



**DEVELOPMENT OF CHEMOMETRICS AIDED ENERGY DISPERSIVE X-RAY
FLUORESCENCE AND SCATTERING (EDXRFS) METHOD FOR RAPID
DIAGNOSTICS OF CANCER**

By

OKONDA J. JUSTUS

I56/70422/2011

B.Ed (Hons)

A thesis submitted for examination in partial fulfillment of the requirements for the award of the degree of Master of Science in Physics of the University of Nairobi.

2015

DECLARATION

"I hereby declare that this thesis submitted for the degree of MSc at Department of Physics, University of Nairobi is my own original work and has not previously been presented to any other institution of higher education."

Okonda J. Justus

156/70422/2011

Department of Physics, University of Nairobi.

Signature: .....Date: 26/6/2015.....

This thesis is submitted with approval of the undersigned supervisors for examination,

Dr. Angeyo H. Kalambuka,

Department of Physics, University of Nairobi.

Signature: .....Date: 26/6/2015.....

Mr. Michael J. Mangala,

Institute of Nuclear Science & Technology, University of Nairobi.

Signature: .....Date: 26/6/2015.....

Prof. Seth M. Kisia

Department of Veterinary Anatomy and Physiology, University of Nairobi.

Signature: .....Date: 26/6/2015.....

ACKNOWLEDGEMENTS

I am indebted to the University of Nairobi for its material support towards the completion of this research. My gratitude goes to my supervisors, Dr. Hudson A. Kalambuka, Department of Physics, University of Nairobi, Mr. Michael J. Mangala, Institute of Nuclear Science & Technology, University of Nairobi and Prof. Seth M. Kisia, Department of Veterinary Anatomy and Physiology, University of Nairobi for their encouragement, support, guidance and advice throughout this research project. Their suggestions, comments and corrections made it possible for me to successfully complete my work in time.

I am also grateful to Dr. John D. Mande, Department of Clinical Studies and Head of Small Animal Clinic and the Surgery Section, University of Nairobi, Mr. Simon Bartilol of the Institute of Nuclear Science & Technology, University of Nairobi together with Mr. Charles Magiri and Miss Martha, of the Immunology and Histopathology laboratories at Kenya Medical Research Institute (KEMRI), for their substantial help during the sample collection and preparation stage of this study. Finally, yet important, I'm deeply touched and indebted to my wife; Micali and my daughter Gloria for being patient and understanding during my absence.

ABSTRACT

The goal of this work was to investigate and exploit the potential of chemometrics-assisted energy dispersive X-ray fluorescence and scattering (EDXRFS) spectrometry towards rapid diagnostics of cancer (and its severity) based on the correlative concentration, speciation and multivariate alterations of trace elements in human body and animal model tissues. This is due to the increasing realization that successful treatment of cancer depends on accurate diagnosis of the disease at the local stage of development based on trace elements as biomarkers.

Utility of EDXRF spectroscopy to direct rapid trace elemental analysis in body tissues is however at the moment limited by the complexity of the samples, extreme matrix effects and problematic recovery of weak fluorescence signals from the normally enhanced spectral background and overlapped fluorescence peaks. Further, analysis of elemental speciation, which is essential in disease diagnostics via trace element biomarkers, is not possible in conventional EDXRF spectroscopy. Energy dispersive X-ray fluorescence and scattering (EDXRFS) spectrometry exploits both X-ray fluorescence and Compton scatter radiations obtained directly and non-invasively from samples for correction of the matrix effects observed in the deconvolution of the fluorescence intensities to concentration and thus has the potential to be developed towards diagnostics of cancer at early stage of development. In this work, a method has been developed for diagnosis of cancer based on the use of paraffin wax ‘standards’ spiked with a multi-element stock solution of Fe, Cu, Mn, Zn and Se in the concentration range 11.5-120 ppm, 5-32.5 ppm, 0.75-9 ppm, 5-195 ppm and 0.5-5.5 ppm respectively. Simulate tissue samples containing Fe, Cu and Mn were also prepared in their respective oxidation states for development of a protocol towards speciation analysis. Spectra were acquired using the Shimadzu tube-excited EDX 800HS spectrometer for a live time of 50 s at 50 Kv and 1mA and analyzed by exploiting multivariate chemometrics. PCA was used for reduction of spectral data dimension and pattern recognition (classification). ANN and PCR were used to calibrate strategies for direct quantitative analysis using fluorescence spectral regions of interest and Compton scatter peaks. K-nearest neighbor (KNN) technique was used for quantitative speciation analysis of Mn, Fe and Cu. The results underscore the role of trace elements in cancer and indicate that Cu, Fe, Zn, Se and Mn can be used as trace biomarkers to identify the presence and stage of cancer development in the human body. ANN calibration strategy gave the good prediction results for trace element concentration analysis in tissues.

The indicative concentration ranges of Fe, Cu and Zn in cancerous tissues were 154.0 ± 4.5 - 191.2 ± 9.5 ppm, 17.4 ± 5.4 - 30.9 ± 2.6 ppm and 74.5 ± 26.4 - 103.2 ± 3.8 ppm respectively with quantifiable multivariate alterations observed in the concentrations of Mn and Se from early to late stage of cancer genesis. The observed elevated concentrations of Fe and Zn in cancerous tissues can be attributed to the significant role of these elements in cell growth and proliferation as a result of rapid cell division in neoplastic tissues. Results obtained from cultured tissue samples indicate that Fe increased gradually from 44.3 ± 4.6 - 169.7 ± 1.1 ppm with staging of cancer, which can be attributed to the role of Fe in blood supply to the neoplastic tissues. The Pearson correlation coefficient between Fe and Zn was high (0.98) in cultured cancerous as compared to 0.93 in cultured normal tissues, which can be linked to the increased metabolic activity in the neoplastic tissues. The results of speciation analysis indicate the dominant role of Fe^{2+} , Cu^{2+} and Mn^{4+} during progression of cancer from early to late stage. The higher speciation of Mn and Cu in cancerous tissues is probably due to the role of these elements in fenton reaction where they generate free radicals essential for cancer carcinogenesis. Levels of Cu, Fe, Se and Zn were thereafter used to successfully characterize real tissues as either cancerous or non-cancerous whereas speciation levels of Mn, Fe and Cu were used to successfully indicate the various stages of cancer. It may be concluded that chemometrics aided EDXRFS spectrometry may be exploited towards developing a cancer diagnostic model for rapid analysis of concentration levels and speciation alterations of trace elements (Fe, Cu, Zn, Se and Mn) in human body tissues which have hereby been evaluated as suitable biomarkers for diagnosis of cancer (especially at the local stage of development).

TABLE OF CONTENTS

DECLARATION	Error! Bookmark not defined.
ACKNOWLEDGEMENTS.....	ii
ABSTRACT	iv
LIST OF TABLES	ix
LIST OF FIGURES	x
LISTS OF ABBREVIATIONS/ACRONYMS AND SYMBOLS	xiii
CHAPTER 1: INTRODUCTION.....	1
1.1 Background	1
1.2 XRF analysis of trace elements in biological samples	5
1.3 Statement of the problem	8
1.4 General objective	9
1.5 Specific objectives	9
1.6 Justification and significance of the study	10
1.7 Scope and limitations of the study.....	11
1.8 Hypothesis of the study.....	11
CHAPTER 2: LITERATURE REVIEW	12
2.1 Overview	12
2.2 Trace element analysis in soft tissue.....	12
2.3 Basic principles of Energy Dispersive X-Ray Fluorescence (EDXRF) Spectroscopy.....	15
2.4 Principles of chemometric analytical spectroscopy for material analysis	16
2.4.1 Applications of multivariate chemometrics in XRF analysis.....	17
CHAPTER 3: THEORETICAL FRAMEWORK.....	20
3.1 Introduction	20
3.2 X-ray fluorescence analysis	20
3.3 Principles of quantitative X-ray fluorescence analysis	22
3.3.1 Fundamental parameters methods.....	22
3.3.2 Influence coefficients	24
3.4 Chemometric analytical spectroscopy	26
3.4.1 Pattern recognition	26
3.4.1.1 Principal Components Analysis (PCA)	27

3.4.1.2 K-nearest neighbor (KNN) technique.....	28
3.4.2 Artificial Neural Networks (ANN)	30
3.4.3 Principal Component Regression (PCR)	33
CHAPTER 4: MATERIALS AND METHODS	36
4.1 Introduction	36
4.2 EDXRF Spectrometric Instrumentation.....	36
4.3 Sampling	38
4.3.1 Simulate samples.....	38
4.3.2 Cultured tissue samples	39
4.3.3 Domestic dog tissue samples	39
4.4 Sample preparation	40
4.4.1 Simulate samples.....	40
4.4.2 Incubated cell cultures	43
4.4.3 Domestic dog tissues	44
4.5 Energy dispersive X-ray fluorescence (EDXRF) analysis.....	45
4.6 Chemometric analysis	45
4.6.1 Principal component analysis (PCA)	46
4.6.2 Artificial neural networks (ANN)	47
4.6.3 Principal Component Regression (PCR)	48
4.6.4 K-nearest neighbors (KNN) technique	48
CHAPTER 5: RESULTS AND DISCUSSION	51
5.1 Introduction	51
5.2 EDXRFS spectra of simulate samples	51
5.3.1 Cultured tissue photomicrographs	55
5.3.2 Detection limits of elements in simulate samples	56
5.4 Principal component analysis (PCA) exploration of simulate samples.....	57
5.5 Quantitative multivariate chemometric techniques	65
5.5.1 Artificial neural networks (ANN) for quantitative analysis of trace elements	65
5.5.2 Principal Component Regression (PCR) of trace elements	72
5.6 Speciation analysis of simulates using K-nearest neighbors (KNN).....	76
5.7 Principal Component Analysis (PCA) of cultured tissues	79

