UNIVERSITY OF NAIROBI

DEVELOPMENT AND VALIDATION OF A LASER INDUCED BREAKDOWN SPECTROSCOPY METHOD FOR CANCER DETECTION AND CHARACTERIZATION IN TISSUE VIA MULTIVARIATE CHEMOMETRICS

BY

EMILY AKINYI OTIENO

BEd (Science)

Thesis Submitted for Examination in partial fulfillment of the Requirements for the Award of Master of Science Degree in Physics, University of Nairobi, Kenya

2018

DECLARATION

This thesis is my own original work and has not been submitted by anyone else for examination for the award of a degree or publication. Other people's work that have been used are appropriately referenced in conformity with the University of Nairobi's guidelines.

Signature…….Date……

Emily Akinyi

I56/83934/2012

We approve the submission of this thesis for examination as the research supervisors.

	Signature	Date
Dr. Angeyo H. Kalambuka	……	……..
Department of Physics		
University of Nairobi		
Dr. Dehayem A. Massop	…….	……
Department of Physics		
University of Nairobi		

DEDICATION

This work is a dedication to Semaya, Mum Florence and Dad George for continually encouraging me throughout the research.

ACKNOWLEDGEMENTS

I acknowledge my supervisors Dr. Hudson Angeyo and Dr. Alix Dehayem-Massop for being my mentors besides being encouraging and insistent on quality work; many thanks to Dr. Alix for immense assistance with sample preparation and analysis. I also thank the Institute of Nuclear Science and Technology laboratory for facilitating the equipment I used to prepare simulate samples, not forgetting the department of Human Anatomy laboratory for allowing me to process the breast, liver, and abdominal tissue samples from there.

A special tribute to Dr. Wambani of Kenyatta National Hospital (KNH) for availing the samples used for this study. My particular thanks to Githaiga, for availing the Hep-2 and Lewis lung cell lines for this study.

Finally, I express gratitude to International Science Program (ISP) of Uppsala University, Sweden under the Thematic Head Dr. Kaduki for awarding me the scholarship to undertake this Master of Science degree.

ABSTRACT

Early and accurate diagnosis of cancer is important for proper management and treatment of the disease. The conventional techniques used for cancer diagnosis are expensive, require specialized training and do not carry out prognosis effectively. This necessitates developing techniques that are rapid, direct, affordable and accurate for cancer detection especially in the early stages.

Biometal analysis in tissue for the purpose of cancer prognosis and diagnosis is important since the trace metals can be utilized as disease biomarkers. LIBS analysis uses pulsed laser for ablation (simultaneous atomization and excitation). Laser Induced Breakdown Spectroscopy (LIBS) has been used to obtain spectral data from the samples under study (simulate, breast, liver, abdominal tissues and cultured cell lines) and multivariate chemometric tools applied for data preprocessing towards quantification of trace elements in human tissues and cancer cell lines. The samples were obtained from Kenyatta National Hospital (KNH) then processed and fixed in paraffin wax to make 2 cm thick blocks. These blocks were sliced to 2 cm thickness, weight of 2 g ready for study.

Simulate tissue samples prepared by embedding known concentrations of Fe, Cu, Zn, Mn and Mg on molten paraffin wax, were used to create a multivariate calibration model by exploiting Artificial Neural Network (ANN) for predicting concentrations of the above named trace elements in body tissues (R^2 value > 0.95). Multivariate chemometric techniques (PCA, ANN and SVM) were used to achieve prognosis and diagnosis of cancer using modeled LIBS spectral data, trace biometal concentrations and multivariate alteration of the biometals (Cu, Mg, Mn, Fe and Zn). These metals were chosen based on the frequent

occurrences of these elements in the tissues. The method was used to identify the trace biomarkers in the tissues. The concentration ranges for the tissues obtained are Fe (51.2 - 137.2 µg/g), Cu (5 - 18.7 µg/g), Zn (36 - 56.8 µg/g), Mg (78.2 - 507.4 µg/g) and Mn (8.8 - 19.5 µg/g) for liver tissue. Beast tissue had Fe (87.7 - 113.9 µg/g), Cu (10.9 - 12.3 µg/g), Zn (49.3 µg/g to 55.7 µg/g), Mg (194.3- 242.3 µg/g) and Mn (14.5 µg/g - 16.1 µg/g). Abdominal tissue had Fe (96.7- 125.7 µg/g) Cu (6.7- 7.5 µg/g) Zn (88.3 - 93.9 µg/g) Mg (467.5 - 583.1 µg/g) Mn (9.5 -10.5 µg/g).

PCA was employed for pattern recognition as it grouped the human tissue samples with respect to the part of the body from which it was obtained based on trace biometal signatures. Besides, it also characterized them in terms of malignant and benign cancer staging. Support Vector Machine (SVM) was used to develop a classification model using simulate samples.

The developed method is rapid and suitable for early diagnosis of cancer and thus can be applied for proper cancer management. The whole process of acquiring data and analyzing to give results takes about 15 min as compared to the other methods, which take approximately 1 hour. This makes the methodology viable for spectral diagnostics of cancer in human tissue including at its early stages.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS, ACRONYMS AND SYMBOLS

AES- Atomic Emission Spectroscopy

ANN- Artificial Neural Network

CCD - Charge Coupled Device

CF-LIBS - Calibration Free Laser Induced Breakdown Spectroscopy

CT - Computed Tomography

DMEM - Dulbecco's Modified Eagle Medium

DNA - Deoxyribonucleic Acid

EDXRF- Energy Dispersive X-Ray Fluorescence

eV- Electron Volt

FBS- Fetal Bovine Serum

FWHM- Full Width at Half Maximum

GFAAS - Graphite Furnace Atomic Absorption Spectrometry

HBRA - High Background Radiation Areas

HEp-2 - Human Epithelial Type 2

ICP-AES- Inductively Coupled Plasma Atomic Emission Spectroscopy

ICP-MS- Inductively Coupled Plasma Mass Spectroscopy

IR- Infra Red

ISP- International Science Program

KEMRI- Kenya Medical Research Institute

KNH- Kenyatta National Hospital

LIBS - Laser Induced Breakdown Spectroscopy

LL- Lewis Lung Carcinoma

LOD- Limit of Detection

MSE- Mean Square Error

MRI- Magnetic Resonance Imaging

Nd-YAG - Neodymium-Doped Yttrium Aluminium Garnet

NIST- National Institute of Standards and Technology

LIBS- Laser Induced Breakdown Spectroscopy

PCA- Principal Component Analysis

PLS- Partial Least Squares

SEP- Standard Error of Prediction

SICA - Spatial Independent Component Analysis

SVM- Support Vector Machine

UV- Ultra Violet

VIS- Visible

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND INFORMATION

Laser induced breakdown spectroscopy (LIBS) is an atomic emission spectroscopy method that is fast, minimally invasive and non-destructive. It is used to analyze the elemental composition of solids, liquids or gases (Miziolek *et al.*, 2006). This method uses a pulsed laser to excite the elemental constituents of the sample under study leading to the production of a micro-plasma. The ablation process generates free atomic species (neutrals and ions). The atoms get excited and emit radiation of given intensity that is detected by a set of spectrometers. The spectral information obtained is used for qualitative and quantitative analysis of the sample under study. The ablation process produces spectra showing different intensity values for the various elements in the sample under study. These intensity values are used for calibration by relating them to concentration.

Unlike other spectroscopic techniques such as Inductively Coupled Plasma Atomic Emission Spectroscopy (ICP-AES), Inductively Coupled Plasma Mass Spectroscopy, ICP-MS and Graphite Furnace Atomic Absorption Spectrometry (GFAAS), the principal advantages of LIBS over them include its simplicity, rapid and direct analysis (Lee et al., 2004).

LIBS has vast applications in various fields. These areas include art (Savastenko and Tarasenko, 2011) (Colao *et al.*, 2002; Melessanaki *et al.*, 2002), space exploration (Colao *et al.*,2004), geomaterials (Mukhono, 2012; Harmon *et al.*, 2009), detection of hazardous materials (Harmon et., al, 2013), (Gottfried *et al.*, 2009; DeLucia *et al.*, 2005) and study of cancer tissues (El-Hussein *et al.*, 2010; Kumar *et al.*, 2004).

In the biomedical field, LIBS has been used for disease prognosis and diagnosis. LIBS offers a simple and real-time technique for disease diagnosis compared to other diagnostic techniques basing on its ability to detect and profile trace elements which are disease biomarkers.

Diseases change the biochemical composition of body tissues and fluids in some distinctive ways particularly changes in concentrations and structure of proteins, carbohydrates, and lipids. These changes when detected, profiled and characterized hold the key towards developing diagnostic techniques for those diseases, including the degree of their severity even before the commonly used histapathological analysis is able to reveal them. In order for the human body to maintain normal physiological functions for growth and development, the body requires trace elements.

The human body needs trace elements, in as little quantities as micrograms per gram, as essential components of biological enzyme systems or of structural portions of biologically active enzyme constituents. Examples of essential trace elements include Fe, I, Cu, Mn, Zn, Co, Cr, Se, Mo, Mg, V, Si and Ni. For this research, Cu, Zn, Mn, Mg, and Fe are under study. (Tehrani *et al.*, 2007).

The physical and chemical properties of the sample can affect the LIBS plasma composition, a phenomenon known as matrix effect (Mohamed, 2007). The dependence of intensity of spectral lines on laser pulse energy is affected by the nature of the matrix such as thermo-physical/ chemical matrix composition, surface reflectivity, conductivity and constituents' melting and boiling points of the sample (Mukhono, 2012).

Detection of trace elements in biological specimens depends essentially on the specimen (blood, urine, tissue, hair or nails), the method for preparation and the detection limit of the trace element of interest (Cremers and Radziemski, 2006). Detection limits of $1\ \mu g/g$ to greater than $100\ \mu g/g$ by mass are common for LIBS (Sarkar, 2010).

The spectral lines obtained from LIBS suffer from spectral overlaps, matrix effects and self-absorption that adversely affect the trace element biomarkers of cancer in the tissues lowering the detection limit. LIBS combined with multivariate chemometric tools such as ANN, SVM and PCA ensures that only the important information is captured from the LIBS spectra. In this research, LIBS with chemometric techniques has been exploited towards early detection and characterization of cancer in breast, liver and abdominal tissue based on trace elements.

Scope and Limitations of study

Study hypothesis

1.2 Statement of the Problem

Early cancer diagnosis is a challenging area in the management of the disease. Most of the patients are diagnosed at the late stage of development of the disease. The current methods of cancer diagnosis are limited by inadequacy in early, direct and rapid detection. This necessitates attempts to improve on the diagnostics methods by developing rapid and non-invasive techniques. Detection of cancer at an early stage using LIBS based on the concentration, multivariate alterations and correlation of trace elements is feasible but is yet to be fully explored.

1.3 Objectives

1.3.1 **Main** Objective

To develop a LIBS method for rapid and non-invasive detection, quantification and characterization of cancer in human body tissue based on the concentration and alterations of the following trace elements: Cu, Mn, Mg, Zn and Fe.

1.3.2 Specific Objectives

i. To qualitatively identify the trace element biomarkers in liver, breast and abdominal human cancerous tissues using LIBS.

ii. To design and validate a multivariate calibration model for determining concentration of the trace elements in the above cancerous tissues using ANNs.

iii. To utilize the ANN calibration model developed above in the prediction of concentration values of the trace elements in liver, abdomen and breast cancerous tissues and cancer cell lines.

iv. To differentiate and characterize cancer type and stage using exploratory analysis of the trace elements in human cancer tissue and cancer cell lines using PCA and SVM.

1.4 Justification and Significance of Study

Traditional diagnostic techniques of cancer such as X-ray radiography, computed tomography (CT) scan, angiogram, magnetic resonance imaging (MRI), and mammography are complex and are not as effective to establish the occurrence of the disease at hyperplasia stages as well as the prognosis of the disease. LIBS is a very powerful and rapid technique that has the ability to differentiate malignant growth from normal and benign tissues based on trace element levels (and possibly early stage of occurrence) of the disease.

The mortality rate of cancer is high despite having techniques such as mammogram, X-ray radiography, MRI, CT scan and angiogram. According to World Health Organization, 8.2 million people worldwide died from cancer in 2012. Of the world's total new annual cases, 60 % occur in Africa, Asia, Central and South America. The annual cases are expected to increase from 14 million in 2012 to 22 million within the next two hundred years. The burden of cancer can be reduced if early detection and identification of the disease is done. This leads to proper management of the patients hence minimizing the death rates. (Globocan, 2012).

There is a need to establish alternative diagnostic techniques of cancer to enhance early detection hence early treatment and proper management. Trace metal bio-analysis is a good indicator of early stage of cancer development. This research utilizes innovative spectroanalytical approaches based on the combination of LIBS and chemometrics that enhances the trace analytical detectability of the biometals in tissues while increasing the information gained via multivariate exploratory analysis capability.

CHAPTER TWO

LITERATURE REVIEW

2.1 Disease Diagnostics

Disease diagnosis refers to the process of determining whether there is presence of disease in an organism. It involves correlating selected pieces of information, recognizing formed patterns and differentiating those patterns. Disease diagnosis uses several techniques to determine the presence of the disease. In this research, laboratory techniques have been used to rapidly detect and directly diagnose and characterize cancer. Cancer is a principal cause of death universally, resulting in 8.2 million deaths in 2012 (Stewart and Wild, 2014). The most common causes of cancer deaths are cancers of the lung (1.59 m), liver (745, 000), stomach (723, 000), colorectal (694, 000), breast (521, 000), esophageal (400, 000) (Stewart and Wild, 2014). Diagnosis of cancer currently is based on various methods for example MRI, CT scan, mammography and X-ray radiography.

Trace elements are chosen for study due to their occurrences in both healthy and cancerous tissues in different levels. They can therefore be used as indicators of early stages of cancer development, for which the conventional analytical techniques such as computed tomography (CT), scintillation scan, ultrasound, biopsy, magnetic resonance imaging (MRI) and mammography screening are incapable of detecting their levels.

2.2 Methods of Cancer Diagnostics

2.2.1 Magnetic Resonance Imaging (MRI)

MRI is an imaging tool that develops thorough, cross-sectional images of the inside of the body. However, there is generation of noise when the MRI system is in operation. A patient is at risk if they have metallic implants that they are not aware of, thus great safety measure is

required to avoid occurrence of accidents. During the process, a patient lies on a table top and slips into an enormous shaft scanner; this is not suitable for claustrophobic people. Although MRI is a very sensitive technique and does not use radiation, undertaking the scan is expensive.

2.2.2 Computed Tomography (CT) Scans

A CT scan is a radiation technique that produces 3-D, cross-sectional images of the organs. It delivers a high dose of radiation, which exposes the patient to carcinogenesis. The dose is about the same radiation exposure that an individual would get in a year. A contrast dye is necessary before the procedure for some organs. Iodine is mostly used but is a source of allergic reaction to most people.

2.2.3 Mammography

This is an X-ray examination of the breast tissue and it creates pictures of the tissue. It then detects and profiles breast changes. The procedure may be uncomfortable and perhaps may hurt depending on a woman's menstruation cycle. The radiologist gives a diagnosis based on the film readings attained. This can create a likelihood of a misdiagnosis. The technique also faces a challenge of not being able to detect breast cancers that do not form visible tumors. If a mammogram reveals an abnormal area, a biopsy may be performed to determine if it is cancer. The disadvantage to this is if the abnormality turns out not to be cancerous.

2.2.4 X-Ray Radiography

X-ray is a type of great energy radiation, used for cancer diagnostics and management by producing images and recording them on a film known as radiograph or digital images. Since different tissues absorb the radiation at varying rates, the images produced appear light or dark; dense materials, such as bone, appear white while muscles show a varying shade of

gray. The ionization radiation used by the X-ray equipment can cause cell damage. Exposure to significant amounts of radiation from X-ray scans may in fact increase a patient's risk of developing more cancer.

2.3 Biomedical Applications of LIBS

Application of LIBS in the study of various tissues to determine their health status of the tissues is widespread. LIBS is a rapid method that gives information instantly. The rapidness of this technique makes it essential in disease diagnostics to obtain relevant information in the medical field (Musazzi and Perini, 2014).

Trace element biomarkers are useful in detecting the presence of certain types of cancer in the body as well as the staging of the disease. LIBS has been utilized in detection of trace biomarkers in the medical field by various researchers. In a study done to characterize malignant tissue cells using LIBS, the spectra of malignant and healthy tissue at different spectral regions indicated a clear distinction of the two tissues. The research showed the intensity of the following elements: Na, Cu, Ca, Al, Fe, K and Mg, associated with concentration of trace elements in these tissues, was dissimilar (Kumar *et al.*, 2004). In this study, several Fe lines were detected in these tissue spectra whose intensities varied in both malignant and benign cells. From this study, comparison of normal and malignant tissues qualitatively and quantitatively using LIBS spectra is feasible. It has been demonstrated that LIBS is feasible for differentiating malignant and benign tissues.

Quantitative LIBS analysis of concentration values of trace element in solidified tissue has been performed and it showed the ability to differentiate between infected and healthy teeth. LIBS can be implemented and utilized in dental boring using laser (Samek *et al.*, 2001). The trace elements were quantified and the difference noted between healthy and calcified teeth.

This also indicates that LIBS can quantify the trace elements from the intensity values of the spectral lines. This study does not consider the nonlinear aspects of the data and hence the need for multivariate techniques.

Calibration- Free LIBS (CF-LIBS) has been employed in the analysis of Hair Tissue Mineral Analysis (HTMA) to determine the concentration levels of major minerals found in the human hair and the results likened with those obtained through Inductively Coupled Plasma Mass Spectroscopy (ICP-MS) (Corsi *et al.*, 2003). It was observed that there was a variation in the concentration of Mg, K, Ca, Na and Al in the eleven people of different age, gender and hair colour under study. These trace elements occur in very low concentrations of 16 ppm-100 ppm for K, 50-200 ppm Ca, 20-580 pm Mg, 60-156 ppm Na and 0.4-14 ppm Al. The ratio of Na/K was found to be 1.3- 4 while that of Na/Mg 0.03-1.6.

The presence of Zn in human skin and evaluation of the efficiency of creams preventing absorption of Zn ions has been explored using LIBS. The investigational results designated an exponential decrease of the concentration levels with skin depth (Sun *et.al*., 2000). In this research, it is evident that LIBS is suitable for study of amount of trace elements in the human body.

In a study to determine the distribution of elements in different parts of kidney stones using LIBS, the elements detected were, Cl, S, Sr, Zn, Fe, Ca, O, K, Mg, Mn, Cu, Na, H, N, P and C. This shows the ability of LIBS to do elemental analysis by spectral analysis. The researchers did optimization of the LIBS spectra by changing the laser energy from 10 mJ to 40 mJ to get the best signal-to-background and signal-to-noise ratios. The calibration curves showed that concentrations of trace elements in the kidney stones decreased from the center to shell and surface. Moreover, the element concentrations in the stones increased as the age

of the patients increased (Singh *et al.*, 2009). Spectral analysis using LIBS is possible as already explored in this study.

LIBS has been used to explore elemental components of urinary calculi. Here, seven stone samples: pale-off white 70 mm, pale cream 22 mm, grey smooth 30 mm, brown crystalline 10 mm, dark brown crystalline 8-10 mm, most small cream calculi of 5 mm and dark brown large calculus 50 mm were analyzed. The absolute concentrations of Ca, K, Na, Se, Mg and Pb obtained before calibration of the system for individual elements were found to be extensively different in separate samples. The conclusion is that LIBS technique has the potential for predictable clinical applications in urological disorder diagnosis (Fang *et al.*, 2005).

The intensity of the spectral lines detected is a function of concentrations of the elements and thermo-chemical properties of the matrix. The composition of plasma does not only rely on the component of the sample but also on the characteristics of the laser, condition of the sample surface and on the optical properties of the sample (Quentmeier *et al.*, 1990).
The spectra obtained give the constituent atoms of the sample and the relative concentrations that can be obtained from the intensities by univariate or multivariate methods (Miziolek *et al.*, 2006)

Due to the nature of vast information involved, combining LIBS with multivariate chemometric techniques such as PCA, ANN, and SVM can help to obtain the information required regardless of the multivariate nature of the spectral data.

2.4 Multivariate Chemometrics Analysis of LIBS Data

Chemometrics comprises of statistical methods that are used to mine information from a large set of variables in a given dataset. Chemometrics is useful for classification of known samples and grouping an unknown sample to either of the distinct groups. SVM classification model is utilized to predict the class of a sample based on closest examples by drawing a hyper plane that divides the groups. Chemometrics helps in normalizing data. The classification models are more consistent and include the aptitude to reveal rare samples in the data. Classification of samples can also be done using unsupervised technique whereby the method reveals any similarities and differences in the dataset. PCA is an example of unsupervised chemometric technique for exploratory analysis (Camo, 2017)

Chemometrics is also useful in regression to predict related variables that are easier to measure. The goal of chemometric regression analysis is to train, validate and test a model, which compares the information in the assortment of known measurements to the target. Artificial neuron networks (ANNs) and support vector regressions are among the techniques for regression. Chemometric regression is used in prediction of concentration values of trace elements in this study.

Exploratory analysis to retrieve existing patterns or trends in the data is done using PCA Besides, it reduces data dimensionality and provides a solid representation of all the variables in the data set. PCA algorithm reduces the complexity of the data making it deduce the meaning of the patterns from the scores and loadings plot. The scores plot shows the group patterns while the loadings plot show the variables responsible for the group patterns.

The above chemometric techniques are used together with LIBS to extract qualitative and quantitative information. SVM has been explored in prediction of prostate cancer based on

Cr, Ca, P, Cu, K, Mn, Mg, Zn Fe, and Se using ICPMS (Guo *et al.*, 2007). The model developed was found to have a prediction ability of 95.8%. LIBS and chemometric techniques namely PCA, PLS, ANNs and SIMCA have been utilized by (Mukhono *et al*, 2012) to quantify trace elements in High Background Radiation Areas (HBRA) geothermal field matrices. PCA and SIMCA were used to classify soil from HBRA and non-HBRA into two distinct classes. While PLS and ANN were used to develop a calibration model used for prediction of the concentration of elements under study in these areas. Given the multivariate nature of most data sets, ANN is therefore a better regression method as it is able to look into all the variables in a given data set simultaneously besides looking into non-linearity of the data as well.

Artificial neural network has proven to be a very useful technique in cases where the problem is ill defined and development of an algorithmic answer is not clear. Data from cancerous tissues has non-linear information necessitating the use of a brain-like technique to sift through the web of available information (Naguib and Sherbet, 2000). ANN has been explored for rapid and direct analysis of soil quality indicators, using EDXRFS. A multivariate calibration curve which was well suited for the analysis of Mg, Fe and Cu with R^2 >0.9 and SEP of 0.08%, 4.02µg/g and 0.88µg/g (Kaniu *et al.*, 2012), has indicated the importance of ANN as a multivariate chemometric calibration technique in elemental analysis. In this research, ANN has also been used to develop a calibration model for prediction of concentration of trace elements in tissues.

Saberkari *et al.*, (2014) used Spatial Independent Component Analysis (SICA) and SVM algorithm on three cancer datasets (leukemia, breast cancer and lung cancer), and compared the results with other existing methods. It demonstrated that there was a higher accuracy and validity in the classification accuracy. A number of researches that have been done on

classification and differentiation of cancerous tissues are based on concentration of trace elements and not on speciation of trace elements though speciation has become important for biomedical studies (Kawakami *et al.,* 2003). Study of trace elements in various types of cancer using LIBS has been successfully applied in detection of malignancy in colorectal and breast cancers (El-Hussein *et al.,* 2010) from which the LIBS spectra showed an obvious increase of the intensity of the spectral lines for calcium in the tumourous samples compared to normal tissues for both breast and colorectal tissues. Thus El Hussein et al., (2010)'s study was based on univariate analysis and this ignores the non-linearity of biological samples which can be addressed by multivariate analysis.

Most studies using LIBS relate intensities with concentration of the trace elements, which is a linear relationship. The complexities of biological samples make them to be non-linear thus multivariate correlation study in this research has looked into the non-linearity of the samples under study.

In summary little progress has been made in the use of LIBS as a cancer diagnosis method though the studies using LIBS have shown progress, the studies have not been done using multivariate analysis which takes into consideration the complexity of biological samples.

CHAPTER THREE

THEORETICAL BACKGROUND

3.1 Laser Induced Breakdown Spectroscopy

Laser induced breakdown spectroscopy is an atomic emission spectroscopy method for elemental analysis. It employs a laser pulse to excite the sample. The excited sample produces a micro- plasma plume that consists of excited atoms and ions (Cremers *et al.*, 2006). This method works through production of spectral lines of the radiation emitted after firing a laser onto a sample.

LIBS is the most preferred spectroscopic method for elemental analysis of solids, liquids and gases as opposed to other spectroscopic techniques such as GFAAS, ICP-AES and ICP-MS. This is due to the fact that it is a fast and non-destructive technique as exciting the sample takes microseconds and only a small sample is removed forming a small crater upon ablation. Besides, with LIBS, both high and low Z elements can be studied since all elements emit radiation when excited to an adequately high-energy state unlike other methods such as EDXRF.

The laser is focused on to a particular part of a sample and this does not affect the surrounding region as opposed to X-ray radiation which is not unidirectional and affects the cells surrounding the one under study despite being focused at a particular one. Therefore LIBS qualifies as a non-invasive technique and stands a chance to be in use in the medical field for diagnostics purposes. However, this method has its shortcomings like matrix effects that affect the spectral lines. The spectral lines of biological samples are affected by noise, which needs to be eliminated. Pre-processing of the spectral data is thus necessary to eliminate the noise, remove spectral overlaps and self-absorbed lines.

3.1.1 Principles of LIBS

LIBS, a high energy pulsed laser is used to ablate a solid, liquid or gaseous sample to create a micro-plasma. This happens by focusing the laser onto the sample causing a small amount of the sample to create super-heated plasma ionizing the elements inside the plasma. This process of plasma formation is known as 'breakdown'. The excited atoms eventually de-excite and give off light (photons) as they decay to the ground state.

Plasma is a general group of atoms, ions, molecules, and free electrons wherein the charged species regularly act together. Plasmas are embodied by a variety of factors such as the degree of ionization, the plasma temperature and the electron density.

Weakly ionized plasma is one whose ratio of electrons to other species is below 10%. Conversely, highly ionized plasmas could have atoms without most of their electrons, leading to extremely high electron to atom or ion ratios. During electron-ion recombination, neutral atoms and then molecules form. All through there is background continuum that decays with time faster than the spectral lines. The continuum is principally due to Bremsstrahlung and recombination events. In the former process, electrons accelerated or decelerated in collisions emit photons. The latter takes place when a free electron is attracted into an ionic or atomic energy level and surrenders its excess kinetic energy in the form of a photon. Resolution time of LIBS therefore allows for identification of the lines in the region of interest.

Identification of spectral lines can be done using the software OOILIBS. It is even easier if understandings of the elements in the sample of concern are well-known. Knowledge of the existing elements allows for selective study of those elements in the sample. NIST database gives relative intensities of neutrals and ions of various elements. These act as a guide to classifying lines while comparing with NIST.

Spectral line identification can be guided by the following; presence of one strong line of a specific element means that all the other strong lines identified by NIST should be present. For instance, if the conspicuous Cu (I) lines at 324.754 nm and 327.396 nm appear, the strong Cu (I) lines at 766.465 and 780.7659 nm should also be detected. The experimental environment whether done in air or vacuum can define the species detected. For example, in air, emissions due to Fe I and Fe II are detected with the ionization potential of Fe I being 7.87 eV (Cremers and Radziemski, 2013).

In addition, in case of spectral overlap of a neutral and an ionized line, it is certain that the line observed is for the neutral element. Singly ionized elements are frequently observed in LIBS plasma, though the observation of higher ionization states in air hardly occurs. (Gaftetal, 2011).

3.2 Multivariate Chemometric Techniques

Multivariate chemometric techniques are statistical, computational and symbolic techniques that are used for classification, pattern recognition, reduction of data sets and recovery of weak signals from high background and overlapped peaks (Micklander, 2002).

These techniques involve statistical analysis of data that arise from more than one variable measurement such as LIBS spectra. The techniques take into account nearly all variables in the entire spectra, remove the unnecessary and correlated information and extract the most relevant information from the original LIBS spectra. This therefore makes LIBS more feasible in determining trace elemental composition and differentiation of samples (Labbé *et al.*, 2008). Information such as speciation, correlation, and classification of trace elements is also possible using multivariate analysis.

3.2.1 Principal Component Analysis (PCA)

PCA classifies data into similar groups from which it retrieves important information. This technique allows prediction of similarity in various sets of data. It looks for a few linear combinations of the variables to summarize the data without losing too much information in the process. PCA is based on a linear transformation represented by equation below

$$X = TP + E \tag{1}$$

The multivariate data matrix X, with n rows and p columns represents objects and variables respectively. PCA reduces the matrix to a lower dimension **T×P** (PCs) and a residual matrix E where T is a matrix score that summarizes the x-variables (sample spectra), P is the loading matrix showing the influence of the variables on each score (intensities or concentrations) and E is the residual matrix due to outliers (Virendra *et.al*., 2011).

According to (Samek et.al., 2001), in PCA analysis, training sets of spectra are decomposed into a series of mathematical spectra called factors which, when added together, reconstruct the original spectrum. A scaling coefficient or score represents the contribution any factor makes to each spectrum, which is calculated for all factors identified from the training set. Therefore, by knowing the set of factors for the whole training set, the scores will represent the spectra as accurately as the original responses at all wavelengths.

PCA extracts the most important information from the data table, reduces data set and keeps only this important information, simplifies the data set, and finally analyzes the structure of the observations and the variables and classifies the data set according to similarities within it (Abdi and Williams, 2010).

3.2.2 Support Vector Machines

Known samples are trained and classified into groups by an SVM model which separates them by a hyper plane. Two different sets of data are classified into two groups which become the basis of grouping other samples. The model then tests its performance using a percentage of the training data set.

In this study, SVM was used to classify data by mapping training set onto the two classes. This is done onto a hyper plane and a separation margin is such that minimal distance between the planes and the training examples is achieved. Unknown data set is then put into the model and it classifies them into the groups of the training data. The support vector machine-training algorithm builds a model that assigns new data into one category or the other (Vance *et al.*, 2010).

SVM algorithm classifies vectors according to the equation below:

$$c = \sum_i a_i k(s_{i,} x) + b \qquad (2)$$

where - s_i represents the support vectors, -a_i the weights, b- bias and k- the kernel function.

This kernel function is the mathematical function with which the original input data is projected onto a higher space. The kernel was used to map the input data onto two categories; with higher and lower oxidation states of the trace elements.

3.2.3 Artificial Neural Networks

Artificial neural networks (ANN)s are “intelligent” systems that have the capacity to learn, memorize, and create relationships among data. It is able to learn significant information patterns within a multidimensional information realm (Kalogirou, 2001). This technique

impersonates the learning process of a human brain and therefore does not need characteristic information about the system; instead, it learns the relationship between the input parameters and the output variables by studying previously recorded data (Kalogirou, 2000).

ANNs consist of a group of interconnected nodes imitating the network of neurons in a brain. In Figure 3.1, each circular node represents an artificial neuron and an arrow represents a connection from the output of one neuron to the input of another.

A neuron is a computational device that calculates the weighted sum of its inputs and calculates the output signal from this using a nonlinear function. The weights are estimated using an appropriate algorithm based on a calibration set using cross validation (Kim *et al.*, 2000). ANNs can be used to build empirical multivariate calibration models of the form

$$y = f(x) + \varepsilon \tag{3}$$

where y is the vector matrix containing sample response (concentration), f is the network function x is the input (intensities/ counts) of LIBS spectra and ε is the error of calibration (Marini *et al.*, 2008).

