance. The plasma parameters, including their temporal evolution, have been estimated.

5.1 Introduction

Identification of isomers plays a vital role in pharmacology where preferential selection of a specific isomer is crucial at nearly all stages of drug discovery, development and production. Isomers of a drug often show drastically different effects owing to significant differences in their pharmacokinetic and pharmacodynamic properties. While evaluation of stereo isomers (especially enantiomers) has received considerable attention, the rapid characterization of structural isomers is equally warranted. For example, contamination of anti-malarial drug primaquine with its structural isomer quinocide leads to adverse side effects due to the latter's higher toxicity [1]. Pentobarbital and amobarbital, which are positional isomers, differ in their potency to increase *in vivo* benzodiazepine receptor binding in cerebellum with the former being more potent than the latter [2]. Similarly, terbutaline is used as a β -adrenergic stimulant bronchial asthma drug whereas the use of its positional isomer N-(tert-butyl)-

norepinephrine is limited by its shorter duration of action owing to its specificity to metabolic enzyme catechol-O-methyltransferase [3].

Standard mass spectrometry (MS), which is the principal method for chemical characterization in most analytical laboratories, employs an electron gun to cause electron impact ionization of the analyte. In many cases, however, this method of ionization is not capable of exhibiting adequate differences for reliable isomer identification. Therefore, instead of using electron gun, Dantus and co-workers proposed femtosecond lasers with pulse shaper and demonstrated successful identification of xylene isomers [4]. Taking a different route, Berman et al. combined MS with multivariate analysis and successfully demonstrated the possibility of distinguishing monosaccharide isomers [5]. Yet, the deployment of the advanced MS methods as process analytical technology is limited by its slow analysis time, high cost and complex operations that necessitate substantive user expertise. Therefore there is an unmet need for a facile, real-time isomer analyses tool that requires minimal or no sample preparation and is cost-effective.

In this milieu, laser induced breakdown spectroscopy (LIBS) offers a powerful method to meet these demands of isomer detection due to its intrinsic advantages of *in situ* monitoring, speed of spectral acquisition, portability and depth profiling. LIBS relies on characterization of the emission from the plasma plume induced by the light-matter interaction and has shown great potential for analysis of high-energy materials [6-9] and pharmaceutical drugs [10-12] and environmental monitoring [13, 14]. In particular, the integration of chemometrics with LIBS hardware has pushed the technique from an esoteric laboratory-based method to routine identification and discrimination of a wide variety of targets on the basis of a spectrum that is acquired in a fraction of a second [15-19]. Investigations by Rehse and others have also revealed the capability of identifying biological specimen, notably bacteria, pollen and proteins, using nanosecond and femtosecond LIBS [20-22]. Intriguingly, the possibility of distinguishing normal and malignant tumor cells from histological sections [23] has also been explored. Motivated by these findings, it is hypothesized that LIBS may provide a suitable isomer segmentation tool, based on the (potential) differences in molecular fragmentation and the subtle, but consistent, variations in plasma conditions for each isomer.

5.2 Materials and methods

The feasibility of using LIBS for isomer classification by selecting three representative isomers of the pyrazole class viz., 1-Nitro pyrazole (1NP), 3-Nitro-1H-pyrazole (3NP) and 4-Nitro-1H-pyrazole (4NP) (Table 28)

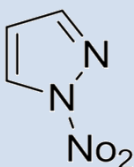
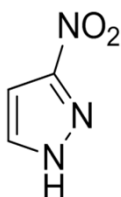
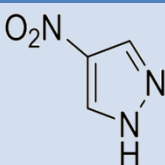
S no	Sample Structure	Sample Name & Chemical Formula	No of LIBS spectra			
			Gated Config.			Nongated Config.
			Set 1	Set 2	Set 3	
1		1-Nitro pyrazole $C_3H_3N_3O_2$	76	101	13	74
2		3-Nitro-1H-pyrazole $C_3H_3N_3O_2$	88	289	78	54
3		4-Nitro-1H-pyrazole $C_3H_3N_3O_2$	101	206	108	80

Table 28. Details of isomer compounds 1NP, 3NP and 4NP : sample name, chemical formula, chemical structure and number of LIBS spectra recorded using two detection systems. Set refers to a trial.

prepared in house [24] have been investigated. The selected class of nitropyrazole isomers have the following molecular formula: $C_3H_3N_3O_2$. Derivatives of such pyrazoles are commonly used for their analgesic, anti-inflammatory, antipyretic and tranquilizing properties in medicine [25]. Specifically, an important application necessitating identification of nitropyrazole isomers is due to their differential effects on post-ischemia retinal function recovery [26]. 1-Nitropyrazole (also known as N-Nitropyrazole) exhibits nitric oxide (NO) formation, which in turn promotes ocular blood flow, whereas the other two isomers, C-Nitropyrazoles, do not [27, 28]. Moreover, 4-Nitropyrazole is a known mutagen [29] and therefore the selection of the correct isomer is critical in designing post-

ischemic insult therapeutics. Out of three, two compounds viz., 3NP and 4NP are positional isomers i.e., NO₂ functional group changes its position. 200 mg of each compounds was ground to nice smooth powder and pelletized using hydraulic pressure machine by applying of pressure about 3-4 tones. Two detection systems have been employed i.e., gated and nongated spectrograph configurations, which will be discussed detailed in the next sections.

5.3 Spectral observations

The details of experimental arrangement and instrumentation have been discussed elaborately in section 2.4. Briefly, the parameters used are: laser energy 22 mJ, repetition rate 0.8 Hz, spectrometer in single scan mode with gate width of 2 μs, initial gate delay 1 μs and MCP gain at 85. The LIBS spectra for all the three samples are shown in figure 70. It is evident that

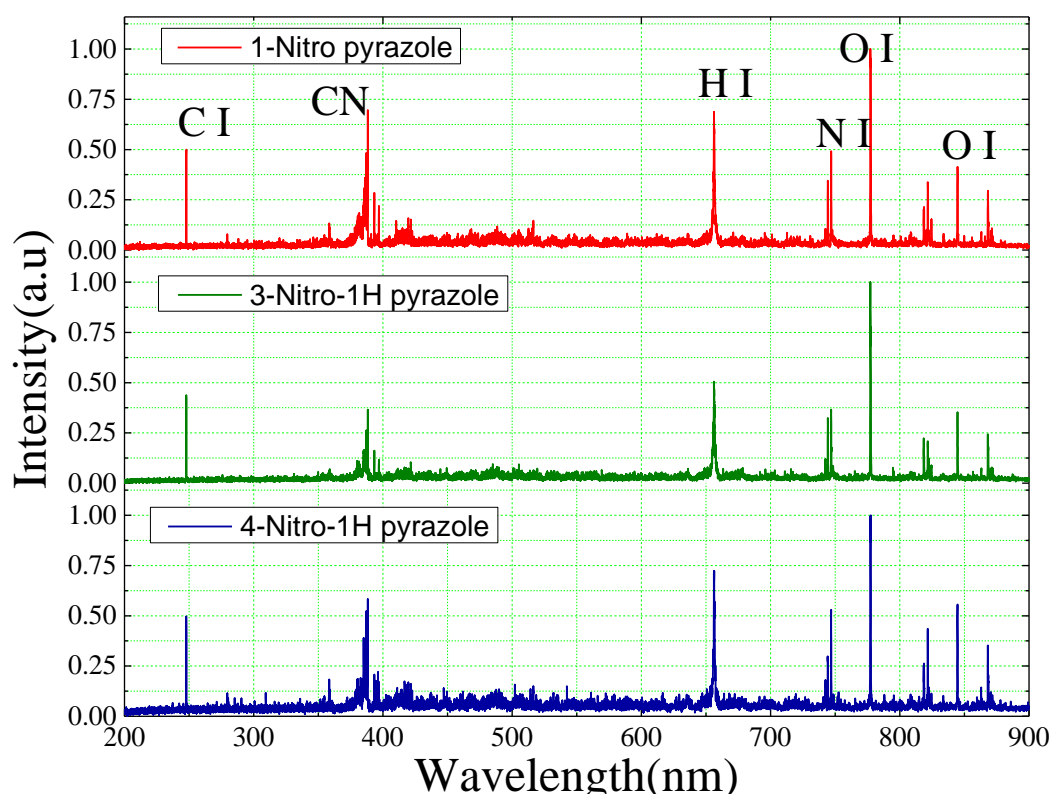


Figure 70. The LIBS spectra of isomers 1Nitro pyrazole, 3-1H-Nitro pyrazole and 4-1H-Nitro pyrazole possessing formula C₃H₃N₃O₂

the spectra corresponds to 1NP, 3NP and 4NP shown in figure 70 look similar and it is hard to be distinguished by visual inspection. The oxygen peak centered at 777.3 nm is

used for normalization. The elemental peaks such as carbon (247.8 nm), hydrogen (656.6 nm), nitrogen (742.8, 744.6, 746.8 and 868 nm triplet) and oxygen (715, 777.3 nm and 844 nm) are observed in the spectra. Along with elemental peaks some molecular peaks, majorly CN band due to $B^2\Sigma$ transitions at $\Delta v = -1, 0$ and $+1$ in between 357- 360 nm, 384 - 389 nm and 414 - 423 nm respectively are also observed. The relative intensities associated with the three compounds have been tabulated in [Table 29](#).

S no	Element	Wavelength (nm)	Relative intensities		
			1NP	3NP	4NP
1	C I	247.80	0.50	0.44	0.50
2	CN	386.17	0.36	0.16	0.39
3	CN	387.13	0.46	0.26	0.52
4	CN	388.29	0.69	0.37	0.58
5	H I	656.60	0.69	0.50	0.72
6	N I	742.40	0.11	0.12	0.18
7	N I	744.60	0.34	0.33	0.29
8	N I	746.80	0.49	0.37	0.53
9	O I	777.30	1.00	1.00	1.00
10	O I	844.60	0.41	0.35	0.56
11	O I	868.20	0.30	0.24	0.35

Table 29. Details of elements present in the isomers and their relative intensities.

These intensities of C, CN, H, N and O peaks are drawn from [figure 70](#).

Due to scarcity of sample, time series spectra could be recorded only for 3NP, 4NP as shown in [figure 71](#) and [figure 72](#). The time evolution of LIBS shown in [figures 71, 72](#) decay faster for atomic lines such as carbon, hydrogen, oxygen and nitrogen lines, but longer for molecular peaks such as CN band. In both the three dimensional graphs CN peak has significant signature for more than 20 μ s. The gate width to be time series was 1

μs , with $0.5 \mu\text{s}$ initial delay and time step of $1 \mu\text{s}$. The time evolution has been fitted using exponential fit and the decay constants are found to be in the order of $2\text{--}40 \mu\text{s}$. Time evolution of peak fitted using exponential decay has been shown in [figure 73](#).