5.8 Analysis of NIST Oyster tissue using ANNs	83
5.9 Prediction of elemental concentrations in cultured tissue samples by ANN	83
5.10 Prediction of concentration of trace elements in domestic dog tissues by ANN	89
5.11 Speciation analysis of cancerous tissues using KNN	92
5.11.1 Results for cultured cancerous tissues	93
5.11.2 Results for domestic dog tissues	96
5.12 Principal Component Analysis (PCA) of selected dog tissue samples	99
CHAPTER 6: CONCLUSION AND RECOMMENDATION	103
REFERENCES	105
APPENDICES	114

LIST OF TABLES

<i>Table 4.4.1:</i> Chemical compounds used in the preparation of simulate samples	40
<i>Table 4.4.2:</i> Elemental concentration of the stock solution	41
<i>Table 4.4.3a:</i> Calibration set design for simulate samples for Cu^{2+} , Zn^{2+} , Fe^{3+} , Mn^{4+} and Se^{4+}	42
<i>Table 4.4.3b:</i> Calibration set design for simulate samples for Cu^{+} , Zn^{2+} , Fe^{2+} , Mn^{2+} and Se^{4+}	42
<i>Table 4.6.1:</i> Selected spectral regions of interest (feature selection) for Mn, Fe, Cu, Zn and Se	46
<i>Table 5.3.1:</i> Detection limits for direct univariate analysis of elements in simulate tissue	57
<i>Table 5.5.1:</i> Analytical performance indices for different ANN models	67
<i>Table 5.5.2:</i> Regression coefficients for ANN and PCR	75
<i>Table 5.8.1:</i> Evaluation of elemental concentrations in Oyster tissue (NIST 1566b)	83
<i>Table 5.9.1:</i> ANN predicted concentration for cultured samples using fluorescence peaks	84
<i>Table 5.9.2:</i> ANN predicted concentration for cultured samples using both fluorescence and Compton scatter	85
<i>Table 5.9.1.1:</i> Correlation coefficients of elemental concentrations in cultured cancer tissues	87
<i>Table 5.9.1.2:</i> Correlation coefficients of elemental concentrations in healthy cultured tissues	88
<i>Table 5.10.1:</i> Concentrations of trace elements in histopathologically classified dog tissue samples analyzed using fluorescence peaks	89
<i>Table 5.10.2:</i> Pearson correlation coefficients of elemental concentrations in prostate tissue samples for early stage cancer	91
<i>Table 5.10.3:</i> Pearson correlation coefficients of elemental concentrations in prostate tissue samples for advanced stage cancer	91
<i>Table 5.10.4:</i> Pearson correlation coefficients of elemental concentrations in mammary tissue samples for advanced stage cancer	91
<i>Table 5.11.1:</i> KNN results for speciation of Fe, Cu and Mn in cultured cancer samples	94
<i>Table 5.11.2:</i> KNN results for speciation of Fe, Cu and Mn in cultured healthy tissues	95
<i>Table 5.11.3:</i> Speciation of Fe, Cu and Mn in cultured cancer tissue	96
<i>Table 5.11.2.1:</i> KNN results for speciation of Fe, Cu and Mn in selected dog tissues	97
<i>Table 5.11.2.2:</i> Speciation concentrations of Fe, Cu and Mn in selected dog tissue samples	98

LIST OF FIGURES

<i>Figure 3.2:</i> Schematic diagram showing the attenuation of X-rays in a target sample.	21
<i>Figure 3.4.1:</i> General purpose of principal component analysis.	27
<i>Figure 3.4.2:</i> Schematic diagram of K-nearest neighbor (KNN)	29
<i>Figure 3.4.3:</i> Natural neural network (After: Fausett, 1994).	30
<i>Figure 3.4.4:</i> Artificial neural network as a black-box.	31
<i>Figure 3.4.5:</i> Internal organization of a BP-ANN for multivariate calibration of EDXRFS spectra (After: Jose et al. 2009).	32
<i>Figure 3.4.3.1:</i> Schematic diagram of Principal Component Regression.	34
<i>Figure 4.2.1:</i> Schematic presentation of EDXRF instrumentation arrangement.	37
<i>Figure 4.6.1:</i> Flow chart for pattern recognition using KNN (Sartoros and Eric, 1997).	49
<i>Figure 4.6.2</i> Block diagram showing the steps involved in for model development.	50
<i>Figure 5.2.1:</i> Typical EDXRFS spectrum of S10H simulate tissue.	52
<i>Figure 5.2.2:</i> Typical EDXRFS profiles of S10H simulate tissue acquired at 50 and 100 s.	53
<i>Figure 5.2.3:</i> Typical EDXRFS spectral overlaps for S10H and paraffin wax (base matrix).	54
<i>Figure 5.2.4:</i> Typical EDXRFS spectrum of incubated Vero SV9 cancerous tissue sample.	55
<i>Figure 5.3.1:</i> Photomicrograph of Lymphocyte cells harvested at day 4 (300×).	56
<i>Figure 5.3.2:</i> Photomicrograph of Cultured Vero cancer cells harvested at day 4 (300 ×).	56
<i>Figure 5.4.1:</i> PC1 (77%) × PC2 (16%) score plot of simulate samples.	58
<i>Figure 5.4.2:</i> PCA loadings plot of simulate samples for 50s live time for a total fluorescence.	59
<i>Figure 5.4.3:</i> PC1 (79%) × PC2 (15%) score plot of simulate samples for 100s live time.	60
<i>Figure 5.4.4:</i> PCA loadings plot of simulate samples for 100s live time for a total fluorescence.	61
<i>Figure 5.4.5:</i> PC1 (75 %) × PC2 (20 %) score plot of simulate samples for 50s live time for the fluorescence peaks plus Compton scatter.	62
<i>Figure 5.4.6:</i> PCA loadings plot of simulate samples for 50s live time using both fluorescence plus Compton scatter.	63
<i>Figure 5.4.7:</i> PC1 (84%) × PC2 (11%) score plot of simulate samples using selected fluorescence region of Cu, Fe and Mn for 50s live time.	64
<i>Figure 5.4.8:</i> PCA loading plot showing the variables of simulate samples for 50s live time using fluorescence region of Cu, Fe and Mn.	64

<i>Figure 5.5.1.1:</i> Performance plots for ANNs training errors for 4 neurons exploiting selected (Fe, Cu, Zn, Mn and Se) fluorescence region plus Compton scatter.	66
<i>Figure 5.5.1.2:</i> Linear regression of ANN outputs and corresponding targets utilizing the selected fluorescence regions plus Compton scatter regions for simulates.	68
<i>Figure 5.5.1.3:</i> Linear regression of ANN outputs and corresponding targets utilizing selected fluorescence regions.	69
<i>Figure 5.5.1.4:</i> ANN regression plots for predicted verses known Zn utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.	70
<i>Figure 5.5.1.5:</i> ANN regression plots for predicted verses known Cu utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.	70
<i>Figure 5.5.1.6:</i> ANN regression plots for predicted verses known Mn utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.	71
<i>Figure 5.5.1.7:</i> ANN regression plots for predicted verses known Fe utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.	71
<i>Figure 5.5.1.8:</i> ANN regression plots for predicted verses known Se utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.	72
<i>Figure 5.5.2.1:</i> PCR regression plots for predicted verses known Zn utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.	73
<i>Figure 5.5.2.2:</i> PCR regression plots for predicted verses known Cu utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.	73
<i>Figure 5.5.2.3:</i> PCR regression plots for predicted verses known Mn utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.	74
<i>Figure 5.5.2.4:</i> PCR regression plots for predicted verses known Fe utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.	74
<i>Figure 5.5.2.5:</i> PCR regression plots for predicted verses known Se utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.	75
<i>Figure 5.6.1:</i> Speciation of Cu for S6L and S6H using KNN technique.	77
<i>Figure 5.6.2:</i> Speciation of Fe for S3L and S3H using KNN technique.	78
<i>Figure 5.6.3:</i> Speciation of Mn for S1L and S1H using KNN technique.	79
<i>Figure 5.7.1:</i> PCA score plot analysis for staging of cancer in cancer tissue cultures using selected fluorescence peaks of Mn, Fe, Cu, Zn and Se.	80

<i>Figure 5.7.2:</i> PCA loadings plot of cultured samples for 50s live time using partial fluorescence peaks.	81
<i>Figure 5.7.3:</i> PCA score plot analysis for staging of cancer in incubated tissue cultures using both fluorescence plus Compton scatter.	82
<i>Figure 5.7.4:</i> PCA loadings plot of cultured samples for 50s live time using both fluorescence plus Compton scatter.	82
<i>Figure 5.9.1:</i> 3-D grouped bar graph showing the variation in concentration of trace elements in cultured cancerous and healthy tissues.	86
<i>Figure 5.11.1.1:</i> Speciation of Fe in typical sample SV1 using KNN technique.	93
<i>Figure 5.11.1.2:</i> Speciation of Mn for sample SV7 using KNN technique.	93
<i>Figure 5.11.1.3:</i> Speciation of Fe and Cu for in sample SV1 using KNN technique.	94
<i>Figure 5.11.2.1:</i> Speciation of Mn for sample SP4B using using KNN technique.	97
<i>Figure 5.12.1:</i> PCA score plot analysis for domestic dog tissue samples using partial fluorescence region.	99
<i>Figure 5.12.2:</i> PCA loadings plot of domestic dog tissue samples using partial fluorescence regions.	100
<i>Figure 5.12.3:</i> PCA score plot analysis of domestic dog tissue samples using selected spectral regions.	100
<i>Figure 5.12.4:</i> PCA loadings plot in domestic dog tissue samples.	101