The neural network consists of an input layer of neurons, one, two, or three hidden layers of neurons and a final layer of output neurons. The lines each connect these neurons and they have a numeric number known as a weight. The output, h_i of neuron i in the hidden layer is given as:

$$h_i = \sigma\left(\sum_{j=1}^{N} v_{ij} x_j + T_i^{hid}\right) \tag{4}$$

Where (σ) is called the activation (or transfer) function and is defined as:

$$\sigma(u) = \frac{1}{1+\exp(-u)} \tag{5}$$

N is the number of input neurons, v_{ij} the weights, x_j inputs to the input neurons and T_i^{hid} the threshold terms of the hidden neurons (Wang *et al*., 2003). The activation key introduces non-linearity onto the network and bounds the neurons so that the network is not affected by contrary neurons. The values given to the input neurons are independent variables whereas those produced by the output neurons are dependent variables to the back propagation algorithm being approximated by the network.

The network consists of three layers of units: the input layer, the hidden layer and the output layer. The input values corresponding to the spectral intensities that characterize a sample are fed into the first layer, which processes the information. The outputs of the first layer feed the neurons of the next layers. The hidden layer then extracts the main prominent features of the input data to obtain a discrete characterization of each sample (Boueri *et al.*, 2011).

Back propagation is the training function that was used in this research to train the feed forward networks whose performance was examined using the performance function, MSE.

Figure 3.1 below illustrates the feed forward back- propagation network model.

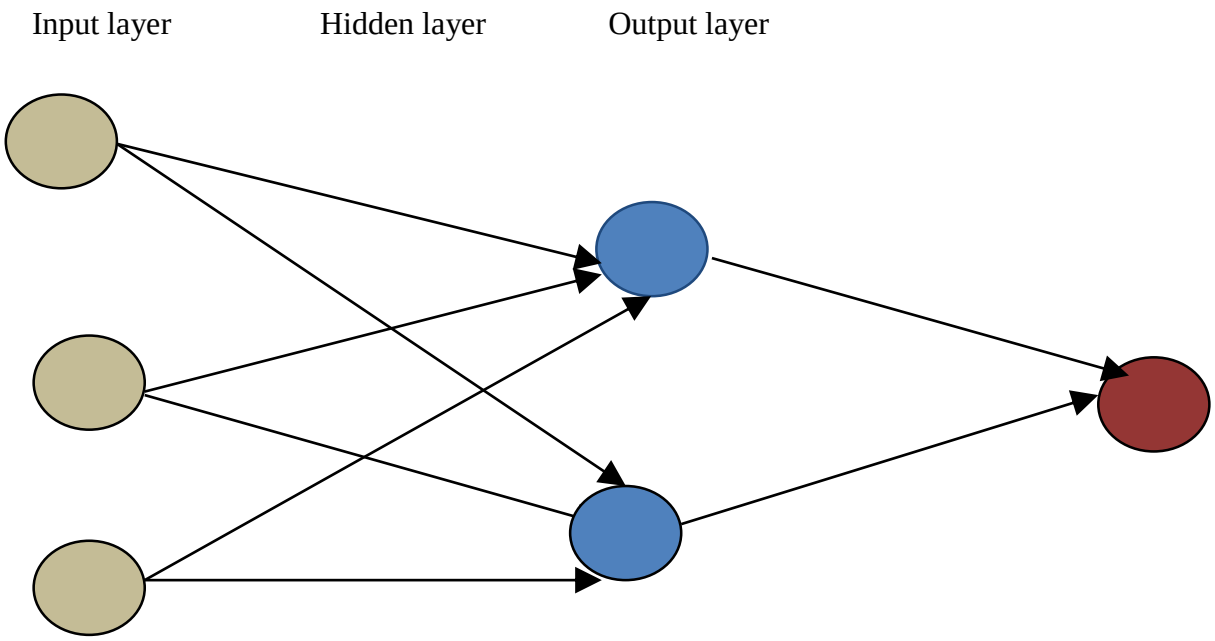

Figure 3.1: Multi-layer Artificial Neural Network model showing feed forward back-propagation 3-layer network model.

CHAPTER FOUR

MATERIALS AND METHODS

4.1 Sample Preparation

The samples under study were human liver, breast and abdominal tissues. Tissue biopsies of the above were obtained from Kenyatta National Hospital and placed in formalin. The tissues were prepared into tissue blocks with paraffin wax as the base matrix. Simulate samples were also prepared to mimic normal and cancerous tissues using the concentration levels reported in literature. These were also made into blocks of paraffin wax embedded with Cu, Mn, Mg, Zn and Fe ions. Twenty samples were prepared; 10 for lower speciation Cu^{+}, Fe^{2+}, Mn^{2+}, Mg^{2+} and the other 10 for higher speciation Zn^{2+} Cu^{2+}, Fe^{3+}, Mn^{4+}, Mg^{2+} and Zn^{2+}. Hep (Hela) stages 1-5 and Lewis lung stages 1-5 cultured cancer cell lines were grown in Dulbecco's Modified Eagle growth Medium (DMEM (1x), Gibco®, by Life Technologies). These stages mimic the different stages cancer tissue undergoes as the disease progresses. Oyster tissue powder was compressed using a hydraulic press to produce pellets of 2 g mass. These were for validation of the predictive model.

4.1.1 Preparation of Simulate Samples

The salts shown in Table 4.1 were used to prepare standard solutions of Cu, Zn, Fe, Mn and Mg by dissolving some amounts of analytical grade salts containing these ions in ethanol. The masses were obtained by first calculating the molar mass of the given salt. Using $FeCl_3.8H_20$ for example, the molar mass is 270.32 g and the molar mass of Fe is 55.848 g. If 55.848 g is in 270.32g of the salt then to find the mass of salt that contains 1g of Fe, we took the molar mass of the salt divided by the molar mass of Fe. Since the purity of the salt is

99%, the value obtained is thus equivalent to this percentage. The value for 100 % is then calculated. To prepare the stock solution, we dissolved 1g of Fe into 1000 ml of ethanol to make 1000 ppm. Since 1 g of Fe is in 4.8892 g of the salt the mass of salt needed to make a solution of 2000 ppm, from 10 ml volume of the stock solution was calculated. This yielded the mass of the salt to be used.

The concentration ranges shown in Table 4.2 were distributed using research randomizer. The volume selected was determined using the formula:

$$C_1V_1 = C_2V_2 \tag{6}$$

Where C_1 concentration of the stock solution

V_1 -Volume of stock solution needed to make the new solution

C_2 -Final concentration of new solution

V_2 -Final volume of new solution

About 2 ml of molten paraffin wax was poured into a mold onto which 5 ml of the mixture was added. The mixture was then stirred to ensure homogeneity. Stirring was done while the mixture was being heated at 78°C to ensure that ethanol and acetone boils off. The mold was covered with an embedding cassette and put in a freezer to cool and form a block. This block was sliced to 2 cm thickness, weight of 2 g ready for study.

Table 4.1 : Mass in grams of the compounds used for making stock solutions

Salt	Chemical formula	Mass in grams
Cupric nitrate	$Cu(NO_3)2.3H_2O$	0.038
Ferric chloride(Hexahydrate)	$FeCl_3.8H_2O$	0.073
Ammonium ferrous sulphate	$NH_4Fe(SO_4)2.6H_2O$	0.106
Zinc nitrate purified (Hydrate)	$Zn(NO_3)2.6H_2O$	0.091
Magnesium chloride hexahydrate	$MgCl_2.6H_2O$	0.213
Manganese II chloride	$MnCl_2.4H_2O$	0.037
Potassium permanganate	$KMnO_4$	0.029

The concentration ranges were selected in the ranges they occur in the soft tissues. The general concentration levels in cancerous tissues are as follows: Fe is 30- 170 µg/g, Mg 62-502 µg/g, Zn 20- 200 µg/g, Cu 1-10 µg/g and Mn 1-30 µg/g. The distribution ranges was done randomly and the values were combined to produce each of the samples. The samples are divided into two sets; 10 for higher speciation and 10 for lower speciation, a total of 20 samples. One set contains the ions of elements with higher energy states while the other has ions of lower energy states.

Table 4.2 Concentration levels µg/g in Mg^{2+}, Fe^{3+}, Zn^{2+}, Mn^{7+} and Cu^{2+} in the stock solution used for preparing higher speciation simulate samples

Salt	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Mg^{2+}	298	501	346	429	354	421	351	474	494	328
Fe^{3+}	105	170	128	163	143	157	139	166	167	127
Zn^{2+}	75	178	113	145	133	141	124	161	173	90
Mn^{7+}	6	30	12	23	18	20	16	25	29	10
Cu^{2+}	1	7	2	9	10	6	5	3	8	4

Table 4.3: Concentration levels in µg/g of Mg^{2+}, Fe^{2+}, Zn^{2+}, Mn^{2+}, and Cu^{+} in the stock solution used for preparing lower speciation simulate samples

SALT	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20
Mg^{2+}	108	153	65	140	247	169	297	378	318	153
Fe^{2+}	53	81	50	45	38	40	51	89	60	37
Zn^{2+}	26	55	48	101	58	47	45	153	49	79
Mn^{2+}	3	5	4	12	6	18	11	9	10	28
Cu^{+}	3	7	6	9	8	5	1	2	4	10

4.1.2 Preparation of Oyster tissue

Oyster tissue powder was placed in a hydraulic press to form pellets of 2g mass each. These pellets were then ablated upon using LIBS.

4.1.3 Preparation of Human Cancer Tissue Samples

Breast, liver and abdominal tissue needle biopsies were taken from the Kenyatta National Hospital from various patients. The tissues were trimmed to about 2 mm, and put in 10% formalin in labeled bottles. Dehydration process was done by soaking the tissues successively in 70%, 80%, 90% and 95% alcohol consecutively for an hour in each solution followed by absolute alcohol, (alcohol that contains 99% pure alcohol and not more than 1% water) at three different stages; I, II and III consecutively for an hour each. The tissue samples were then cleared of alcohol by soaking them in 50:50 alcohols for an hour followed by toluene for another 1 hour and finally toluene in two stages I and II each for 30 minutes.

The tissues were then dipped into molds filled with molten paraffin wax of temperature 58°C. The molds were then placed in the oven overnight for infiltration of wax to fill up the spaces left in the tissue after clearing of alcohol.

The following day the tissues were embedded in fresh molten wax at 58°C in molds. It was left to cool at room temperature. The blocks formed were labeled and stored in a cool place. The blocks were trimmed on the surface until the tissue was exposed. They were then made into blocks of 2 cm thickness and 2g in mass ready for study. A few sections of $3\mu m$ were also prepared and placed on Mylar film.

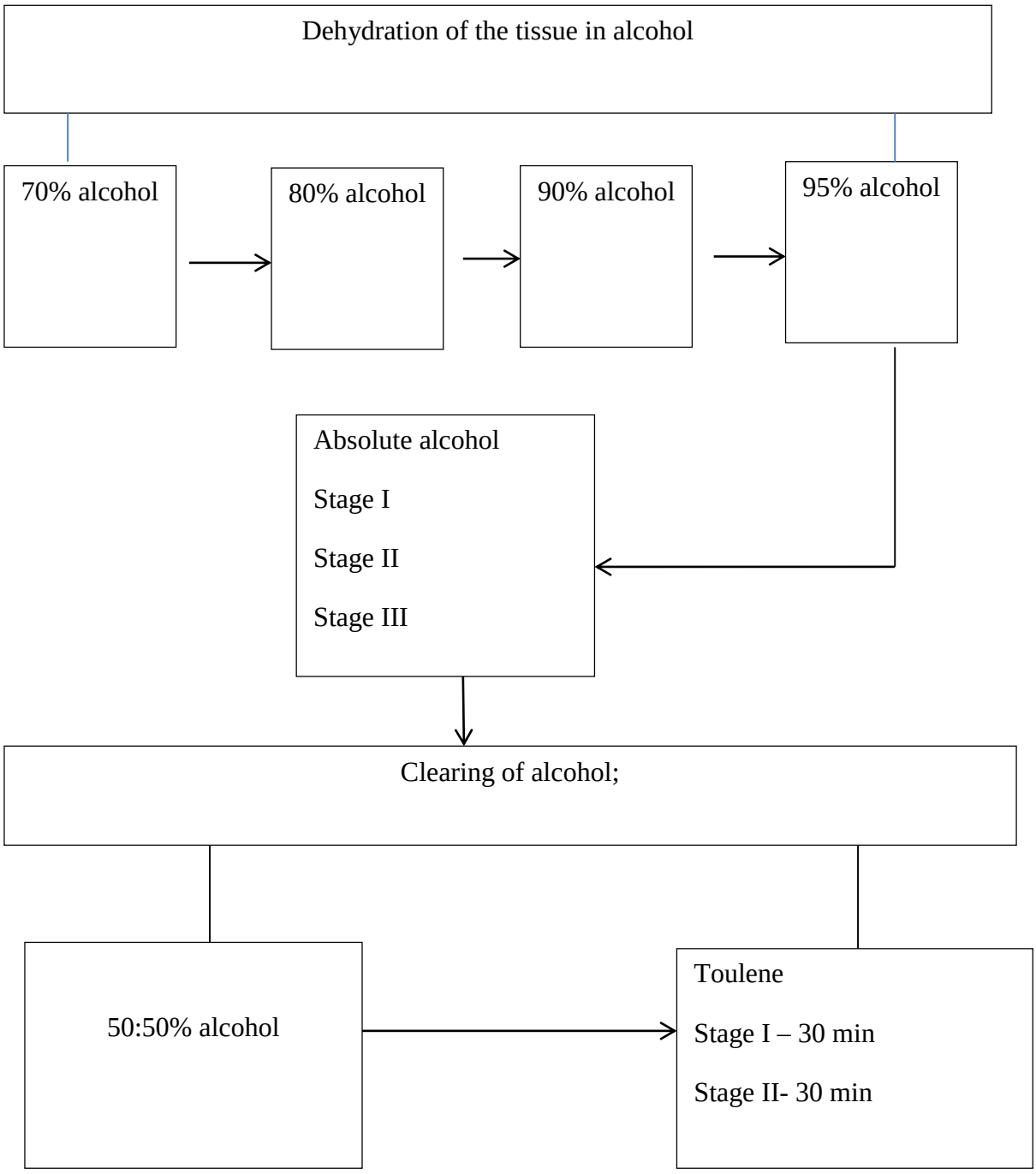

Figure 4.1: A block diagram illustrating the systematic stages carried out while processing breast, abdominal and liver cancer tissues.

4.1.4 Preparation of Cultured Cancer Cell Lines

Human epithelial type 2 cells from the human cervical carcinoma Hela (HEp-2) and Lewis lung carcinoma (LL) cells were acquired from Kenya Medical Research Institute (KEMRI), Nairobi. Sterilized T-25 culture flasks (Fischer Scientific), containing a chemical inside to facilitate adhering of the cells onto the surface, were used to culture the cells in growth media prepared using Dulbecco's Modified Eagle Medium (DMEM (1x), Gibco®, by Life Technologies). The growth medium contains complemented amino acids, carbohydrates, vitamins, minerals and salts that provide a favorable medium for growth of the cells. This basal medium lacks protein and growth providing agents. It is therefore complemented with 10%, by volume, fetal bovine serum (FBS, ATCC), to provide the growth factor and hormones as well as proteins, vitamins, lipids, minerals and gases.

All these procedures were undertaken inside a biosafety cabinet. 1% Fungizone Amphotericin B 250µg/ml (Gibco®, by Life Technologies), 1% penicillin-streptomycin, 1% L-glutamine, and 0.1% Gentamycin (all from Sigma Life Science) were added to prevent contamination by probable growth of fungi and bacteria. The cells were then incubated for 24 hours under a 35% relative humidified atmosphere of 5% CO_2 and at a temperature of 37 ^{o}C. All cells were within 7-8 passages of the primary cell line; divided into subcultures of up to a maximum of 8 sub- cultures.

Preceding spectroscopic measurements, the cultured cells were rinsed using 3 ml of Hanks Balanced Salt Solution 0.25% trypsin - EDTA (1mM) (Gibco®, by Life Technologies), they were rinsed again using 2 ml of the same solution, incubated for 3-5 minutes at 37 ^{o}C until cells separated completely, then followed by adding 8 ml of prepared growth media to deactivate trypsinization process. The obtained cell suspension was centrifuged at 1200 rpm

for 5 min (eppendorf – centrifuge 5810R), supernatant removed, and vortexed (Vortex Genie 2, Scientific Industries) to ensure uniform cell distribution. The resultant cells were then washed twice in 1.5 ml phosphate-buffered saline solution - PBS (Sigma-Aldrich), and centrifuged at 1200 rpm for 5 min after each wash. After removing remaining supernatant, cells were vortexed again, and 25 µl suspension drops plated on 12 well – 6 mm diameter-autoclavable well plate (Thermo Scientific Cel-Line ® Brand). The cell pellets were then allowed to dry at room temperature (approximately 25^{o}C) for not less than 48 hours in a chamber. After preparation of the samples, LIBS was used to obtain the spectra of the simulate samples, breast tissues, liver tissues, abdominal tissues and the cell lines.

4.2 LIBS Set Up

The set up involves a pulsed Nd-YAG laser of maximum energy 50 mJ operating at a fundamental wavelength of 1064 nm and 8 ns pulse width .The laser is fired onto a sample, directed by the focusing lens of focal length 10.16 cm, exciting it to produce a micro-plasma that is a characteristic of the sample. The optical to sample distance was maintained at 30 mm as the optimal distance. The micro-plasma is then collected through a lens by the fiber optic cable LIBS 2500 PLUS (Ocean Optics, Inc) of 0.22 Numerical aperture and 101 mm focal length into a set of seven HR 2000 atomic emission spectrometers of resolution 0.1 nm (in the range 200 nm to 980 nm), which spectrally disperses the radiation.

A CCD camera in each spectrometer acquires data simultaneously and displays the spectrum in the computer with the help of OOILIBS software. The camera has 2048 pixels and an optical resolution of 0.065 nm. Spec line software package by Ocean Optics analyses the emitted light to reveal the elemental composition of the sample by automatically identifying the peaks, comparing the corresponding wavelengths with a data base of atomic and

molecular lines and availing a list of possible elements present in a sample at particular wavelengths. The data is encoded into a chip of each spectrometer plus the wavelength calibration coefficients (LIBS 2500 PLUS Operation Manual, 2008). Simulate samples, real samples and cell lines were ablated upon in the sample chamber by the laser under the conditions mentioned above. Q-switch delay of 0.42 µs was used for data collection.

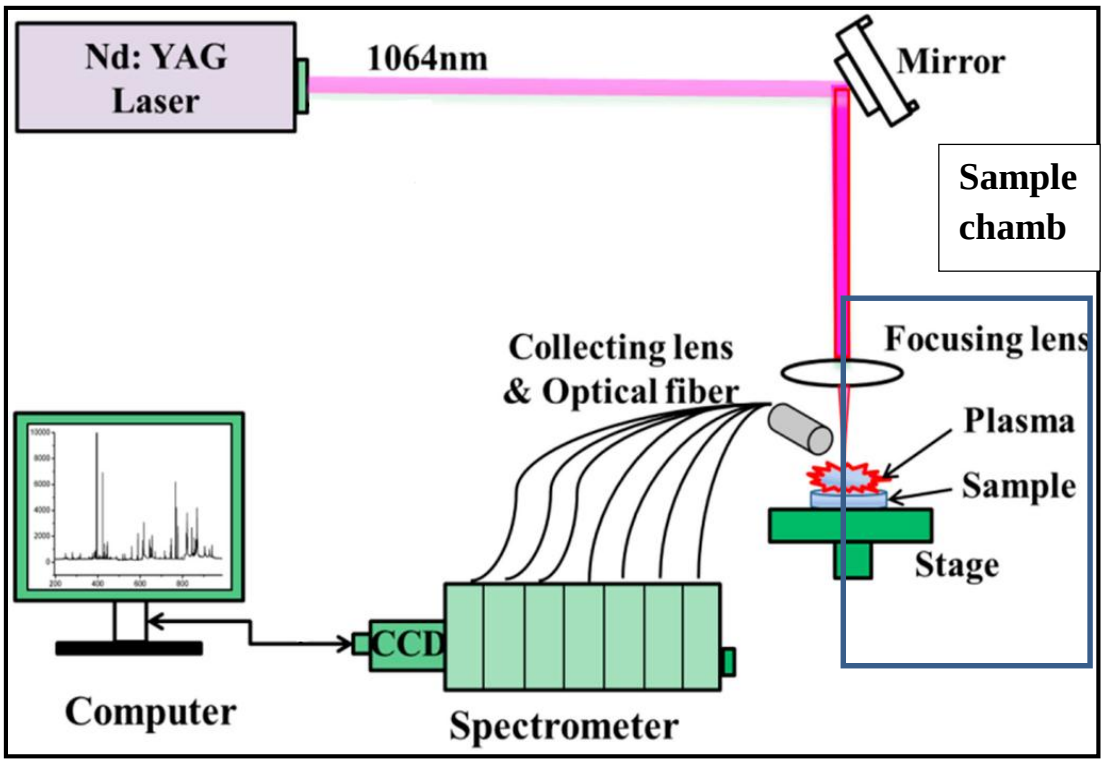

Figure 4.2: Schematic diagram of LIBS (Liu et. al., 2014). This figure shows the major components of LIBS; ND-YAG laser, CCD camera, computer, spectrometers, focusing/ collecting lens, optical fiber, mirror and the stage.

Table 4.4 below shows the specifications of the seven spectrometers used in the LIBS set up.

Table 4.4: LIBS 2500 PLUS spectrometer specifications of the seven spectrometers working in unison from wavelength 200 nm to 980 nm; UV, visible and near infra-red regions

Model	Region	Gratings (lines/mm)	λ Range (nm)
HR + C0463	Ultraviolet	2400	200- 305
HR + C0464	Ultraviolet	2400	295- 400
HR + C0465	Visible	1800	390- 525
HR + C0466	Visible	1800	520- 635
HR + C0467	Visible-Near	1800	625- 735
HR + C0468	Near infra-Red	1800	725- 820
HR + C0469	Infra-Red	1800	800- 980

Figure 4.3: Photograph of the LIBS 2500 PLUS (Ocean Optics) used while carrying out the research.

4.3 Spectral Calibration of the Seven Spectrometers

Calibration was done using HG-1 Mercury- Argon lamp, prior to using LIBS set up to take data, as a source of light and an optical fiber cable for transmitting the light to the spectrometers. The following equation, which shows the relationship between the pixel number and wavelength as a third order polynomial, was used to do regression.

$$\text{Lambda} = I + C_1 p + C_2 p^2 + C_3 p^3 \quad (7)$$

In equation 7, I is the wavelength of pixel 0, C_1= the first coefficient, C_2= second coefficient and C3 the third coefficient. These values were obtained for the seven spectrometers. The calibration yielded spectrometer resolution of 0.1nm.

4.4 Optimization of LIBS Parameters

After calibration of the system, it was necessary to optimize the parameters before taking data. Optimization was done by taking spectra at different energy values while keeping integration time, delay time and optical to sample distance constant. This was done by varying one parameter, as the others remained constant. The values with the highest SNR were selected as the optimal values achieved for each of the parameters were used for data acquisition as shown in Appendix IV. The optimized LIBS conditions are shown in Table 4.5. They are further discussed below.

4.4.1 Energy of the Laser

Laser energy provides the excitation energy which aids the ablation process. This in turn affects the features of the spectra achieved from the sample. Different samples have different ionization energy of the elements. Very low laser energy may not be sufficient to ionize the elements. The Nd-Yag laser produces energy of up to 50 mJ. Spectral data was collected for different sets of energy and SNR values compared.

4.4.2 Q-Switch Delay Time

Q- switching is a method for obtaining pulses of high energy by storing energy in the optical cavity leading to a buildup of population inversion until the Q-switch is turned on. Once the Q-switch is turned on, the stored energy is released in a single pulse of high energy and peak power. The terminology comes from the Q factor of the laser resonator, which is given as

$$Q = \frac{\text{Energy stored in the cavity}}{\text{Energy loss per cycle}} \tag{8}$$

Adjusting the pulse duration is necessary since the laser peak power depends on the duration as well as the energy of the pulse. The shorter the duration of the pulse, the higher its peak power.

In this research, the Q-Switch value was varied between 0.83 µs, 0.42 µs and 1.67 µs upon which 0.42 µs was chosen as the best as it gave spectra with less interference (self-absorption and overlapping) from matrix effects. This was selected after observing the effect of these other time values on the spectra.

4.4.3 Fiber to Sample Distance

The distance of the sample to the optic fiber cable through which the laser is fired is of great importance since beam divergence takes place leading to different amounts of the laser pulse being used for sample ablation. An appropriate distance should therefore be determined while observing the profile of the spectral lines. Upon optimization, a distance of 0.3 cm was preferred since it is at this point that interference free emission lines with a greater SNR were generated despite having low concentration values. At this distance, more emission lines were seen over and above, most light was emitted by the sample plasma and collected by the optical fiber cable.

4.4.4 Number of Ablations per Scan

Increasing the number of ablations per laser shot results in a substantial decrease in intensities of emission lines due to increase in the crater depth that inhibits amount of plasma viewed by the optical fiber (Tognoni et.al., 2002; Sneddon, 2002) hence reduction in line intensities. Moreover, it enables the study of the entire sample and not just the surface. Sample homogeneity can be studied by comparing the results from the surface and depth. Single

ablation shots were compared to 5 ablations and 10 ablations. Table 4.5 below shows a summary of the optimized LIBS conditions.

Table 4.5: Optimized LIBS conditions showing laser energy, fiber to sample distance, Q-switch delay and number of ablations per second.

Number of ablations per scan	1scan
Optical to sample distance	3 mm
Q- Switch delay time	0.4 us
Energy of the laser	50mJ

4.5 LIBS Spectral Data Acquisition

A pulsed Nd-YAG laser of 50 mJ energy, at a Q-switch delay time of 0.42 μs was used to ablate the simulate samples, liver, breast, abdominal samples, oyster tissue and cell lines. The spectrum of oyster tissue was used for validation of the ANN concentration prediction model. The optical to sample distance used was 30 mm. The samples were placed in the sample chamber on the stage and then the laser was fired at 50 mJ. Data was taken at 50 different points for simulate samples to ensure homogeneity by getting the average of several data points.

The liver, breast, renal and abdominal tissues were ablated upon in the sample chamber at the same parameters and the spectra obtained were analyzed for the elements present in the sample. Data was taken at 20 different points to average out errors due to homogeneity. The results are shown in Chapter 5. Oyster tissue was also ablated and the results used to validate

ANN model. Finally, the stage 1-5 of Hep-2 and lung cell lines were ablated upon and twenty sets of data were acquired from each stage and average values obtained.

4.6 Spectral Preprocessing

Preprocessing of spectra is essential to eliminate matrix effects. To overcome these, denoising, smoothing, baseline correction and mean centering preprocessing techniques are employed. The intensity of the emission lines observed is a function of both concentrations of the elements of interest as well as the thermo-chemical properties of the matrix that contains them. LIBS suffers from the matrix effects and therefore the spectra require preprocessing (Quentmeier *et al.*, 1990). The LIBS spectral data was subjected to these techniques to remain with the essential information. Smoothing was done using Savisky Golay technique to get clear spectral line profiles from which information can be retrieved while wavelet transforms were used for denoising. This was done to eliminate noises that do not otherwise behold any significant information from the sample Trace metal identification was then done for the spectra of the samples. LIBS spectra have a background which needs to be eliminated by carrying out baseline correction which ensures that the lines emanate from the same level. Mean centering is a preprocessing technique that enables all the data across the spectral region to be involved in all processes. The data acquired from LIBS runs from 200 nm to 980 nm and the counts can be as low as below 1 and as high as in value of thousands.

This was then followed by multivariate chemometrics analysis, which involved predictive model development using ANN, exploratory analysis using PCA and classification using SVM.

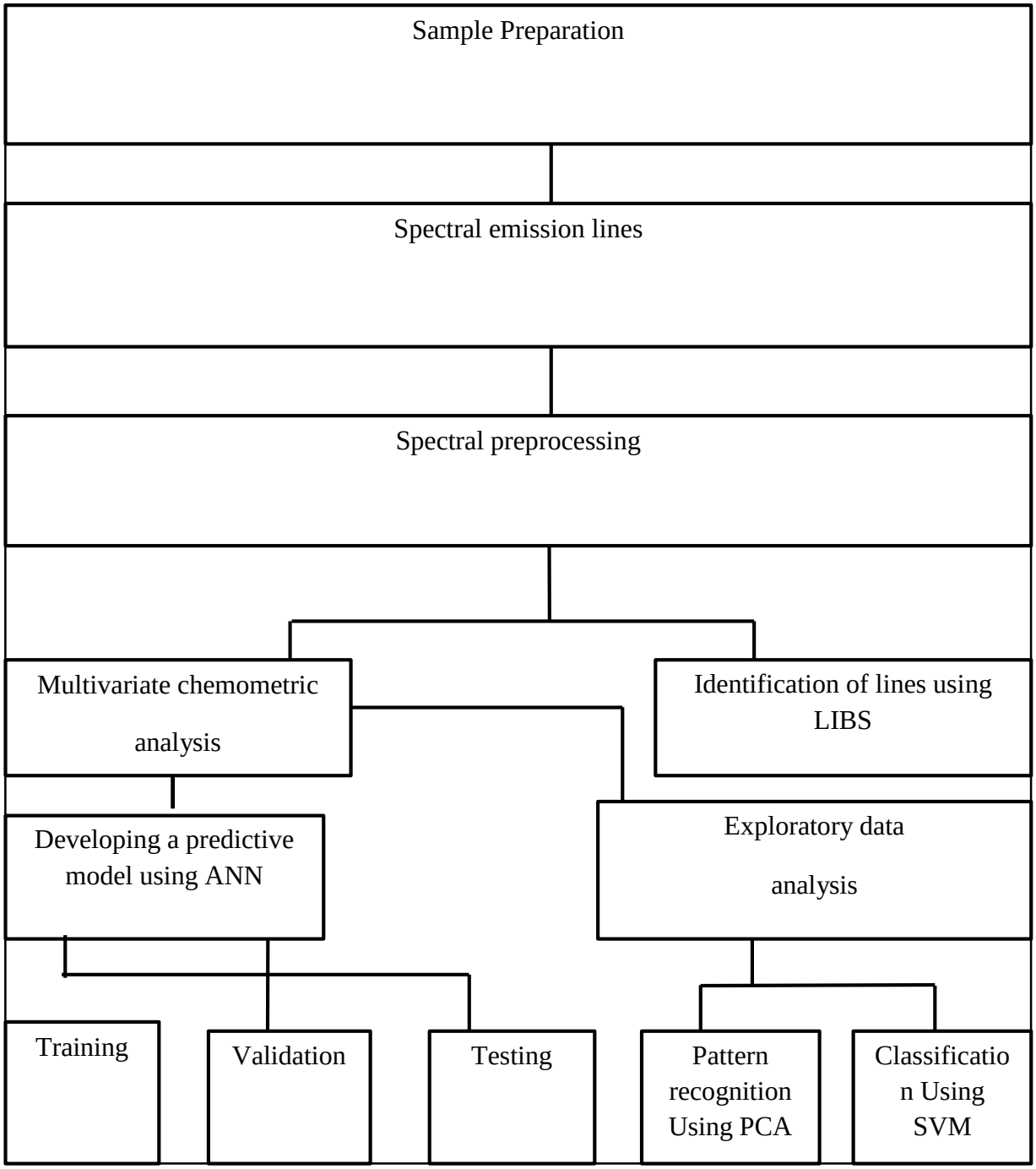

Figure 4.4: Flow chart of methodology and analysis. The methods include preparation of samples, preprocessing techniques, regression supervised classification and non-supervised classification.

4.7 LIBS spectral emission lines

LIBS system was used to acquire spectral emission lines under the optimized conditions. This was done by firing a laser onto a sample placed on the stage in the sample chamber and saving the spectrum for further analysis.

4.8 Spectral Analysis (Identification of Cu, Mn Zn, Mg and Fe lines)

OOILIBS software enables identification of lines. These are then counterchecked from NIST data base.

4.8.1 Multivariate Calibration

Interference free lines of Cu, Mn, Mg, Zn and Fe of 20 simulate samples were used to train the ANN model. The training process involved mapping the intensity values onto their known corresponding concentration values. While training the model, different number of neurons and functions were used to achieve the best training parameters. The best conditions achieved were 3 neurons and feed forward back propagation algorithm using matlab software. Out of 20 samples, the model was trained using 60% of the data, 20% was used for validation while the remaining 20% for testing. The model was trained a number of times until one with the least RMSEP and R^2 value closest to 1 was achieved.