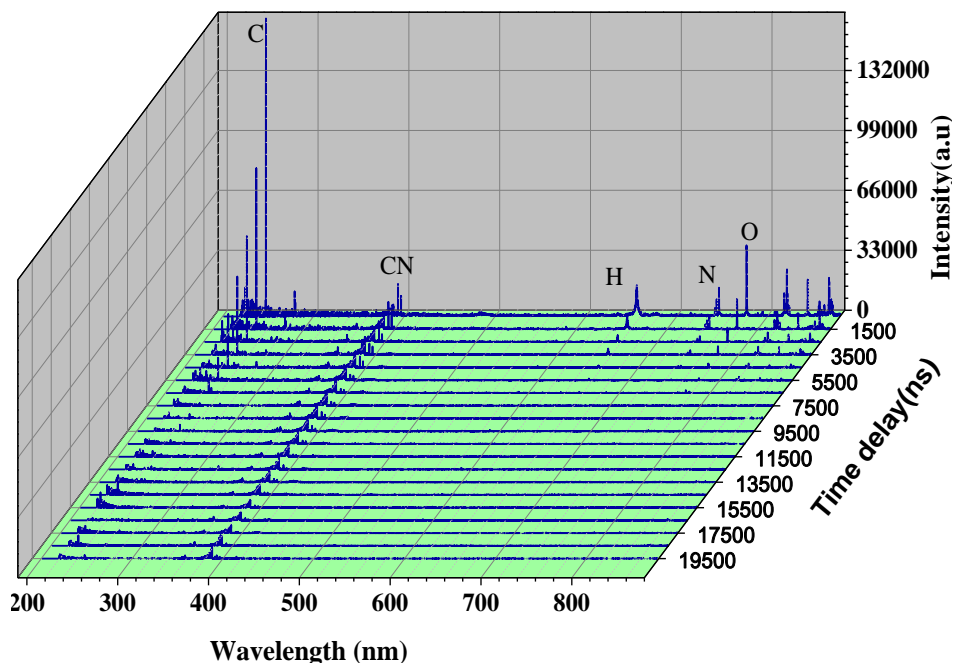


Figure 71. Time evolution of LIBS signal using 3NP with laser energy 92 mJ and step, gate width are fixed to $1 \mu\text{s}$ and initial delay chosen $0.5 \mu\text{s}$.

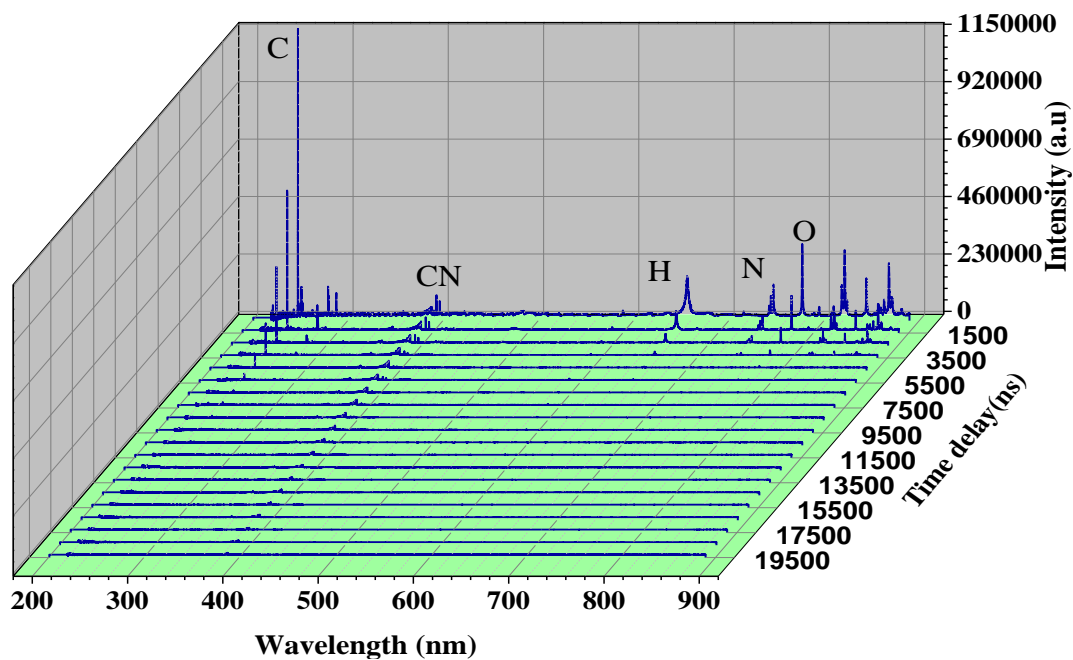


Figure 72. Time evolution of LIBS signal using 4NP with laser energy 92mJ and step, gate width are fixed to $1 \mu\text{s}$ and initial delay chosen $0.5 \mu\text{s}$

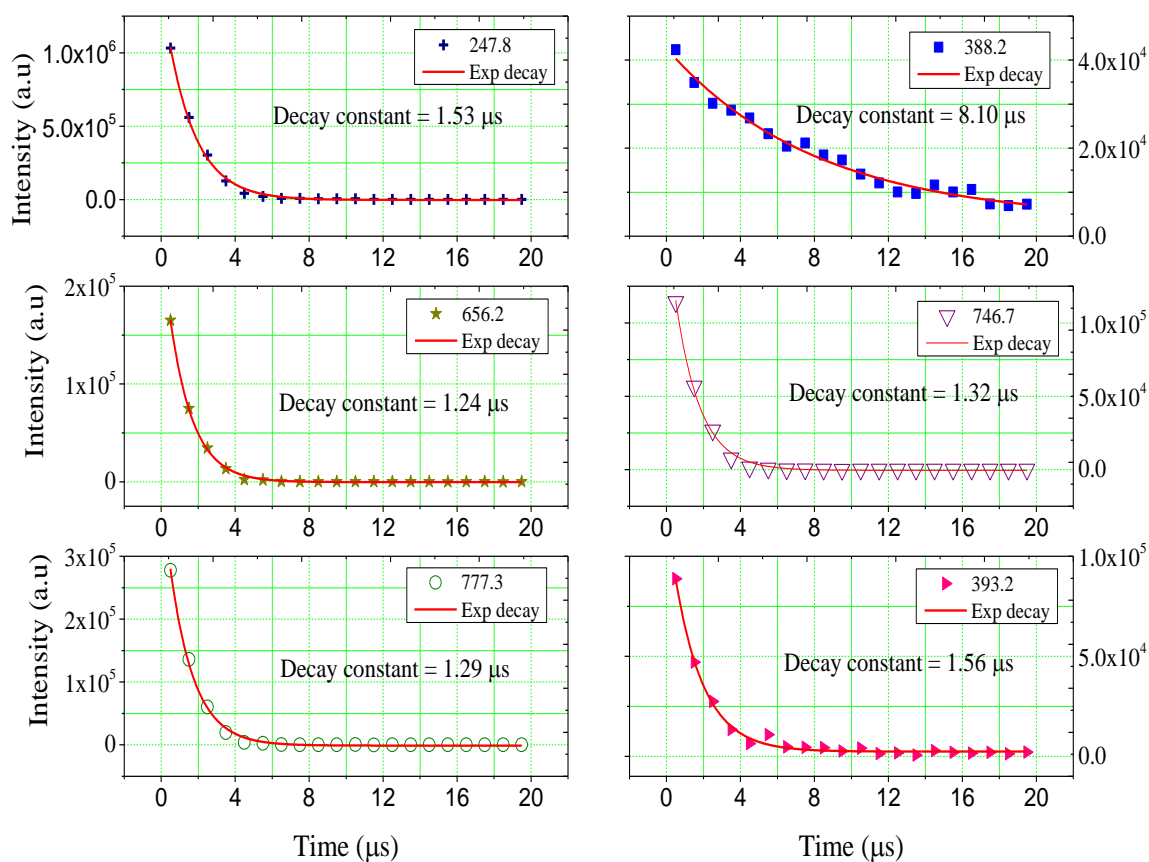


Figure 73. Time evolution of individual peaks and fit using exponential decay, j refers to decay constant.

The peak intensities are almost similar and using ratiometrics the differentiation between these compounds cannot be achieved. The results with ratiometrics approach in the case of drugs identification guides to use statistical multivariate methods and discussed in subsequent sections elaborately.

5.4 Results and discussion

5.4.1 GATED SPECTROGRAPH CONFIGURATION

This chapter also employs the same methods which were used in the case of gated explosive data viz., PCA and ANN. Figure 74 shows the score plot with first 3PCs. Careful observation of the three dimensional PC plot shows that there are only two separate clusters of red squares and blue filled circles combined with green asterisks.. Blue circles and green asterisks are indistinguishable, which hints the data pertaining to

those are closely correlated. The fluctuation of CN peak height centered at 388.2 nm has been evaluated. RSDs of 1NP, 3NP and 4NP are found to be 20.02%, 23.07% and 20.45% respectively.

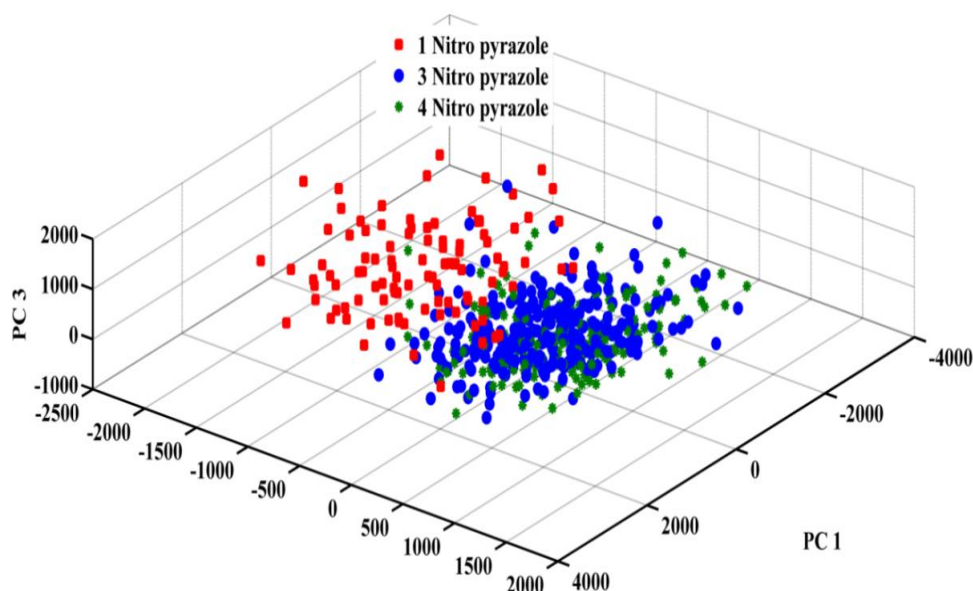


Figure 74. The score plot of 1NP, 3NP and 4NP with LIBS data collected by single scan