LISTS OF ABBREVIATIONS/ACRONYMS AND SYMBOLS

CT	Computed tomography
MRI	Magnetic resonance imaging
KEMRI	Kenya Medical Research Institute
XRF	X-ray fluorescence
EDXRF	Energy dispersive X-ray fluorescence
EDXRFS	Energy dispersive X-ray fluorescence and scattering
ROS	Reactive Oxygen Species
WD-XRF	Wavelength Dispersive X-ray Fluorescence
PBMCs	Peripheral Blood Mononuclear cells
FP	Fundamental Parameters
SNR	Signal-to-Noise Ratio
Rmm	Relative molecular mass
Ram	Relative atomic mass
PCA	Principal Component Analysis
ANNs	Artificial Neural Networks
PCR	Principal Component Regression
KNN	K-Nearest Neighbor
TXRF	Total Reflection X-ray Fluorescence
μ -XRF	Micro X-ray Fluorescence
SR-XRF	Synchrotron Radiation X-ray Fluorescence
SRIXE	Synchrotron Radiation Induced X-ray Emission
XANES	X-ray Absorption Near-Edge Structure
ROC	Receiver Operating Characteristic
PLS	Partial Least Squares
QA	Quality Assurance
CRM	Certified Reference Material
BDL	Below Detection Limit

CHAPTER 1: INTRODUCTION

1.1 Background

Energy dispersive X-ray fluorescence (EDXRF) spectroscopy is an analytical method for materials analysis, which identifies and quantifies the element contents of samples by measurement of the spectral emission energy and intensity of elements (Silva *et al.*, 2012; Karen *et al.*, 2007). However, due to enhanced spectral background and low signal to noise ratio as well as fluorescence yield of elements in low-Z matrices, there is a draw back in the quantitative analysis of trace elements from the above EDXRF spectral data in biomedical samples. In practice, EDXRF spectra consist of fluorescence peaks and line overlaps, spectral artifacts, scatter peaks and weak signals of low-Z elements. This presents major challenges in the determination of low analyte concentration masked in low-Z complex matrices. These challenges may be partly overcome by utilizing chemometrics (Jose and Neil, 2007) to extract latent relationships in the analytical data. Chemometrics is a relatively new discipline for application of mathematical, univariate and multivariate statistical methods for extracting comprehensive physical and chemical information from the spectroscopic measurements in complex multivariate spectra (Brereton, 2003).

X-ray fluorescence (XRF) spectra when analyzed by chemometrics can yield information on among other sample properties, the concentrations and speciation of trace elements in low-Z matrices. This is following training of the spectral data with appropriate calibration standards which makes XRF an ideal method for biomedical analysis.

Development of an EDXRF technique for cancer diagnosis based on the concentration, speciation and alterations of trace elements in tissues and body fluids has the potential to yield the most viable biomarkers for early diagnosis of diseases. The available conventional methods

for cancer diagnosis include; mammography, X-ray computed tomography (XCT), scintillation scan, ultrasound, biopsy and magnetic resonance imaging (MRI). In general, these methods are not effective for early diagnosis of cancer, they are expensive, and require specialist skills. Innovative spectroscopic techniques have potential to be rapid, mostly non-destructive, affordable and reliable for diagnosis of cancer at the local stage of development and to quantify and characterize its severity and type.

Trace elements are those that occur in body tissues in the order of a few micrograms per gram (Henry *et al.*, 1995): they constitute about 0.01% by weight of the total body composition. Trace elements get into the body through air, food, water, or dermal exposure and cross the plasma membrane to enter the cells. Some trace elements are regarded as nutritionally essential to humans and are present in proteins, enzymes and cellular membranes for normal physiological functions. In general, trace metals in human tissue are classified as; nutritionally essential (Co, Cr, Cu, Mn, Se, Zn, and Fe), those with possible beneficial effects which include B, Ni, Si and V while those with unknown beneficial effects include Al, Sb, As, Ba, Pb, Ag and Hg (Rose, 1983; Prasad, 1978).

The essential trace metals have four main functions which include acting as (i) stabilizers, (ii) elements of structure, (iii) essential elements for hormonal function, and (iv) co-factors in enzymes (Mehmet *et al.*, 2007). They are associated with chemical reactions in the body; hence are essential for the control and maintenance of body functions (Silva *et al.*, 2009). Inadequate or lack of trace elements will not only affect the structure but also its structural function due to lack of stabilization and change of charge properties (Feinendegen and Kasperek, 1980). However, when in abnormal proportions, essential trace metals contribute to several pathological processes, including tumor growth, invasion and metastasis (Yaman, 2006; Sigel, 1980; Rizk and

Skypeck, 1984). The elements contribute to development of cancer based on their concentration and/or speciation as they are either activators or inhibitors of enzymatic reactions which thus affects directly or indirectly the carcinogenesis of cancer.

Iron (Fe) is the main regulator for the creation of blood supply sites to growing malignant tissue acting as a regulatory factor in angiogenesis (Richardson *et al.*, 2002). It also acts as a catalyst in the conversion of hydrogen peroxide to free radicals that attack cellular membranes, leading to inactivation of enzymes and breaking of DNA strands (Konemann *et al.*, 2005). Copper (Cu) plays an important role in angiogenesis as it is the main essential enzyme regulator and a co-factor for Fe metabolism where it is required in human body for the conversion of the body's Fe to haemoglobin (Nasulewicz *et al.*, 2004). It is also involved in oxidative enzymatic reactions which produce hydroxyl radicals that modify protein, lipids and DNA thereby leading to carcinogenesis (Huang *et al.*, 1999). Zinc (Zn) is a versatile trace element required as a co-factor by more than 200 enzymes (Grodner *et al.*, 2000); it acts as a protector of cellular growth including growth of neoplastic cells. It is essential for protein, DNA and RNA synthesis and thus plays an important role as an activator of matrix metal proteinase which contributes to tumor cell invasion and metastasis (Hambidge *et al.*, 1986).

Speciation of trace elements in human blood occurs in two forms: protein associated species, which play significant roles in enzyme activities and in transportation and storage of trace elements; and protein non-associated species, which are related to membrane transport or excretion of the elements (Guardiner, 1987). When normal cells transform into cancer cells, their protein-expression profiles drastically change. Proteins require metals to carry out their functions, and enzymes use a particular metal as a cofactor; thus metals are crucial for life and are indispensable for a series of biological processes.

Carcinogenesis is considered to occur in four stages: *initiation, promotion, progression, and metastasis* (Geraki *et al.*, 2002). The mechanism of metal-induced carcinogenesis is believed to be involved in all stages of cancer development. Consequently, the roles of trace metals in cancer development and inhibition are complex and have raised many questions because of their essential and toxic effects on people's health. Susceptibility to cancer is characterized by extensive DNA damage which is thought to result from decreased repair capacity and/or by the direct carcinogenic interaction of metallic ions with DNA (Hartwig *et al.*, 2002). The carcinogenic capability of these metals depends mainly on factors such as concentration, speciation and chemical structures. The oxidative concept in metal carcinogenesis proposes that complexes formed by these metals are *in vivo* within the vicinity of DNA where they catalyze redox reactions which in turn oxidize DNA (Kryston *et al.*, 2011). The most significant effect of reactive oxygen species (ROS) in the carcinogenesis progression is DNA damage, which results in isolated base lesions or single-strand breaks, complex lesions like double strand breaks as well as some oxidatively generated clustered DNA lesions and the sister chromatid exchange (Stephen, 2013; Valko *et al.*, 2005). Nutritionally essential trace elements individually and collectively appear to contribute to various pathological processes; however, there is still lack of information concerning their function in staging (initiation, promotion, progression and metastasis) and inhibition of cancer (Banas *et al.*, 2001).

There has been previous study interest using several complementary analytical methods to establish trace elements differences between healthy and malignant tissues (Edger *et al.*, 1989). This can be attributed to changes in concentration, speciation and distribution of the trace elements in biological samples (Kubala-Kukus *et al.*, 2004). However, there is hardly an analytical method that is capable of simultaneously furnishing with speed, reliability and accuracy the concentration, speciation and multivariate alteration of trace elements in human

body tissues. The impact of metals in disease is not necessarily due to total element content alone (Caruso *et al.*, 2003); the metal speciation is additionally important (Szupunar and Lobinski, 1999; Rojas *et al.*, 1999). Lately, there has been increasing interest in chemical speciation of elements which can be attributed to alterations in the chemical and biological properties of the elements depending on their oxidation state (Kallithrakas and Moshohoritou, 1998).

The changes in concentration, speciation and correlation of cancer associated trace elements in response to metastasis when detected, measured and characterized, may be considered as a “blueprint” of a particular cancer capable of constituting a diagnostic for the presence and level of severity of the cancer (Kuehnelt *et al.*, 2005). There is therefore need to rapidly and non-invasively obtain information on the concentration, speciation and multivariate alterations of trace metals in tissues in relation to cancer.

1.2 XRF analysis of trace elements in biological samples

X-ray fluorescence (XRF) spectrometry, is a proven analytical technique widely used for both qualitative and quantitative determination of most heavy Z elements in various materials with detection limits down to a few parts per million. The technique is a very convenient tool for the study of trace elements and their relationship in body organs. Conventional XRF spectral data is confined in the region between 0.5 keV to 50 keV with four modes of radiation interaction; photoelectric absorption, radiationless transition (the Auger effect), elastic scattering (Rayleigh scattering) and inelastic scattering (Compton scattering). The desired mode of radiation interaction in XRF spectrometry is photoelectric absorption where the incident X-ray radiation is entirely absorbed by the atoms to excited state. De-excitation results in the emission of characteristic X-rays from the atom. The energy of the emitted X-ray radiation is distinctive for a

certain element while the number of X-rays detected is a measure for the elemental concentration.