Validation of the model was done after it was properly trained using spectral data of oyster tissue. The model was used to predict the concentration values of Cu, Fe, Zn, Mn and Mg elements in oyster tissue. The predicted values were compared against the standard reference values. The breast, liver, abdominal tissues and cell lines were then tested using the model that was trained and validated. Correlations and ratios were calculated so as to study the trends and alterations of these elements in the tissues.

Exploratory analysis using PCA was done on simulate samples. Although there was prior knowledge of the nature of these samples; higher speciation and lower speciation, PCA was done to find out any possible clustering. Pattern recognition was also done on liver, breast and abdominal tissues. Clustering was also observed as shown in Chapter Five. PCA was carried out on the Hela (hep-2) and Lewis lung cell lines separately using all spectra from 200 nm to 980 nm and using feature selected lines of the five elements only. The scores plots and the loadings plots were studied to observe the patterns formed and the lines of elements responsible for the clusters respectively

4.8.2 Classification using SVM

A classification model was developed using simulate samples; higher speciation and lower speciation into two groups separated by a hyper plane. Three models were developed using the three elements that exist in different oxidation states; Cu, Mn and Fe. The liver, abdominal and breast tissues were each classified by the three models. The model produced an output for the given tissue and grouped to either one of the two groups.

This was done using the package e1071 in R studio developed by (David et., al, 2017)

CHAPTER FIVE

RESULTS AND DISCUSSION

5.1 Spectral Analysis of Human Tissue Samples

Samples under study were breast, liver and abdominal tissues. The samples were prepared and studied using LIBS. The spectral data were used to do spectral analysis of the unique and viable biomarkers of cancer. The focus was on five trace elements; Cu, Fe, Mg, Mn and Zn. The role of these trace elements in cancer diagnostics was explored in this study by checking for their occurrences and the quantity levels.

The trace biomarkers in the samples under study were identified as shown in spectra of intensity versus wavelength listed below. The wavelength lines at which the elements were identified are tabulated as well. Figure 5.3 shows a spectrum of a breast tissue for 200 nm to 980 nm spectral region. Most of the trace elements of concern are in the ultraviolet and visible regions. The Infra- red region mostly has the macro elements. The intensity values for these elements are in arbitrary units, they are not therefore the true values in the tissue. However, these values signify the ratios in which the elements occur. A prediction model was therefore used to calculate the true concentration values based on a well-trained and validated ANN prediction model.

There are several lines with very low intensity values which are of great importance in the study. Most of them are Fe lines as will be observed in the zoomed out regions of the spectrum. Fe is a very essential element in the growth of a tissue. It is therefore present in a tissue that has replication of its cells which require constant supply of nutrients by the blood and its components. PCA was done on the simulate samples together with the breast, liver

and abdominal tissues. Two clusters were formed: one with all the human tissues and higher speciation simulate samples and the other with simulate samples of lower speciation. These are clusters of cancerous and non-cancerous respectively as shown in Figure 5.1 scores plot. The figure is able to illustrate the essence and role of trace elements in the development of tissues.

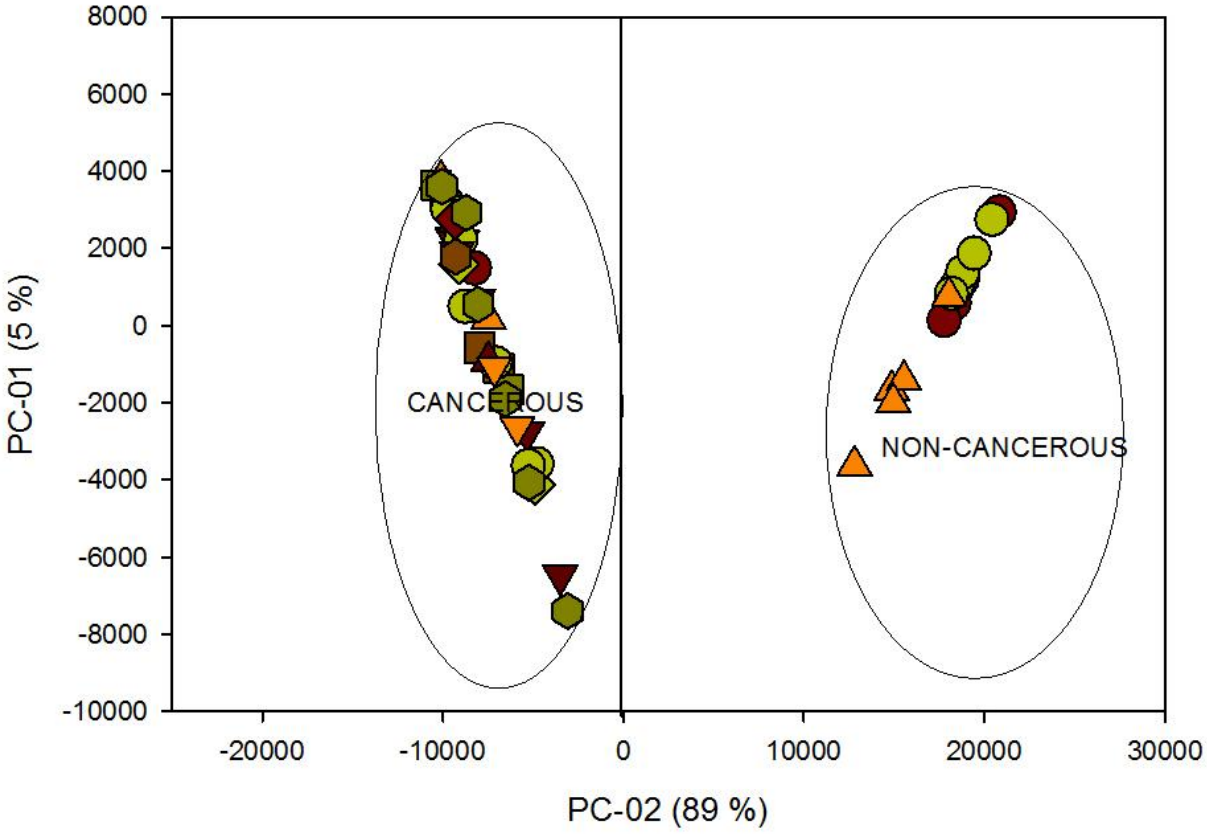

Figure 5.1: PCA clustering of cancerous and non-cancerous tissues. The diagram shows the two clusters formed using real sample tissues as well as simulate samples of both higher speciation and lower speciation.

Fig 5.1 shows a cluster of all the cancerous tissues together with higher speciation simulate samples while lower speciation samples are clustered on their own. The lines responsible for the grouping are shown in the loadings plot in figure 5.2.

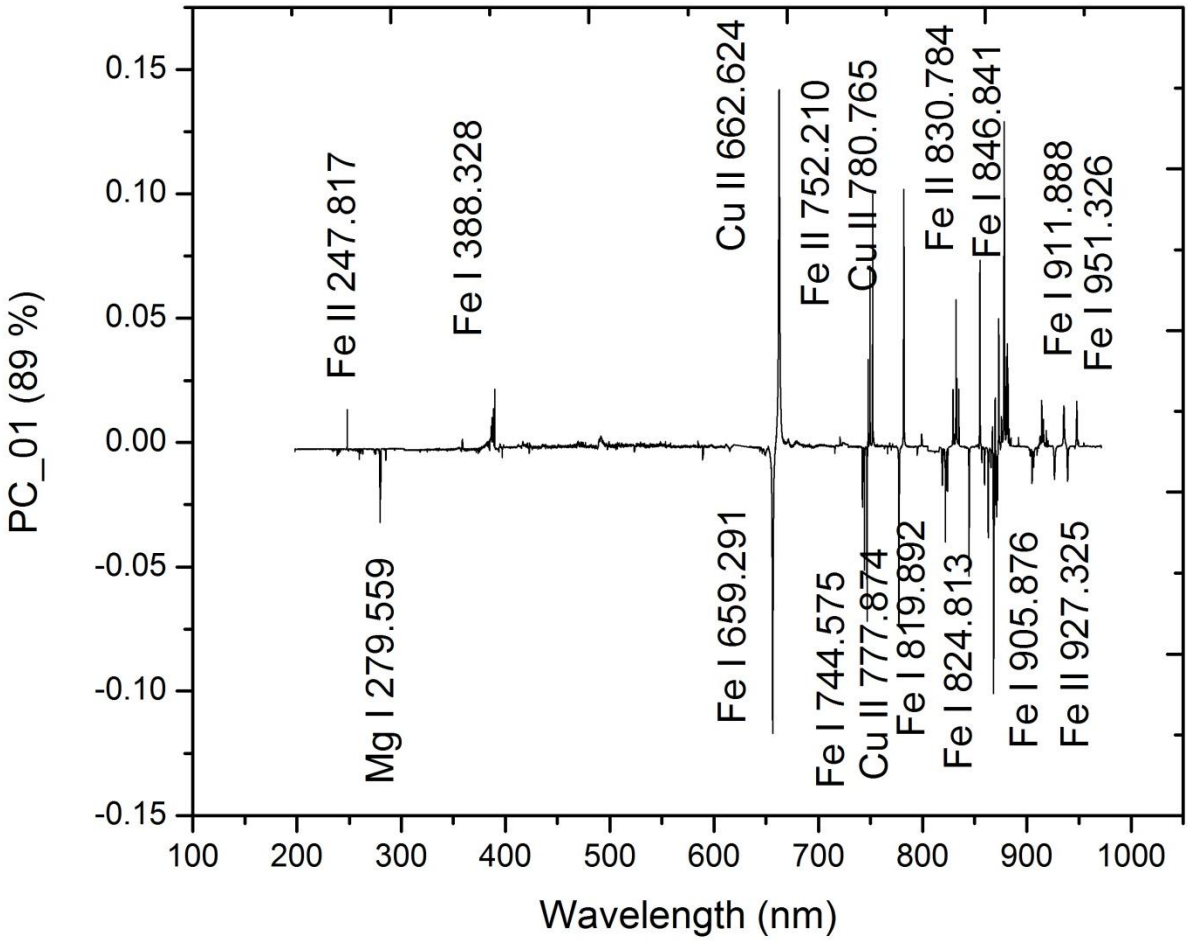

Figure 5.2: A figure showing loadings plot of the clusters of cancerous and non -cancerous tissues. These are the lines that are responsible for the clusters. PC 1 contributed 89 %.

The loadings plot of this score plot shows the elements responsible for this grouping as shown in figure 5.2. These elements, which are also identified in the spectra of liver, breast and abdominal tissues, are identified as trace element biomarkers. Table 5.1 shows the

elements identified in the breast, liver and abdominal tissues. Table 5.2 on the other hand shows the trace element biomarkers identified.

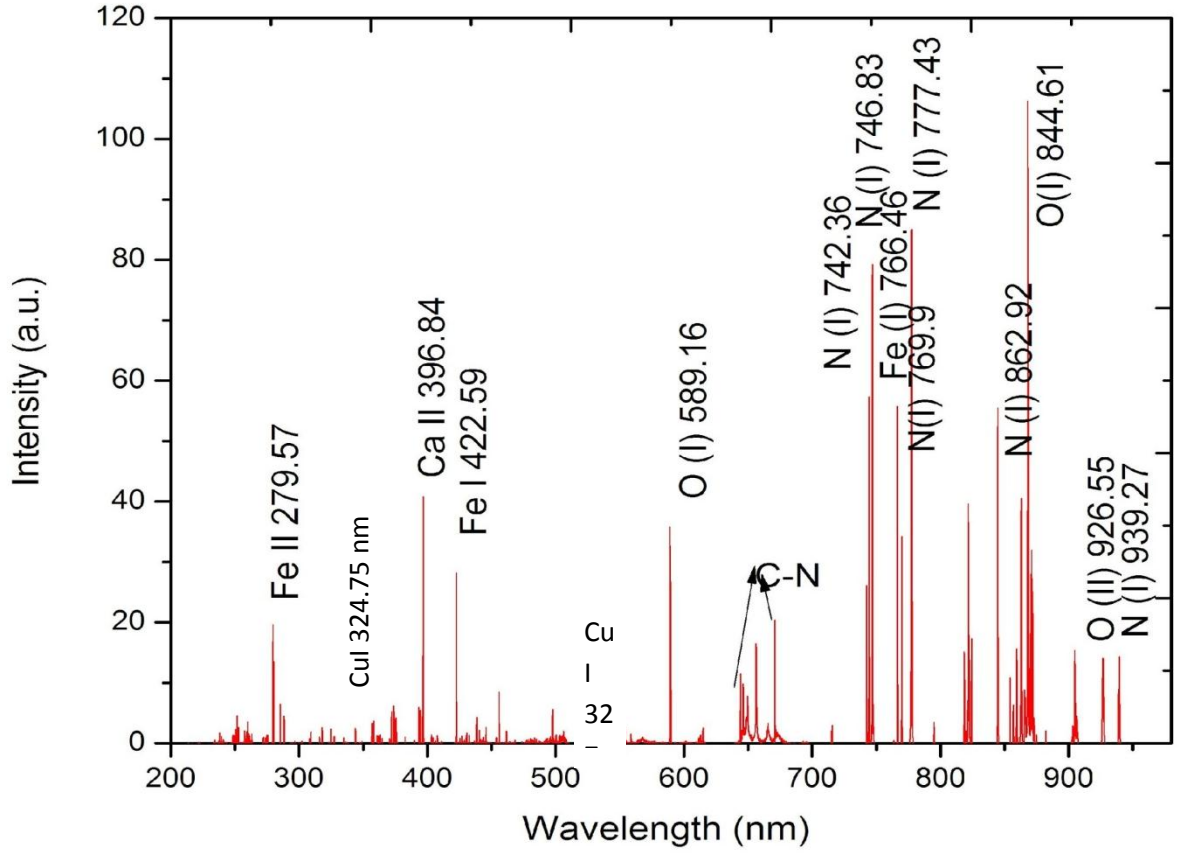

Figure 5.3: LIBS spectrum of breast cancer tissue showing wavelength values of trace element lines observed for the entire spectral range from 200 nm to 980 nm. The figure shows

The lines that were observed in the breast tissue within the ultra violet region are shown in Fig 5.4. It is evident from the spectrum that the majority of the biomarkers are in this region. The lines observed are quite a number of Fe, Cu and Mn lines. These elements appear in higher speciation as the multiplying cells change form. Ca lines were also identified in this tissue. This can be attributed to the role it plays in a breast tissue. Presence of Zn element

signifies the important role it plays of being a co-factor of enzymes. The comparison of the normal tissue and cancerous tissue is usually the occurrence of these elements in higher oxidation states. The biomarkers present in a cancerous tissue and absent in the healthy tissue are viable biomarkers and give a great lead towards the diagnostics process. The magnitude of the role an element plays in the proliferating tissue is eminent in it being observed in the spectrum as well as the number of lines from the same element.

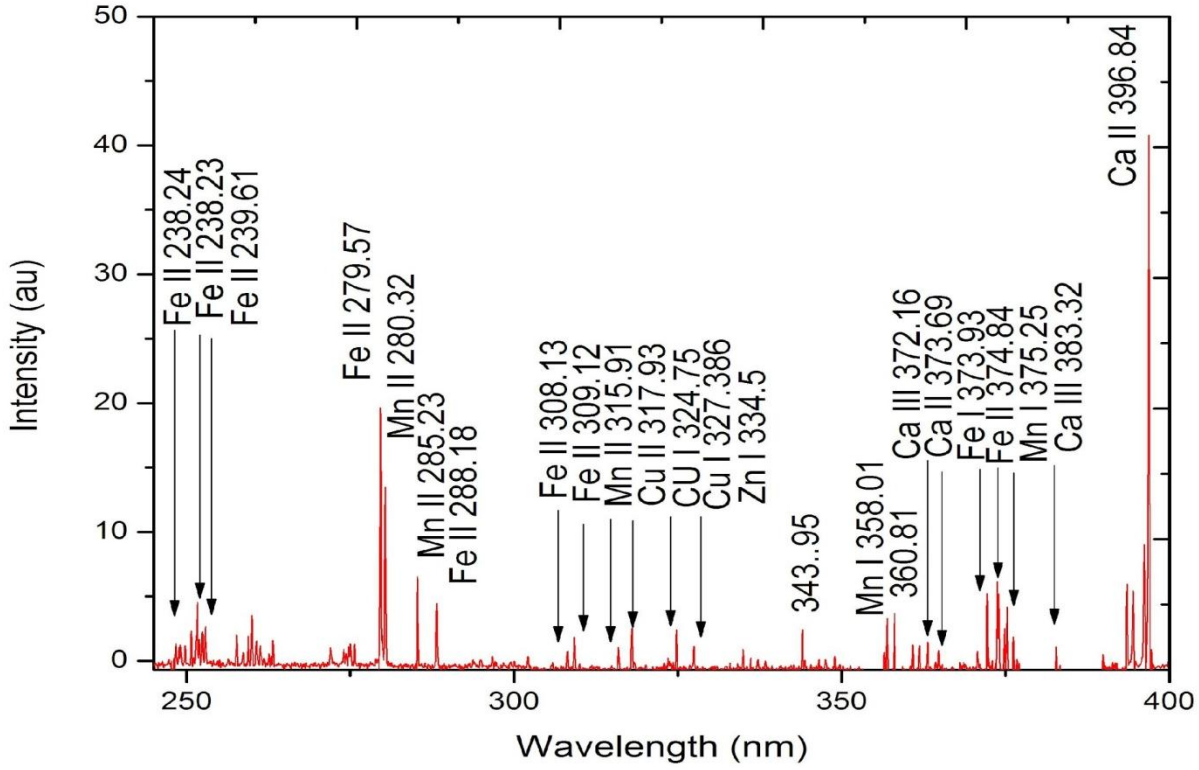

Figure 5.4: LIBS spectrum of breast tissue for the spectral region 250 nm to 400 nm showing lines of Fe, Mn, Cu, Zn and Ca observed in this region.

The visible spectral region of the breast tissue has very little information as compared to the UV region as shown in Figure 5.5. There are few Fe lines and the macro elements. The presence of the latter elements shows the entire composition of a cell is not only made up of

the trace elements. Few but important lines are found within this region. The lines are listed in Table 5.1.

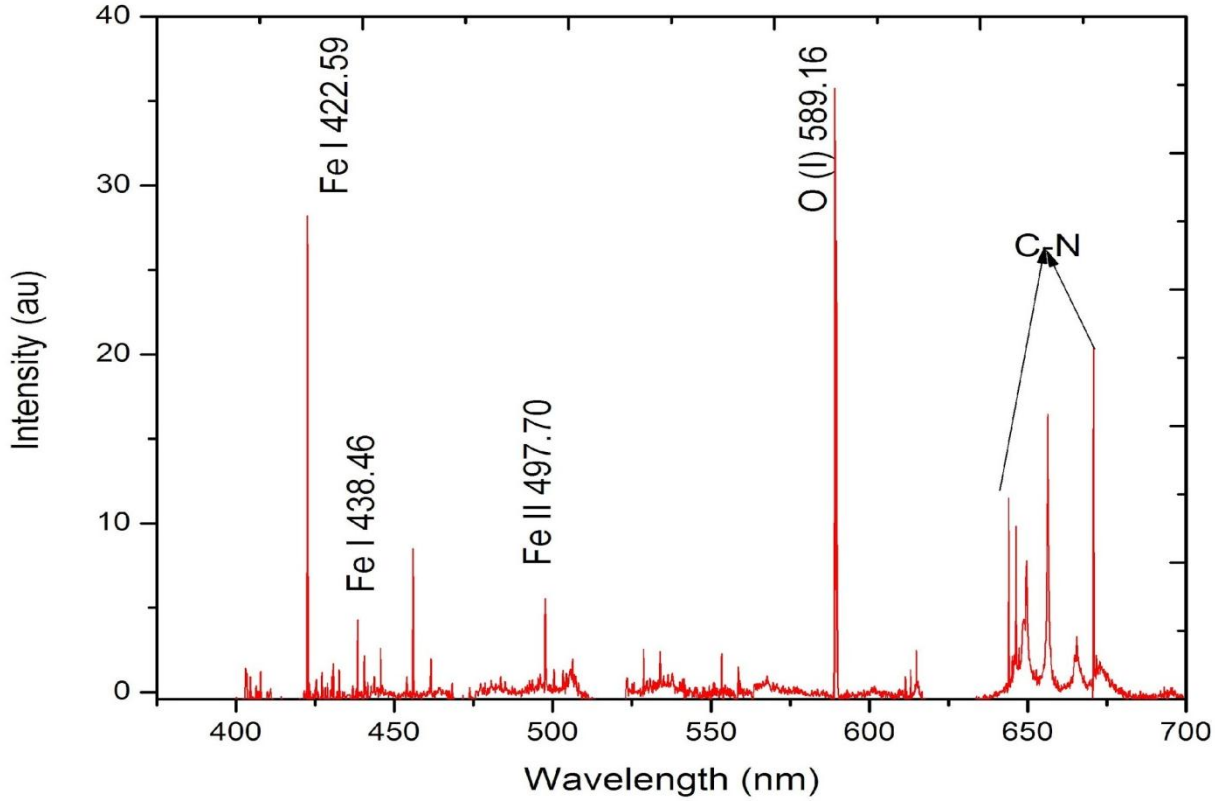

Figure 5.5: LIBS spectrum of breast cancer tissue showing lines of Fe, C-N and O in the visible spectral region from 400 nm to700 nm.

The IR region of the spectrum of the breast tissue is prominently filled with lines of N, O and the C-N bands. Only one Fe line at 786.48 nm was visible. These lines have relatively high intensity values as opposed to the trace elements, which are the elements of interest.

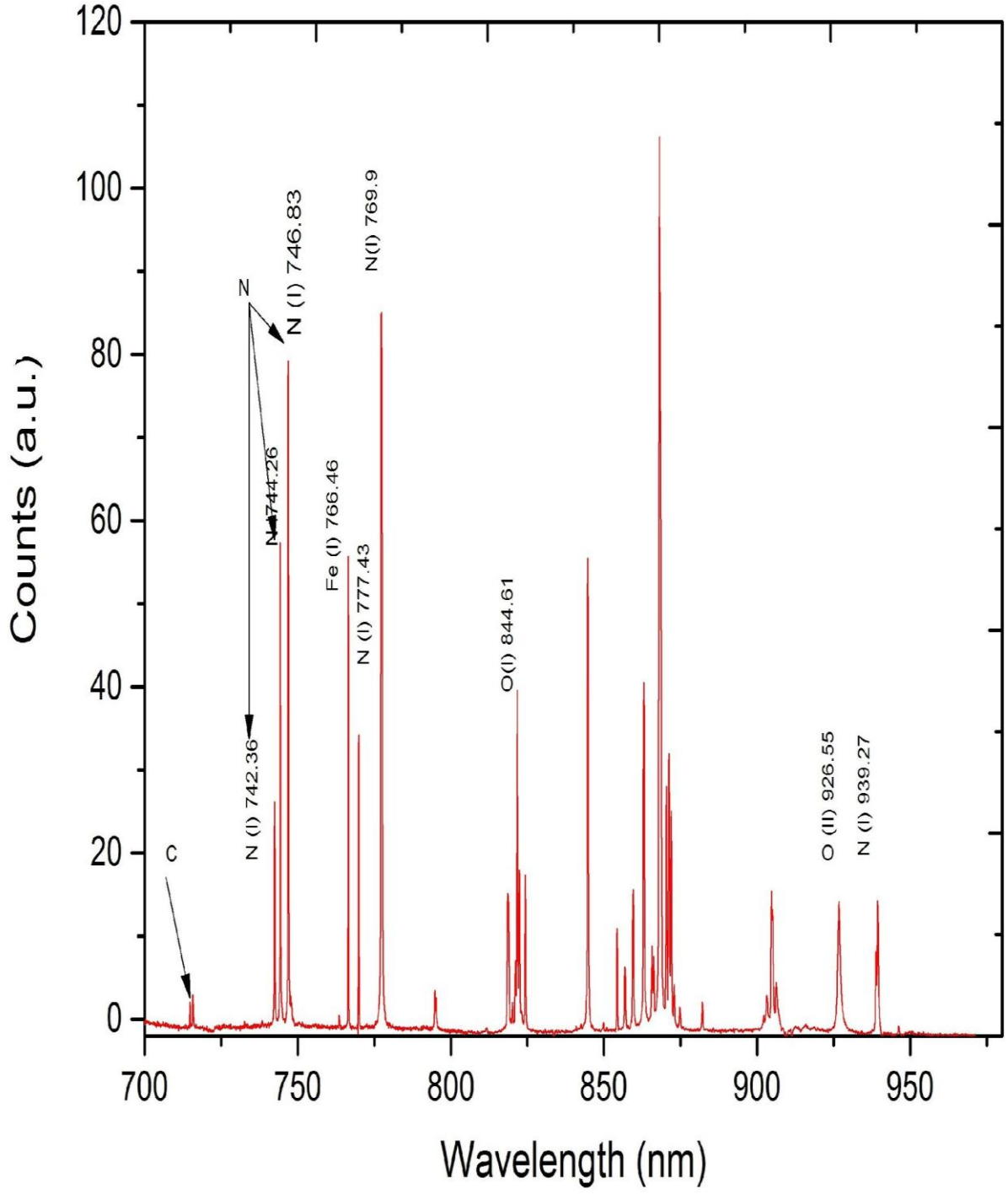

Figure 5.6: LIBS spectrum of breast cancer tissue from 700 nm to 980 nm showing majority of lines as bands and trace elements.

Table 5.1: Spectral lines of Fe, Cu, Mn, Zn and Mg identified in breast, liver and abdominal tissue using LIBS

Element	Breast tissue	Liver tissue	Abdominal
Cu I	324.96 nm	324.96 nm	
	327.36 nm	327.36 nm	
Cu II	317.93 nm	317.93 nm	317.93 nm
Fe I	373.93 nm	373.93 nm	
		372.16 nm	
		327.45 nm	
Fe II	238.23 nm	238.23 nm	
		241.07 nm	
	239.61 nm	239.61 nm	
	279.57 nm	279.57 nm	279.57 nm
	288.18 nm	288.18nm	288.18nm
	308.13 nm		
	309.12 nm	309.18 nm	
	373.93 nm		
	374.84 nm		
Mn I	375.25 nm	375.25	
Mn II	280.32 nm	280.32 nm	280.32 nm

	285.23 nm	285.23 nm	285.23 nm
	315.43 nm	315.43 nm	315.43 nm
		315.92 nm	
		302.10 nm	
		271.99 nm	
		263.20 nm	
		257.61 nm	
Zn	334.501 nm		
	213.857 nm	213.857 nm	213.857 nm
		255.795 nm	
		481.053 nm	
Mg II		279.773 nm 280. 271 nm	
Mg I		285.215 nm	285.21 nm

The spectra obtained are listed in Table 5.1. The spectral regions were subdivided into visible, ultraviolet and the entire spectral region as was the case with the breast tissue. Fig. 5.7 shows the spectrum of the entire wavelength region. The spectrum shows the presence of the trace biomarkers as well as the macro elements. The intensity values of the trace elements are lower than those of the latter elements. The lines observed are for Fe, Cu, Zn, Mn and Mg, the majority of which were those that belong to elements occurring in more than one

oxidation state. In comparison to the breast tissue, the liver tissue has more of Fe lines in the visible region. This is attributed to the activities that take place in the liver tissue. The liver has a task of detoxification in the body. It therefore is more likely to feature more of the heavy metals as it tries to clear the blood of any contaminant. It also synthesizes proteins and carries out metabolism. It therefore requires a lot of oxygen supply by blood whose major component is Fe.

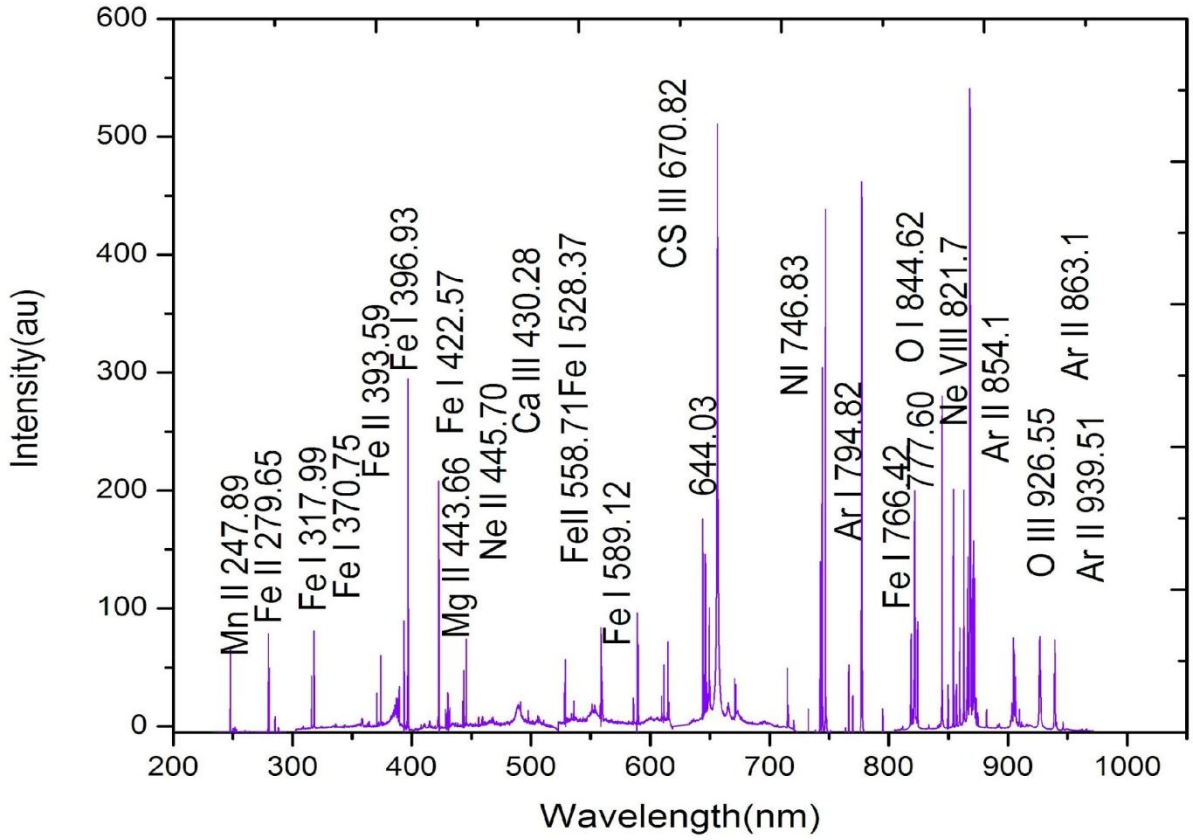

Figure 5.7: LIBS spectrum of liver cancer tissue showing Fe, Mn, Ca and major elements in the entire spectral region from 200 nm to 980 nm.

The bulk of the lines identified in the UV region of the liver tissue belong to Mn and Fe lines. The presence of elements that occur in more than one oxidation state shows the importance of

these elements in cancer diagnostics. Regardless of the type of the soft tissue, LIBS is capable of mining important information through the ablation process.