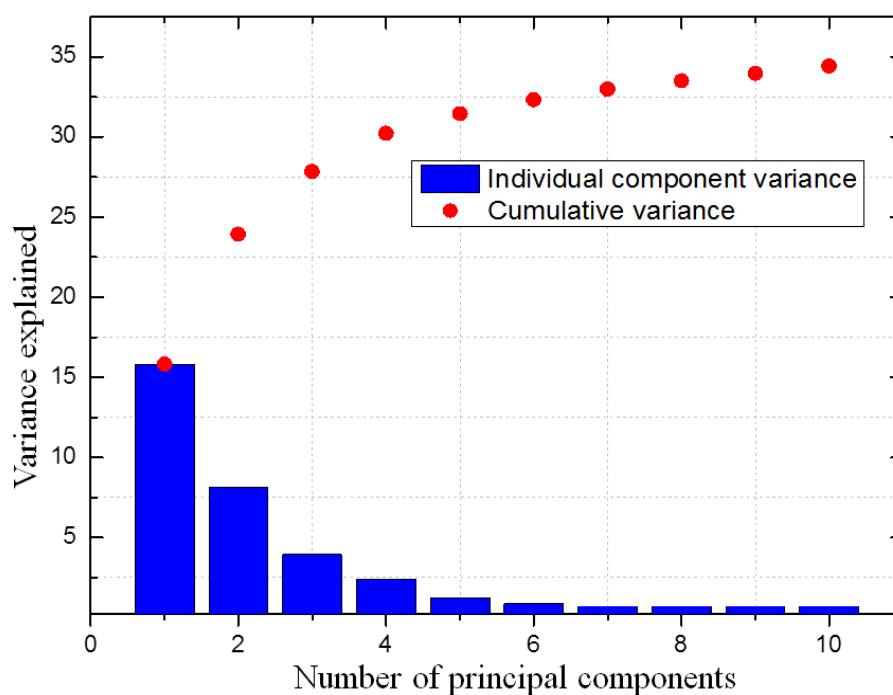


Figure.75: Scree plot shows the individual components versus corresponding variance associated with it, while the red colour circle denoted cumulative variance with increasing order with number of components

The major goal to choose PCA is to find directions that corresponds to highest variance. The first dimension of score plot corresponding to maximal variance, second dimension has the next maximum and so on. The results of individual variance are plotted in blue bar, while cumulative explained variance is depicted with red color filled circle dots as shown in figure 75. The first five individual components' cumulative explained variance is about 31.45% of the original data.

In order to reduce the fluctuations, data has been recorded with different excitation conditions and collection parameters. The same experiment was performed by accumulating the signal for three pulses, keeping other experimental parameters fixed. PCA has been applied on this data and a improved separation is observed between the data points in the score plot as shown figure 76.

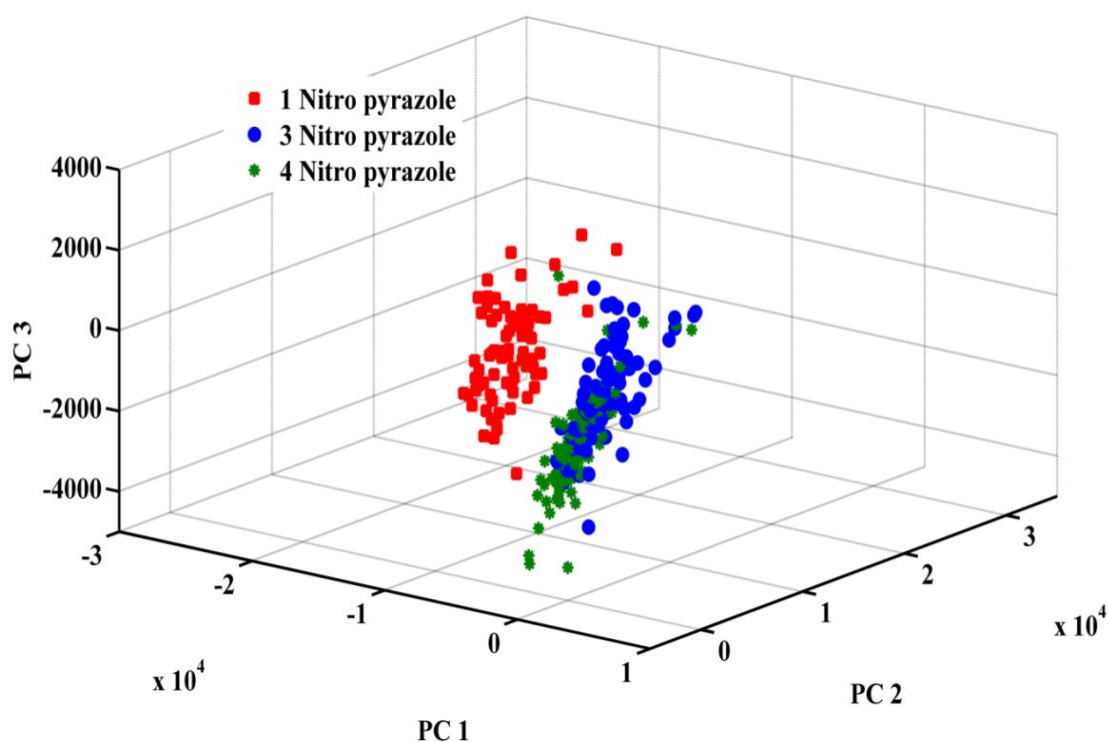


Figure 76. The score plot of 1NP, 3NP and 4NP with LIBS data acquired using accumulation of three pulses

Now two sets of more compact clusters are observed, where the first one corresponds to 1NP, the second to both 3NP and 4NP. The clusters are now more dense and close, when compared with the single scan LIBS data. This could be attributed to decrease in RSD which are now 12.08%, 15.83% and 16.02% respectively. To account the trend of

separation, the correlations between the original data and each principal component have been observed carefully.

The loadings plot shown in figure 77 gives the correlation with that original data set; more precisely it tells which of the variables are connected and impact the separation. This is the powerful tool to understand the structure of the data. The first PC loadings have positive

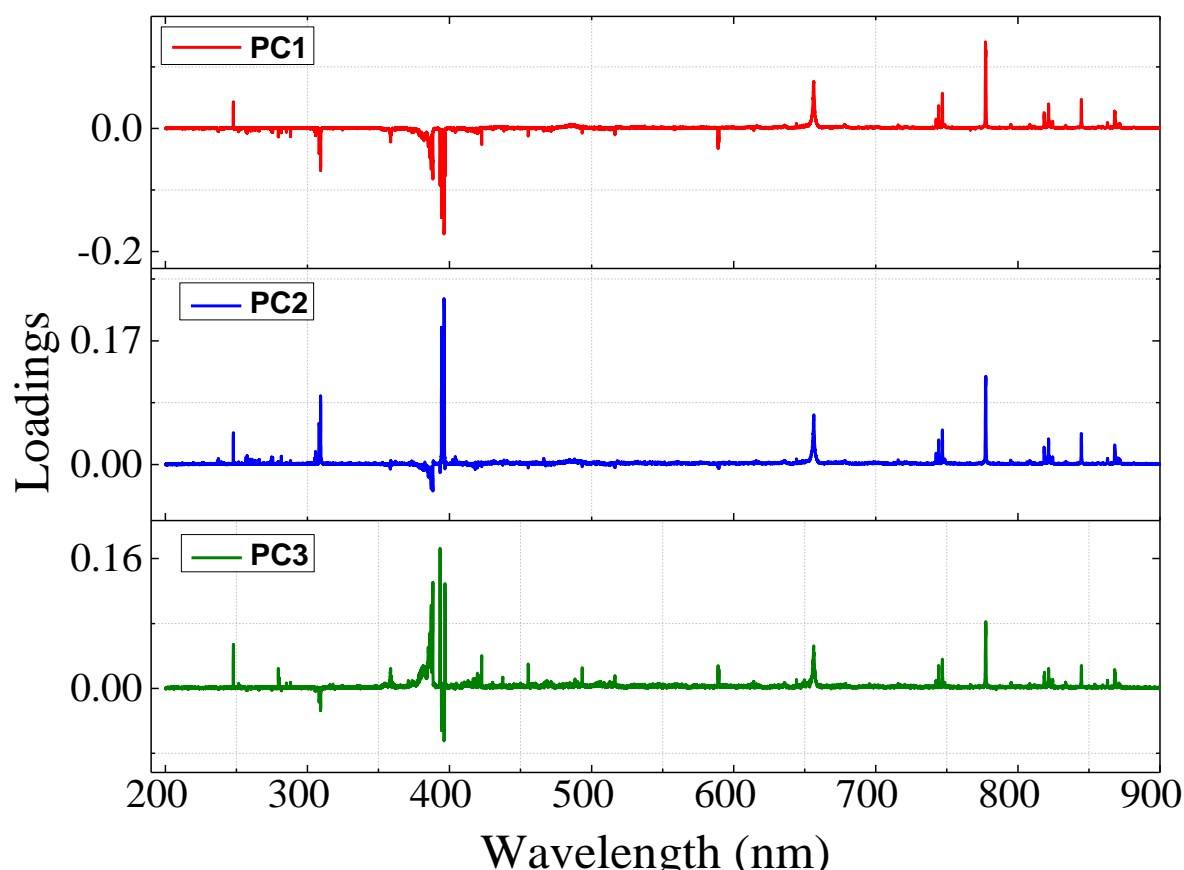


Figure 77. Loadings plot of Principal components: This figure explains the correlation of original data with that of principal components

correlation of C, H, N and O, negative correlation with peaks CN, Ca and Na. The second PC has correlation with all the peaks except CN. Most of the variance information of the data was not explained by starting three principal components.

As observed in sec 5.3, the signal from the elements such as C, H, N and O lasts upto 5 μ s and CN signal up to 30-40 μ s. Hence from this observation, utilization of the information of total features from 1 – 40 μ s may be a good choice for multivariate analysis. But it is observed that the contribution of signal after 20 μ s is almost at the noise

level. Hence, the choice of gate width of 1- 20 μs could be better choice. With this longer gate width and laser energy of 90 mJ, gate delay 1 μs , MCP gain 140 and accumulation of 2 pulses third set of LIBS data has been collected. 12 spectra for 1NP, 76 for 3NP and 106 for 4NP have been collected. 1 NP sample developed cracks and broken down into small pieces and hence only limited number of spectra could be recorded. Again PCA is applied to view the similarity in LIBS data. In spite of very less number of spectra for 1NP, a fine cluster can be observed in the score plot as shown [figure 78](#). The inter sample distance between the clusters has also increased when compared with previous score