Two common modes of X-ray spectrometry exist, namely; Wavelength Dispersive X-ray Fluorescence (WD-XRF) and Energy Dispersive X-ray Fluorescence (ED-XRF). The main distinctive feature between the two methods is the detection and measurements of spectral data; WD-XRF uses an analyzing crystal to diffract the different X-ray wavelengths for detection of X-rays at various angles while Energy Dispersive X-ray Fluorescence (ED-XRF) uses a solid state detector or proportional counter to measure all energies and X-ray radiation intensities simultaneously. To accurately determine the X-ray energy, all the X-ray radiations resulting from each photoelectric effect are analyzed following storage of digital calibration.

The complexity of organic sample spectral data remains a major limitation in the EXDRF spectrometry application for analysis of tissues. In general, biomedical samples are composed mainly of low-Z atoms i.e carbon, hydrogen and oxygen atoms with heavy elements at trace concentration (Banas *et al.*, 2001). This results to poor signal to noise ratio and large background level due to back-scattered radiation leading to low detection limits of the analytes. The X-ray photo peak (fluorescence) intensities have been used to evaluate the elemental concentration using mostly the Fundamental Parameters (FP) and Influence Coefficient methods (Criss and Birks, 1968) based on the assumptions of sample as being homogeneous and *a priori* knowledge of the sample matrix composition.

The analytical challenges in direct analysis of biomedical samples with complex matrices; extreme matrix effects, enhanced Compton scatter, spectral overlap, poor signal-to-noise ratio (SNR) and elevated background by EDXRF spectrometry has stimulated development of alternative techniques for biomedical sample analysis (Fabiola *et al.*, 2005). The work presented

herein utilizes Energy Dispersive X-ray Fluorescence and Scattering (EDXRFS) spectroscopy for simultaneous non-invasive acquisition of both fluorescence and scatter spectra from samples in order to build a model for determination of trace elemental content in complex matrix materials. In this thesis, we have developed the method to analyze cancer tissues. The EDXRF method is of great analytical use when combined with chemometrics (Kaniu *et al.*, 2012) to derive information about the overall matrix composition using appropriate multivariate chemometric calibration techniques (Kessler *et al.*, 2002; Schram, 2000). For instance, Principal Component Analysis (PCA) has ability to reduce complexity of EDXRFS spectral data so as to mine out the most important information in relation to the physical and chemical structure of the samples. Artificial Neural Networks (ANN) and Principal Component Regression (PCR) enable direct relation of spectral data to analyte concentration through multivariate calibration in which peak overlaps, weak signals and matrix effects are resolved (Lemberge *et al.*, 2000). ANNs has especially proved to be a powerful tool for extracting useful information and revealing inherent relationship in multivariate data analysis and thus offers potential for speciation analysis through pattern recognition techniques such as K-nearest neighbor (KNN).

Other than the above approach, qualitative (speciation) information may also be acquired through other pattern recognition techniques such as PCA and KNN (Massart *et al.*, 1988). Chemometrics techniques facilitate extraction of important information (e.g. underlying biological, chemical and physical phenomena) from complex EDXRFS spectral data sets by reducing the data complexity while increasing the information gained which are otherwise difficult to analyze with univariate methods. EDXRFS spectroscopy coupled with chemometrics is hereby demonstrated as potential diagnostic tool for cancer diagnosis based on the concentration, speciation and multivariate alterations of trace elements (Mn, Fe, Cu, Zn and Se) in human body tissues.

1.3 Statement of the problem

X-ray fluorescence (XRF) spectroscopy enables simultaneous multi-elemental analysis of trace elements in samples. However, information on the trace concentrations and speciation of elemental constituents of body tissue contained in the XRF spectrum suffers from poor detection limits for complex low-Z matrices; XRF is not suited for light elemental analysis and in principle, inapplicable for speciation analysis. In general, EDXRF encounters analytical challenges in the direct, rapid, non-invasive analysis of biomedical samples due to the extreme matrix effects, spectral overlap and poor signal-to-noise ratio (SNR) for the trace analytes. This limits its potential applicability to simultaneous multivariate analysis of trace elements' concentration, speciation and alterations in low-Z matrices (soft body tissues) which calls for an alternative non-invasive spectroscopic technique for disease diagnostics based on the trace signatures as cancer biomarkers.

1.4 General objective

The goal of this study was to develop a chemometric-aided energy dispersive X-ray fluorescence and scatter (EDXRFS) spectrometric method for direct rapid diagnosis and characterization of cancer and its severity (staging) in human body tissue.

1.5 Specific objectives

- i. To develop and validate multivariate calibration models (based on ANN and PCR) for determination of trace element (Zn, Se, Fe, Cu, and Mn) concentrations in model tissue using EDXRFS spectrometry.
- ii. To predict and quantify the trace element (Se, Zn, Fe, Cu, and Mn) contents in tissue cultures using EDXRFS spectroscopy utilizing the better calibration model for each element developed in (i) above.
- iii. To investigate and exploit PCA and KNN in conjunction with the information in (i) and (ii) above to determine the speciation of Fe, Cu and Mn in cultured cancer tissues.
- iv. To propose a diagnostic model using domestic dog tissues for detection, quantification of cancer and its severity (staging) based on the above elemental concentrations, spectral signatures, speciation and their multivariate correlations and alterations using PCA, KNN, ANN and PCR.

1.6 Justification and significance of the study

Cancer, especially breast and prostate are the leading cause of death in the world (WHO, 2012). The disease accounted for 7.9 million deaths (about 13% of all deaths worldwide) in 2009. In Kenya, cancer is the third highest cause of death, with about 7% of the total national deaths every year and about 28,000 cases reported resulting to over 22,000 deaths (Ministry of Health, 2011). The overall burden of cancer in the world is projected to continue increasing particularly in developing countries (Mathers and Loncar, 2006). The available methods for diagnosis are often possible at an advanced stage which is expensive and delays in treatment resulting to high mortality rate than the survival rate. Most of the cancer patients seen in Kenya are diagnosed at an advanced stage where management, monitoring and treatment is difficult if not impossible (Nairobi Cancer Registry). Successful diagnosis of cancer at the local stage of development can be based on the combined concentration of trace elements (Mn, Fe, Cu, Zn and Se) and speciation of Cu, Fe and Mn as cancer biomarkers. The quest for early diagnosis of cancer is based on the fact that early intervention translates to higher survival rate and successful treatment of the disease.

EDXRF analytical method is based on the fact that determination of concentration of all the elements of interest present in the sample must be known so as to correct for matrix effects, and only relies on the fluorescence X-ray radiations of heavy elements where possible. The method is not suited and applicable to low-Z elements and for direct trace metal as well as speciation analysis. Reliable analysis of biomedical samples is only possible when EDXRF is combined with chemometric techniques such as PCA, KNN, ANN and PCR that yield multivariate information which can be interpreted and used to determine trace element concentration, speciation and multivariate alterations in cancer tissues after appropriate training with calibration standards. This has the ability to provide necessary diagnostic information about cancer, its type

and staging following application to speciation analysis of biomedical samples as a basis for early diagnosis.

1.7 Scope and limitations of the study

The study was limited to five trace elements; Fe, Cu, Mn, Zn and Se. This is based on the fact that the elements are part of essential trace metals in body tissues and play major roles in carcinogenesis process. In addition, speciation analysis of trace elements was limited to three trace metals, namely; Fe, Cu and Mn which are actively involved in generation of free radicals in Fenton reaction that initiates carcinogenesis.

1.8 Hypothesis of the study

Carcinogenesis is greatly influenced by concentration of trace elements; Fe, Cu, Mn, Zn and Se and also by speciation of Fe, Cu and Mn.

CHAPTER 2: LITERATURE REVIEW

2.1 Overview

This chapter contains five sections that review the basis of this study. Section 2.2 highlights conventional methods of trace element analysis in tissues while section 2.3 outlines the basic principles of EDXRF spectrometry, the technique used in this study. Section 2.4 introduces analytical chemometric spectroscopy while section 2.5 reviews the utility of multivariate chemometrics in XRF analysis.

2.2 Trace element analysis in soft tissue

Elemental concentrations of Ca, Fe, Cu and Zn in healthy and cancerous breast tissues have been studied using mono-energetic Synchrotron Radiation X-ray Fluorescence (SR-XRF) method (Silva *et al.*, 2009). Scattered radiation from each sample was used as an internal standard for construction of calibration curves. Elemental concentrations were obtained by comparing the ratio between fluorescence peak areas and scattered peak areas, determined from tissue spectra, to the equivalent ratio in calibration curves. The levels were significantly higher ($p < 0.001$) in neoplastic tissues than in healthy breast tissues with concentration of Fe and Cu positively correlated in malignant breast tissues as well as Ca and Zn levels. These results were in agreement with the hypothesis that there is a connection between Fe and Cu with increased cellular activity and blood supply in the formation of cancer in breast tissue. In a similar study, Geraki *et al.* (2002) investigated the correlations between the breast tissue levels of Fe with the presence of breast cancer by using 40 healthy and 40 tumor samples. In the study, mean concentrations of Fe in specimens obtained from healthy individuals differed significantly from the mean concentrations of normal tissue adjacent to breast tumors (31% difference from the average) which was statistically significant. This apparent difference in Fe concentrations between the two groups needs to be replicated using a larger number of samples. In another

similar study (Poletti *et al.*, 1998), the concentrations of Fe, Cu and Zn were determined in breast tissues in which a decrease in concentration levels of Fe and Zn compared to normal concentration in breast tissue samples was evident. This indicates the anti-carcinogenic role of Fe and Zn which cannot explain the increased blood supply and cell proliferations in tumors.