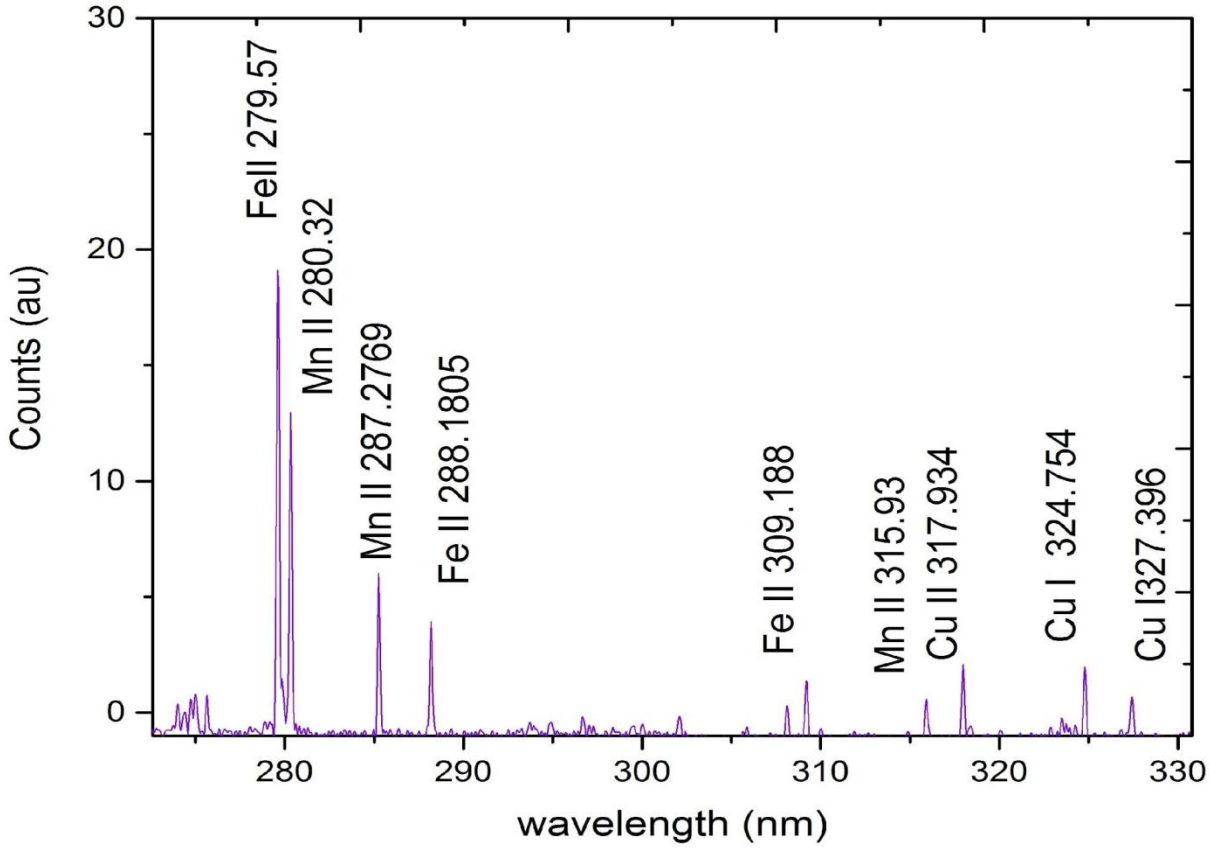

Figure 5.8: LIBS spectrum of liver cancer tissue showing few lines of Cu, Mn and Fe identified in the UV region from 275 nm to 330 nm

The spectrum in the visible region shows a number of Fe and Cu lines. This region does not have a lot of information as compared to the UV. The role of these elements in the liver tissue is similar to the tasks they do in a breast tissue.

The IR region of the liver tissue has very little information with regards to determining the presence of the biomarkers. It however shows the other components of the liver tissue. These elements are present due to the structural difference of a tissue.

Spectral analysis of abdominal tissue was also done using LIBS as shown in the results of the lines identified tabulated in Table 5.1. This tissue had uniquely featuring elements as compared to the ones identified prior in the liver and breast tissues. Besides, the tissue has fewer lines as compared to the two tissues discussed above. Notable are the Fe and Mn lines identified across the spectral region, Fig 5.9 shows the spectrum for the entire spectral region; 200 nm to 980 nm. The lines are listed down in Table 5.1. The occurrence of these lines signifies that a reproducing cell requires vital nutrients supplied by blood which has Fe as a bulk component.

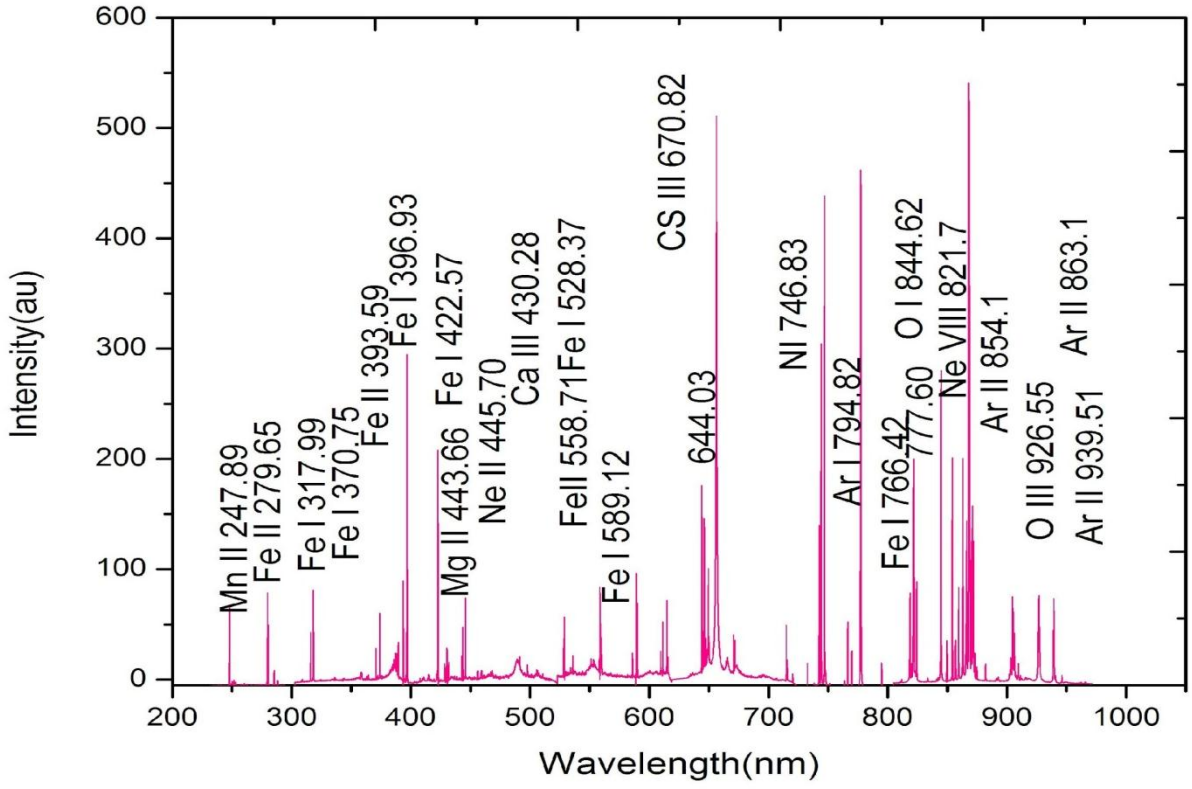

Figure 5.9: LIBS spectrum of abdominal tissue from 200 nm to 980 nm. This figure shows Mn, Mg, Fe, Cu, Ca and macro element lines present in the tissue.

The spectrum in the UV region has the following lines of interest: Fe II at 279.57 nm, 288.18 nm, Mn II at 280.32 nm, 285.23 nm and 315.93 nm and Cu at 317.94 nm. The elements observed in this region are Fe, Cu and Mn. They have very few lines.

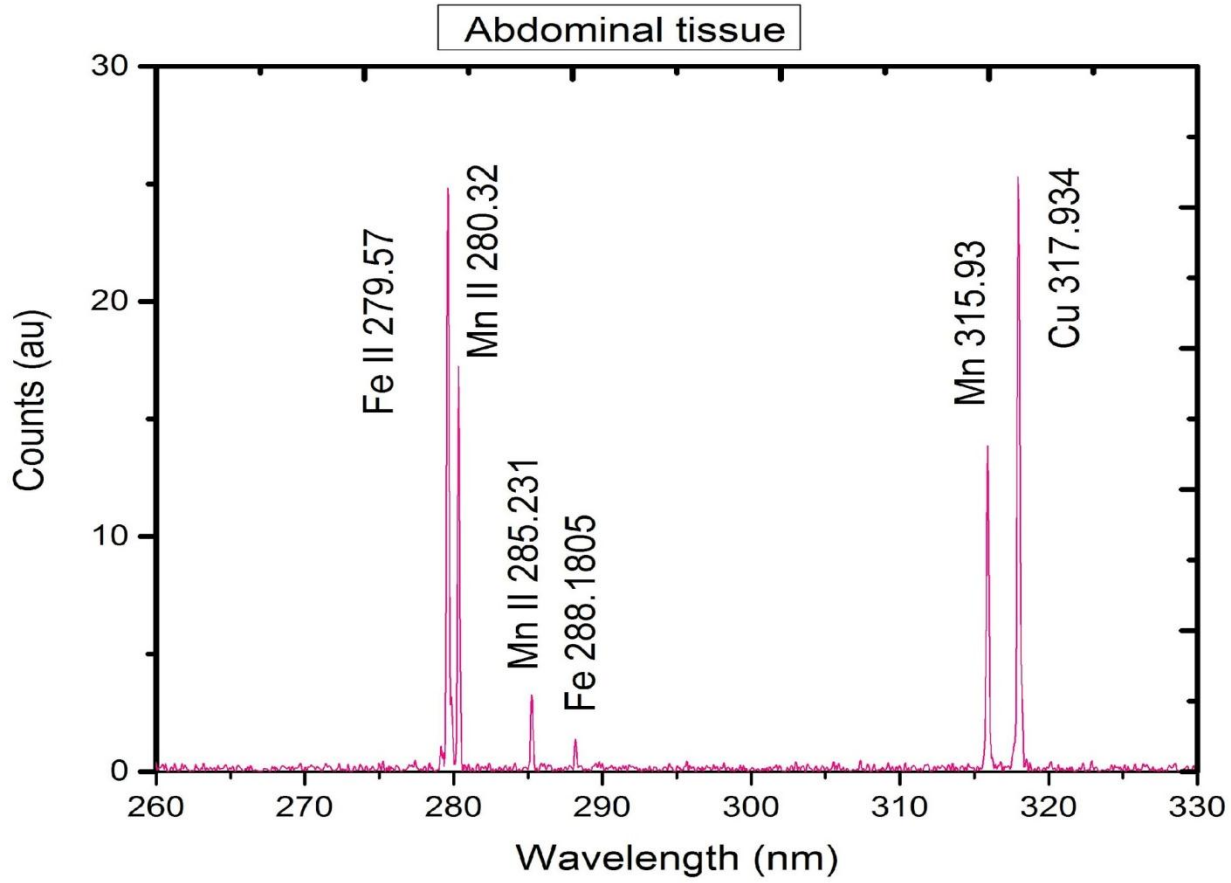

Figure 5.10: LIBS spectrum of abdominal tissue showing Fe, Mn and Cu lines at respective wavelength values from 200 nm to 400 nm.

The VIS region of the spectrum had a few Fe lines, Mn lines Cu lines and nearly no Mg and Zn lines. This region does not have as much information as the UV.

The spectrum of this abdominal tissue within the IR region has C-N bands and the other macro elements. The C-N bands show a possibility of molecular analysis using LIBS. This is

an area being explored and can be used to find important information especially the components of proteins which are important in disease diagnostics.

Table 5.2: Table showing the trace element biomarkers responsible for development of cancer in breast, liver and abdominal tissues.

Lines	Wavelength (nm)
Fe II	248.256 nm
Fe II	
Fe II	373.935 nm
	374.837 nm
Mg II	279.559 nm
Mn II	280.325 nm
Mn II	
Mn II	285.231 nm
	375.25 nm
Cu II	662.537 nm
	777.874 nm
Zn	334.501 nm
	213.857 nm

The following lines were common to the four tissues; Cu lines at 324.96 nm, 327.36 nm and 317.93 nm in breast and liver tissues. Cu 317.93 was present in abdominal tissue. The Fe lines identified in the breast and the liver tissues that are similar are 373.93 nm, 238 nm, 239.61 nm, 279.57 nm, 288.18 nm and 309.12 nm. These two tissues have several Fe lines.

The presence of Fe lines signifies its importance in these tissues that are replicating. Fe is found in blood and since these tissues require nutrients for their growth, blood readily supplies them. Fe is a major component of blood and thus explains its abundance in these tissues. Mn lines observed in the tissues are at 375.25 nm, 280.32 nm 285 nm and 315.43 nm in both liver and breast tissue. These tissues share the last three lines with the abdominal tissue. The abdominal tissue generally has fewer lines. Some of these lines were present in the simulate samples and were used to develop the ANN predictive model; Fe II 279.88 nm, Mn II 280.32 nm and Mn II 285.37 nm. This is an indication that these lines are biomarkers of cancer. They were chosen as the biomarkers from literacy study and upon study of the tissues using LIBS for spectral identification, it was proven that they are present thus making them biomarkers of cancer.

5.2 Design and Validation of ANN Multivariate Calibration Model

5.2.1 Artificial Neural Network Model Development

Simulate samples were used to develop the ANN model using the matlab script [see Appendix II]. Out of 20 samples 60% were used for training, 20 % for prediction and 20% for validation. The spectra signatures of selected biomarkers of Cu, Zn, Fe, Mn and Mg were used as the input data. Fig 5.11 shows an overlay of the spectra to compare the choice of lines as per their sensitivity to change in concentration levels. The lines show an increase in intensity with an increase in the concentration values. This indicates a good choice of lines used in development of the ANN prediction model. The spectra of the lines were obtained using optimal LIBS parameters.

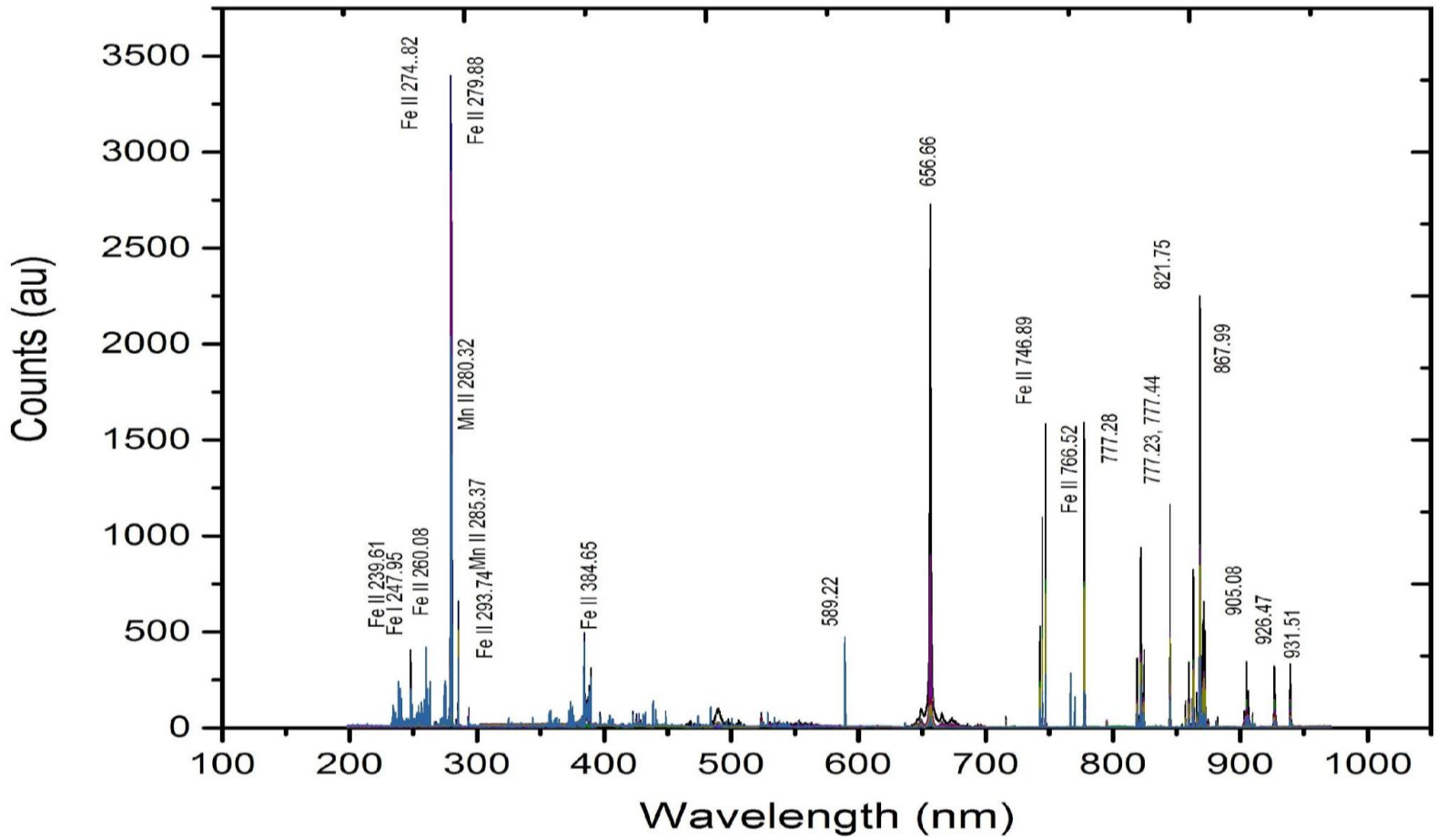

Figure 5.11: LIBS spectral overlay of simulate samples showing the comparison between the concentration values and different elements from which the sensitive lines used to develop the ANN model

Paraffin wax was used as the base matrix for embedding the elements. A spectrum of the matrix was taken and overlaid against the simulate sample spectrum to compare the presence of the elements. In the spectral region 228 nm to 248 nm, the base matrix does not have most of the elements identified in this region. The ones available are in very low concentration values. Figure 5.12 shows an overlay of the spectra with the blank matrix; paraffin wax. The base matrix clearly has no trace of the elements present in the samples. This shows the percentage purity of the base matrix that was used in the research as well as the sensitivity of the lines chosen as seen in the curve.

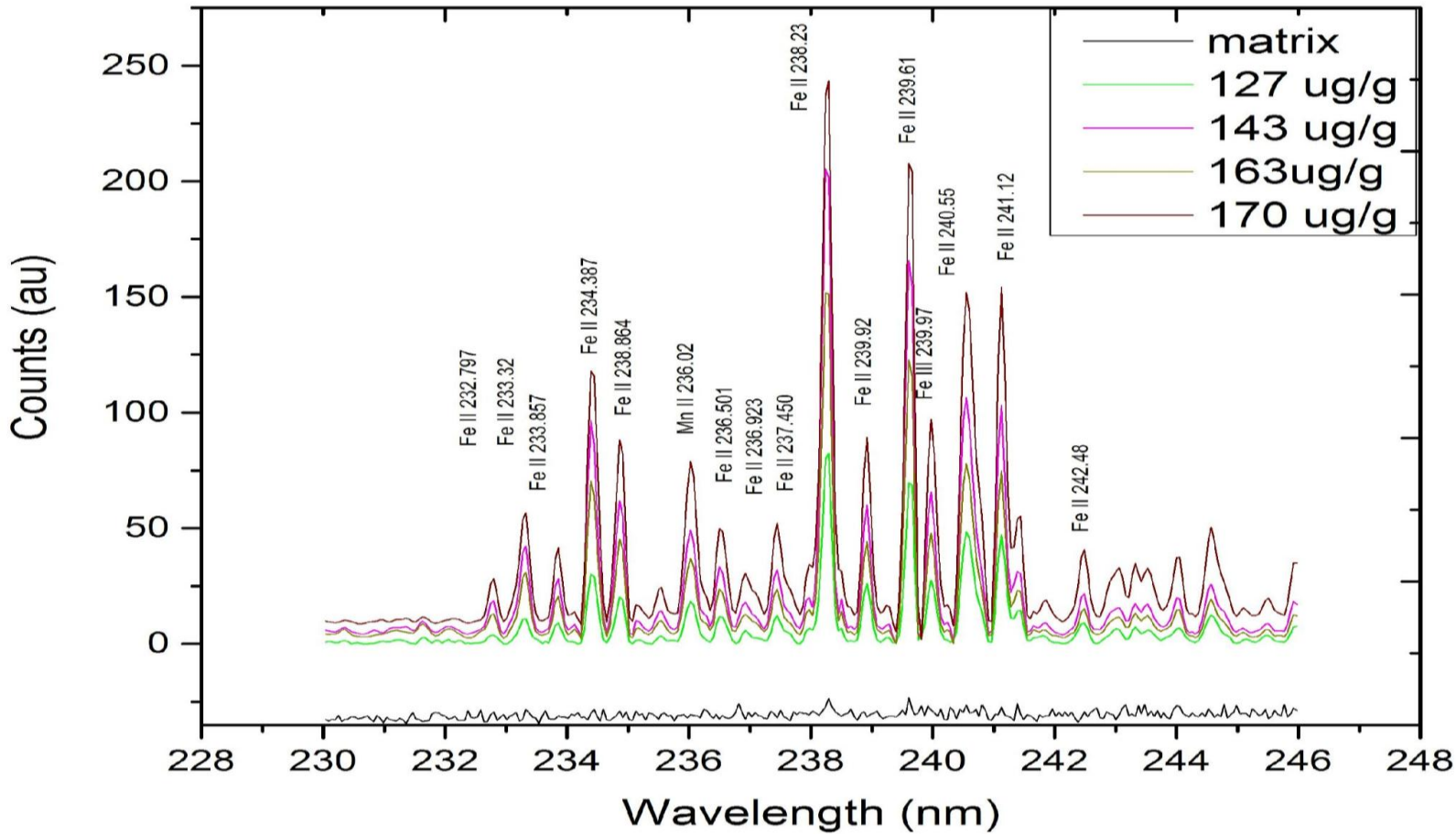

Figure 5.12: LIBS spectral overlay of simulate samples with the blank matrix showing the difference in concentration levels or absence of these elements in the base matrix.

The known concentration that was spiked in the simulate samples was fed into the neural network as the target. The network was then trained with 60 % of the data, 20 % for validation and the remaining 20 % used for testing.

The network mapped the values of intensity on to the corresponding concentration value after

which it validated its training then tested the performance of the model. The model was trained several times to minimize the error in prediction. The standard reference material, oyster tissue, was used to validate the calibration model. The concentration values of these trace elements in oyster tissue are known. LIBS was used to obtain the spectral data then the model was used to predict their concentration values then a comparison made against the known values.

The ANN model trained for prediction of concentration of the trace biomarkers and validated using oyster tissue, as the standard reference material gave the best validation performance of MSE of 13.63 as shown in Figure 5.13. The figure shows the value of MSE reducing with as training, validation and testing was being done.

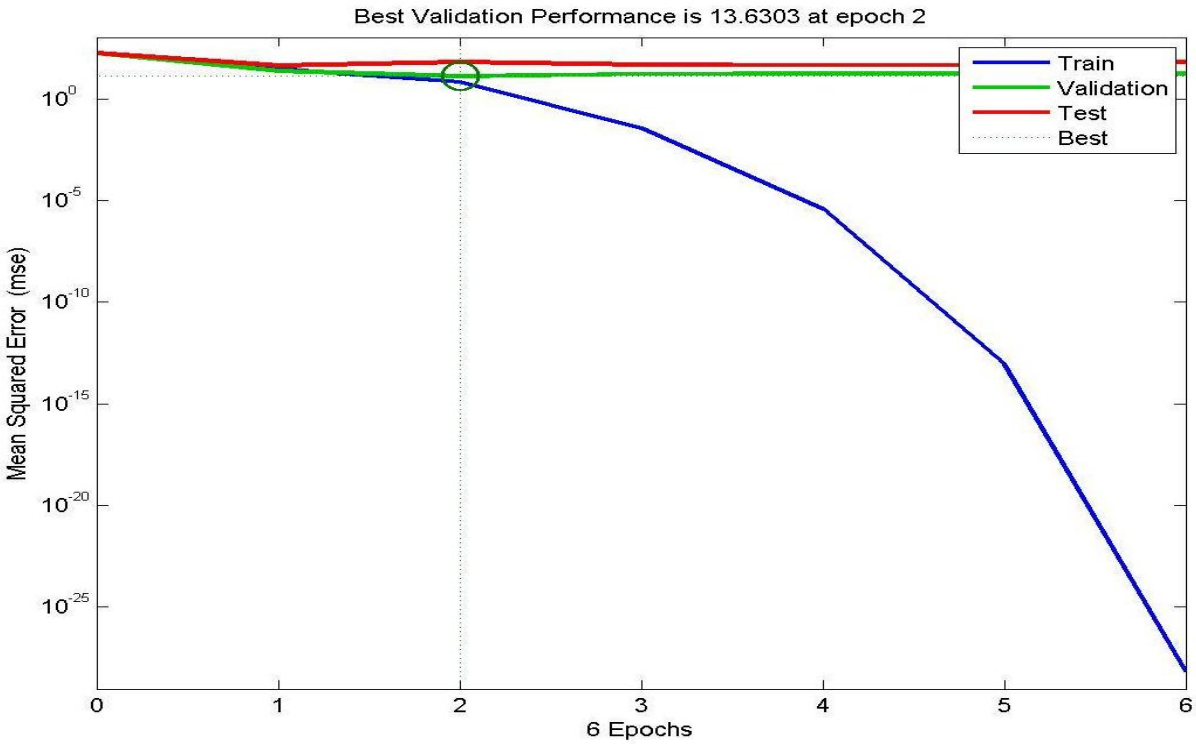

Figure 5.13: ANN performance plot of simulate samples showing the best validation of RMSE 13.6306 at epoch 2.

The model developed had regression curves for training, validation, testing and overall curve as shown in Figure 5.14. The model trained very well showing R^2 values of 0.9999 for training, validation 0.948, testing 0.950 and overall 0.985. These values signify the square of the error in the difference between the measured concentration and the predicted concentration. Since the value of this error is close to one, it means that the measured concentration value was mapped onto a value close to itself.

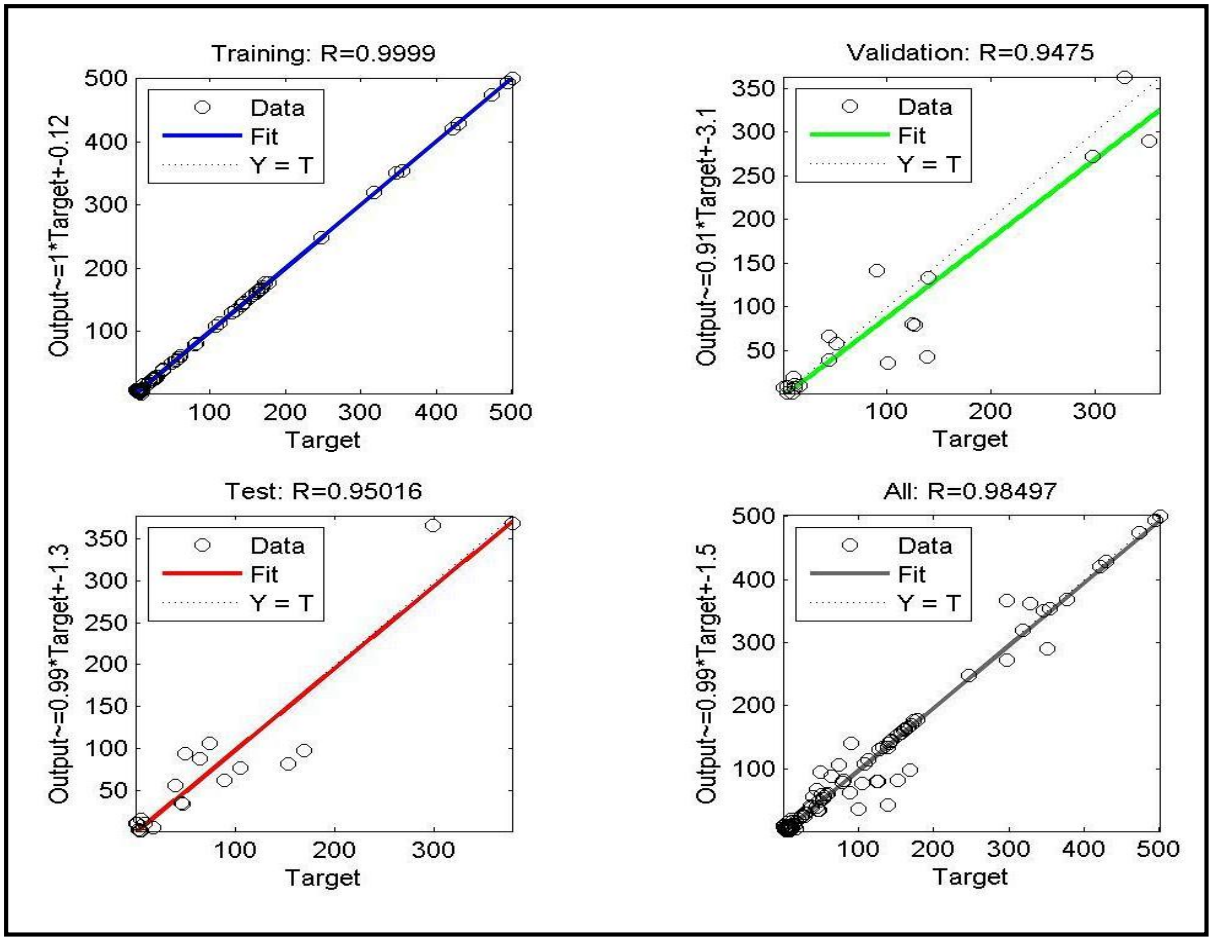

Figure 5.14: Regression curves of the ANN prediction model showing training curve of R^2 value of 0.9999, validation curve of R^2 value of 0.9475, testing curve of R^2 value of 0.9502 and overall regression curve of R^2 value of 0.985.

The model was well trained and used for prediction of concentration values in the simulate samples. The model predicted the concentration values of the elements of interest giving R^2 values closer to 1 as possible as shown in Appendix III Table I, II and III. The ANN prediction model was then used to predict the concentration values in the human tissue samples and the cell lines (Hep- 2 and Lewis Lung) then compared with the values obtained

in the literature. These values were also used to find the ratios between the elements as well as the correlations between them.

The regression curves for Cu, Fe, Mg, Zn, Mn and Mg, showing the relationship between predicted concentration and known concentration were plotted as shown below in figures 5.15 to 5.19.

Regression curve for Fe has R^2 value of 0.993 and an MSE of 13.418. R^2 correlation coefficient is an indicator of a very close relationship between the known concentration and the predicted concentration of the 50 % of the samples used for prediction. Figure 5.15 shows the regression curve of predicted concentration against known concentration of Fe.

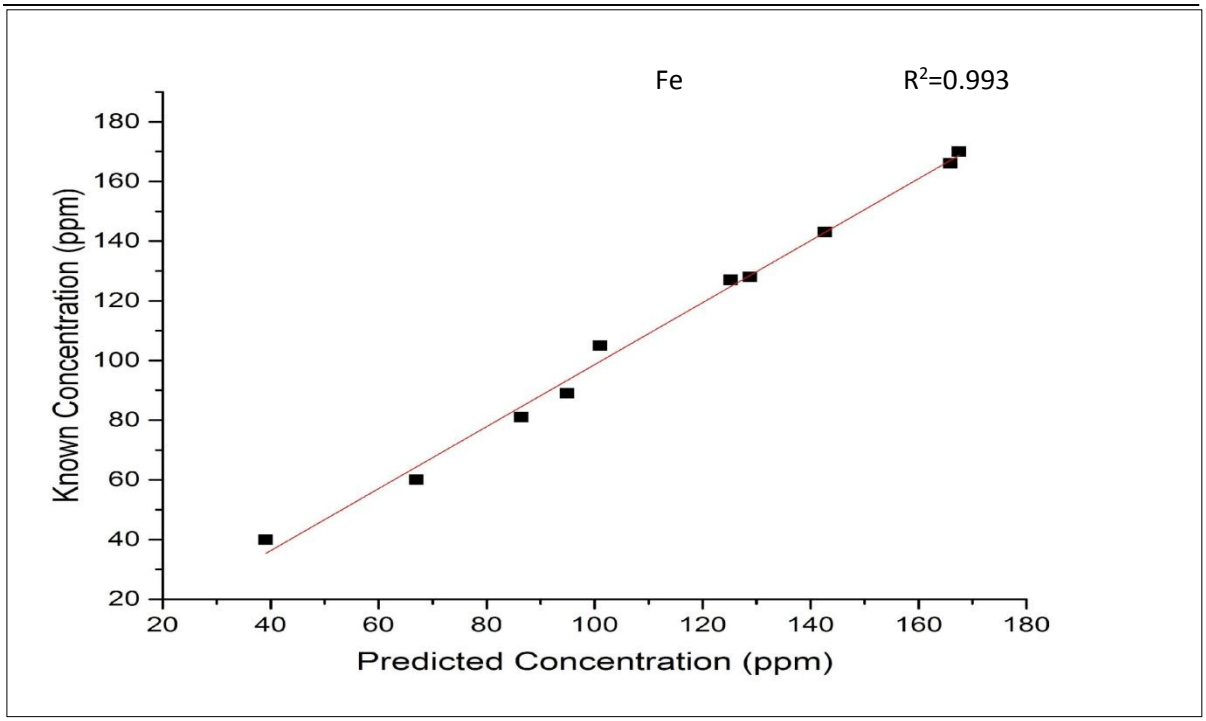

Figure 5.15: ANN regression curve of predicted concentration versus known concentration of Fe of R2 value of 0.993.