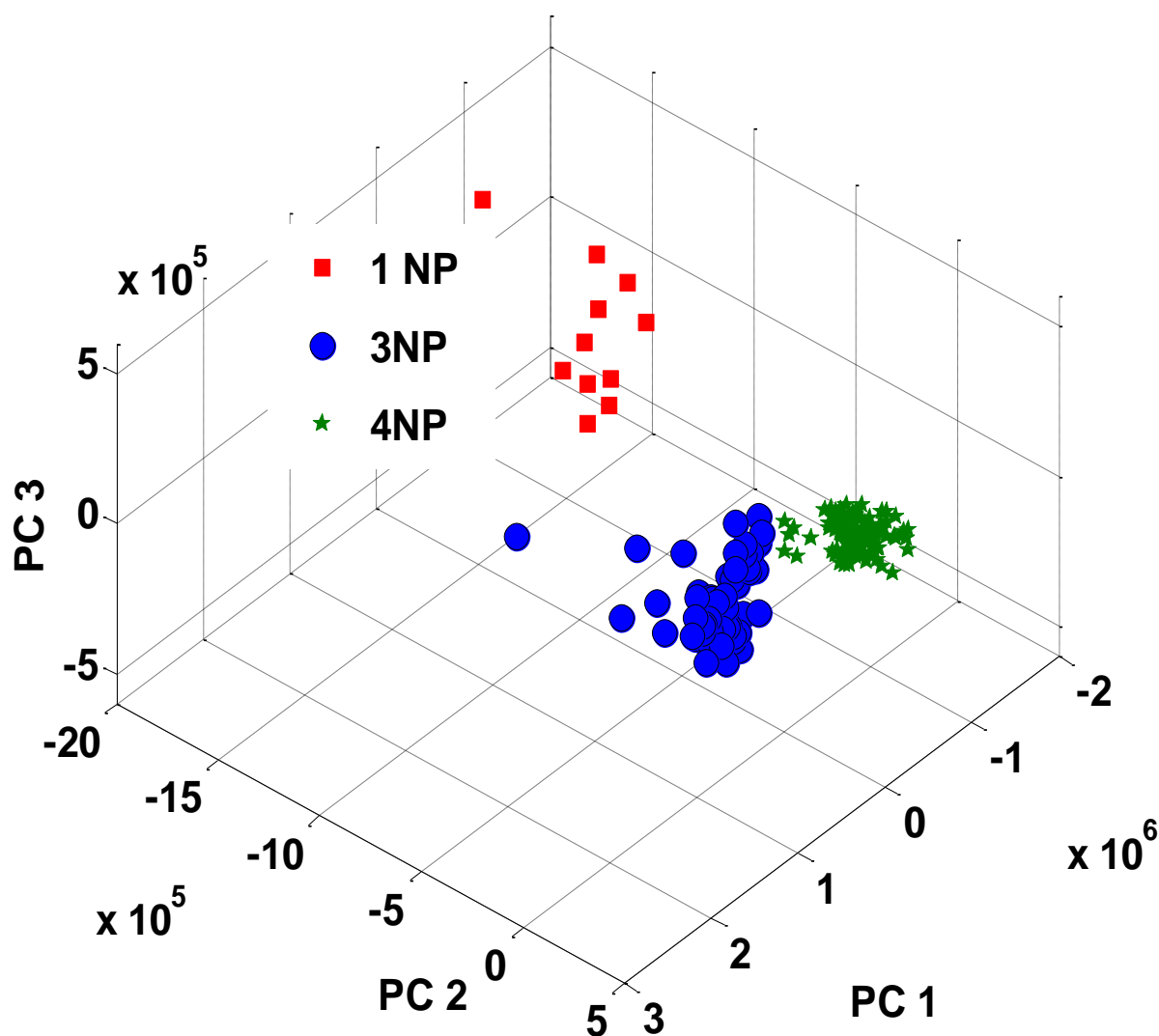


Figure 78. The score plot of isomer compounds using LIBS data obtained with 19 μs gate width. The variation between 3NP and 4NP has found to be increased and cluster of all three compounds are closely packed.

plots. The first 5 PCs explain most of the variance information associated with the data. Cumulatively these 5 PCs explain variance of 91.44 % of total LIBS data. The scree plot is shown [Figure 79](#).

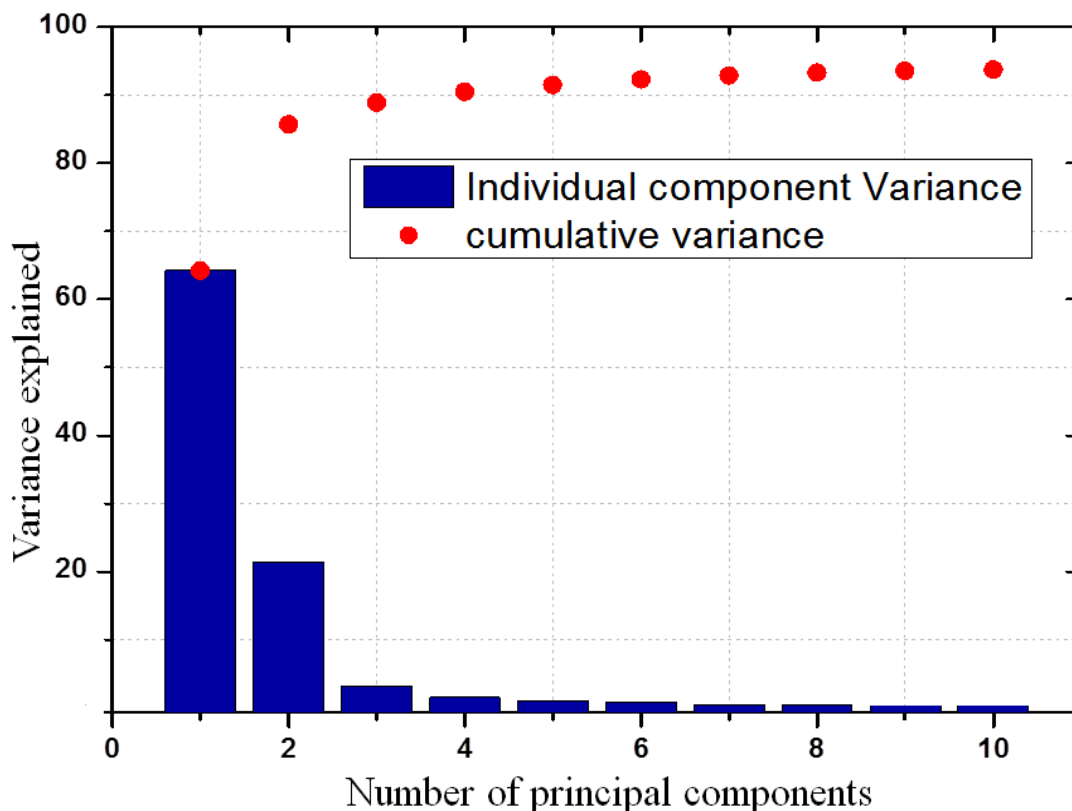


Figure 79: Scree plot of number of components versus associated variance. Bar plot shows the individual components versus corresponding variance associated with it, while the red color circles denotes cumulative variance

It is well known that PCA is useful to discover whether samples belong to same class or different. It helps to find out the variables responsible for discrimination, but it is not adequate for classification of unknown LIBS sample. Though PCA is a very good exploratory tool, it cannot perform the task of identifying an unknown spectrum. An actual application scenario demands the identification of a given unknown isomer by recording its LIBS spectrum. For a given application, the knowledge of the number of isomers is a prerequisite to enable the process of training and validation. The LIBS spectrum of the unknown isomer is recorded and tested with the model created by data analysis methods referred to as supervised methods.

Neural network is mostly adopted when the data set contains a large degree of nonlinearity. It understands the pattern, based on the training datasets and adjusts its

neurons when test data set is applied for testing. Neural networks are self learning classifiers, where user intervention is less. The method uses two layer feed forward network for classification model. Sigmoid functions were used as neurons for input and hidden layers. The final output consists of binary encoding used as target, where 1 stands for correct assignment and 0 for absent. The Network was trained with scaled conjugate gradient back propagation algorithm. The data division control was inbuilt in the Matlab toolbox. Out of the total data, 70% is presented to the network during training, and the network is adjusted according to its error. Another 15% is used to measure network generalization and to halt training, indicating an increase in the mean square error of the validation samples. This step has no effect on training and provides an independent measure of network performance during and after training.

To analyze the network response, the confusion matrix is examined by considering the outputs of the trained network and comparing them to the expected results. The results with all three sets of data are shown in Table 30. Identification accuracies calculated using ANN for the three sets of LIBS data are 78.39%, 87.79% and 84.41% respectively. The misclassification rate for 1NP is decreased from set 1 to set 2 and increased in the

	Set 1	Set 2	Set 3
1NP	75.22	89.16	55.42
3NP	90.26	86.42	98.97
4NP	69.69	87.79	98.85
Average	78.39	87.79	84.41

Table 30: Correct classification results using ANN for three sets of experiments

case of set 3. This is a direct consequence of very less number of spectra. However, for 3NP and 4NP there is significant increase in the classification rates. It is well known that all statistical classification methods require multiple number of datasets that capture all the possible variations. Hence with a longer gate delay and number of spectra of the order of ~100 the pyrazole isomers can be classified with very high classification rates.

The concept of identification of isomer compounds using a common handheld Czerny turner spectrograph can be of practical use in the industrial sectors where identification of isomers is a routine task. This can lead to reduction of overall cost and size of the instrumentation. As detailed in chapter 4, this spectrograph used in the experiment does not have provision of time gating. The signal is collected with minimum of 6 ms gate width. Hence, the LIBS spectrum not only contains the atomic and molecular information but also has a strong continuum. Hence the outcome resultant spectrum consists of sharp peaks riding over continuum background. However it has been shown in chapter 4 that a better performance of the identification can be achieved with a non-gated signal collection [24]. LIBS spectra of the pyrazole isomers collected with Maya 2000 spectrometer are shown in Figure 80. Elemental peaks of Hydrogen (656.6 nm),

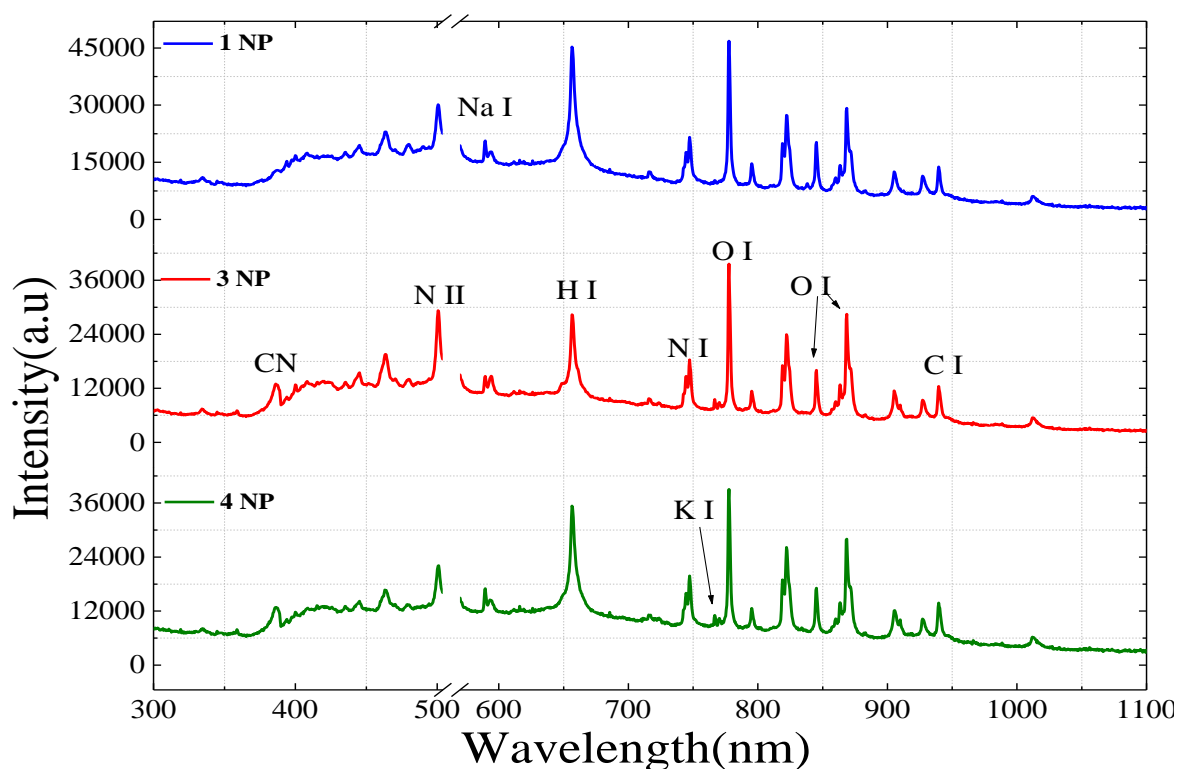


Figure 80: Non gated LIBS spectra of 1NP, 3NP and 4NP

unresolved triplets of oxygen (centered at 777.5, 794.8 and 844.4 nm), unresolved triplets of nitrogen (742.6-746.8 nm), carbon (940.0 nm), sodium (589.0 and 589.6 nm) and potassium (766.5 nm) have been observed. In addition to above atomic emissions, a broad CN molecular peak (384.0-389.0 nm) is also observed. Because of poor resolving power