Synchrotron Radiation Induced X-ray Emission (SRIXE) technique has been applied to determine elemental concentrations in 20 mm thick kidney and prostate tissue sections, taken from cancerous and healthy organs (Banas *et al.*, 2001). The spectra were fitted using GNUPLOT program which deconvoluted low and poorly resolved peaks. Significant concentration differences in elemental concentration between cancerous and non-cancerous tissues were evident. Mn, Se and Fe were at elevated levels in cancerous part of the tissues in comparison to non-cancerous parts while the concentrations of V, Cr, Zn and Cu were suppressed. The results for Mn and Fe are in agreement with the fact that free radicals generation is a precursor to cancer initiation. SRIXE data analysis in biomedical samples was quite a challenge due to matrix interference from scattered radiation and poor signal to noise ratio due to low-Z elements in the samples.

Healthy and cancerous prostate tissue sections have also been analyzed using Synchrotron Radiation-Induced X-ray Emission (SRIXE) in conjunction with X-ray Absorption Near-Edge Structure (XANES) to determine the oxidation state of Fe (Kwiatek *et al.*, 2004). The XANES spectra were recorded for selected spectral regions with high concentrations of Mn, Fe, Cr and Zn. The above elements were quantified and oxidation states of Fe in prostate tissues (cancerous and non-cancerous) determined. The results indicated high levels of Fe and Mn in the cancerous sections of the analyzed tissues than the corresponding non-cancerous sections. The increased concentration of Fe in cancerous tissue sections is attributed to its catalytic role in the Fenton

reaction (Halliwell *et al.*, 1984), which facilitates the production of the free radicals responsible for supply of oxygen for cancer growth. On the other hand, the concentration of Zn was suppressed in cancerous tissues. From the results of XANES spectra analysis, oxidation state determination of Fe was achieved based on pre-edge peak in selected standards that were assumed to be linearly dependent on the energy of the emitted X-rays which was also assumed to be dependent on the oxidation state of Fe. Based on the above assumptions, the oxidation state of Fe was found to be +3 in cancerous prostate tissues.

In EDXRF analysis of Ca, Fe, Cu, and Zn, the trace element concentrations were determined in healthy and breast cancer tissues (Silva *et al.*, 2012). Cut-off values for each trace element were determined through receiver operating characteristic (ROC) analysis from the elemental distributions. These values were used to set the positive or negative expression that was subsequently correlated with clinical prognostic factors through Fisher's exact test and chi-square test. Concentrations of the trace elements were found to be higher in neoplastic tissues (malignant and benign) compared to normal tissues. Results from ROC analysis indicated that trace elements can be considered as tumor biomarkers. Based on established cut-off value, it was possible to classify different tissues as healthy or cancerous, as well as to differentiate types of cancer.

In a similar study of trace element concentrations in human tissues by EDXRF spectrometry (Carvalho *et al.*, 1998), the concentration of Fe, Cu and Zn in the liver were found to be higher than in kidney tissues, this is attributed to the fact that the former is a store for the excess of these elements in the organism. For hair tissue, a significant difference between male and female was established for Zn. The Pearson correlation matrix was used to investigate the correlation between elemental concentrations and the dependence of these concentrations on age which

clearly indicated a weak negative correlation with age for Cu and Sr. No significant correlations existed among the studied elements in hair, except for Fe and Cu. For bone samples, Zn, Sr and Pb indicated a significant positive correlation with age. The correlation among elements in bone was significant only for Fe and Cu.

Geraki and Farquharson (2001) and Siddiqui *et al*, (2006) analyzed Zn concentration in breast tissue biopsies, mastectomies and breast reduction surgeries. The results suggested significant elevation of Zn in the pathological tissues compared to the benign and normal tissues. The above studies tend to suggest that Zn accumulates in diseased samples compared to healthy samples and it is not clear if the uncontrolled growth of cells in the tumor tissues accumulates more Zn to cope with the demands of the excessive dividing of the cells.

2.3 Basic principles of Energy Dispersive X-Ray Fluorescence (EDXRF) Spectroscopy

Energy Dispersive X-Ray Fluorescence (EDXRF) spectrometry (Beckhoff, 2006) is a widely used spectro-analytical method for simultaneous elemental analysis in a wide range of samples presented as solids, powders and particulates (Angeyo *et al.*, 2012). It is based on measurement of X-ray intensities emitted by the elements contained in a sample, when irradiated with X-ray or other radiation sources (Bertin, 1991). EDXRF quantitative analysis of samples requires application of adequate empirical or theoretical methods that enhance conversion of the measured fluorescent intensities into the concentration of the analytes. The evaluation of EDXRF spectra begins with the determination of net intensities (Schramm, 2000) which are inputs in empirical methods and Fundamental Parameters (Sherman, 1958), for the calculation of element concentrations. The elemental analysis of thick specimen is rather complicated because the measured intensities not only depend on the analyte concentration but also on accompanying non-detectable matrixes of low-Z elements like in the case of C, H and O in biomedical samples.

The main challenge in EDXRF spectrometry is the determination of the elements of interest at low concentrations in presence of absorbing and interfering matrix elements. The quantitative determination of the elements achieved by methods such as fundamental parameters (FP) and empirical models do not rely on a large set of standards. In addition, it is not necessary to have matrix and concentration levels matching those of the samples to be analyzed. These features of fundamental parameters FP method and empirical coefficients are, however, based on the assumptions of homogeneity, negligible particle size effects and complete definition of the sample matrix (Kaniu *et al.*, 2012). These assumptions limit EDXRF in the analysis of complex especially biomedical matrices.

2.4 Principles of chemometric analytical spectroscopy for material analysis

The correlations between trace elements and their speciation in human tissues and body fluids together with corresponding spectral patterns are complex. This generates a multivariate relationship which can be resolved by use of chemometric techniques. Chemometrics is a relatively new discipline that utilizes mathematical and statistical methods by using formal logic for design of optimal experimental procedures. These provide maximum relevant quantitative and qualitative information following analysis of multivariate data (Kurt *et al.*, 2008; Massart *et al.*, 1988). Chemometric evaluation of XRF spectra involves combining conventional methods for elemental identification and quantification and application of multivariate tools which enhances separation of non-characteristic information of the spectra.

Chemometrics plays an important role in data reduction and enabling for establishing indirect function relationships among data, and building multiple dimensional structures (Liqiang and Luo, 2006). It is capable of handling noisy and incomplete data, and self-adaptive learning to a certain degree to develop descriptive models for data analysis. Among the best valuable

chemometric techniques are; artificial neural networks (ANNs), Principal component analysis (PCA), K nearest neighbors (KNN) and principal component regression (PCR).

In general, chemometric analysis mainly deals with the following broad problem formulations (Kowalik and Einax, 2006):

- (i) reduction of large data sets into fewer dimensions;
- (ii) elimination of “noise” and redundancy information;
- (iii) visualization of multidimensional data sets;
- (iv) extraction of relevant information; and
- (v) hypothesis formulation.

One of the benefits of using chemometrics is its ability to represent multivariate data in a graphical interface: data reduction permits visualization of the relationship between samples characterized by multiple measured variables. Chemometrics also affords exploratory analyses, which opens up possibilities for investigation of unknown aspects of analyzed samples. The typical chemometric strategy consists of the following steps; (i) collection of data sets; (ii) generation of a mathematical model; (iii) interpretation of model parameters in terms of underlying latent relationships in the data sets and (iv) application of the model to new cases, or often the search for a better model or more appropriate variables (Kaniu, 2011).

2.4.1 Applications of multivariate chemometrics in XRF analysis

Chemometrics continues to be applied mostly for matrix corrections and spectrum processing in XRF spectrometry. Multivariate chemometrics techniques are also applicable to investigating influences of multiple parameters on the results of quantitative XRF analysis. The methods do not rely on predefined peak shape and can deal with considerable peak overlaps in the presence of interfering matrices. Chemometrics reduces the dimensionality of multivariate data and

extracts vital information from the spectra such as disease biomarkers besides being able to extract subtle trace element signatures from noisy spectra which would otherwise be impossible to obtain with classical spectroscopic analysis. To overcome the challenges of the classical EDXRF technique as discussed in Section 2.3, multivariate chemometrics techniques viz. PCA, ANNs and PCR have been applied to analyze XRF spectra through multivariate calibration and pattern recognition.

Multivariate chemometrics necessitated interpretations of chemical data to maximise the information extracted from EDXRF data of fossils (Daniel and Anusuya, 2011). Fossil samples from two localities were distinguished using principal component score values, and coefficient loadings allowed chemical interpretation of the score clusters. PCA analysis of untreated (raw) spectra classified two clusters along PC1 which separated all teeth samples from bone samples. PC1 explained spectral intensity: raw spectra of teeth samples had substantially higher counts than spectra from bone samples because bone was more porous.

In a similar study, Angeyo *et al.* (2012), partial least squares (PLS) used in conjunction with energy dispersive X-ray fluorescence and scattering (EDXRFS) spectrometry was used to rapidly analyze low-Z elements such as Na, Mg and B in lubricating oils which could not be determined conventionally by Fundamental Parameter (FP) methods. The method involved spectrum evaluation and quantification in one single step thus was successfully applied to direct rapid quality assurance (QA) characterization of lubricating oils. Significant differences in PLS analyte quantification capability with respect to SNR in the interval 50 s–1000 s; 200 s gave the best results, indicating the potential of the PLS–EDXRFS method for rapid non-invasive characterization in QA analysis of complex matrix liquids.