The regression curve of predicted concentration against known concentration for Mn has R^2 value of 0.918 and MSE of 5.118 (Figure 5.16). The latter is a good indicator of minimal error in the developed prediction model. The results imply that ANN model can be used to predict the concentrations of Mn in unknown samples with an accuracy of 92 % as indicated by the R^2 and MSE values.

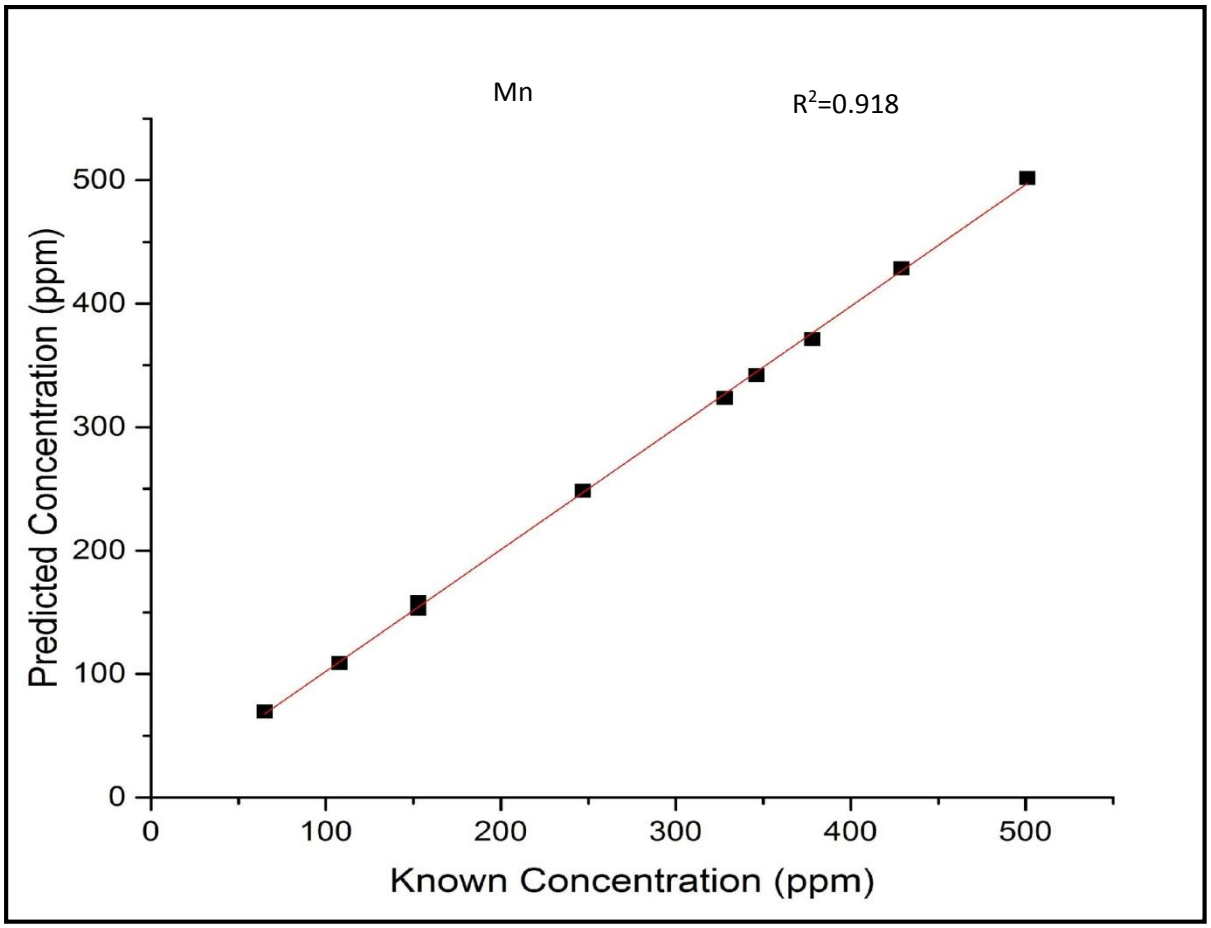

Figure 5.16: ANN regression curve of predicted versus known concentration of Mn with R^2 value of 0.918

The regression curve of predicted against known concentration of Mg shows R^2 value of 0.999 and a mean square error of 11.326 (Figure 5.17). The model predicted the

concentration values of Mg well. The correlation coefficients show a close relationship between the known and predicted concentration values while the MSE error shows the error in relating the two variables. The R^2 value is an indicator of close relationship between predicted and known concentration. This proves the effectiveness of the model that has been developed.

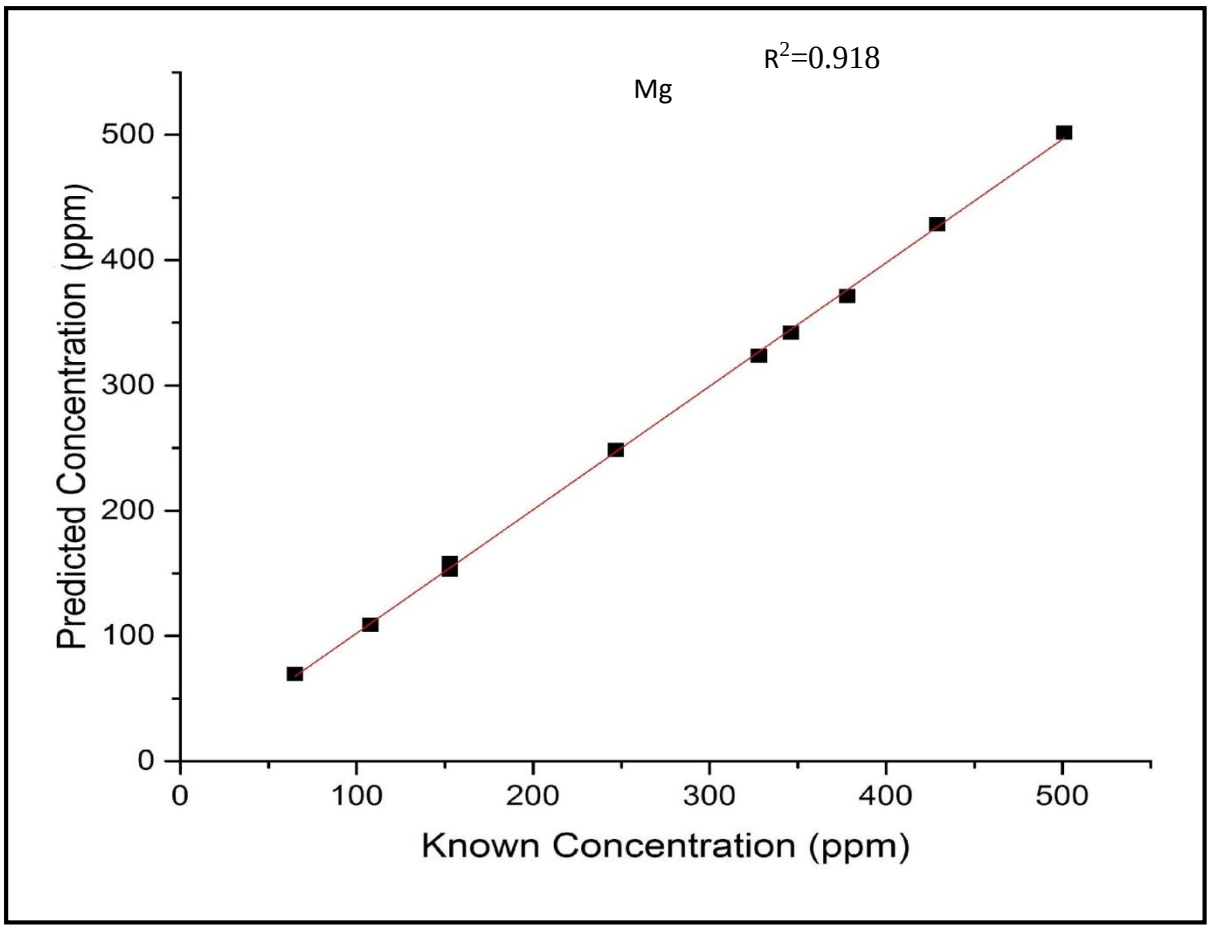

Figure 5.17: ANN regression curve of predicted concentration versus known concentration of Mg of R^2 value of 0.999.

The regression curve for Cu had R^2 0.94742, MSE of 5.887. These values are summarized in table 5.3. The model was also utilized in producing the regression curve of Cu and the results show R^2 value of 0.94742. The MSE achieved with the model is 5.887.

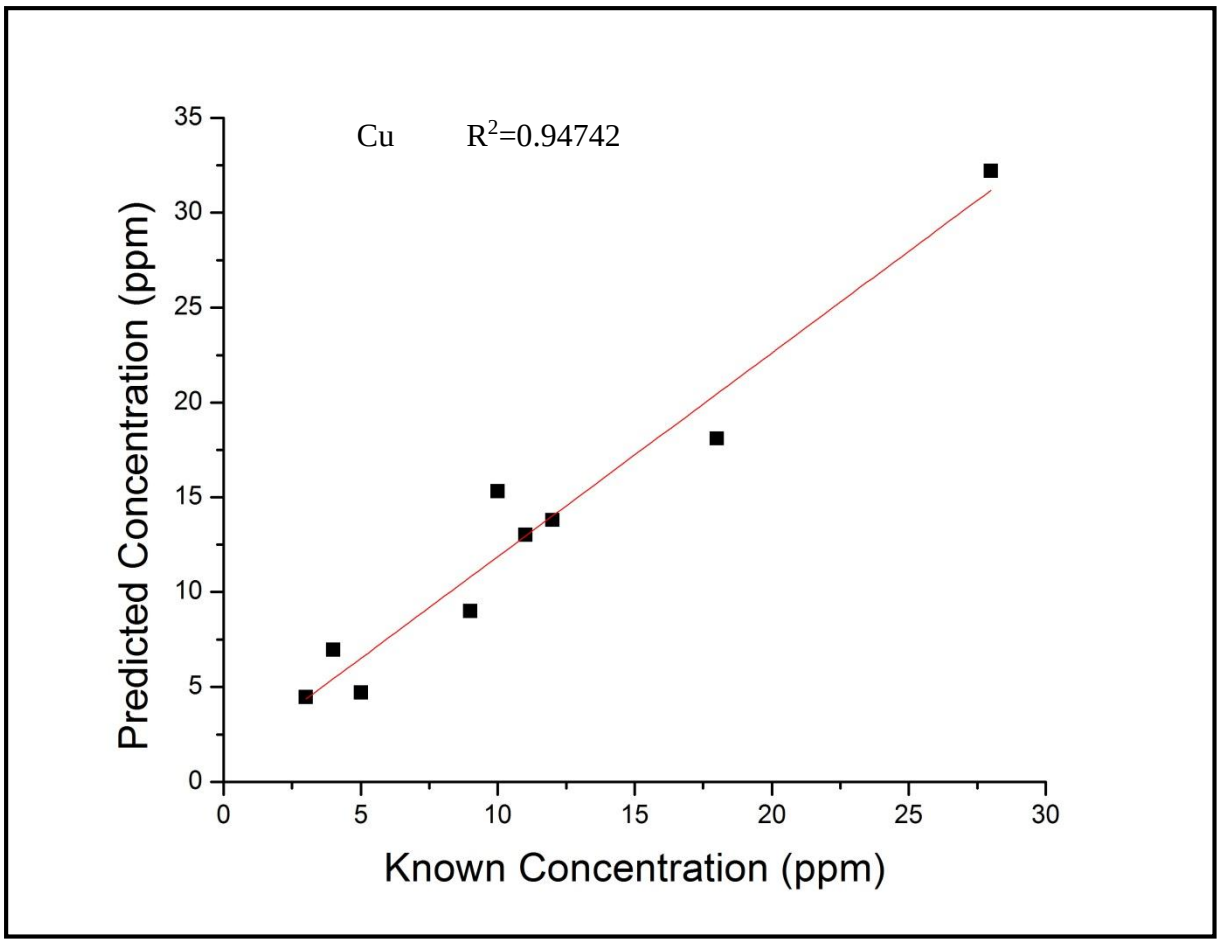

Figure 5.18: ANN regression curve of predicted concentration against known concentration of Cu with R2 value of 0.94742 using simulate samples

The regression curve of Zn is shown in Figure 5.19. The MSE was 9.082, the Pearson's correlation 0.989 and 0.971 R^2 value. Furthermore, the model when used to produce a regression curve for Zn, the R^2 value achieved was 0.971 and an MSE of 9.082. This is very positive of the model developed. It can therefore be relied on to predict with more than 80 % confidence the concentration levels of trace elements in a tissue.

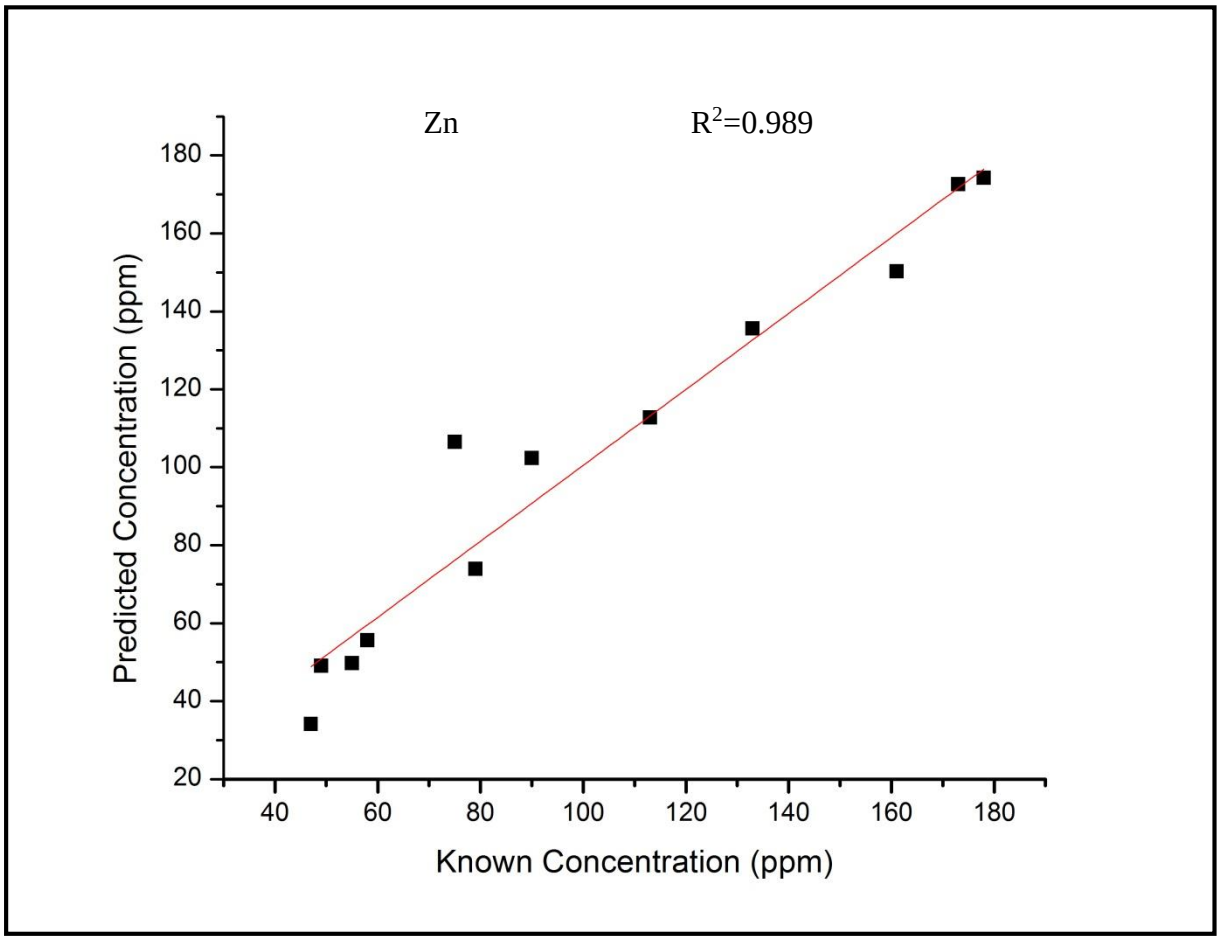

Figure 5.19: ANN Regression curve of predicted concentration against known concentration of Zn with R^2 value of 0.989 using simulate sample.

Table 5.3 below shows a summary of the RMSE, Pearson's correlation coefficient and the R^2 values for the five elements as per the model developed.

Table 5.3: ANN model performance of Mn, Mg, Fe, Cu and Zn regression curves in terms of RMSE and R^2 values.

Element	RMSE %	R^2
Fe	13.418	0.993
Mn	5.118	0.918
Mg	11.326	0.999
Cu	5.887	0.871
Zn	9.082	0.971

A summary of the performance of the ANN regression curves for Cu, Fe, Mn, Mg and Zn showing the respective values for MSE and R^2 in Table 5.3. The model is suitable for prediction of concentration values of these elements in tissues as evident in the good performance it portrays.

5.2.2 Validation of ANN model

The model, once developed, was validated using a standard reference material, NIST 1566B, oyster tissue. This was done by making pellets from the oyster tissue powder using a hydraulic machine and ablating it using LIBS. The spectral values were then fed into the network for prediction of the concentration values of Cu, Mn, Mg, Zn and Fe. These values were fed into the network as input and simulated by the network to predict the concentration values that were compared against the standard values as shown in Table 5.4. The model predicted the concentrations of Fe, Mn, Cu, Zn and Mg. The percentage error is less than 5% for all the elements under study as shown in table 5.4.

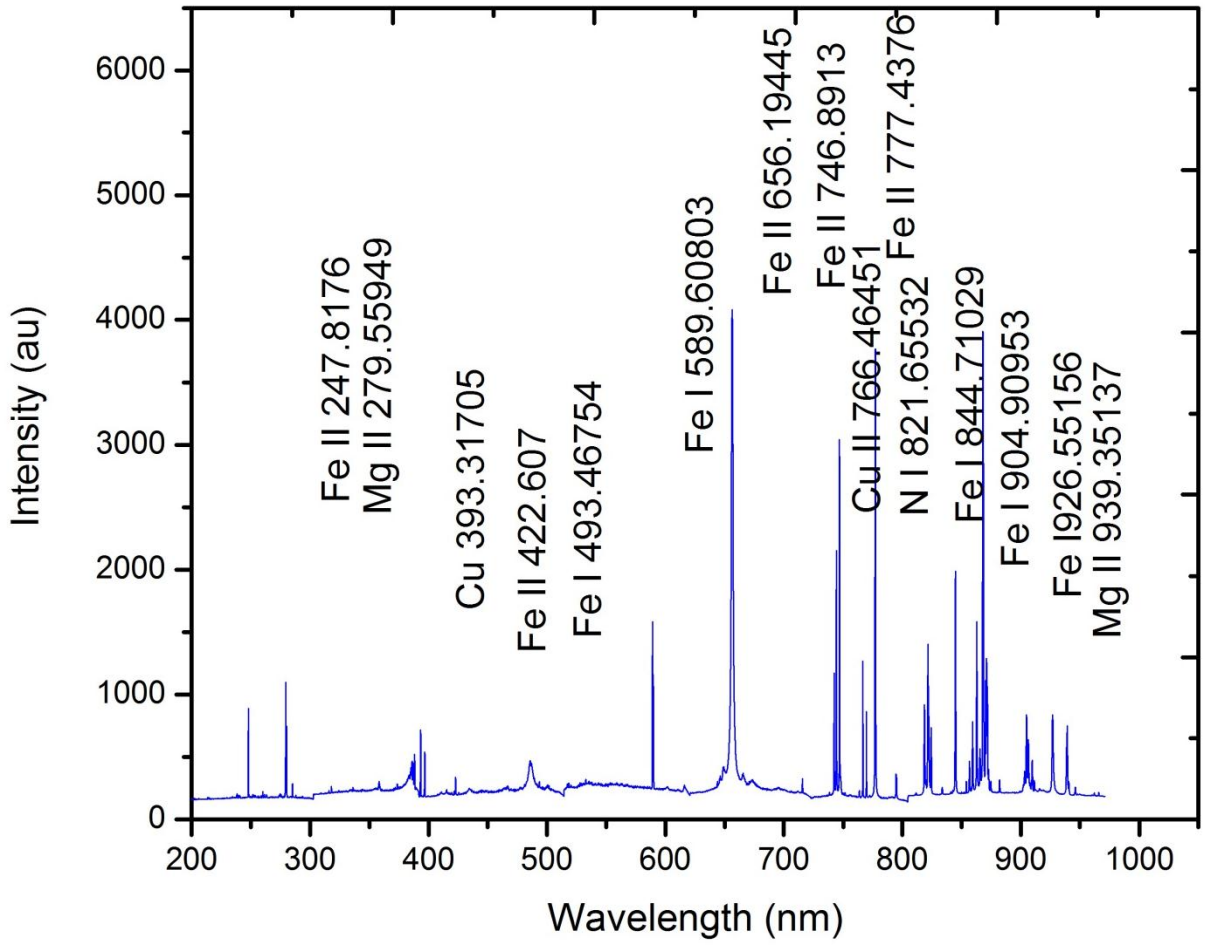

Figure 5.20: Spectrum of oyster tissue showing the respective lines of Cu, Fe, Mn and Mg lines. Oyster tissue is the standard reference material used for validation of the ANN prediction model created. It has known concentration values of the trace elements.

Figure 5.21 shows the occurrence of the trace elements whose concentration values were predicted using the ANN prediction model developed. The prediction was accurate since all the values were within the range of prediction error given from NIST.

Table 5.4: ANN calibration model showing certified and predicted concentration values of Fe, Mn, Mg, Zn and Cu in oyster tissue (NIST 1566B).

Element	Oyster Tissue Standard Concentration (µg/g)	Predicted Concentration (µg/g)	Percentage deviation
Fe	205.8 ± 6.8	213.8 ± 28	-3,89
Mn	18.5 ± 0.2	18.1 ± 0.9	2.16
Mg	1085 ± 23	1068 ± 117	1.57
Cu	71.7 ± 1.6	70.16 ± 4.2	2.15
Zn	1424 ± 46	1418 ± 127	0.42

Oyster tissue was used to validate the ANN prediction model by studying the concentration levels of the elements using this ANN predictive model.

5.3 Prediction of Concentration Values of Cu, Zn, Mn, Mg and Fe using the ANN Model

The ANN model developed was used to predict the concentration values of the trace elements under study. Prediction of concentration values of Cu, Fe, Mn, Mg and Zn using the model in breast, liver and abdominal tissues was done. The results are tabulated in Table 5.5 below.

Table 5.5: Predicted concentration values of Cu, Mn, Mg, Fe and Zn in breast, liver and abdominal tissues using ANN

Tissue	Mg	Mn	Fe	Cu	Zn
Liver 1	87.9 ±9.7	18.6 ± 0.9	58.9 ± 7.7	17.6 ±1.1	39.6 ± 3.6
Liver 2	457.1 ±50.3	9.3 ± 0.5	121.4 ±15.8	5.3 ± 0.3	52.1 ±4.7
Abdominal	525.3±57.8	10.0 ± 0.5	111.2±14.5	7.1 ± 0.4	88.6 ± 5.3
Breast	218.3 ±24.0	15.3 ± 0.8	100.8 ±13.1	11.6 ± 0.7	52.5 ± 3.2

Liver 2 and abdominal samples, which were classified under malignant tissues, have higher concentrations of Fe and relatively low concentrations of Cu as compared to the other liver tissue, which is benign. Moreover, they have a lower concentration of Mn as compared to their counterparts. These alterations in the concentration values can explain the increased need of Fe in a highly proliferating tissue due to the constant demand of supply of the nutrients.

5.3.1 Elemental Ratios Analysis

Ratios of these trace elements were calculated to determine the relationship between the five elements for all the tissues. The results are presented in Table 5.6.

Table 5.6: Ratios of concentration levels of Cu, Mn, Mg, Fe, and Zn in liver, breast and abdominal tissues

Ratio	Liver 1	Liver 2	abdominal	Breast
Mg/Mn	4.717749	49.29347	52.55037	14.29169
Mn/Fe	0.316397	0.076372	0.089912	0.151533
Fe/Mg	0.669935	0.265629	0.211645	0.461751
Mg/Cu	4.980876	86.90273	74.34809	18.77795
Mg/Zn	2.216026	8.771981	5.926272	4.157763
Mn/Cu	1.055774	1.762966	1.414797	1.313907
Mn/Zn	0.469721	0.177954	0.112773	0.290922
Cu/Fe	0.299683	0.04332	0.063551	0.11533
Zn/Fe	0.673586	0.429167	0.797278	0.520873
Cu/Zn	0.444907	0.10094	0.07971	0.221417

From the ratios the concentration of Cu: Mn is 1:1 this shows that they are in equal amounts in liver, breast and abdominal tissues whereas Fe is more than Cu, Mn and Zn in all the tissues showing the significance of Fe in a proliferating tissue as compared to the other elements. The ratio of Cu to Fe indicates presence of large amounts of Fe in the tissues. Iron has vital functions in tissues. It enables the function of vital iron- and hemoglobin -containing enzymes, including mitochondrial enzymes that are involved in respiratory complexes, enzymes involved in DNA synthesis and the cell cycle, detoxifying enzymes such as peroxidase and catalase, and many more. Therefore, iron is essential for cell replication, metabolism and growth. However, the ability to gain and lose electrons - the very attribute

that makes iron useful enzymatically- also enables iron to participate in potentially detrimental free radical-generating reactions.

Zinc is known to be an essential component of proteins that bind DNA, is an enzyme co-factor and is an antioxidant and is involve in DNA repair. Lack of enough zinc can contribute to single- and double-strand DNA breakage and oxidative alterations to DNA that elevate the risk for cancer development. (Emily, 2004)

5.3.2 Correlations of the Concentration of Elements

The correlation values in Table 5.8 show the relationship of the trace elements with regards to their contribution to development of cancer in tissues. There is a very strong negative correlation between Cu and Fe which means that a large amount of Fe leads may result in a decrease in Mn and vice versa. The positive correlation between Mn and Fe and Mg and Fe is very weak.

The ANN model developed was used to predict the concentration values of these elements in the cell lines and the results are shown in Table 5.8. The concentration of Fe is required more by a cell line in advanced stage as compared to one in the early stages. This is evident in the increasing amount of Fe as the cancer progresses.

Magnesium concentrations do not show any particular trends along the stages of development. On the other hand, concentration values of Fe increase from stage one to five non- significantly. Manganese on the other hand has the same value for stage 1 and 2. This shows that the model could not distinguish between the two. The variation in the concentration is very minimal such that the two stages are almost a replica of one another.

Concentration levels of Cu are generally dropping from stage 1 to the last stage. Finally, the concentration values of Zn do not show any significant trends.

Table 5.7: Predicted concentration values of Cu, Fe, Mg, Mn and Zn in Hep-2 cell lines using the ANN prediction model

Element	hep stage 1	hep stage 2	hep stage 3	hep stage 4	hep stage 5
Mg	365.0 ± 40.2	364.6 ± 40.1	372.0 ± 40.9	300.6 ± 33.1	360.0 ± 39.6
Fe	63.1 ± 8.6	64.6 ± 8.4	70.3 ±9.1	71.9 ± 9.3	73.2 ± 9.5
Mn	18.7 ± 09	18.7 ± 09	20.8 ± 1.0	18.1 ± 0.9	29.3 ± 1.5
Cu	18.3 ± 1.1	19.6 ± 1.2	7.0 ± 0.4	4.0 ± 0.2	7.4 ± 0.4
Zn	33.6 ± 3.0	83.2 ± 7.5	49.8 ± 4.5	50.2 ± 4.5	49.8 ± 4.5

Table 5.8: Pearson's correlation coefficient in Hep 2 Cell Lines

element	Mg	Mn	Cu	Fe	Zn
Mg	1	0.277	0.541	-0.481	0.0856
Mn		1	-0.339	0.593	-0.126
Cu			1	**-0.938**	0.311
Fe				1	-0.133
Zn					1

5.4 Prediction of Concentration Levels of the Trace Elements in Lewis Lung Cell Line

The Lewis lung cell line was also subjected to the ANN predictive model. Prediction of Fe seems to be increasing from stage 1 to stage 5. Cu and Zn on the other hand have relatively the same amount. Since concentrations of Cu are very low, the model detected and displayed

the small differences in the concentration values. Manganese levels increase during the first three stages and then decrease in the last two stages. The levels of magnesium increase from stage 1 to stage 4 and then drops in the last stage. The results are shown in Table 5.9 below.

Table 5.9: Predicted Concentration values of Cu, Fe, Mg, Mn and Zn in Lewis Lung cell lines using ANN prediction model

	lung stage 1	**Lung stage 2**	**lung stage 3**	**lung stage 4**	**lung stage 5**
Mg	332.3 ± 36.6	358.7 ± 39.5	39 1.4 ± 43.1	427.9 ± 47.1	346.2 ± 38.1
Fe	74.6 ± 9.7	77.9 ± 10.1	80.2 ± 8.8	78.9 ± 10.3	78.4 ± 10.2
Mn	36.8 ± 1.8	42.3 ± 2.1	47.6 ± 2.4	34.0 ± 1.7	31.8 ± 1.6
Cu	7.4 ± 0.4	7.4 ± 0.4	7.4 ± 0.4	7.4 ± 0.4	7.4 ± 0.4
Zn	49.8 ± 19.4	49.8 ± 19.4	49.8 ± 19.4	49.8 ± 19.4	49.8 ± 19.4

Table 5.10: Pearson's correlation coefficient in Lewis Lung Cell Lines

element	**Mg**	**Mn**	**Cu**	**Fe**	**Zn**
Mg	1	0.116		0.683	
Mn		1		0.349	-0.126
Cu			1	**-0.938**	0.000
Fe				1	
Zn					1

The relationship between Cu and Iron shows a strong negative correlation while that of Mg and Fe is a weak positive correlation.

5.5 Multivariate Chemometric Exploratory Analysis

Exploratory data analysis was done on the simulate samples [see Appendix III], liver, abdominal, breast tissues and the cell lines. Characterization and differentiation of cancer in liver, breast, abdominal and cancer cell lines was performed using PCA and SVM. This analysis is essential in determining the relationship if any between respective samples. PCA was used for unsupervised grouping of the tissues and the cell lines and it produced results clusters which are discussed below.

5.5.1 Classification **of Liver, Breast and Abdominal Cancer Tissues Using PCA**

In this research, PCA was done on the different clusters between the breast, liver and abdominal tissue and analysis of the clusters done. The tissues consisted of breast, abdominal and liver tissues from different patients. The tissues B1, B2 and L2 in Figure 5.23 are leaning more towards PC1 which contributes 97 % and the lines responsible are shown in the loadings plot. The numerical values of the loadings plot were thus plotted and the lines influencing the classification identified as Mg, Fe, Mn, Cu and Zn as illustrated in Figure 5.24. The lines observed from the loadings plot are Fe lines at 386.55 nm, 386.55 nm, 387.29, Cu 388.42 and Mn 247.89 and 389.76 nm. The lines identified are of those elements that occur in higher speciation states. This implies that speciation has a role to play in the classification of these tissues as evident in the presence of these lines in the loadings plot. These elements can thus be identified as biomarkers of cancer.