of detector, most of the atomic triplets and molecular bands are not resolved explicitly. A notch filter was placed in front of the spectrometer collection optics to remove the scattered 532 nm laser radiation. Multiple spectra of 74, 54 and 80 for 1NP, 3NP and 4NP respectively were recorded to incorporate the fluctuations in the signal arising from different sources like laser energy fluctuation, sample in homogeneity and stochastic nature of the plasma.

Expectedly, the spectra acquired for all the three isomers are almost identical and gross visual inspection does not offer substantive clues to their identity. The entire dataset of 208 spectra, with each spectrum consisting of intensities at 1961 wavelengths, was then subjected to principal component analysis (PCA) [30, 31]. Figure 81 shows the first three PCs, which

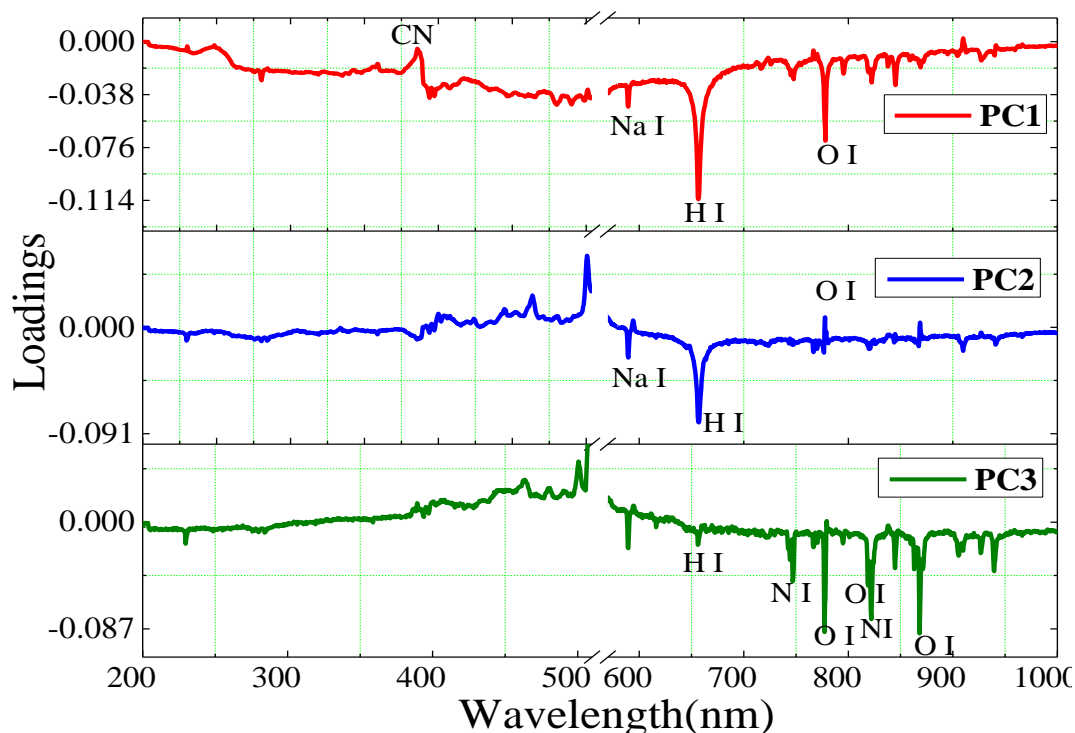


Figure 81. Plot of the first three principal component loading profiles corresponding to the dataset constituted by spectra recorded from the isomers 1NP, 3NP and 4 NP

explain 88.8% of the variance in the dataset and the first five PCs account for nearly 95% of the variance in the data. The first PC consists of the CN peak as well as the elemental peaks of hydrogen, oxygen and nitrogen. PC 2 spectral profile is dominated by the hydrogen peak whereas the principal features of PC 3 are the oxygen and nitrogen features. Although the PC loadings themselves are abstract mathematical entities, they

can be well correlated to the chemical content of the samples and, as such, form the basis for identifying the crucial differences between the isomer samples.

The corresponding scores plot for the 1st two PCs is shown in Figure 82. It is evident from this figure that there is sufficient elemental/molecular information in the spectra that can

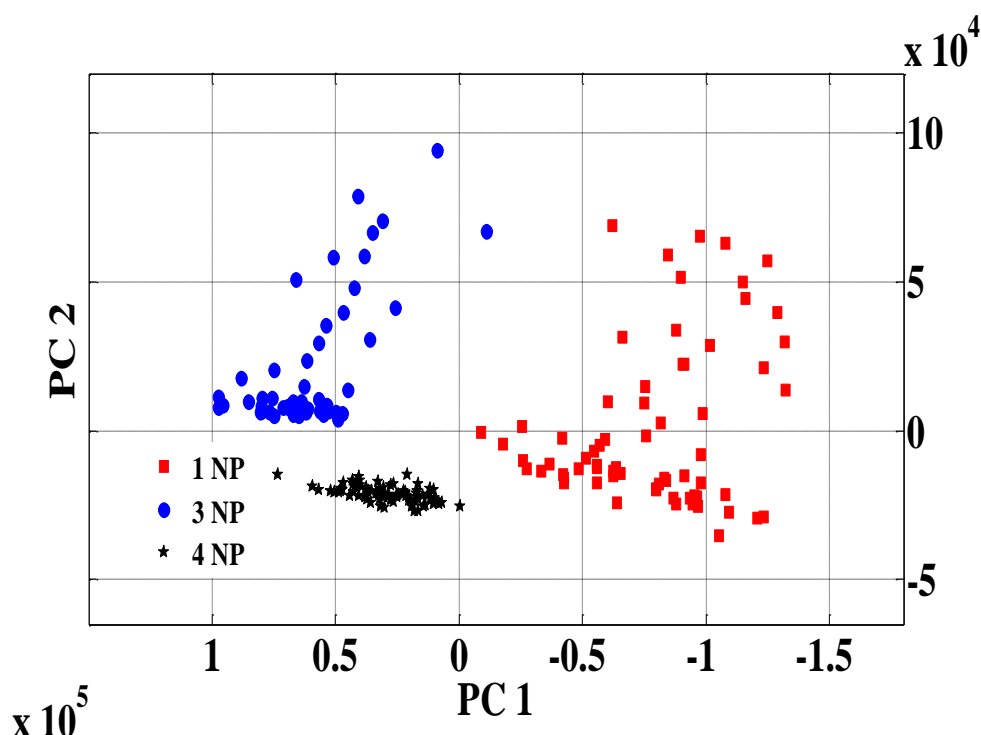


Figure 82. Clustering of 1NP, 3NP and 4NP samples based on principal component analysis

of the recorded LIBS spectra. Here the scores corresponding to PC 1 are plotted against those of PC 2 are plotted

be utilized for the purpose of differentiating the isomers, as each isomer forms a distinct cluster. The inter-class differences exceed the intra-class variability lending credence to the original hypothesis of LIBS based accurate isomer differentiation. The consistency of separation based only on scores of 2 PCs would also appear to preclude the possibility of differentiation relying on spurious factors such as instrumentation system drift. Further inspection reveals that the 1NP cluster possesses larger scatter than the clusters belonging to the other two isomers. Additionally, the 3NP and 4NP clusters are closer to each other in relation to the 1NP cluster. This proximity of the 3NP and 4NP clusters can be attributed to the similar attachment of NO₂ group to the pyrazole ring at its C atom in the 3NP and 4NP isomers. The classification results obtained by the application of ANN are

shown Table 31, which shows the resultant confusion matrix, where the diagonal elements denote the percentage of correctly identified samples for each isomer. The results represent the performance of neural network analysis averaged over 1000 iterations, where in each iteration the training and validation sets were randomly chosen. However, while 1NP gets misclassified almost equally as 3NP and 4NP, 3NP is principally misclassified as 4NP that is in agreement with the observation of relative proximity of the PC scores of 3NP and 4NP in Figure 82. Also, near perfect classification is achieved for 4NP, which is consistent with the compactness of the corresponding PC score cluster (Figure 82).

To confirm the veracity of these findings, a control study was performed where arbitrary labels were assigned to the spectral data regardless of their true identity. In these circumstances, the average correct rate of classification was found to be ca. 38.5% thereby reinforcing the minimal (or no) impact of spurious and chance correlations on the ANN-derived decision algorithm performance. Additionally, one of the frequently cited challenges in LIBS is the possibility of the classification algorithm being based on non-sample analyte specific variance, notably the differences in Na and K lines that may be present in the recorded spectra. To directly address this concern, we re-analyzed the data using the original protocol but eliminating spectral regions (589 – 595 nm and 765- 772 nm) surrounding the respective Na (589.0 and 589.6 nm) and K (766.5 and 769.9 nm) lines. We observed that the average classification accuracy in such a case is ca. 98%, which is within the error margin of the previously obtained values with the full complement of spectral data.

		Predicted Labels (%)		
		1NP	3NP	4NP
Reference Labels	1NP	96.12	1.61	2.27
	3NP	0.16	96.90	2.94
	4NP	0.70	0.30	99.00

Table 31. Confusion matrix for the three isomers using the Artificial Neural Network Algorithm.