X-ray spectrometry allied to chemometrics has been utilized in direct Cr speciation analysis with no prior separation or pre-concentration of Cr (III) and Cr (VI) species (Oliveira *et al.*, 2010). The spectra of several solutions containing different concentrations of Cr (III) and Cr (VI) were analyzed by principal component analysis and partial least squares regression. PC1 and PC2 were able to separate Cr (III) and Cr (VI) samples which were influenced by the Cr signals (Cr K_{α} and Cr K_{β}) and by Rh K_{α} and K_{β} scatter signals. The PLS regression vector plot was used to check independent variables that were most important to the model, in this case, Cr signals and X-ray scattering signals had significant contribution to model development. The results obtained after applying PCA and PLS to the set of working solutions of Cr (III) and Cr (VI) clearly indicated that direct speciation analysis of chromium is possible. Therefore, the direct speciation of chromium without using high-intensity X-radiation (synchrotron radiation) is possible but the study could not be done simultaneously for more than one element.

Chemometric aided energy dispersive X-ray fluorescence has been used for the determination of matrix parameters from scattered radiation (Schramm, 2000). Principal component regression (PCR) calibration model (fundamental parameters combined with matrix correction by PCR) was used for the determination of carbon, hydrogen and oxygen from water and oil-based samples using coherent and incoherent scattering lines. This information was used in combination with fundamental parameters to determine the concentration of Zn in liquid samples. An artificial neural network model has been used in multivariate matrix calibration, compared with cross-validation and partial least-squares methods, which were combined with the fundamental-parameters in X-ray fluorescence analysis (Luo *et al.*, 1997).

CHAPTER 3: THEORETICAL FRAMEWORK

3.1 Introduction

In this chapter, we outline the basic principles of X-ray fluorescence and scattering spectroscopy.

In addition, we also outline the principles of multivariate chemometric techniques i.e. PCA, PCR ANNs and KNN for their utility in analytical modeling and characterization of XRF spectra.

3.2 X-ray fluorescence analysis

The X-ray fluorescence analytical technique is based on principle of radiation interaction with matter in which the interaction can be through the process of;

i) Photoelectric absorption: incident radiation is totally absorbed by the atoms that constitute the sample to generate characteristic X-rays whose energies identify the atom in the sample (West *et al.*, 2007). This interaction process is dominant over other processes in the low energy regime.

The probability of absorption is described by the Beer-Lambert law under the assumption of a monochromatic and parallel beam (Schramm, 2000);

$$I_T = I_o e^{\left(-\frac{\mu m}{A}\right)} \quad (1)$$

where I_T is the intensity after transmission through matter, I_o , the initial intensity, μ the mass attenuation coefficient, m the mass of the sample and A the irradiated area.

The sample atoms are in excited states following irradiation by incident radiation in which vacancies created in the inner shells orbits are filled by higher level electrons during the de-excitation process of atoms in order to regain its stability. In the process, X-ray characteristic radiations for the element are emitted whose energy is the difference in energy between the two shells. The X-ray characteristic radiations are subsequently detected by semi-conductor detector such as Si (Li) detector) in the form of electrical signals for conversion into pulses of the same

heights. These enable for identification of the element qualitative analysis and for the determination of the concentration from the intensity measurements.

ii) Scattering of the photon: applying processes include elastic and inelastic scattering (Rayleigh and Compton, respectively).

The energy of coherently scattered X-ray remains unchanged, while the energy of an incoherently scattered X-ray shifts to lower values. Figure 3.2 shows the attenuation of X-ray beam in target sample in which photoelectric and scattering effects are evident.

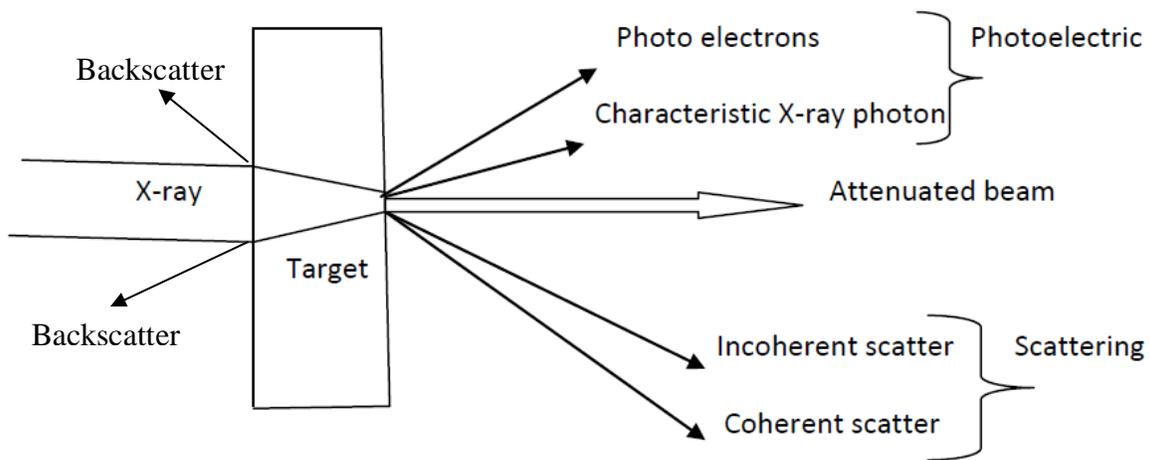


Figure 3.2: Schematic diagram showing the attenuation of X-rays in a target sample.

The energy of an inelastically scattered photon is given by Compton equation: (Compton and Allison, 1967).

$$\Delta E = \frac{E_o \cdot E_c}{m \cdot c^2} (1 - \text{Cos} \theta) \quad (2)$$

where ΔE is the energy difference (keV) between the incident and the scattered radiation, $m \cdot c^2$ is the rest-mass energy of an electron (510.996 keV) and θ is the scattering angle, E_o is the energy of the photon before the scattering takes place.

Scattering contribution of non-detectable X-ray radiation from elements such as C, H and O in biomedical samples can provide an additional information pertaining the overall sample composition: Incoherent (Compton) scattering peak can be used for matrix correction through normalization of the fluorescence line to the incoherent scattering intensity of a secondary target (Feather and Willis, 1976) or for the determination of low-Z elements via multivariate calibration utilizing the scatter peaks (Schramm, 2000)

3.3 Principles of quantitative X-ray fluorescence analysis

Quantitative X-ray fluorescence (XRF) analysis requires the conversion of measured spectral intensities of characteristic X-ray radiations following photoelectric absorption of incident radiation. Several compensation methods to account for the minimization for variations in matrix effects resulting from various sample compositions are applied. In practice, special sample preparation methods are required and more often only one or few elements can be quantitatively determined. Therefore, the compensation methods are less popular for multi-elemental determination than matrix correction methods. In general, radiation intensity of analytes can be calculated from theory of principles of radiation interaction with matter thereby enabling for semi-empirical elemental determination. No other analytical technique allows such a combination of theoretical calculations and experimental measurements. A brief description of these methods is presented in the following section.

3.3.1 Fundamental parameters methods

The fundamental parameters (FP) methods are based on theoretical approach to calculate the intensities of the characteristic radiation of the analytes from a sample of known composition (Debertin and Helmer, 1988). It involves use of fundamental physical parameters of X-ray emission process as well as sample parameters which include; mass absorption coefficients, absorption jump ratios, fluorescence yields and emission line probabilities (Nielson, 1977;

Beckhoff *et al.*, 2006). Applications of these methods involve two steps; calibration based on the measurements of pure element samples to calculate the relative intensities for each analyte, and calculation for comparison of the theoretical results with the experimental data, for determination of the elements.

In general, the fundamental parameter methods are based on the Sherman equation (3), which considers both primary and secondary fluorescence. The tertiary fluorescence and the effects caused by scattered radiation are usually neglected (Sherman, 1955).

$$I_i = \frac{d\Omega}{4\pi \sin \phi_1} Q_i q_i W_i \int_{\lambda_{\min}}^{\lambda_{edge}} \tau_i(\lambda) I_0(\lambda) \frac{1 - e^{[-\chi(\lambda, \lambda_i)\rho t]}}{\chi(\lambda, \lambda_i)} \left\{ 1 + \sum_j W_j S_{ij} \right\} d\lambda \quad (3)$$

where $d\Omega$ is the differential solid angle for detection of the characteristic radiation; i, j are the subscripts for the analyte and matrix element, respectively; Q_i is the instrumental sensitivity of the characteristic radiation of analyte i ; W_i, W_j are weight fractions of the analyte i and matrix element j , respectively; λ_{\min} and λ_{edge} are short-wavelength limit and wavelength of analyte absorption edge, respectively; τ_i is the photoelectric absorption coefficient for analyte i and primary radiation of wavelength λ ; $I_0(\lambda)$ is intensity of the primary radiation, ρ is the density of the sample; t is the sample thickness; q_i is sensitivity of the analyte i . $\chi(\lambda, \lambda_i)$ is total mass-attenuation coefficient of the sample for the incident and fluorescent radiation respectively (Shiraiwa and Fujino, 1966).

Sherman equation enables to calculate radiation intensity of analyte in a sample of known composition. In practice, the goal is to calculate analyte concentration from the actual measurements of intensities. However, the Sherman equation cannot be used to determine analyte concentration directly since analyte concentration is dependent on total mass-attenuation

coefficient $\chi(\lambda, \lambda_i)$ and enhancement term S_{ij} for thick samples. Therefore, the determination of elemental concentrations is performed by iterative calculation which involves comparison of the theoretical results with the experimental data. The fundamental parameter methods have several advantages in that they can be applied in analysis of thick samples, thin films and multi-linear simultaneous determination of elemental composition and thickness is possible. However, application of the FP method is limited by the assumption that, the intensity of the fluorescence radiation is linearly dependent on its concentration (Van Espen and He, 1989; Debertain and Helmer, 1988) and the requirement for homogeneous flat samples (Rousseau, 1984). The challenges presented by “dark matrix” further complicate its practical applicability and uncertainties in mass absorption coefficients and fluorescence yields of the individual elements contribute to overall inaccuracy of the method (Beckhoff *et al.*, 2006). Application of the method is limited to determination of heavy Z elements in sufficient amounts in low Z matrices such as biomedical samples.