Classification of liver, breast and abdominal cancer tissues shows a liver tissue, breast tissue and abdominal tissue clustered together while the other two breast tissues and a liver tissue in one group. Some of the tissues are benign while others malignant.

Breast, liver and abdominal samples from different patients were distinguished using principal component score values, and coefficient loadings plot allowed chemical interpretation of the score clusters. These results were compared with the histopathological results which agree that the group with abdominal tissues is malignant while the other group benign. The clustering of the tissues was influenced by PC1 which contributed 92 %. There was one breast tissue and it was divided into three (i.e. tissue from the same patient). A section of the breast tissue that was tarnished was labelled as B3. In the clusters, B1 and B2 have clustered while B3 is on the other end. B3 was had tarnish in colour and was reported to be malignant by the histopathological report.

The PCA was able to distinguish the tissues and classify the replicates of each of the liver, abdominal and breast tissues. The breast, liver and abdominal tissues were clustered separately based on the part of the body from which it is obtained. When the technique is carried out on concentration values of the trace elements, it yields results similar to these.

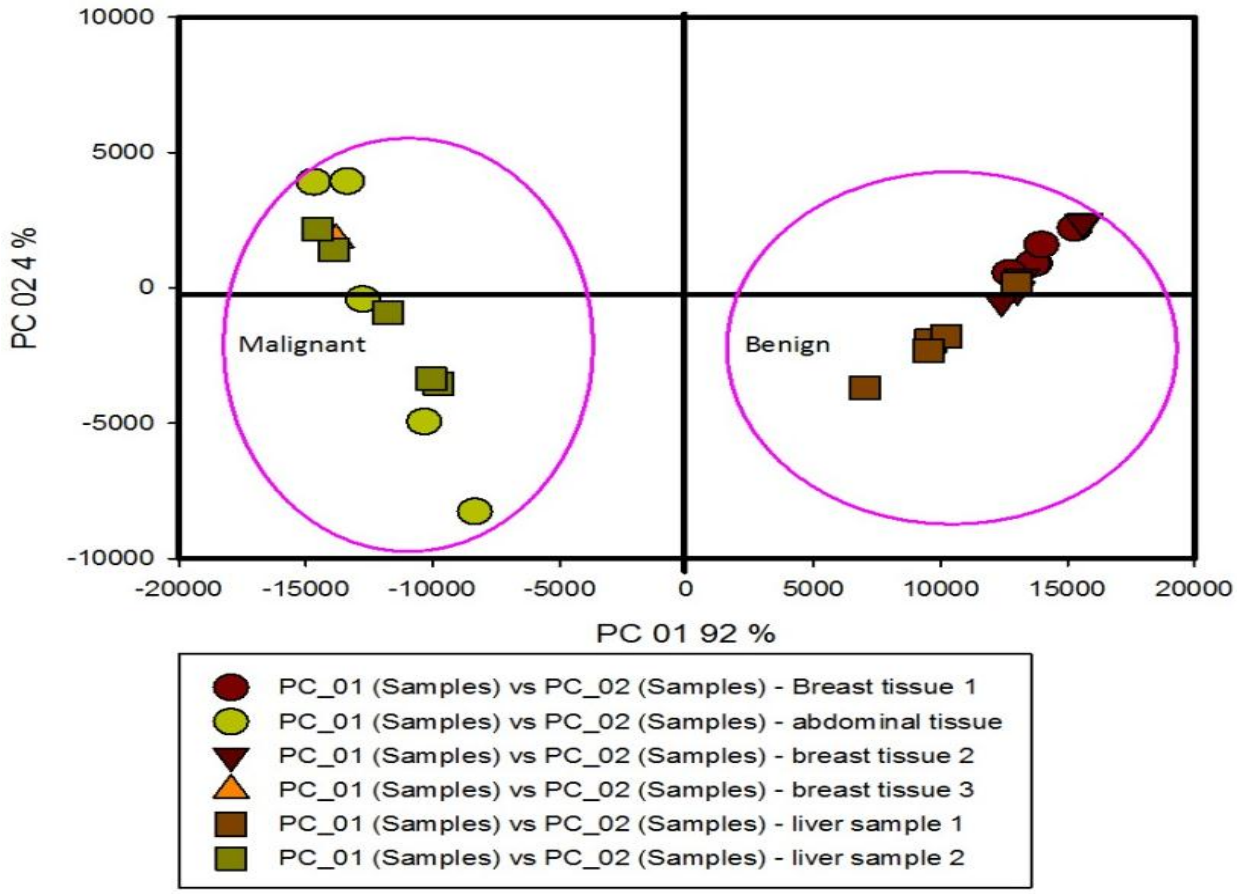

Figure 5.21: Classification of liver, breast and abdominal tissues using PCA. The figure shows the scores plot displaying clustering of breast, liver and abdominal tissues into two groups that can be identified as benign and malignant

This figure 5.21 shows unsupervised learning carried out using exploratory analysis. It shows the existing patterns; the ability of PCA to differentiate the tissues with regards to the part of the body it was obtained as well as differentiating the health status.

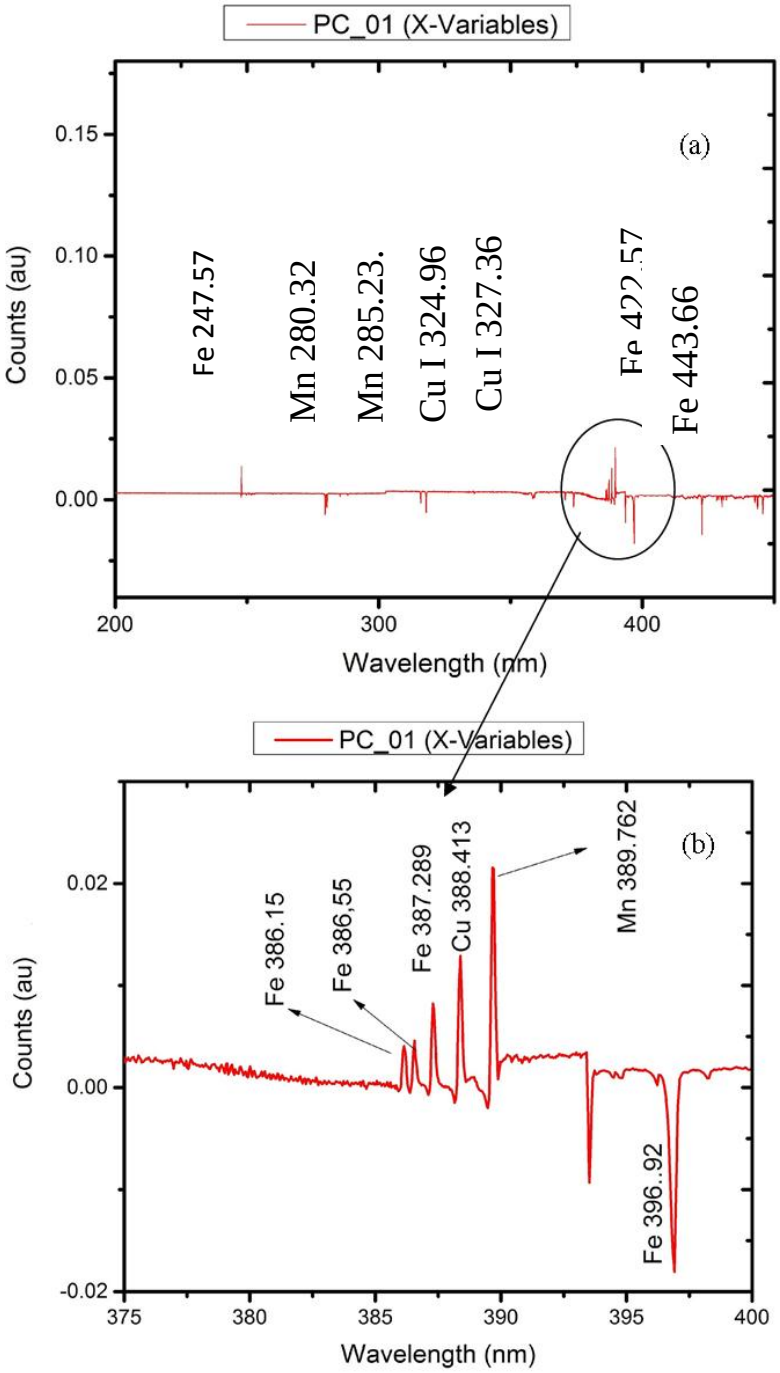

Figure 5.22: (a): PCA loadings plot for breast, liver and abdominal tissues displaying the spectral lines responsible for clustering and (b) zoomed out region indicating the lines at a closer look.

This figures 5.22 (a) and (b) show the chemical interpretation of the scores plot identifying the lines responsible for the clustering. PC1 shows Fe, Cu and Mn lines at various wavelength values. The lines identified are Fe at 386.15nm, 386.55 nm and 387.29 nm, Cu 388.41 nm and Mn line at 389.76 nm and 247.89 nm.

Unsupervised classification using PCA is able to classify different types of tissues depending on the part of the body it is obtained. The tissues are grouped according to the part of the body from which it was obtained. This concludes that PCA can be used for exploratory analysis to characterize types of tissues regarding the region it is obtained as well as the health status of the tissue. The lines identified above can be used as biomarkers.

5.5.2 Differentiation of Cancer Cell Lines Using PCA

Staging of cancer was done by studying Hep- 2 and Lewis lung cell lines were to compare the stages in terms of growth of cancer, concentration levels of the trace elements and the ratios and the correlations between the trace elements. The PCA score plot, Figure 5.23, for the Hep 2 cell lines shows the three stages; stage 1 and stage 2 classified into one group which we could say represent early stage of cancer, stage 3 and stage 4 as the benign stages while stage 5 as the malignant stage. The loadings plot gives the lines responsible for the classification. PC 1 contributes 97 % which consists mostly of Fe, Mn and Cu lines while PC2 contributes 3%.

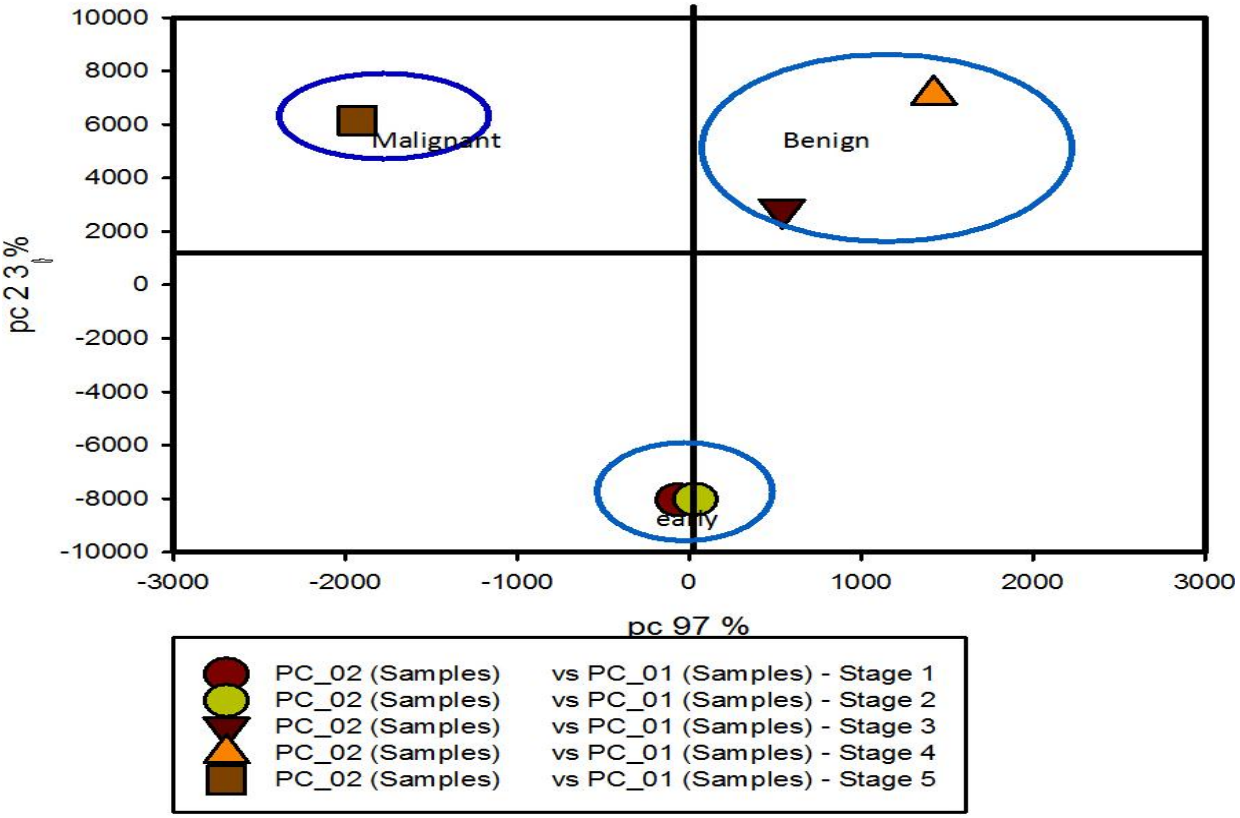

Figure 5.23: PCA scores plot of cultured Hep-2 cancer cell line. The stages have been clustered into early, benign and malignant stages

The scores plot is showing the clusters formed between stages 1 to 5. Stages 1 and 2 are grouped together, stage 3 and 4 as well while stage 5 is on its own. The first cluster can be identified as early stage, stage 3 and 4 benign stage while stage 5 as malignant stage of cancer development.

The study of the loadings plot for the PCA scores plot of the Hep-2 cell lines, figure 5.24, reveals that the lines responsible for the grouping are Fe, Cu and Mn lines. They are lines of elements of higher oxidation states. This suggests that speciation has a great role to play in cancer development. There are several Fe and Cu lines as opposed to Mn lines. The score plot

is showing the clusters formed between stages 1 to 5. Stage 1 and 2 can be identified as early stage, stage 3 and 4 are classified as benign stage and stage 5 as the malignant stage.

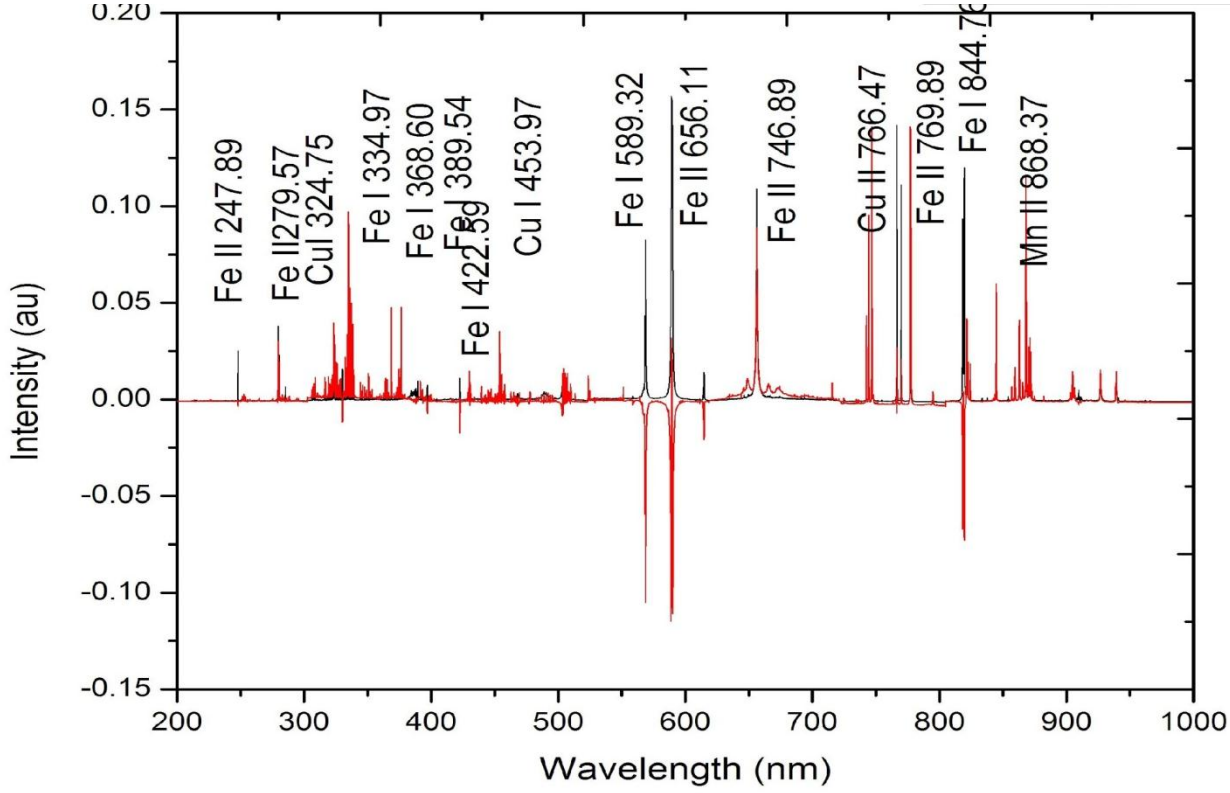

Figure 5.24: PCA loadings plot showing Fe, Cu and Cu lines identified at different wavelength regions to be responsible for the clustering of Hep-2 cell lines. PC1 contributed 97 % while PC2 3 %

PCA was also carried out for Lewis Lung tissue for the whole spectral region as shown in Figure 5.25. It is evident that the stages of development have clustered into three distinct groups. Stage 1 and 2 could be said to belong to the same group of early stage of cancer development, while stage 3 and 4 in the benign and stage 5 malignant. Stage 1, stage 2 and stage 5 are clustered where they are due to these lines. Stages 3 and 4 are influenced more by PC2 which show the presence of the Fe, Mn and Cu as observed in Figure 5.26.

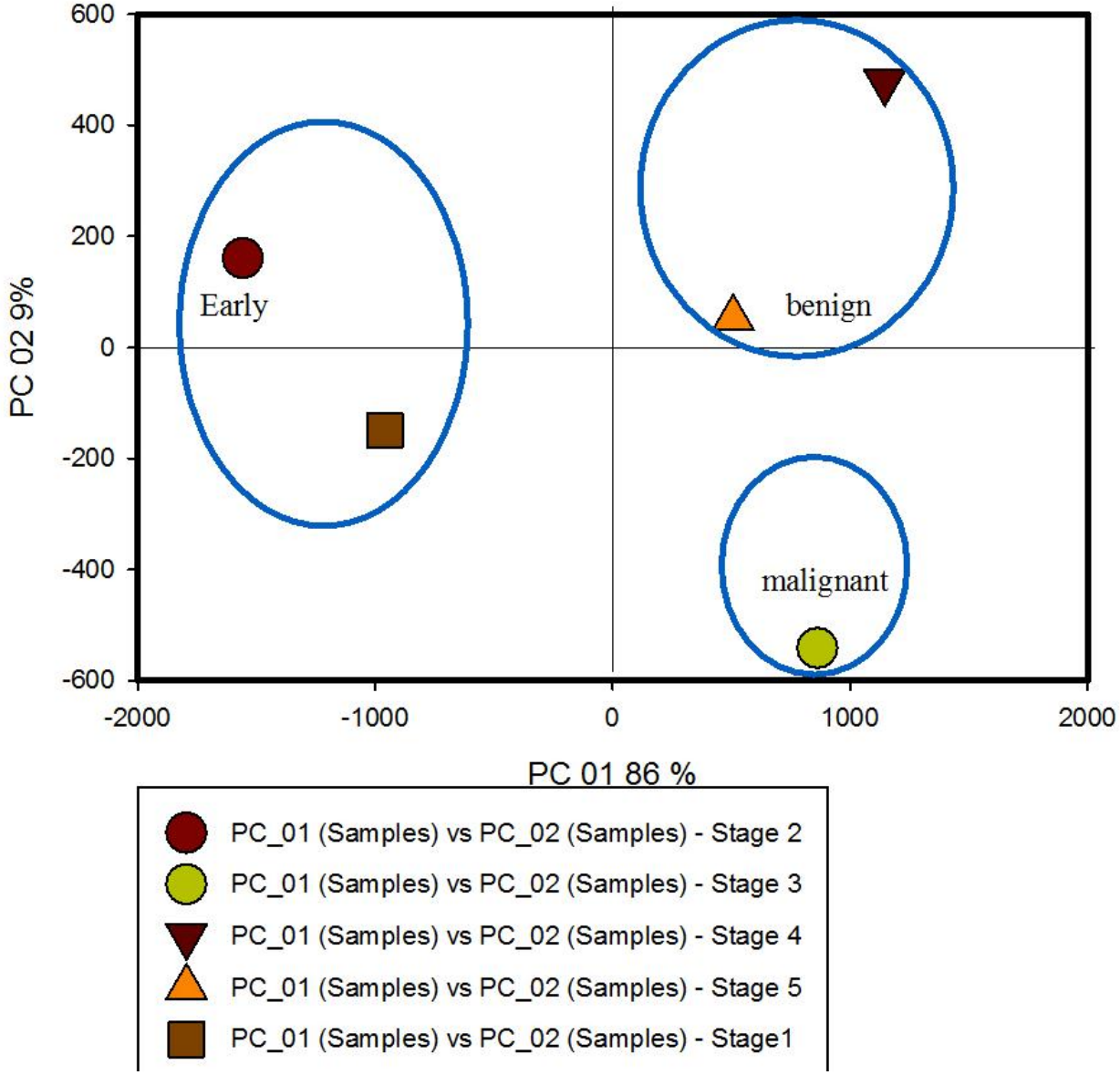

Figure 5.25: PCA scores plot of cultured Lewis Lung cancer cell line. The stages 1 and 2 have been clustered into early, 3 and 4 into benign and 5 as malignant stage

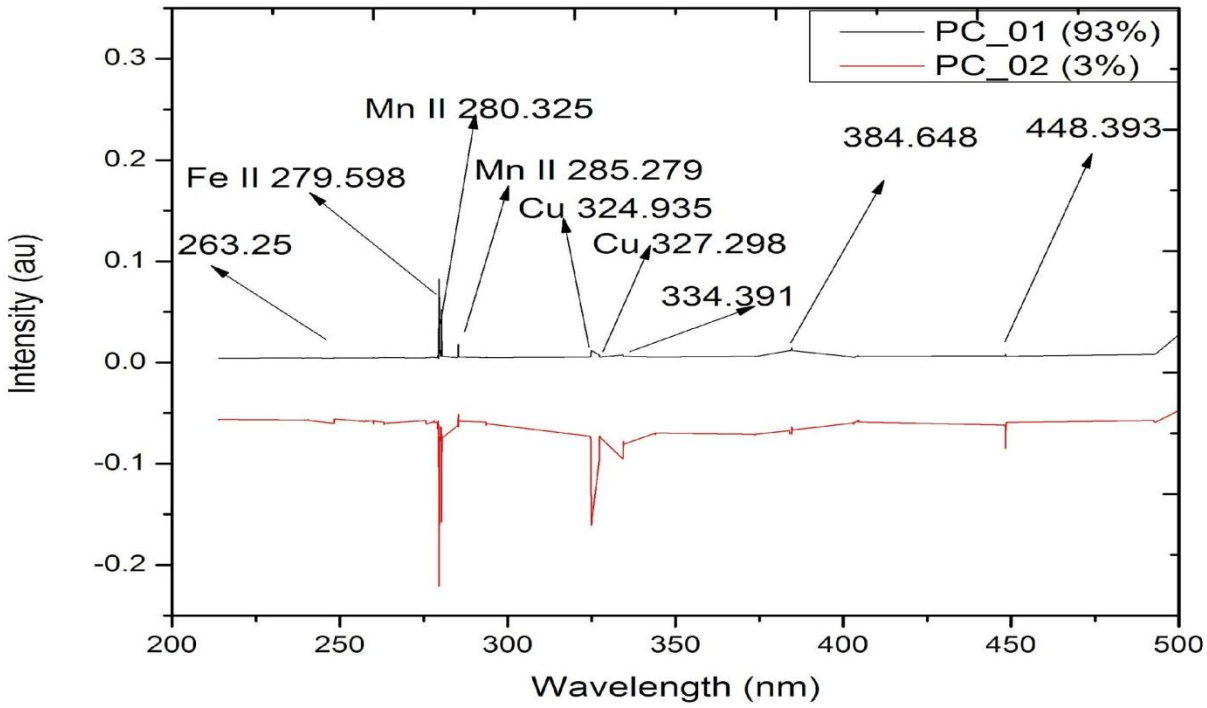

Figure 5.26: PCA loadings plot for Hep 2 cancer cell line using feature selected spectral lines

PCA was carried out on the real samples using the concentration values predicted using the ANN prediction model. The results do not agree with the ones obtained using the intensity values obtained from the spectra. Stages 1, 3 and 5 are clustered together unlike while using intensity values; stages 4 and 5 are clustered together and identified as malignant. This shows that concentration alone cannot be used to determine the classification of cancer cell lines.

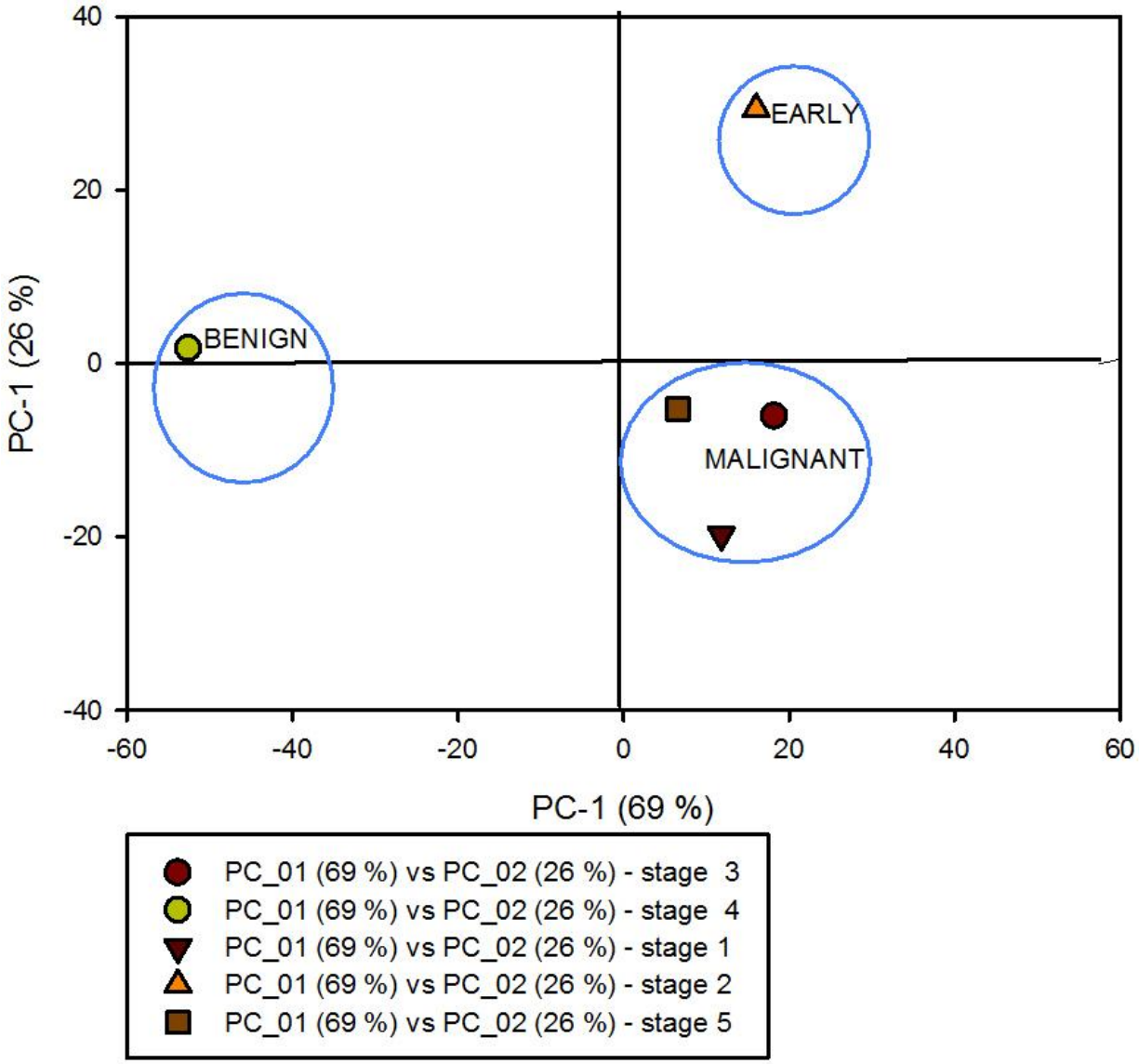

Figure 5.27: PCA scores plot for Hep 2 cancer cell line using predicted concentration values.

Figure 5.28 on the other hand shows the grouping of Lewis Lung tissues using PCA. The concentration values were used and the groupings are diverse when compared to the ones obtained using intensity values. The malignant stage is grouped with stages 1 and 5 together, unlike in the one with intensity where stages 4 and 5 are grouped together. Concentration

values alone cannot be depended on for cancer staging. Other variables such as correlations and ratios are also important.