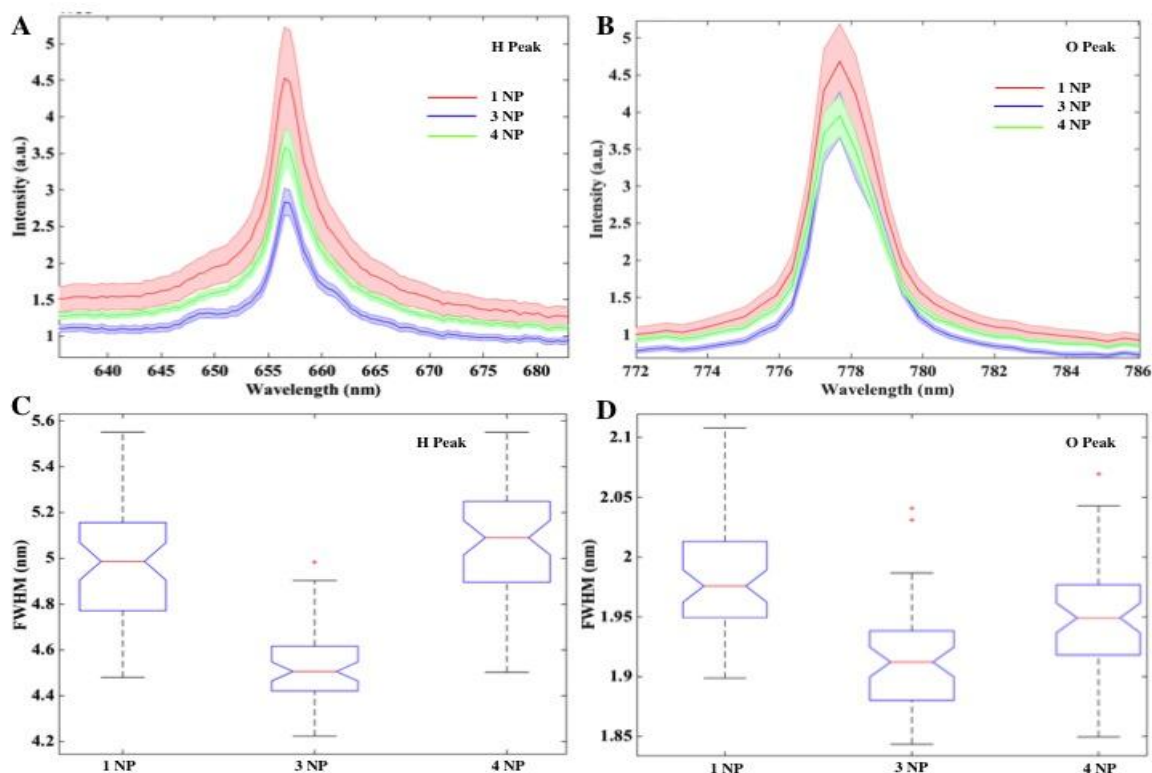


Figure 83. Plot of the spectral profiles of (A) hydrogen and (B) oxygen emission lines from the three classes of isomers, 1NP (red), 3NP (blue) and 4NP (green). The solid profile depicts the mean spectrum of each sample group and the shadow represents ± 1 S.D. Box plot of the FWHM of the corresponding Lorentzian fits of (C) hydrogen and (D) oxygen.

In addition, examined the causality of the classification performance by exploring the presence of differences (or the lack thereof) of the principal LIBS spectral markers. Inspection of the difference spectra between the different isomer classes reveals that the residuals show substantive structure, particularly centered on the H (657 nm) and O (778 nm) emission lines (Figure 83A)above. To further explore these variations, we fitted Lorentzian profiles on these peaks for each of the spectra and extracted the relevant parameters - the intensity, the area under the peak, center position and the width of the band measured by the full width at half-maximum (FWHM). While the first three parameters either show no change (e.g. the center wavelength) or could be attributed to extraneous factors (i.e. the intensity could be influenced by the subtle laser variations), the width of the Lorentzian peaks provides an unexplored context to the underlying spectral differences. Figure 83(B) shows a box-and- whisker plot of the FWHM of each of the spectral classes. For the H emission line, the differences between 1NP & 3NP and 3NP & 4NP were computed to be extremely statistically significant ($p < 0.0001$), whereas the differences between the 1NP & 4NP was almost significant ($p = 0.06$). For the O

feature at 778 nm, the differences between each pair of classes exceeded statistical significance thresholds, with the minimum p-value of 0.003 computed for 3NP & 4NP. Taken together, these results provide compelling evidence of differential peak broadening being one of the, if not the, primary driver(s) of the excellent classification performance seen here. These differences in peak FWHM can be attribute to Stark effect and local plasma temperature-induced variations between the different positional isomers, which in turn stems from the small (yet critical) variations in molecular structure. Coupled to this, we also expect that structure dependent molecular fragments and/or radicals (e.g. CN band) that were created during ablation may also play a role in enabling differentiation between the isomers. While differences due to such fragments have previously been discerned only in spectra recorded using femtosecond lasers, low concentrations of these radicals in nanosecond LIBS may present latent information that is exploited by the classifier building method (ANN).

5.5 Investigation of Plasma parameters

Three emission peaks of oxygen centered at 777.5, 794.8 and 844.4 nm are used to draw the Boltzmann plot. Boltzmann plot constructed for multiple spectra shown in Figure 84.

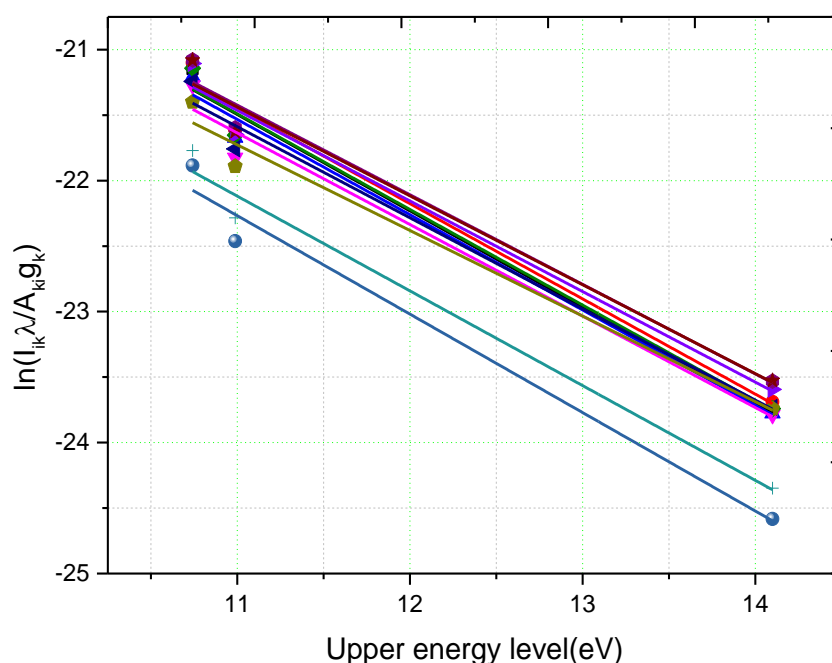


Figure 84: Electron temperature obtained using Boltzmann plot method The Electron temperature is calculated using oxygen peaks at 777.5, 794.8 and 844.4 nm

The mean temperatures for 1NP, 3NP and 4NP are estimated to be 16412, 16184 and 17517 K respectively. Figure 85 shows the statistics of plasma temperature. The relative standard deviations for the three samples are 4.1%, 8.4% and 4.6% respectively, indicating the fine reproducible data. In the case of 3NP outliers are present where as for 1NP and 4NP outliers are not significant.

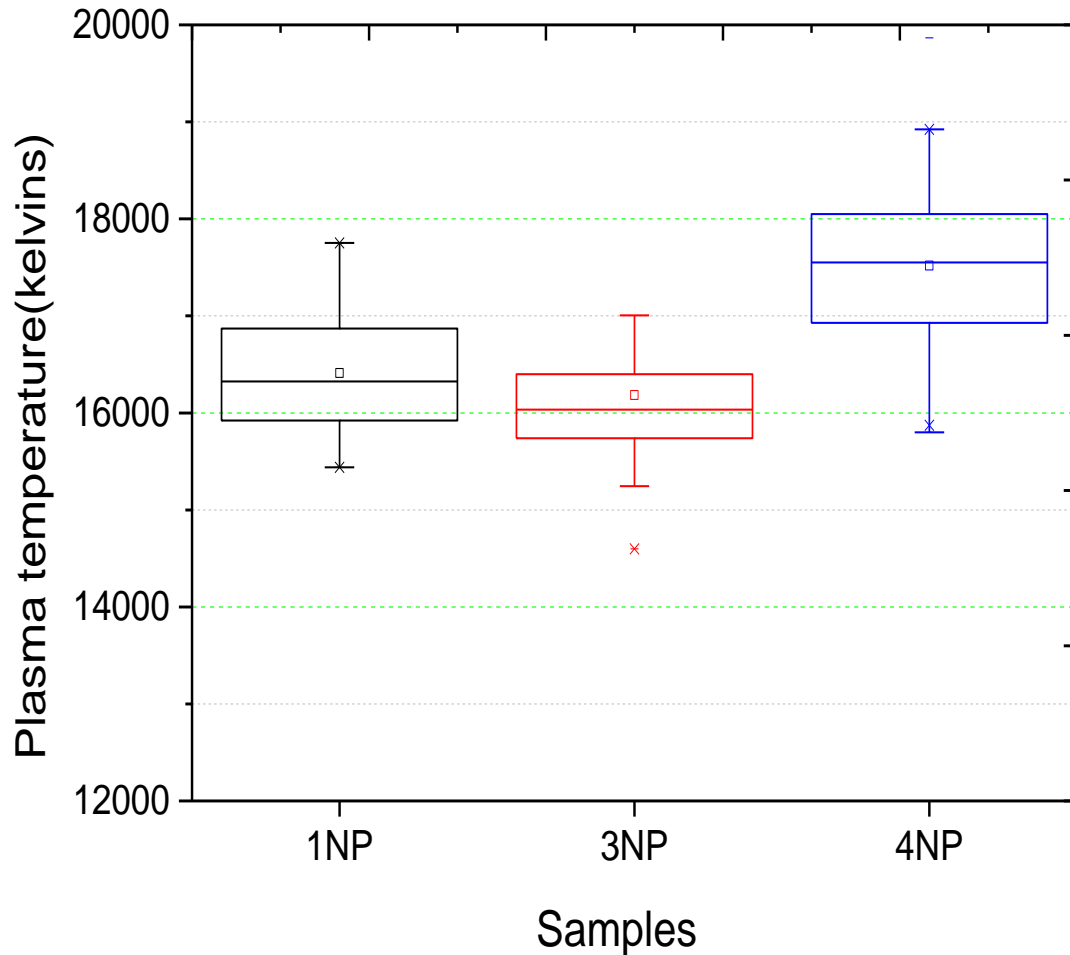


Figure 85. Box plot of plasma temperatures of 1NP, 3NP and 4NP. Small square box denotes mean and 'x' represents outliers.

Electron density is estimated using stark broadening of oxygen line at 844.4 nm. The equation pertaining to the calculation is given in chapter 1 section 1.4; it should be observed that other mechanisms are also prevailing but contribution to the broadening are negligible. The mean electron densities for 1NP, 3NP and 4NP are estimated to be 6.56×10^{15} , 5.31×10^{15} and $6.72 \times 10^{15} \text{ cm}^{-3}$ respectively. The statistics of the electron density is shown in the box plot in Figure 86.