3.3.2 Influence coefficients

The influence coefficient method (Markowicz, 2011) has the ability to quantify matrix effects individually based on the matrix effect of element j and that of element k , ... on analyte i . The concentration of an element i can then be determined from its characteristic spectral line intensity, corrected for the concentrations of other sample matrix elements by influence coefficients, using the expression shown in equation 4 (Lachance, 1999).

$$C_i = R_i \left[1 + \sum_j \alpha_{ij} C_j \right] \quad (4)$$

where C_i is the concentration of the analyte, C_j the concentration of other elements in the sample, R_i the linear regression coefficient and α_{ij} the influence coefficient of element j for

analyte *i*. Influence coefficients can take into account both absorption and enhancement effects but require more standards than the fundamental parameters method.

Analysis of biomedical sample analysis by XRF is a challenge due to matrix effects that result into spectral complexity which complicates spectral evaluation and deconvolution for the determination of net fluorescence intensities. Other complimentary methods used in biomedical sample analysis, for instance TXRF require sample digestion in order to isolate the analytes of interest in solution form which leads to loss of spectral information for volatile elements. However, EDXRF spectral scatter profiles can be utilized to derive information about the overall matrix composition including the determination of low-Z elements, dark matrix and mostly chemical property information about the sample. Biomedical samples are organic in origin and have high proportions of light elements which substantially contribute to scattering and as such their proportions can semi-quantitatively be determined from measurements of coherent and incoherent X-ray scatter peaks (Nielson, 1977).

The proposed method (chemometric aided EDXRFS) for this study, for the analysis of biomedical samples (human body tissues) uses the fluorescence radiation and scatter peaks of the EDXRF spectra to develop multivariate calibration strategies. This enables determination of qualitative and quantitative information in low-Z matrix samples for trace elements. Several multivariate chemometric techniques such as PCA, PCR, ANN and PR are used when applied to resolve overlapping spectral lines and detection of low SNR analytes in the complex matrices.

This study demonstrates the potential of chemometric-aided EDXRFS spectroscopy for determination of concentration, speciation and alteration of trace elements (Cu, Fe, Zn, Mn and Se) for subsequent diagnostic of cancer in human tissue samples.

3.4 Chemometric analytical spectroscopy

Chemometric methods such as PCA, ANNs, PCR and KNN may be used for; (i) analysis of spectral data to extract vital information inherent in the data and (ii) to design of experiments for optimal acquisition of information about the stated problems (Everitt, 1981; Lennart *et al.*, 2014). Chemometric analysis is mainly used for:

- (i) Data description (exploratory data analysis).
- (ii) Classification and discrimination of samples.
- (iii) Correlation and regression analysis.

3.4.1 Pattern recognition

Pattern recognition is a tool used to classify data (patterns) based on either supervised or unsupervised approaches; unsupervised pattern recognition requires no *a priori information* of classification of the objects for analyses of both linear and non-linear spectral data. The patterns are achieved through discrimination among the different groups by division of hyperspace in as many regions as the number of classes so that a sample falls in the region of space corresponding to its category (Massart *et al.*, 1988). The classification is either based on discrimination between classes or similarity within a class where each sample is always assigned to one and only one class (Vandeginste *et al.*, 1998). Among such methods (unsupervised) are cluster analysis and principle component analysis (Liqiang and Luo, 2006). As a powerful tool for classification and identification of sample classes, PR has the potential for speciation analysis of trace elements in samples, for instance using supervised pattern recognition technique, K-nearest neighbor (Sharaf *et al.*, 1986).

3.4.1.1 Principal Components Analysis (PCA)

PCA is the modern mathematical and statistical method of reducing the dimensions of multivariate data. It is a well established multivariate chemometric tool for exploratory data analysis. The method summarizes the information of the initial data set into a new set, which has a smaller dimension than the former (Mathias, 1999; Jolliffe, 1982) as shown in Figure 3.4.1.

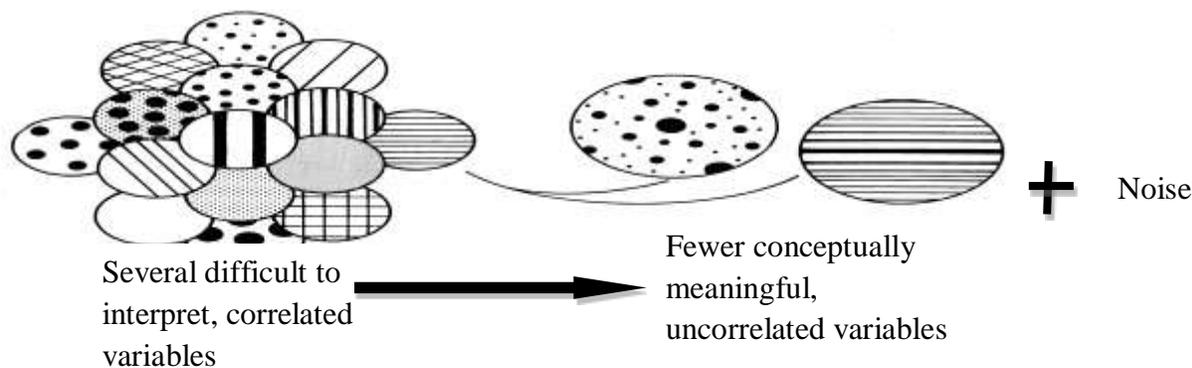


Figure 3.4.1: General purpose of principal component analysis.

The new data set contains only the relevant statistical information of the new system of axes that represent the samples, making it possible to visualize the multivariate characteristics of the data in a few dimensions. The redundant information is eliminated through projection of the data onto a lower dimensional subspace thus maximizing variance, which makes the data treatment and interpretation easier (Zhai *et al.*, 2003; Zomer *et al.*, 2003). It also enhances modeling of data which includes selection of main variables and classification of samples where cluster analysis extends this process to examine any natural groupings existing in the samples (Bueno *et al.*, 2005).

The first PCA score (PC1) is the linear latent variable with the maximum possible variance followed by subsequent PCs in decreasing variance. During PCA analysis, the original data matrix X is mean-centered and decomposed to a score matrix T and loading matrix P (Equation

5), by consecutive orthogonal subtraction of the largest variation with separation of residual in the matrix E (Brereton, 2002).

$$X = T \cdot P^T + E \quad (5)$$

where T represents the scores matrix, calculated as $(n \times A)$ where n is the number of samples and A the number of variables; this expresses the relation among the samples and shows the sample coordinates in the new system of axes with T describing the position of the samples in the new coordinate system (Virendra *et al.*, 2011).

3.4.1.2 K-nearest neighbor (KNN) technique

K-nearest neighbor method is one among other powerful distance oriented classification techniques (Mahalanobis, Minkowski, Hamming and square distances) in which an unknown sample in the validation set is classified according to the majority of its K nearest neighbors in the training set (Kurt and Peter, 2008). The Euclidean distance between two samples i and j is first computed from the sample to every other point in the data set based on a Pythagorean formula:

$$d_{i,j} = \sqrt{\sum_{l=1}^k (x_{i,l} - x_{j,l})^2} \quad (6)$$

where k is the number of dimensions, $x_{i,l}$, and $x_{j,l}$, represent the values of l coordinates for a given sample pair i and j (Thanh *et al.*, 2004).

These distances are arranged from smallest to largest. The K neighbors of an unknown sample with the lowest Euclidean distances are the training samples. The “prediction class” of a sample is the class having the largest number of objects among the K neighbors of the sample. If the

predicted class and the actual class of the sample match, the prediction is considered a success as illustrated in Figure 3.4.2.

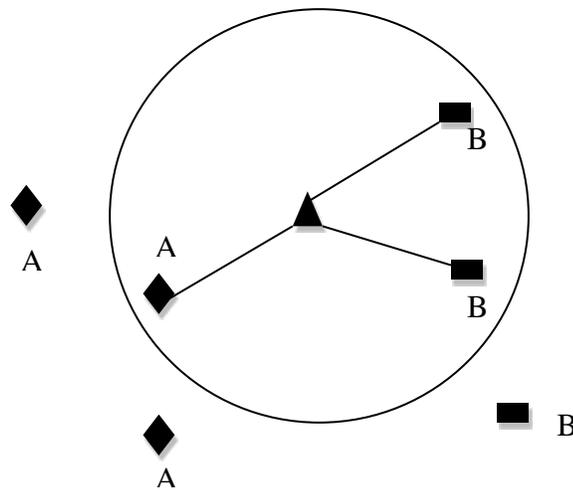


Figure 3.4.2: Schematic diagram of K-nearest neighbor (KNN).

The unknown data point is classified as belonging to the B class when $K = 3$ is the adopted rule. The best choice of K depends on the specific data set, a larger K is considered to reduce the effect of noise on the classification. The algorithm of implementation of the K-nearest neighbor classification used is described by the following sequence of steps:

- 1) Collection of all the samples; training samples and the “unknown” samples.
- 2) Selection of suitable value for K neighbors based on the number of samples.
- 3) Determination of suitable distance metric such as the Euclidean distance.
- 4) Location of K nearest neighbors of the sample based on the selected metric.
- 5) Location of “plurality class” of the nearest neighbors by voting based on the class labels of the nearest neighbors.
- 6) Assigning the sample to be classified to the “plurality class”.