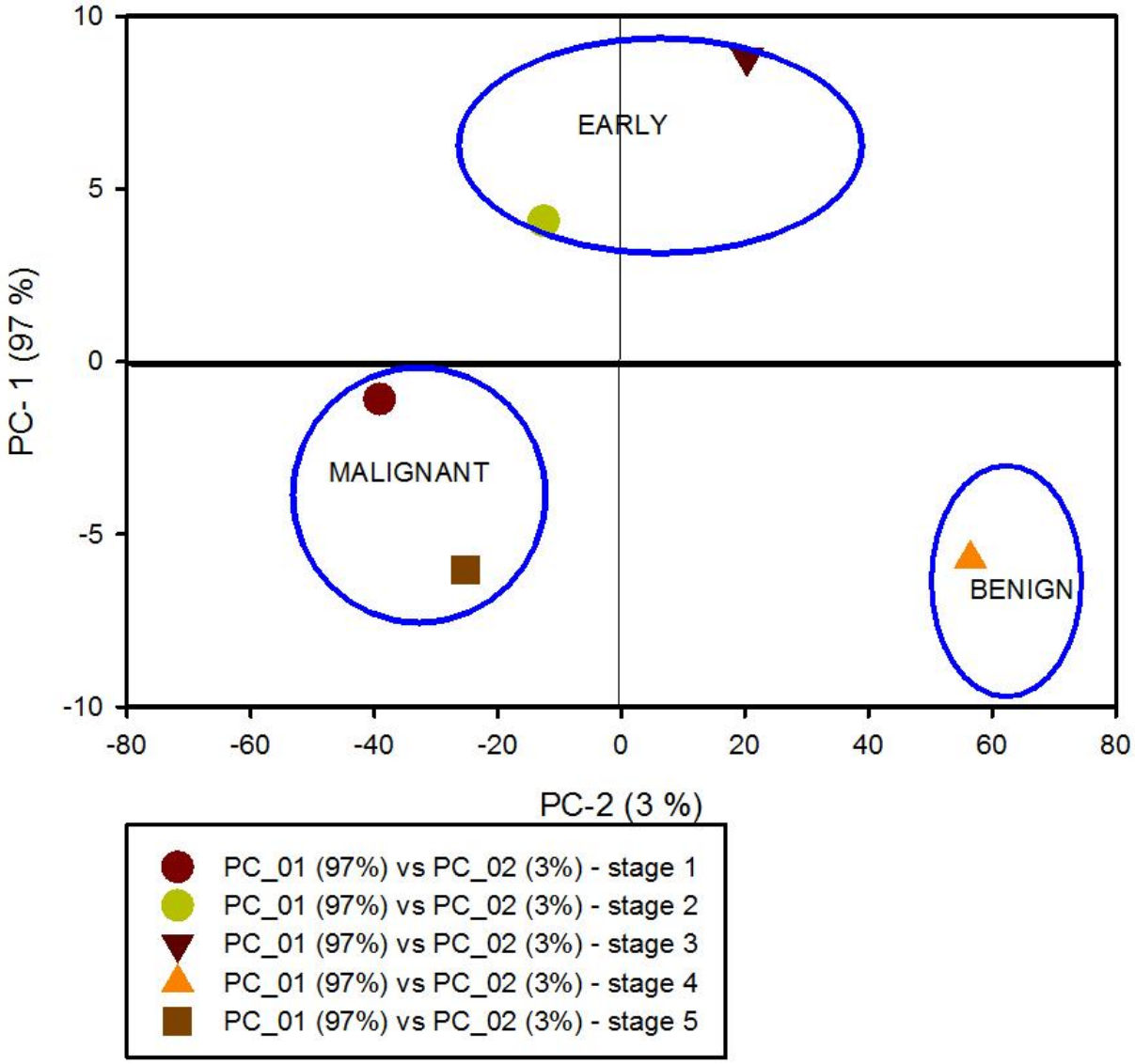

Figure 5.28: PCA scores plot for Lewis lung cancer cell line using predicted concentration values

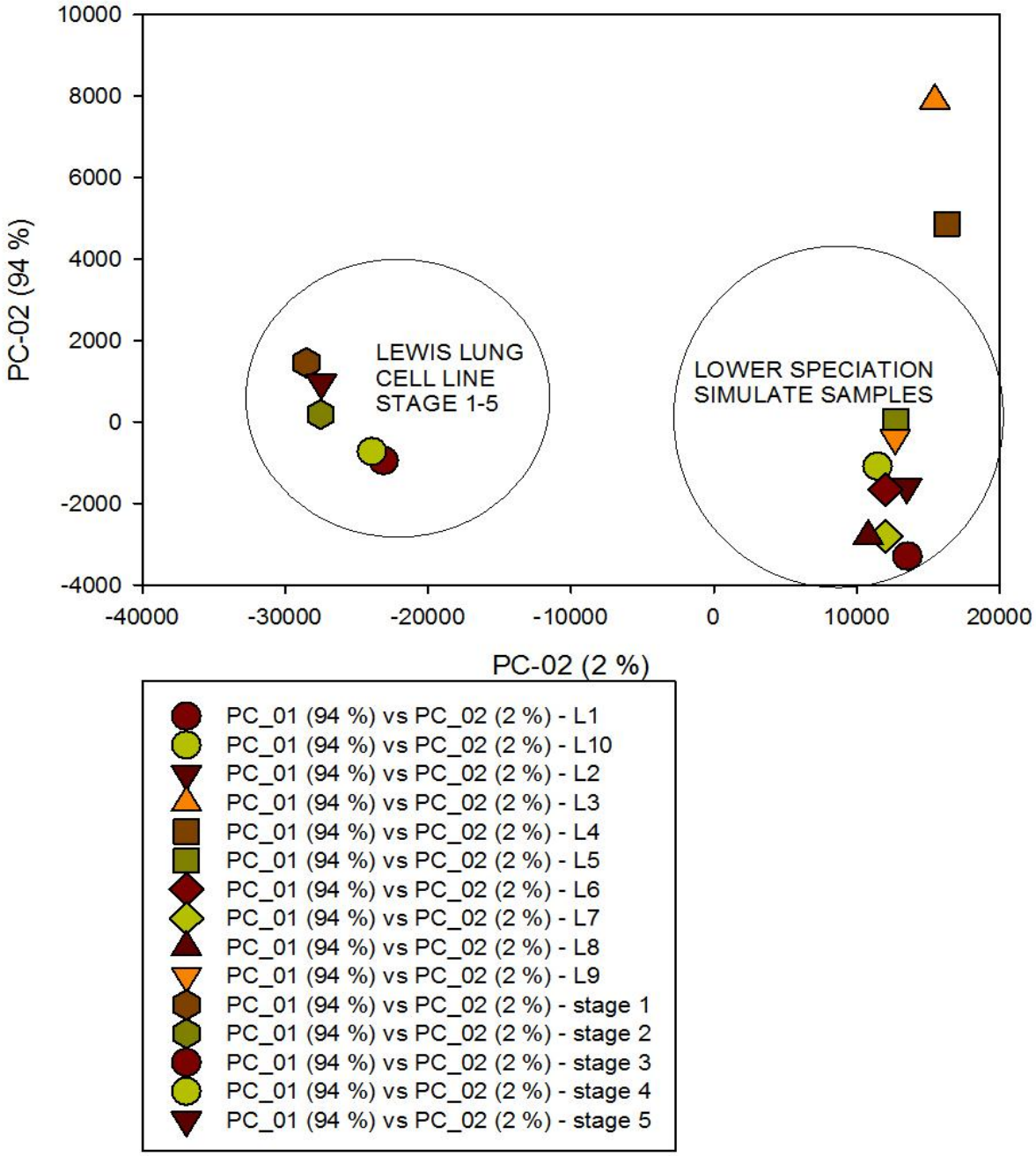

Figure 5.29: Differentiation of cancer cell lines based on speciation. The figure shows the Lewis Lung tissue and lower speciation simulate samples.

In figure 5.29, the cell lines are classified aside from the simulate samples. This shows that the stages of cancer cell lines do not have lower speciation of the trace elements present.

Speciation is a key aspect of cancer differentiation as can be seen by this scores plot.

5.5.3 Classification of Breast, Liver and Abdominal Tissues Using SVM

SVM was used to classify the liver, breast and abdominal tissues by first training the model using the simulate samples into two classes. The simulate samples were classified into higher speciation and lower speciation separated by a hyperplane. The hyper plane shows the boundary of the two. Separate models of Cu, Mn and Fe were developed in to which each tissue was tested to have either lower speciation or higher speciation of these elements. The real tissue intensity values were then fed into the model and classified into either of the groups on the side of higher speciation and lower speciation. The models therefore represented as 1 and 0 for higher and lower speciation levels respectively. The breast, liver and abdominal tissues were classified as either belonging to 0 or 1.

Figure 5.30 shows the SVM plot for Fe. The breast tissue tested using this model produced an output as 1. The model's class performance achieved is 0.7125. The tissues were exposed to a model developed by SVM for classifying them based on the speciation of the elements. The tissue was classified as having Fe element of a higher speciation. This demonstrates the success of SVM classification of tissues based on speciation.

Classification was done using support vector machine to develop a model for classifying sim ulate samples based on speciation. Radial basis function was used for classification.

The best model had a cost of 0.25 and gamma of 10. The miscalculation error for radial basis

function for the copper model obtained was: 0.235434. Manganese had a miscalculation error of 0.1973684 for radial basis function and cost of 1 gamma of 100.

Table 5.11: Table summarizing the cost, gamma and miscalculation errors of the SVM classification model developed using radial basis function. The model was based on speciation of trace elements

	Mn	Cu	Fe
Cost	1	1	0.25
gamma	100	10	10
Miscalculation error	0.1973684	0.2354347	0.2875

Radial basis function was preferred to linear model since the miscalculation error achieved by the linear models was quite high as opposed to the one for radial basis function.
The miscalculation error for the best Fe classification model was 0.4372883, which is quite high. This gave a model with a cost function of 0.03125 and gamma of 0.001.
Moreover, the miscalculation error achieved using the best Mn linear model yielded 0.412359 3, cost function of 0.5 and a gamma of 0.001. The best Cu linear model had a cost function of 1, gamma of 0.001 and miscalculation error of 0.3250304.

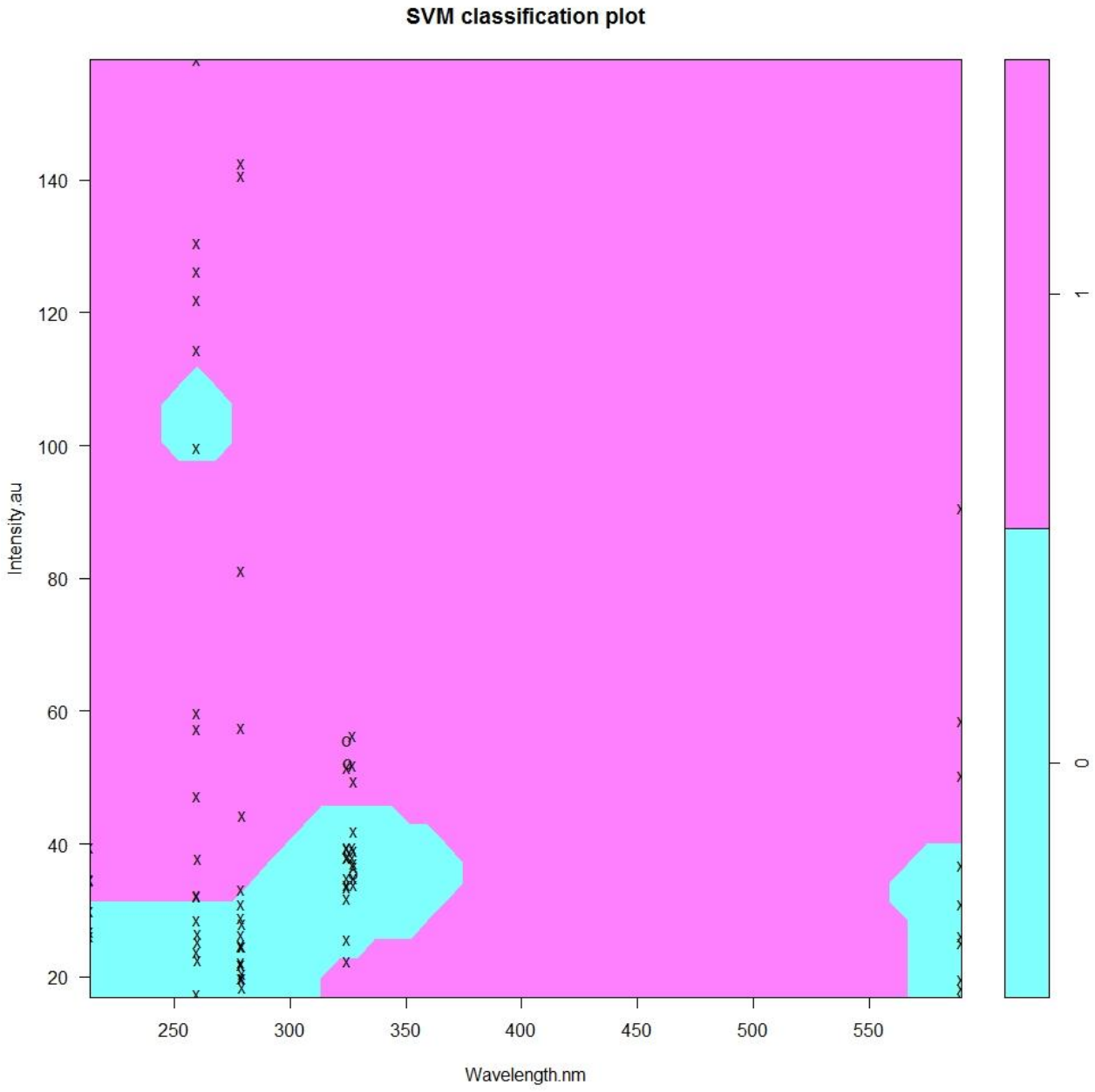

Figure 5.30: SVM classification model based on speciation for Cu. The model shows the simulate samples classified into either having higher speciation and lower speciation of Cu., Cu I or Cu I ions.

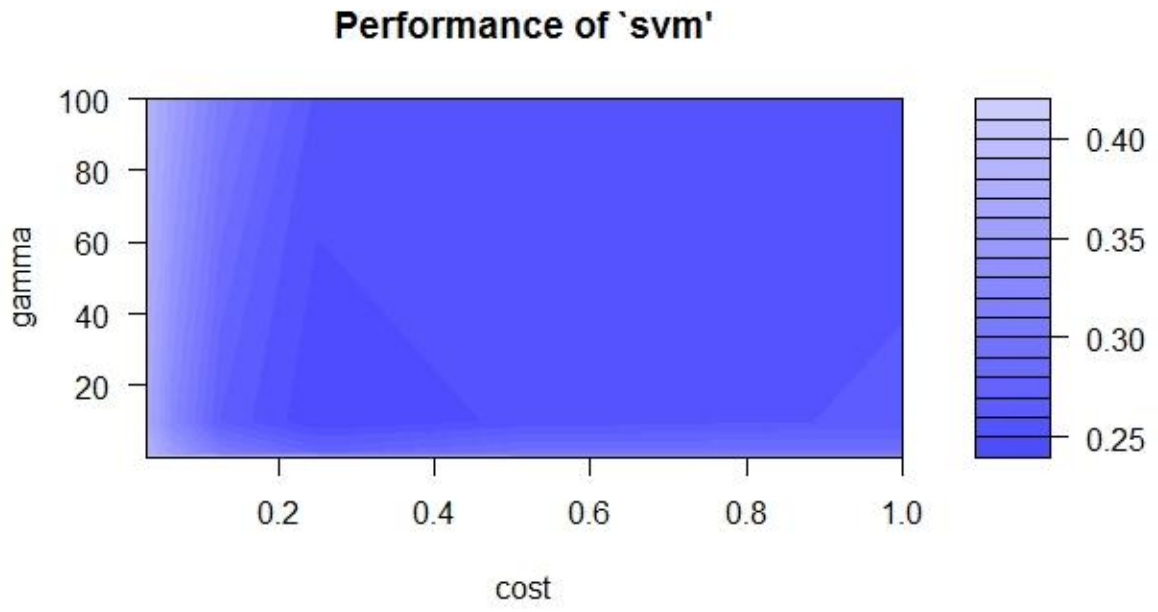

Figure 5.31: SVM classification performance model based on speciation for Cu. The model shows the best cost and gamma at 1 and 10 respectively of the classification model developed.

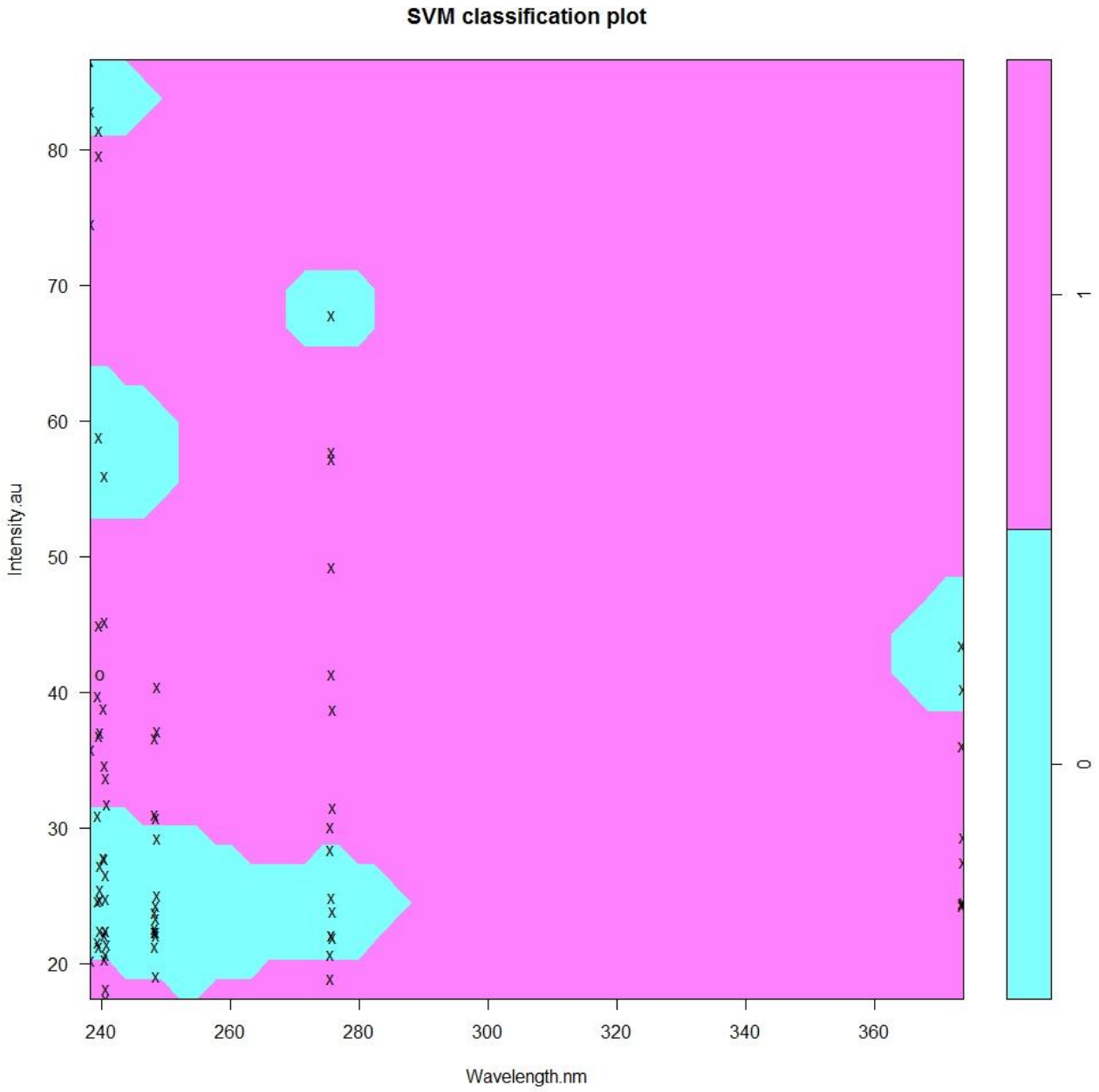

Figure 5.32: SVM classification model based on speciation for Mn. The model shows the simulate samples classified into either having higher speciation and lower speciation of Mn II or M IV ions

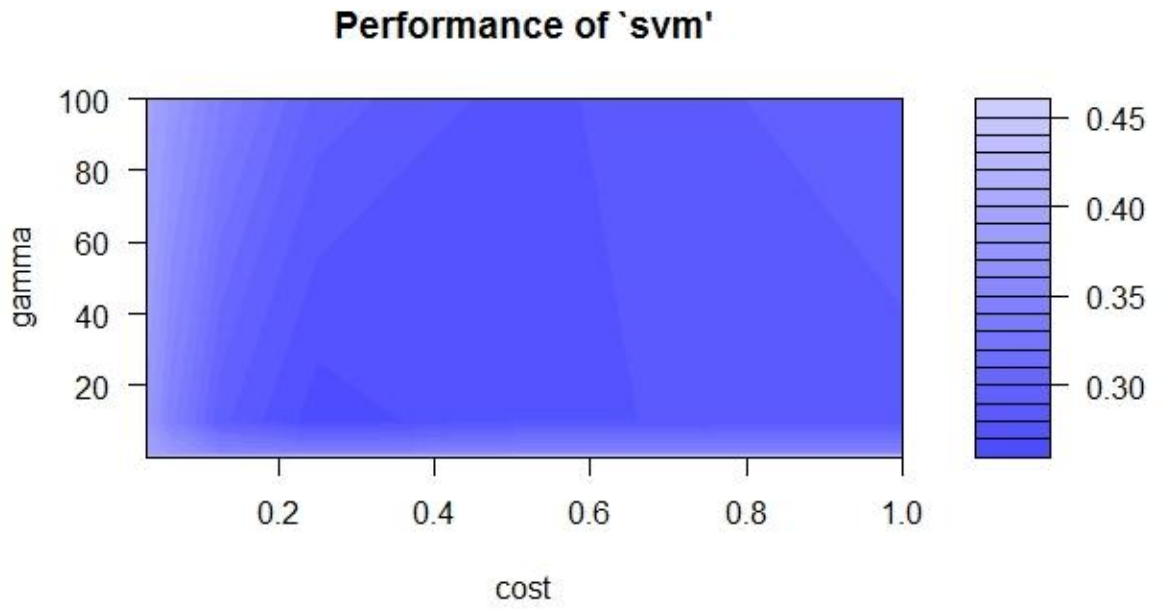

Figure 5.33: SVM classification performance model based on speciation for Mn. The model shows the best cost and gamma of 1 and 100 respectively of the classification model developed

Table 5.12 below illustrates the prediction of the speciation of Cu, Mn and Fe trace elements in the liver, breast and abdominal tissues. This is an indication of how important speciation is in classification of cancer tissues. Aside from concentration values, speciation is also an important aspect in classification.

Table 5.12: Table representing the predictions of the speciation of Mn, Cu and Fe using the SVM classification models.

	Liver 1 tissue	Liver 2 tissue	Abdominal tissue	Breast 1 tissue	Breast 2 tissue	Breast 3 tissue
Cu	Lower	Higher	Higher	Lower	Lower	Higher
Fe	Higher	Higher	Higher	Higher	Higher	Higher
Mn	Lower	Higher	Higher	Lower	Lower	Higher

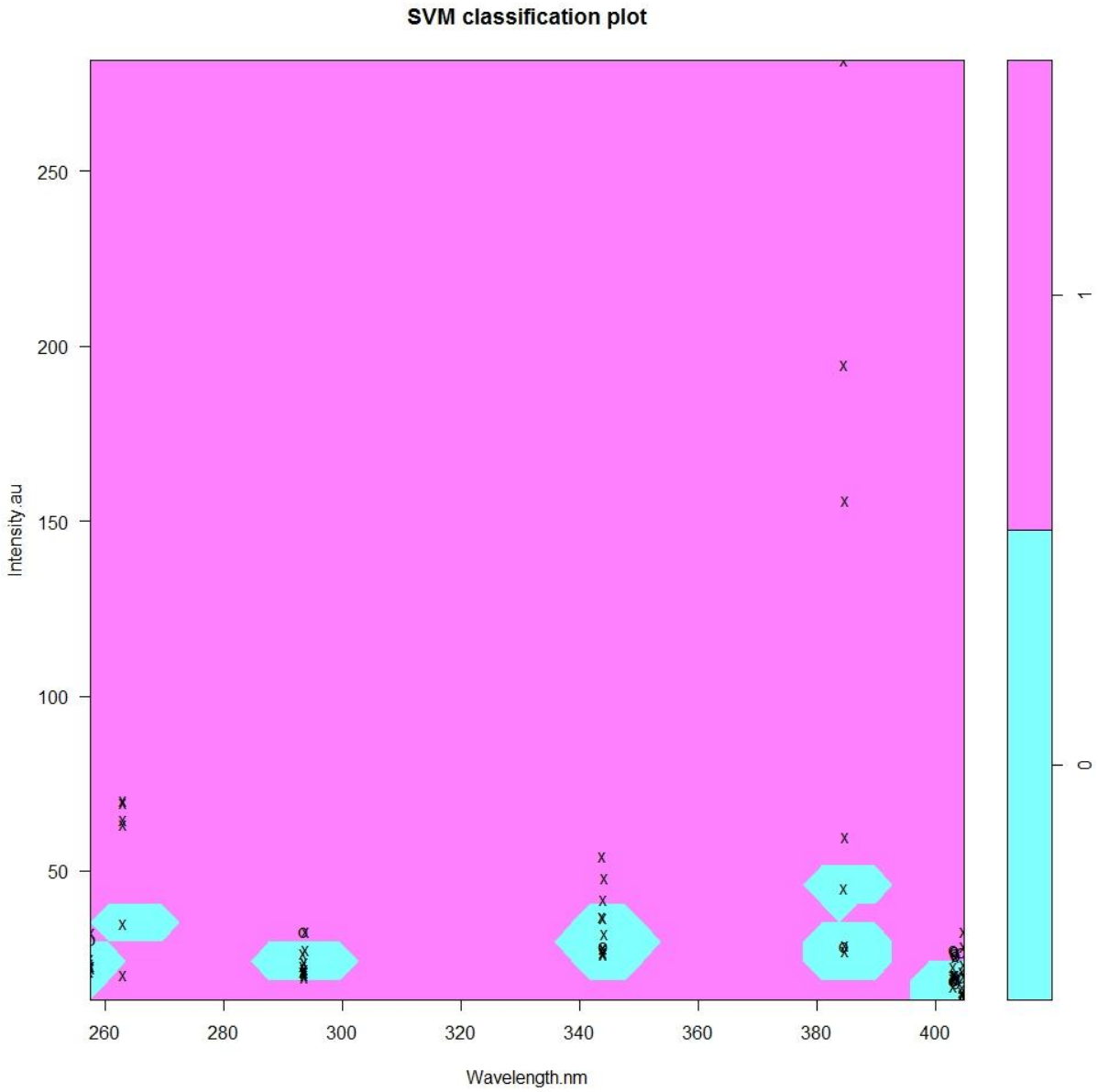

Figure 5.34: SVM classification model based on speciation for Mn. The model shows the simulate samples classified into either having higher speciation and lower speciation of Fe II or Fe III ions

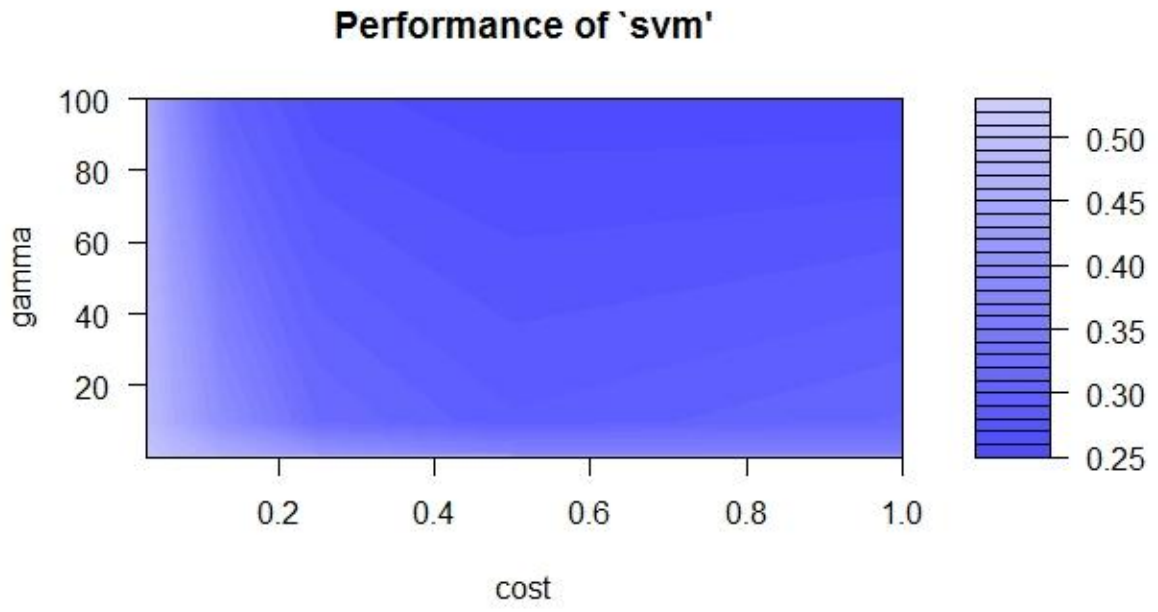

Figure 5.35: SVM classification performance model based on speciation for Fe. The model shows the best cost and gamma of 0.25 and 10 respectively of the classification model developed

All the tissues were classified using SVM by training a model to classify the higher speciation and lower speciation into two classes. These were separated by the hyper plane using radial basis function. The breast tissue was then fed into the model to predict the classification. The three modes predicted the tissue as belonging to the higher speciation side of the hyperplane.

Table 5.12 shows the classification of liver, breast and abdominal tissues using the SVM model developed. It can be concluded that speciation has an important role to play in development of cancer.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

This work involved development of a rapid technique for cancer detection at early stage utilizing LIBS and Chemometric techniques. This was done by identifying the trace element biomarkers in liver, breast and abdominal tissues using LIBS. This objective was achieved by identifying these biomarkers from the spectra obtained when these tissues were ablated upon using LIBS. The possible biomarkers associated with the disease were obtained in these spectra. The following trace elements were under study: Cu, Mn, Fe, Mg and Zn and specific lines were present different quantities in these tissues. Quantification of these trace elements was also be used in determining other aspects of the disease. A calibration model was designed using ANN, a chemometrics tool, for determination of the quantities of these trace elements.

The model was successfully validated using oyster tissue as the standard reference material by using it to predict the concentration values of these elements in the tissue using the designed model. The predicted values obtained were within the range with less than 5% prediction error. The validated model was used to predict the concentration values of the biomarkers. These are tabulated in chapter 5. The ratios of the concentration values are useful in disease diagnostics as their occurrence indicate the health status of the tissue as well as the stage of development of the tissue as seen in the quantity values of the cell lines. The predictive model was used to get the quantities of the trace elements in both tissues and cancer cell lines.

Exploratory analysis, to differentiate and characterize the cancer tissues, was done successfully using PCA as seen in chapter 5. This technique created patterns of the spectral

data obtained using LIBS. The tissues were separated into cancerous and non-cancerous. PCA was able to identify existing similarities and differences between the variables as seen in the scores plot of figure 5.1. The loadings plot on the other hand indicates the information regarding the trace element biomarkers responsible for the classification of the tissues as shown in figure 5.2. PCA was further applied to Hep 2 and Lewis Lung cell lines and it yielded scores plots showing the existing patterns in the stages of the cell lines. Stages 1 and 2 for Hep 2 in figure 5.23 can be identified as early stage, stages 3 and 4 as benign while stage 5 as malignant stages. The same analogy is applied to the score plot of Lewis Lung tissue in figure 5.25. The corresponding loadings plot thus identifies the trace element lines responsible for the grouping. The lines identified are Mn, Fe and Cu lines. These are speciation elements. This indicates that when tissues replicate the elements exist in higher order speciation state. This makes speciation a useful aspect of cancer diagnostics based on trace elements. It was used as a basis of classification using SVM as supervised method for grouping tissues to either higher speciation or lower speciation elements. Cu, Mn, and Fe were used to create models that was used to classify the tissues belonging to either the lower or higher speciation states of these elements.

This research shows a LIBS method developed for rapid and non-invasive detection, quantification and characterization of cancer in human body tissue based on the concentration levels and alterations of these elements. The model gives good results if well and validated using a standard reference, as was the case in this study. Exploratory analysis is very useful in determining the existing similarities and differences in the dataset. The research could be carried forward to examine the performance of the prediction model on other reference standard values to evaluate the robustness of the model. Subjecting more tissue samples under study leads to having a more conclusive method for early detection and staging.

REFERENCES

Abdi, H. and Williams, L. J. (2010). Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2**, 433–459.

Boueri, M., Motto-Ros, V., Lei, W.-Q., Ma, Q.-L., Zheng, L.-J., Zeng, H.-P., and Yu, J. (2011). Identification of polymer materials using laser-induced breakdown spectroscopy combined with artificial neural networks. *Appl. Spectrosc.* **65**, 307–314.

Colao, F., Fantoni, R., Lazic, V., Paolini, A., Fabbri, F., Ori, G. G., Marinangeli, L., and Baliva, A. (2004). Investigation of LIBS feasibility for in situ planetary exploration: An analysis on Martian rock analogues. *Planet. Space Sci.* **52**, 117–123.

Colao, F., Fantoni, R., Lazic, V., and Spizzichino, V. (2002). Laser-induced breakdown spectroscopy for semi-quantitative and quantitative analyses of artworks—application on multi-layered ceramics and copper based alloys. *Spectrochim. Acta Part B At. Spectrosc.* **57**, 1219–1234.

Corsi, M., Cristoforetti, G., Hidalgo, M., Legnaioli, S., Palleschi, V., Salvetti, A., Tognoni, E., and Vallebona, C. (2003). Application of laser-induced breakdown spectroscopy technique to hair tissue mineral analysis. *Appl. Opt.* **42**, 6133–6137.

Cremers, D. A., Yueh, F.-Y., Singh, J. P., and Zhang, H. (2006). "Laser-Induced Breakdown Spectroscopy, Elemental Analysis." Wiley Online Library.

David, M., Evgenia, D., Kurt, H., Andreas, W., and Friedrich, L. (2017). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.6-8. https://CRAN.R-project.org/package=e1071

DeLucia, J. F.C., Samuels, A. C., Harmon, R. S., Walters, R. A., McNesby, K. L., LaPointe, A., Winkel, J. R.J., and Miziolek, A. W. (2005). Laser-induced breakdown spectroscopy (LIBS): a promising versatile chemical sensor technology for hazardous material detection. *IEEE Sens. J.* **5**, 681–689.