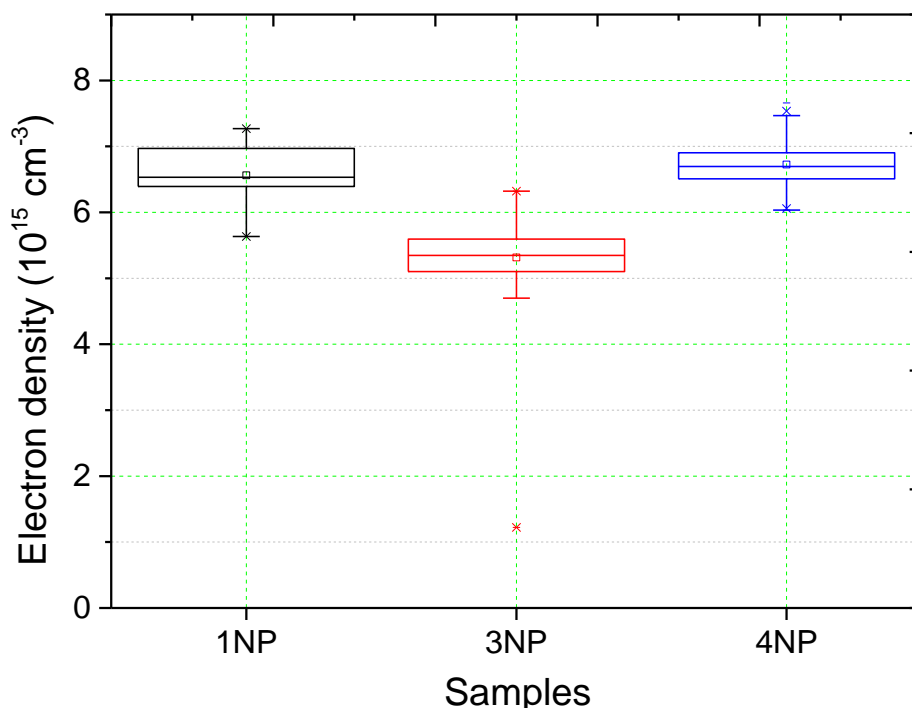


Figure 86. Box plot for electron density of 1NP, 3NP and 4NP.

The temporal evolution of plasma temperature and electron density is estimated by recording the spectra as a function of time delay after laser pulse interaction with the material. The spectra are collected by moving the collection gate in steps of 1 μs and gate width 1 μs . Following the same protocol the parameters have been estimated at each time delay. The time evolution of plasma temperature and electron density of 3P and 4P is shown Figure 87.

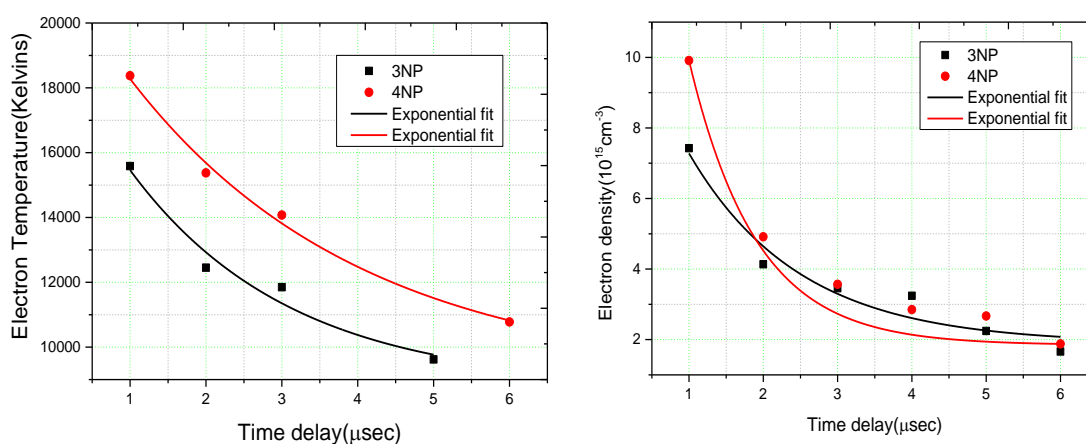


Figure 87. Temporal decay of electron temperature and electron density of 3NP and 4NP isomer compounds.

5.6 Conclusions

In summary, the accurate identification of isomers in the current study provides strong evidence to support recent evolution of LIBS as a tool that can elucidate features based on molecular structure in addition to elemental composition. Primitively, the optimization of gated data collection using analysis is shown. The feasibility of application using a non-gated, compact, inexpensive LIBS spectrometer on nitropyrazoles as a model system has been demonstrated. This promising pilot study paves the way for a new set of materials characterization applications where routine positional isomer identification in a label-free, real-time manner is urgently desired. The plasma parameters of electron density and temperature are estimated, including their temporal evolution.

5.7 References

1. Winefordner JD, Gornushkin IB, Correll T, Gibb E, Smith BW, et al. (2004) Comparing several atomic spectrometric methods to the super stars: special emphasis on laser induced breakdown spectrometry, LIBS, a future super star. *Journal of Analytical Atomic Spectrometry* 19: 1061-1083.
2. Meyerand RG, Haught AF (1963) Gas Breakdown at Optical Frequencies. *Physical Review Letters* 11: 401-403.
3. Radziemski LJ, cremers DA (1989) *Laser-induced plasmas and applications*: Marcel Dekker Inc.
4. Krishan V (2014) *Plasmas*: Cambridge University Press.
5. Bekefi G, Allis WP (1976) *Principles of Laser Plasmas*: Wiley.
6. Lucena P, Doña A, Tobaría LM, Laserna JJ (2011) New challenges and insights in the detection and spectral identification of organic explosives by laser induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* 66: 12-20.
7. De Lucia FC, Gottfried JL (2012) Classification of explosive residues on organic substrates using laser induced breakdown spectroscopy. *Applied Optics* 51: B83-B92.

8. St-Onge L, Kwong E, Sabsabi M, Vadas EB (2002) Quantitative analysis of pharmaceutical products by laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* 57: 1131-1140.
9. Balzer H, Hoehne M, Sturm V, Noll R (2005) Online coating thickness measurement and depth profiling of zinc coated sheet steel by laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* 60: 1172-1178.
10. De Lucia FC, Harmon RS, McNesby KL, Winkel RJ, Miziolek AW (2003) Laser-induced breakdown spectroscopy analysis of energetic materials. *Applied Optics* 42: 6148-6152.
11. Gottfried JL, De Lucia Jr FC, Munson CA, Miziolek AW (2007) Double-pulse standoff laser-induced breakdown spectroscopy for versatile hazardous materials detection. *Spectrochimica Acta Part B: Atomic Spectroscopy* 62: 1405-1411.
12. Gottfried JL, De Lucia JFC, Munson CA, Miziolek AW (2008) Strategies for residue explosives detection using laser-induced breakdown spectroscopy. *Journal of Analytical Atomic Spectrometry* 23: 205-216.
13. De Lucia JFC, Gottfried JL, Munson CA, Miziolek AW (2008) Multivariate analysis of standoff laser-induced breakdown spectroscopy spectra for classification of explosive-containing residues. *Applied Optics* 47: G112-G121.
14. Bohling C, Scheel D, Hohmann K, Schade W, Reuter M, et al. (2006) Fiber-optic laser sensor for mine detection and verification. *Applied Optics* 45: 3817-3825.
15. Anzano J, Lasheras R-J, Bonilla B, Casas J (2008) Classification of polymers by determining of C1:C2:CN:H:N:O ratios by laser-induced plasma spectroscopy (LIPS). *Polymer Testing* 27: 705-710.
16. Sattmann R, Monch I, Krause H, Noll R, Couris S, et al. (1998) Laser-Induced Breakdown Spectroscopy for Polymer Identification. *Applied Spectroscopy* 52: 456-461.
17. Anzano JM, Gornushkin IB, Smith BW, Winefordner JD (2000) Laser-induced plasma spectroscopy for plastic identification. *Polymer Engineering & Science* 40: 2423-2429.

18. Boueri M, Motto-Ros V, Lei W-Q, Qain L, Zheng L-J, et al. (2011) Identification of Polymer Materials Using Laser-Induced Breakdown Spectroscopy Combined with Artificial Neural Networks. *Applied Spectroscopy* 65: 307-314.
19. Baudalet M, Guyon L, Yu J, Wolf J-P, Amodeo T, et al. (2006) Femtosecond time-resolved laser-induced breakdown spectroscopy for detection and identification of bacteria: A comparison to the nanosecond regime. *Journal of Applied Physics* 99: -.
20. Rehse SJ, Diedrich J, Palchaudhuri S (2007) Identification and discrimination of *Pseudomonas aeruginosa* bacteria grown in blood and bile by laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* 62: 1169-1176.
21. Multari RA, Cremers DA, Dupre JM, Gustafson JE (2010) The Use of Laser-Induced Breakdown Spectroscopy for Distinguishing Between Bacterial Pathogen Species and Strains. *Applied Spectroscopy* 64: 750-759.
22. Samuels AC, DeLucia FC, McNesby KL, Miziolek AW (2003) Laser-induced breakdown spectroscopy of bacterial spores, molds, pollens, and protein: initial studies of discrimination potential. *Applied Optics* 42: 6205-6209.
23. Kumar A, Yueh F-Y, Singh JP, Burgess S (2004) Characterization of malignant tissue cells by laser-induced breakdown spectroscopy. *Applied Optics* 43: 5399-5403.
24. Myakalwar AK, Dingari NC, Dasari RR, Barman I, Gundawar MK (2014) Non-Gated Laser Induced Breakdown Spectroscopy Provides a Powerful Segmentation Tool on Concomitant Treatment of Characteristic and Continuum Emission. *PLoS ONE* 9: e103546.
25. Larina L, Lopyrev V (2009) Nitroazoles: Synthesis, Structure and Applications: Synthesis, Structure and Applications: Springer.
26. Ulery AL, Drees LR, America SSSo (2008) Methods of Soil Analysis: Mineralogical methods: Soil Science Society of America.

6. CONCLUSION AND FUTURE SCOPE

The research work has focused on the development of conceptualized LIBS field device and suitable algorithms for identification of organic compounds. It involves instrument interface, extensive data analysis using the huge data bank of LIBS, plasma physics, atomic and molecular spectroscopy. The main drawings of the work been carried out are twofold –one to understand the role of the laser pulse in determining the properties of the plasma and second is to use the spectral information of the laser created plasma for analytical purposes.

6.1 Conclusion

As part of analytical spectroscopy interest, LIBS of organic compounds, particularly pharmaceutical drugs, high energy materials and isomers are utilized in the thesis work. Organic compounds mostly contain C, H, N and O. Hence, LIBS spectra of corresponding compounds look similar. Avoiding spectral interference from surrounding atmosphere is a major problem, when the experiments performed in the ambient air. However, use of argon gas purging or double pulse LIBS could solve the problem of interference. But for the field detection applications, the above methods are impractical. Various approaches like ratiometrics and chemometrics are being explored for the data analysis to accomplish this task. Ratiometrics has been explored to analyze analytical formula of given organic compound. However, ratiometrics did not yield a full proof method of classification of similar looking materials. To overcome this difficulty, multivariate analysis based approaches have been developed. Various methods like PCA, SIMCA, PLS-DA, SVM and ANN are explored for this purpose.

Using time resolved data of LIBS of HEMs, identification problem has been addressed considering different features of the spectra. Along with full spectrum analysis, sub spectrum and specific peaks were chosen to accomplish discrimination of explosive/non-explosive and as well as specific explosive identification. The results of this work, particularly the region based outcome suggests small length of spectral range spectrometer for onboard chip based devices where storage of data is the main concern.

The specific peak concept opens a path towards possibility of use of filter based detectors for specific applications such as field detection of HEMs.

Furthermore concrete steps have been taken forward for finding solutions towards identification of materials using a small, compact, handheld spectrometer. With aid of advanced machine learning method neural network, perfect classification has been demonstrated. Neural networks are self learning classifiers where user intervention is very less, compared to conventional approaches. All the nongated results were explored using the data of HEMs and pharmaceuticals. Most important findings of this work is continuum can be a good source of information, which is overlooked by LIBS community due to elemental feature requirements. The real time data can be used directly without using preprocessing tools, which mimics finding the solutions of real world problems of identification of materials in pharmaceutical industries, explosive identification and also for planetary exploration.

Finally the work is taken towards identification of isomers using both experimental approaches i.e., gated and nongated detectors. Experiments are optimized to get reasonable accuracy for prediction of the compounds. It has been proven the possibility of isomer compounds, whose spectra is over similar by visual inspection. The outcome of this work can be used for drug discovery application as a routine step in pharmaceutical industry.

As a conclusion following are the major achievements of research.

Identification and classification of organic compounds is achieved by multivariate analysis tools instead of calculating ratios. This has a massive application in process analytical technology for pharmaceutical compounds and on-field detection of explosives. [Talanta, 2011]

Bremsstrahlung dominated spectra can be a good spectral candidate for the classification of similarly looking organic compounds. This finding recommends the use of, a very low compact, rugged and small size CCD based spectrometer, instead of ICCD based detector, which is bulkier, heavy and required skilled manpower. (Innovation; usually considered useless in measurements) [PLOS ONE, 2014]

Proposed important spectral regions in laser induced plasma spectra with prior spectral knowledge of data and wavelengths selection using genetic algorithm. This will serve the solution of identification of organic compounds in ambient atmosphere and the development of discrete filter-based detector units (Innovation) [Scientific Reports, 2015]

Identification of structural isomer compounds using LIBS data with aid of chemometrics [Innovation; under review, 2015]

6.2 Future Scope

Experiments can be conducted with explosives in trace quantities on substrates mimicking real time scenario. All the experiments reported in this thesis were performed on pellets. This strategy is quite well suited for the pharmaceuticals and isomer applications or any other real time application like inline monitoring in a factory environment. However, an application scenario involving high energy materials, also, demands the identification of trace quantities of material. Hence, it is of great practical importance to develop the methods of identification where the trace amounts of samples are involved. The experimental strategy remains the same with the bulk samples replaced by the trace amounts of samples. Further, the concept of LIBS system based on handheld CCD spectrometer can be combined with the above experiments as any application invariably demands a cheaper and easy to use alternative. The use of CCD based handheld spectrometer greatly reduces the cost and at the same time is low on resources. Additionally, a remote/standoff LIBS system is a natural step to take the laboratory results into the field applications. The trace amount based experiment with the handheld spectrometer can be ideal choice for the standoff applications.

Exploring the possibility of multiplexing Raman spectroscopy with LIBS will be an additional milestone in this research. Laser-induced breakdown spectroscopy (LIBS) combined with Raman spectroscopy can offer several advantages over the individual techniques towards the identification of High Energy Materials. The spectral information obtained from these techniques is complimentary as LIBS gives information about the elemental composition and Raman spectroscopy enables the molecular information. Also Both the techniques have some similarities in instrumentation. A combined spectral information from both the techniques can obtain a better identification performance. The advantages of a combined technique include -

1. Elemental mapping of the sample
2. Molecular mapping of the sample
3. Impurity detection
4. Atomic + molecular finger print may be serve as a better input for identification.

Also depending on the need of the application, this instrument can work solely as Raman set up or solely as a LIBS set up OR a hybrid that combines the advantages of these techniques. While LIBS gives atomic information of the sample, Raman spectroscopy gives the molecular information. While LIBS techniques gives more light signal compared to Raman, it is less precise. However, LIBS based classification has been very successfully implemented by the aid of chemometrics. Combining the advantages offered by these techniques could lead to a better detection of materials as this is based on complimentary information about the sample.. One of the possibilities is to identify the elemental constituents with LIBS and use the molecular information given by Raman for material identification. The elemental information can also throw useful insights into the nature of impurities and the possible matrix. In the next step, it may be possible to adopt data fusion and create a unique finger print of the material that has both atomic and molecular features. The combined finger print may be a better candidate for classification of materials.

The isomer detection with LIBS has great potential and can be validated with more sets of isomer compounds and can also extended to optical isomers as well. Furthermore, a large number of unexplored data analysis methods which can possibly lead to a design of a more robust LIBS system can also be examined. The data analysis plays a crucial role in the identification problem. In spite of the spectra of organic compounds being very similar and for isomers far more over whelming identical, a clear separation and identification is possible only with the application of suitable analysis protocol. The analysis methods adopted in this thesis are very established and used routinely in many areas of applications like face recognition, bio-metrics or Raman/FTIR spectroscopy. However, the analysis of LIBS data with these methods is still very limited in terms of which algorithms are used. There is great potential for the collaboration of

LIBS scientists with the scientists working in the field of pattern recognition. This partnership can unravel many more beautiful applications of LIBS in the near future.

7. PUBLICATIONS

PUBLICATIONS RELATED TO THESIS

1. **Myakalwar, A.K.**, Sreedhar, S., Barman, I., Dingari, N.C., Rao, S.V., Kiran, P.P., et al. Laser-induced breakdown spectroscopy-based investigation and classification of pharmaceutical tablets using multivariate chemometric analysis. *Talanta*. 2011, 87, 53-9.
2. Dingari, N.C., Barman, I., **Myakalwar, A.K.**, Tewari, S.P., Kumar Gundawar, M. Incorporation of support vector machines in the LIBS toolbox for sensitive and robust classification amidst unexpected sample and system variability. *Analytical chemistry*. 2012, 84, 2686-94.
3. **Myakalwar, A.K.**, Dingari, N.C., Dasari, R.R., Barman, I., Gundawar, M.K. Non-Gated Laser Induced Breakdown Spectroscopy Provides a Powerful Segmentation Tool on Concomitant Treatment of Characteristic and Continuum Emission. 2014.
4. **Myakalwar, A.K.**, Spegazzini, N., Zhang, C., Anubham, S.K., Dasari, R.R., Barman, I., et al. Less is more: Avoiding the LIBS dimensionality curse through judicious feature selection for explosive detection. *Scientific reports*. 2015, 5.

OTHER PUBLICATIONS

5. Rao, S.V., Sreedhar, S., **Myakalwar, A.K.**, Kiran, P.P., Tewari, S.P., Kumara, G.M. Laser induced breakdown spectroscopy of high energy materials using nanosecond, picosecond, and femtosecond pulses: challenges and opportunities. *Proc. of SPIE Vol. 2010*, 8173, 81731A-1.
6. Sreedhar, S., Kumar, **Myakalwar, A.K.**, Kumar, G.M., Kiran, P.P., Tewari, S.P., Rao, S.V. Laser Induced Breakdown Spectroscopy of RDX and HMX with nanosecond, picosecond, and femtosecond pulses. *Proc. of SPIE Vol. 2010*, 7665, 76650T-1.
7. Sunku, S., **Myakalwar, A.K.**, Gundawar, M.K., Paturi, P.K., Tewari, S.P., Soma, V.R. Stoichiometric Analysis of Inorganic Compounds Using Laser-Induced Breakdown Spectroscopy with Gated and Nongated Spectrometers. *ISRN Optics*. 2012, 2012.
8. Sunku, S., Gundawar, M.K., **Myakalwar, A.K.**, Kiran, P.P., Tewari, S.P., Rao, S.V. Femtosecond and nanosecond laser induced breakdown spectroscopic studies of NTO, HMX, and RDX. *Spectrochimica Acta Part B: Atomic Spectroscopy*. 2013, 79, 31-8.

AWARDS

1. Best poster in ISAMP topical conference on Interaction of lasers on atoms, molecules & clusters at University of Hyderabad, Hyderabad, India, 2012
2. Senior Research Fellowship by Council of Scientific & Industrial Research, New Delhi, India, 2013

Laser Induced Breakdown Spectroscopy Studies with Pharmaceuticals , High Energy Materials and Isomers Towards Unfolding The Device Requirements

ORIGINALITY REPORT

32*

%

SIMILARITY INDEX

5%

INTERNET SOURCES

29%

PUBLICATIONS

4%

STUDENT PAPERS

PRIMARY SOURCES

- 1

Myakalwar, Ashwin Kumar, Narahara Chari Dingari, Ramachandra Rao Dasari, Ishan Barman, and Manoj Kumar Gundawar. "Non-Gated Laser Induced Breakdown Spectroscopy Provides a Powerful Segmentation Tool on Concomitant Treatment of Characteristic and Continuum Emission", PLoS ONE, 2014.

10%

Publication
- 2

Myakalwar, A.K.. "Laser-induced breakdown spectroscopy-based investigation and classification of pharmaceutical tablets using multivariate chemometric analysis", Talanta, 20111215

7%

Publication
- 3

Kumar Myakalwar, Ashwin, Nicolas Spegazzini, Chi Zhang, Siva Kumar Anubham, Ramachandra R. Dasari, Ishan Barman, and Manoj Kumar Gundawar. "Less is more: Avoiding the LIBS dimensionality curse through judicious feature selection for

7%

*Sources 1-3 are related to thesis work published by Ashwin Kumar Myakalwart as first author.

4

Submitted to University of Hyderabad,
Hyderabad

Student Paper

3%

5

Dingari, Narahara Chari, Ishan Barman,
Ashwin Kumar Myakalwar, Surya P. Tewari,
and Manoj Kumar Gundawar. "Incorporation
of Support Vector Machines in the LIBS
Toolbox for Sensitive and Robust
Classification Amidst Unexpected Sample
and System Variability", Analytical Chemistry,
2012.

Publication

1%

6

Springer Series in Optical Sciences, 2014.

Publication

<1%

7

www.science.gov

Internet Source

<1%

8

Schröder, Susanne. "Laser-induced
breakdown spectroscopy for planetary
research: Analysis of salts and frozen salt
solutions under Martian conditions",
Technische Universität Berlin, 2013.

Publication

<1%

9

Marcos-Martinez, D.. "Identification and
discrimination of bacterial strains by laser
induced breakdown spectroscopy and neural
networks", Talanta, 20110515

Publication

<1%