El-Hussein, A., Kassem, A. K., Ismail, H., and Harith, M. A. (2010). Exploiting LIBS as a spectrochemical analytical technique in diagnosis of some types of human malignancies. *Talanta* **82**, 495–501.

Emily, H. Zinc deficiency, DNA damage and cancer risk (2004). *The Journal of nutritional biochemistry, 15*(10), 572-578.

Fang, X., Ahmad, S. R., Mayo, M., and Iqbal, S. (2005). Elemental analysis of urinary calculi by laser induced plasma spectroscopy. *Lasers Med. Sci.* **20**, 132–137.

Gottfried, J. L., Jr, F. C. D. L., Munson, C. A., and Miziolek, A. W. (2009). Laser-induced breakdown spectroscopy for detection of explosives residues: a review of recent advances, challenges, and future prospects. *Anal. Bioanal. Chem.* **395**, 283–300.

Guo, J., Deng, W., Zhang, L., Li, C., Wu, P., and Mao, P. (2007). Prediction of prostate cancer using hair trace element concentration and support vector machine method. *Biol. Trace Elem. Res.* **116**, 257–271.

Hardley, W.(2009) ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*, 978-0-387-98140-6. http://ggplot2.org.

Harmon, R. S., Russo, R. E., & Hark, R. R. (2013). Applications of laser-induced breakdown spectroscopy for geochemical and environmental analysis: A comprehensive review. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *87*, 11-26.

Harmon, R. S., Remus, J., McMillan, N. J., McManus, C., Collins, L., Gottfried Jr., J. L., DeLucia, F. C., and Miziolek, A. W. (2009). LIBS analysis of geomaterials: Geochemical fingerprinting for the rapid analysis and discrimination of minerals. *Appl. Geochem.* **24**, 1125–1141.

Kalogirou, S. A. (2001). Artificial neural networks in renewable energy systems applications: a review. *Renew. Sustain. Energy Rev.* **5**, 373–401.

Kalogirou, S. A. (2000). Applications of artificial neural-networks for energy systems. *Appl. Energy* **67**, 17–35.

Kaniu, M. I., Angeyo, K. H., Mwala, A. K., and Mwangi, F. K. (2012). Energy dispersive X-ray fluorescence and scattering assessment of soil quality via partial least squares and artificial neural networks analytical modeling approaches. *Talanta* **98**, 236–240.

Kim, J., Mowat, A., Poole, P., and Kasabov, N. (2000). Linear and non-linear pattern recognition models for classification of fruit from visible–near infrared spectra. *Chemom. Intell. Lab. Syst.* **51**, 201–216.

Kumar, A., Yueh, F.-Y., Singh, J. P., and Burgess, S. (2004). Characterization of malignant tissue cells by laser-induced breakdown spectroscopy. *Appl. Opt.* **43**, 5399–5403.

Labbé, N., Swamidoss, I. M., André, N., Martin, M. Z., Young, T. M., and Rials, T. G. (2008). Extraction of information from laser-induced breakdown spectroscopy spectral data by multivariate analysis. *Appl. Opt.* **47**, G158–G165.

Lee, W. B., Wu, J., Lee, Y. I., & Sneddon, J. (2004). Recent applications of laser-induced breakdown spectrometry: a review of material approaches. *Applied Spectroscopy Reviews*, *39*(1), 27-97.

Liu, X., Zhang, Q., Wu, Z., Shi, X., Zhao, N., and Qiao, Y. (2014). Rapid Elemental Analysis and Provenance Study of Blumea balsamifera DC Using Laser-Induced Breakdown Spectroscopy. *Sensors* **15**, 642–655.

Marini, F., Bucci, R., Magrì, A. L., and Magrì, A. D. (2008). Artificial neural networks in chemometrics: History, examples and perspectives. *Microchem. J.* **88**, 178–185.

Melessanaki, K., Mateo, M., Ferrence, S. C., Betancourt, P. P., and Anglos, D. (2002). The application of LIBS for the analysis of archaeological ceramic and metal artifacts. *Appl. Surf. Sci.* **197–198**, 156–163.

Miziolek, A. W., Palleschi, V., and Schechter, I. (2006). "Laser induced breakdown spectroscopy." Cambridge University Press.

Mohamed, W. T. Y. (2007). Study of the matrix effect on the plasma characterization of six elements in aluminum alloys using LIBS with a portable echelle spectrometer. *Progress in Physics*, *2*, 42.

Mukhono, P. (2012). Chemometrics-assisted laser induced breakdown spectroscopy of high background radiation area (HBRA) geothermal field matrices.

Musazzi, S. and Perini, U. (2014). “Laser-Induced Breakdown Spectroscopy: Theory and Applications.” Springer.

Naguib, R. N. and Sherbet, G. V. (2000). “Artificial neural networks in cancer diagnosis, prognosis, and patient management.” CRC Press, Inc.

Quentmeier, A., Sdorra, W., and Niemax, K. (1990). Internal standardization in laser induced fluorescence spectrometry of microplasmas produced by laser ablation of solid samples. *Spectrochim. Acta Part B At. Spectrosc.* **45**, 537–546.

Saberkari, H., Shamsi, M., Joroughi, M., Golabi, F., and Sedaaghi, M. H. (2014). Cancer Classification in Microarray Data using a Hybrid Selective Independent Component Analysis and υ-Support Vector Machine Algorithm. *J. Med. Signals Sens.* **4**, 291.

Samek, O., Telle, H. H., and Beddows, D. C. (2001). Laser-induced breakdown spectroscopy: a tool for real-time, in vitro and in vivo identification of carious teeth. *BMC Oral Health* **1**, 1.

Savastenko, N. A., and N. V. Tarasenko. "Optical Emission Spectroscopy of C2 and C3 Molecules in Laser Ablation Carbon Plasma." Spectroscopy, Dynamics and Molecular Theory of Carbon Plasmas and Vapors: Advances in the Understanding of the Most Complex High-Temperature Elemental System. 2011. 167-198.

Singh, V. K., Rai, A. K., Rai, P. K., and Jindal, P. K. (2009). Cross-sectional study of kidney stones by laser-induced breakdown spectroscopy. *Lasers Med. Sci.* **24**, 749–759.

Stewart, B. W. and Wild, C. P. (2014). “World Cancer Report 2014.” World Health Organization.

Sun, Q., Tran, M., Smith, B. W., & Winefordner, J. D. (2000). Zinc analysis in human skin by laser induced-breakdown spectroscopy. *Talanta*, *52*(2), 293-300.

Tehrani, H., Walls, J., Cotton, S., Sassoon, E., and Hall, P. (2007). Spectrophotometric intracutaneous analysis in the diagnosis of basal cell carcinoma: a pilot study. *Int. J. Dermatol.* **46**, 371–375.

Vance, T., Reljin, N., Lazarevic, A., Pokrajac, D., Kecman, V., Melikechi, N., Marcano, A., Markushin, Y., and McDaniel, S. (2010). Classification of LIBS protein spectra using support vector machines and adaptive local hyperplanes. *In* “The 2010 International Joint Conference on Neural Networks (IJCNN),” pp1–7.

INTERNET SOURCES

http://www.camo.com/rt/Resources/chemometrics.html retrieved on 31 march,2017.

APPENDICES

APPENDIX I: Pictograms of Samples

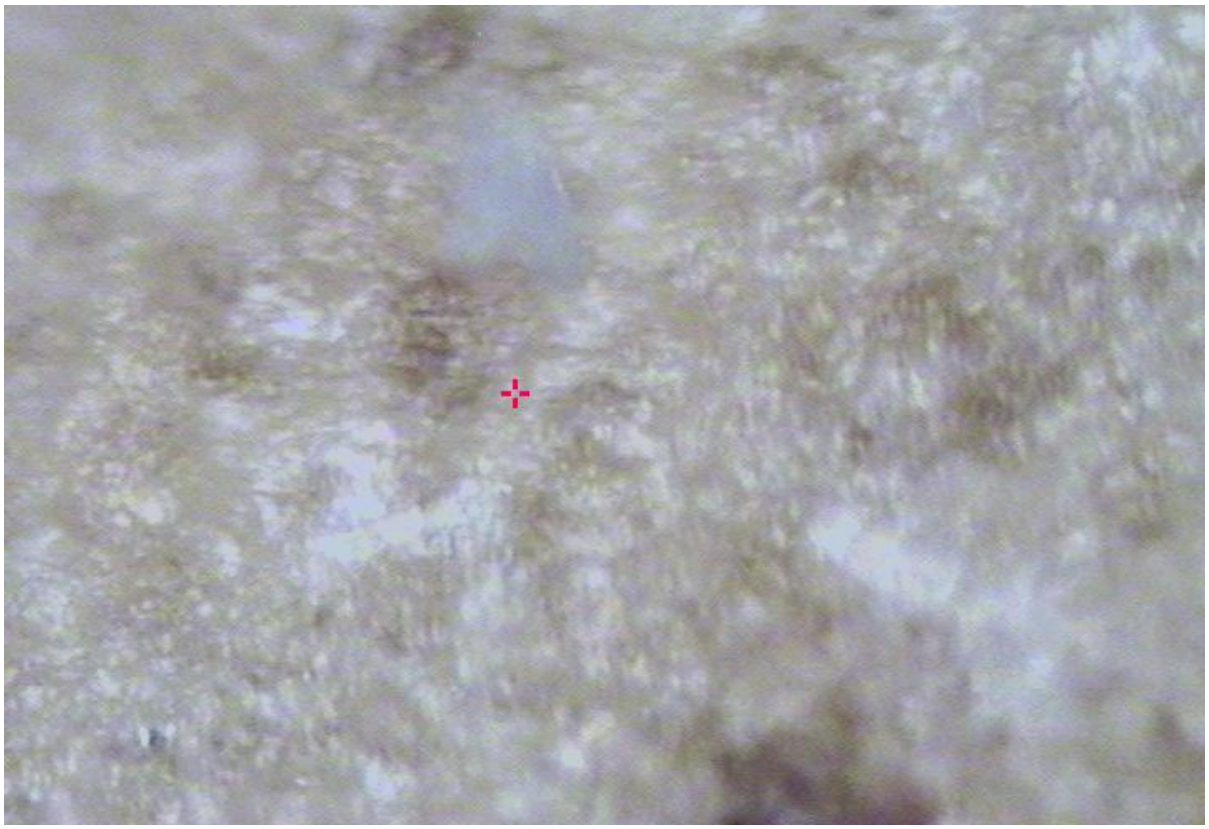

Figure A 1.1**: Microscopic image of an oyster tissue X 50**

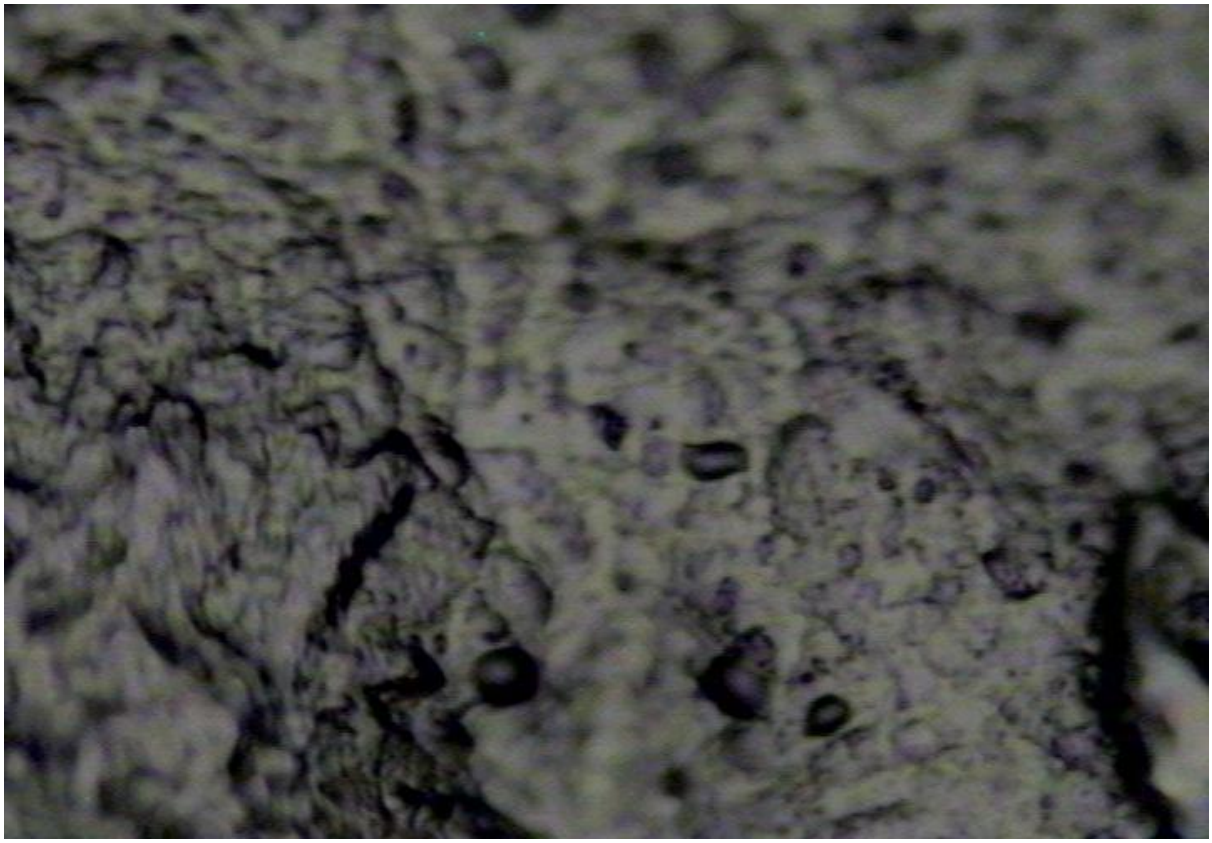

Figure A.1.2**: Microscopic image of simulate sample X50**

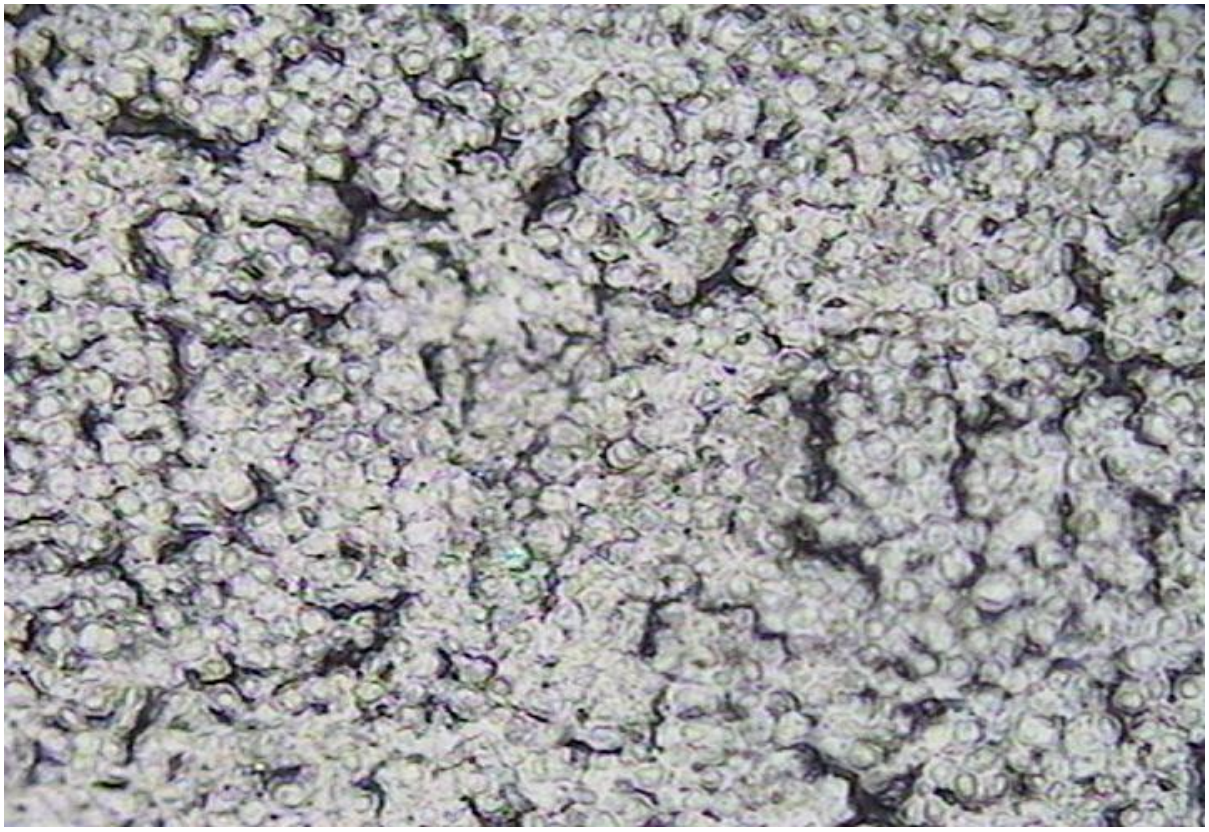

Figure A.1.3: **Pictogram of Hep-2 cell line X 50**

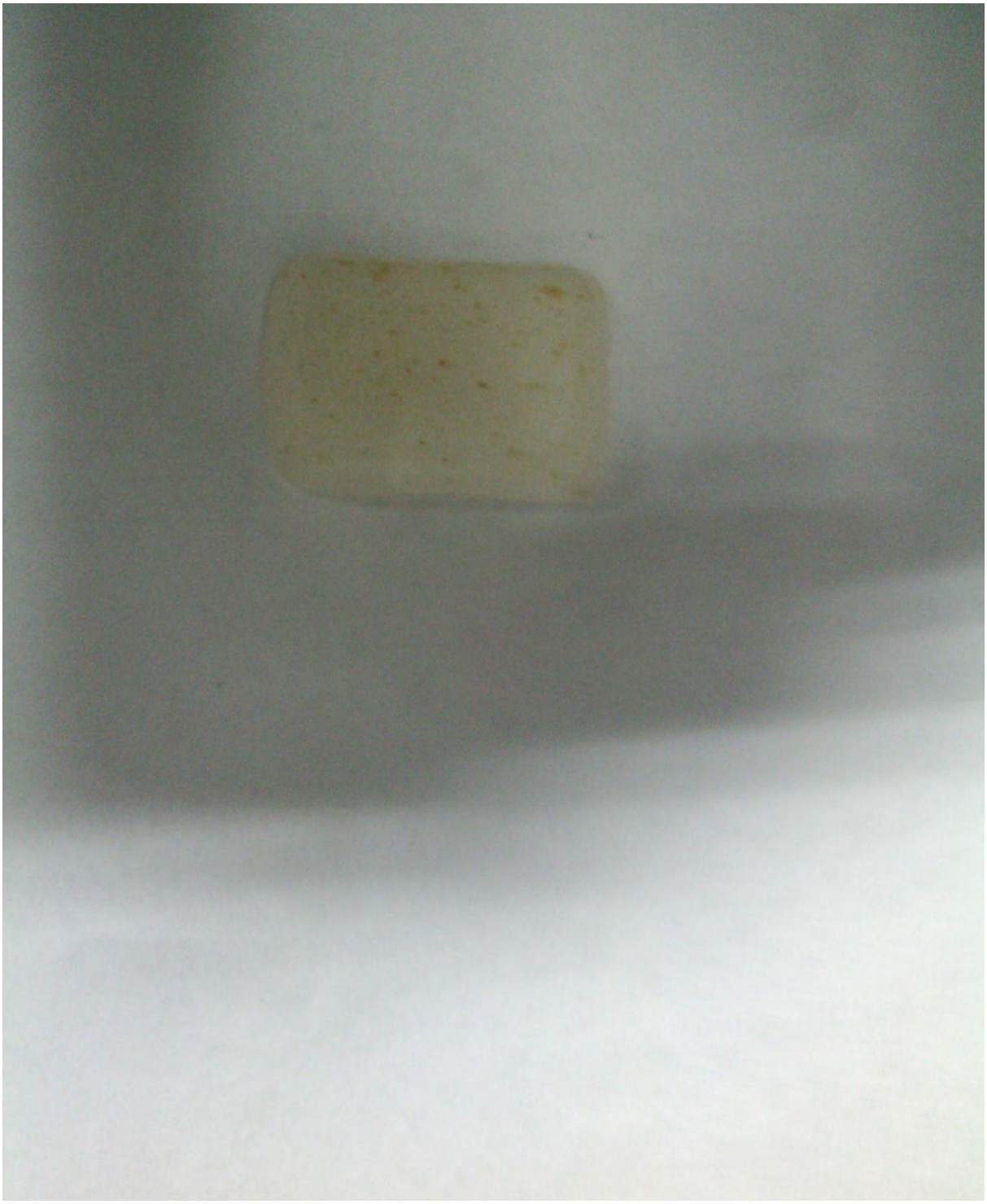

Figure A.1.4**: Pictograph of a simulate sample**

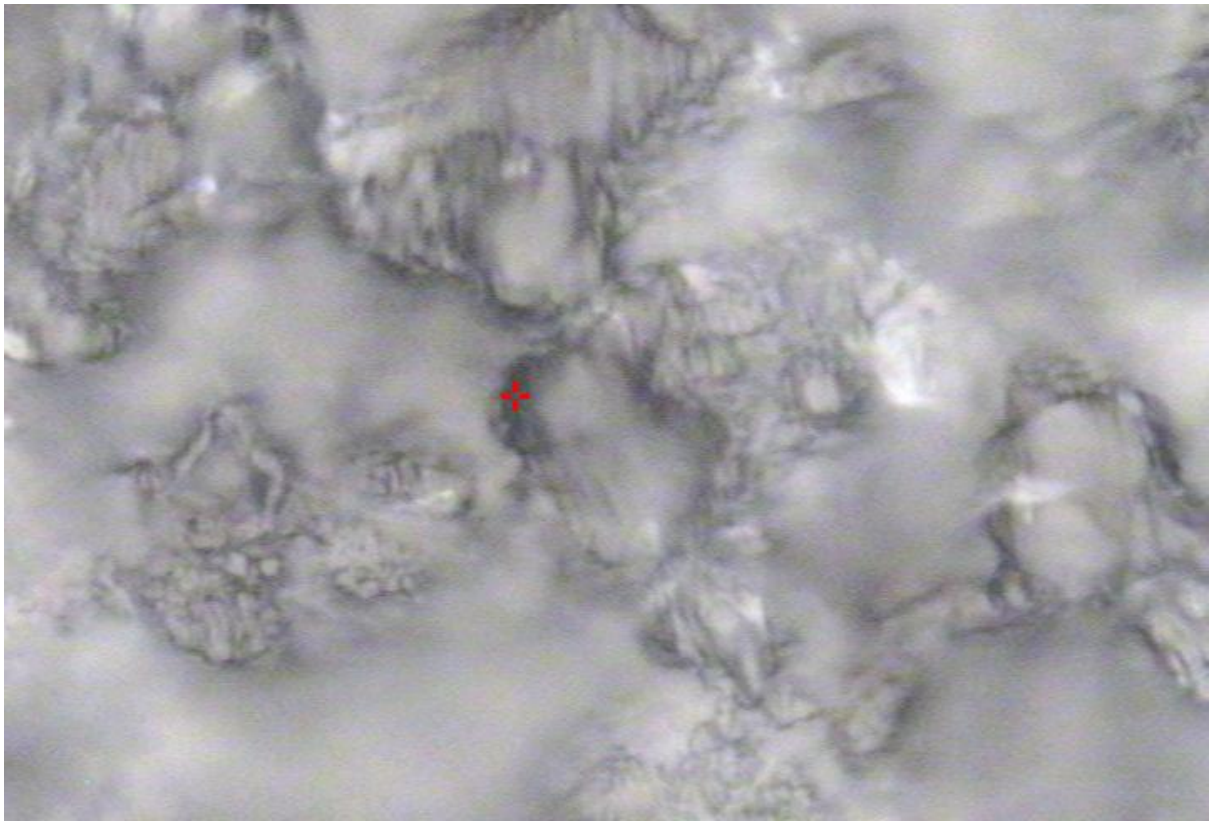

Figure A.1.5**: Microscopic image of Lewis Lung cell line X 100**

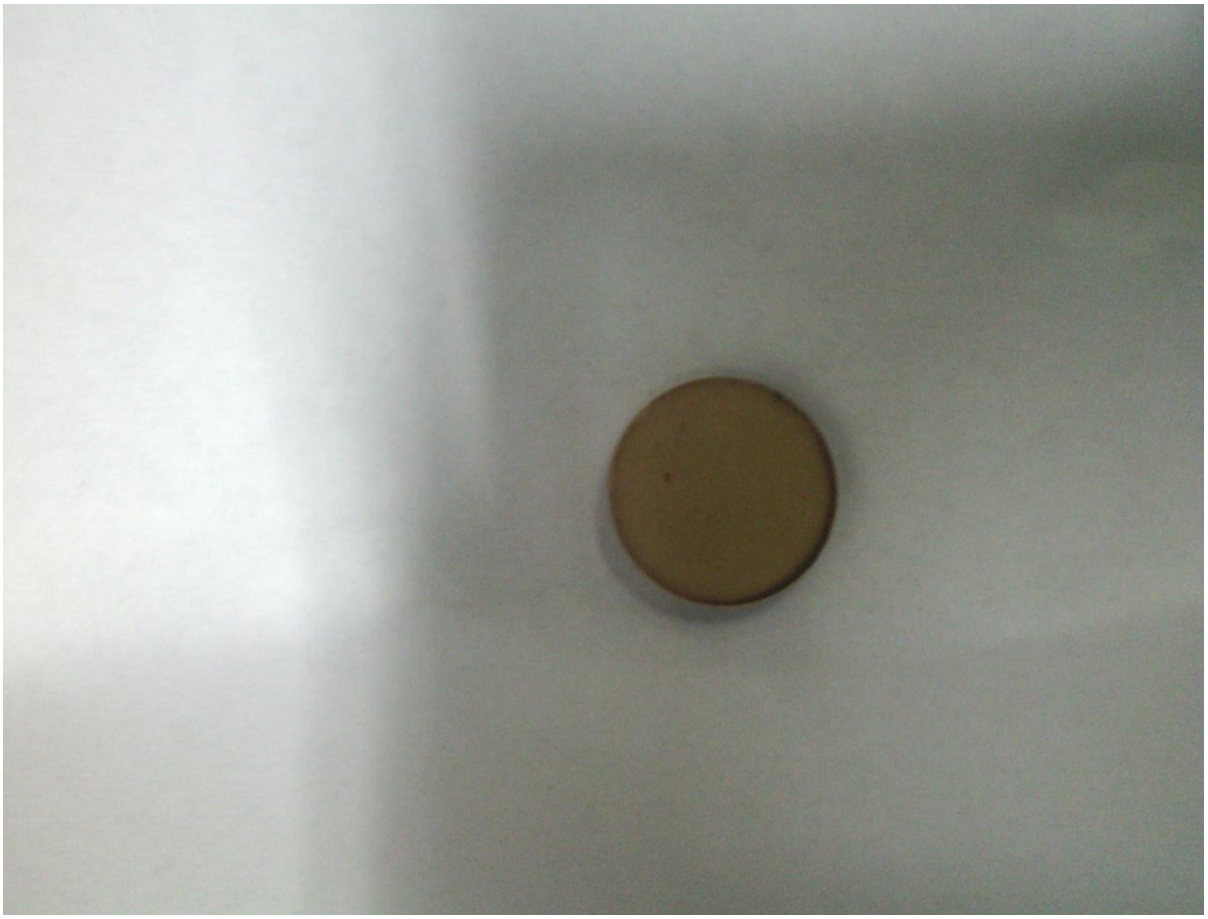

Figure A.1.6**: Pictograph of standard reference material; oyster tissue**

APPENDIX II: ANN and SVM Scripts

The ANN script used

net= newff (Intensities,Concentrations,3); (creating a new feed forward back propagation network of 3 neurons)

net.divideParam.trainRatio=.5; (Dividing the input into 60% for training set)

net.divideParam.valRatio=.3; (Dividing the input into 20% for validation set)

net.divideParam.testRatio=.2; (Dividing the input into 20% for test set)

[net,tr]=train(net, Intensities, Concentrations); (Applying the network on the Intensities and Concentrations values).

Output=net(Prediction Sample); (Predicting concentrations of an unknown sample)

The SVM script used

```
#Data
library(readxl)
TrainFe <- read_excel("R/TrainFe.xlsx")
View(TrainFe)
str(TrainFe)
library(ggplot2)
qplot(Wavelength.nm,Intensity.au,data=TrainFe,
    color=speciation)
#Support Vector
library(e1071)
```

```
mymodel<-svm(speciation~ .,data=TrainFe,type="C",
        kernel="radial")
summary(mymodel)
plot(mymodel,data=TrainFe,
    Intensity.au~Wavelength.nm)
#Confusion Matrix and Miscalculation Error
pred<-predict(mymodel,TrainFe)
plot(pred)
tab<-table(Predicted=pred,Actual=TrainFe$speciation)
tab
1-sum(diag(tab))/sum(tab)
#Tuning
set.seed(123)
tmodel<-tune(svm,speciation~.,data=TrainFe,kernel="radial",
        ranges=list(cost=2^(-5:0),gamma=c(0.001,0.01,.1,1,10,100)))
plot(tmodel)
summary(tmodel)
#best Model
mymodel<-tmodel$best.model
mymodel<-svm(formula=speciation~.,data=TrainFe,type="C",
kernel="radial",cost=0.25,gamma=10)
```

```
plot(mymodel,data=TrainFe,
    Wavelength.nm~Intensity.au)
Summary(mymodel)
Output=svm(RealFe)
plot(output)
Output<- svm(mymodel~ ., data=RealFe, type="C",
                    Kernel="radial")
plot(output,data=RealFe, type="C", kernel="radial", cost=0.03125,gamma=0.001)
```

APPENDIX III: Predicted Concentration Value of Trace Elements in Simulate Samples

Table A.3.1: Predicted versus Known concentration values of Fe and Mg

Predicted Concentration for Fe (ppm)	**Known concentration for Fe (ppm)**	**Predicted Concentration for Mg (ppm)**	**Known Concentration for Mg (ppm)**
100.9972	105	**323.7996**	328
125.2261	127	**342.0571**	346
128.7567	128	**428.6617**	429
142.6931	143	**501.85**	501
165.9006	166	**108.7174**	108
167.5066	170	**158.1268**	153
86.42138	81	**69.70388**	65
39.1122	40	**248.6395**	247
94.89202	89	**371.2531**	378
66.98382	60	**152.6011**	153

Table A.3.2: Predicted and known concentrations for Mn

Predicted Concentration	**Known Concentration**
9.222801	6
8.120721	10
18.47042	12
19.05044	16
18.75164	18

20.5412	20
22.6503	23
27.51827	25

Table A.3.3: Predicted and known concentrations for Cu

4.472516	3
4.711671	5
6.968894	4
13.81117	12
18.09768	18
13.0204	11
8.993601	9
15.31567	10
32.1976	28

APPENDIX IV: SNR for getting optimized LIBS features

Table A.4.1: Table showing the optimized conditions for LIBS apparatus. The SNR increases with energy, Q- switch delay optical to sample distance and number of ablations per scan

	Mg II 279.55 nm			Cu I 324.396 nm			Cu I 766.483 nm		
Energy	30 mJ	45 mJ	50 mJ	30 mJ	45 mJ	50 mJ	30 mJ	45 mJ	50 mJ
SNR	0.435	2.67	3.89	0.44	1.87	2.57	1.54	1.90	2.65
Q-switch delay	0.29 μs	0.33μs	0.49 μs	0.29 μs	0.33 μs	0.49 μs	0.29 μs	0.33 μs	0.49 μs
SNR	0.65	1.98	2.13	0.45	1.98	3.03	0.34	1.99	3.45
Optical to sample distance	10 mm	20 mm	33 mm	10 mm	20 mm	33 mm	10 mm	20 mm	33 mm
SNR	0.98	1.55	2.87	2.67	4.87	6.90	3.01	3.98	4.90
Number of ablations per scan	1	2	5	1	2	5	1	2	5
SNR	3.51	2.32	0.56	4.43	4.67	0.34	2.90	1.00	0.12