These distances are then used to determine the closest neighbors (i.e., smallest distance) to the test samples; nearest K neighbors are selected and the frequency of each is determined. The class to which the majority of training samples belong is assigned to the test sample. If a tie should

occur, the class with the closest neighbors is selected. K Nearest-Neighbor technique has the advantage that it is easy to implement and provide remarkably accurate results if the features are chosen carefully in the computation of the Euclidian distance. This enables classification and identification of samples and species which gives qualitative information about the samples being investigated.

3.4.2 Artificial Neural Networks (ANN)

A neural network consists of a large number of processing units (neurons) connected together to simulate the biological structure of the brain and nervous system (Liu *et al.*, 1993). Normally, the dendrites receive electrical signals from other neurons into the cell body of the neuron; the soma (or the cell body) sums the incoming signals and fires a signal depending on the threshold value; the axon then transmits the resulting signal from soma to other neurons as shown in Figure 3.4.3.

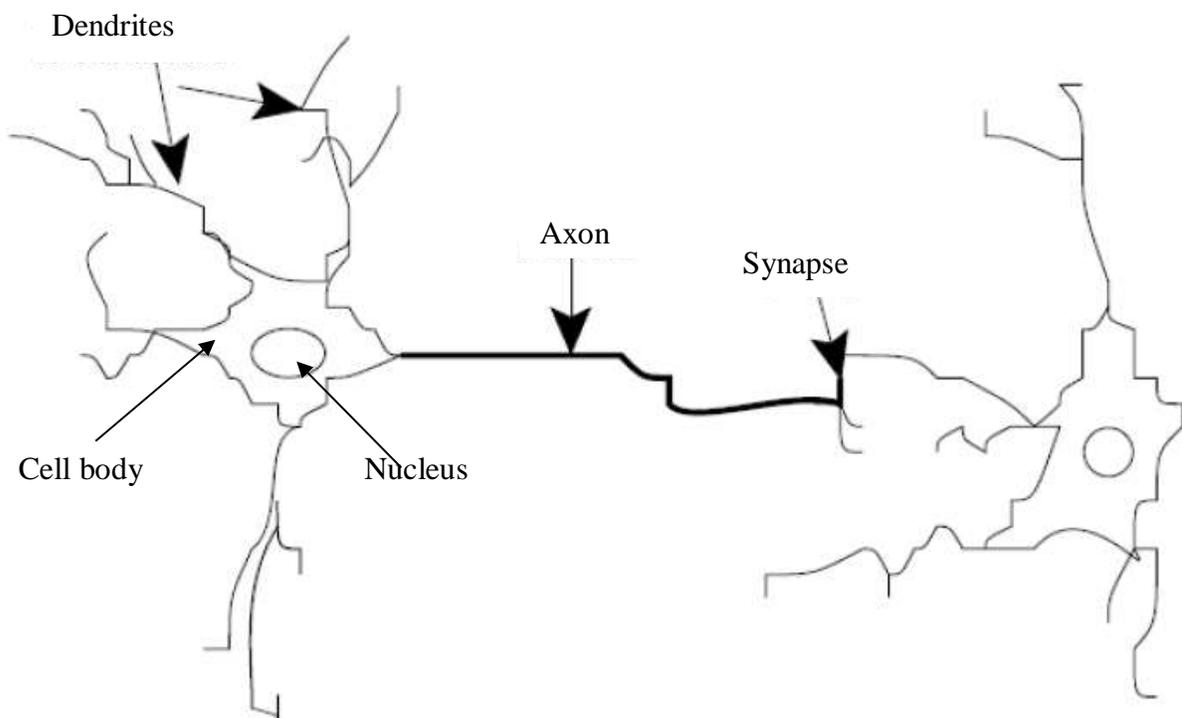


Figure 3.4.3: Natural neural network (After: Fausett, 1994)

From a computational point of view, the simplest representation of an artificial neural network is that of a black-box which receives multiple inputs and produces multiple outputs (Zupan and Gasteiger, 1999) as shown in Figure 3.4.4. In order to ascertain the uniqueness of this technique, it is necessary to open the black-box and look inside, introducing a greater level of detail in our description of how ANNs work.

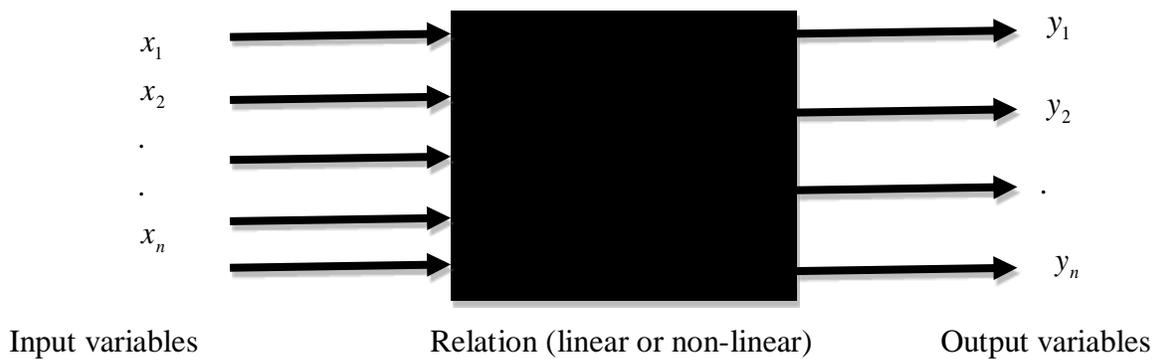


Figure 3.4.4: Artificial neural network as a black-box.

Artificial neural networks (ANNs) simulate human cognitive processes and thus are suitable for processing noisy and incomplete data. These characteristics cannot be expressed through a classical well-defined algorithm; rather, they are based on experience. ANNs is aimed to build a multivariate regression which establishes a relationship between two matrices X and Y . The neural network consists of three layers of nodes: an input layer, a hidden layer and an output layer as shown in schematic diagram in Figure 3.4.5 in the case of XRF analysis.

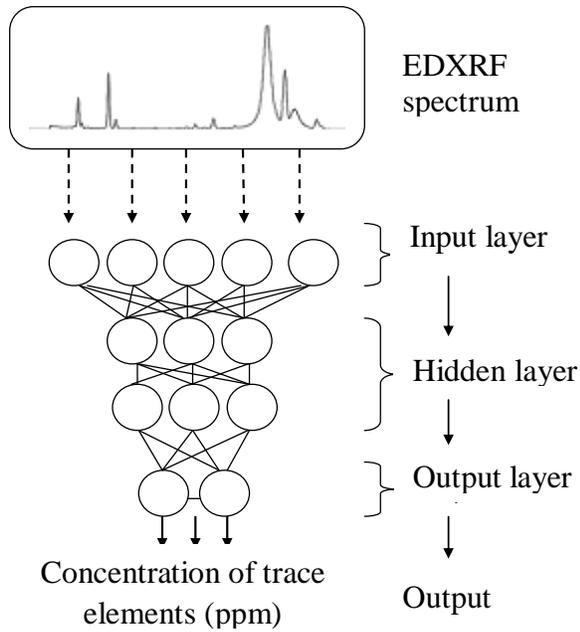


Figure 3.4.5: Internal organization of a BP-ANN calibration utilizing EDXRF spectra (After: Jose *et al.*, 2009).

The training functions in back propagation (BP) are used to train the feed forward networks for function approximation (non-linear regression). The training process consists of four stages namely; (i) collection of training data sets, (ii) creation of new network, (iii) training of the network, and (iv) simulating the network response to new unknown inputs. The number of neurons in the hidden layer is chosen based on the least value of minimum square error (MSE) and network training performance for each network.

For instance, XRF intensities of the elements in the training data set are input into the first layer in the network and the input nodes have the linear outputs X_i . The sum of X_i multiplied by weight W_{ij} , with a threshold parameter θ_{ij} , is used as the input (Net_j) of the node j in the hidden layer (Luo *et al.*, 1997):

$$Net_j = \sum (x_i W_{ij}) + \theta_{ij} \quad (7)$$

A sigmoid function $f(u)$, is chosen as the transfer function:

$$f(u) = 1/(1 + e^{-u}) \quad (8)$$

The output (OUT_j) of the node j is produced after the transformation by the sigmoid function:

$$OUT_j = f(Net_j) \quad (9)$$

After the transformation by the sigmoid function again,

$$OUT_k = f(Net_k) \quad (10)$$

The output (OUT_k) of the node k in the last layer is obtained and then compared neuron by neuron with the corresponding component concentrations (Y_k) in the training data set and the error.

$$Y_k = F(OUT_k) + \varepsilon \quad (11)$$

where F is the network training function and ε , the error of calibration.

Artificial neural networks (ANNs) are therefore non-linear computational tools suitable for practical application due to their flexibility and adaptability (Marini *et al.*, 2008).

3.4.3 Principal Component Regression (PCR)

PCR is a two-step multivariate calibration technique which considers the more descriptive principal components (PC) obtained by Principal Component Analysis (PCA) as independent variables instead of adopting original variables (Pires *et al.*, 2008). The first step involves principal component analysis of the data matrix \mathbf{X} where the measured variables (intensities at different channel numbers) are converted into new ones (scores on latent variables). This is then followed by a multiple linear regression step, MLR, between the new variables (scores) obtained from the PCA and the characteristic \mathbf{Y} (concentration) to be modeled. The principal component scores obtained from the PCA are taken as the independent variable in the multiple linear regressions which resolves collinearity problems by providing stable solutions plot for

interpretation of scores and loadings. PCR as a bilinear calibration method performs the regression of Y on selected principal components of X (Martens and Nas, 1991).

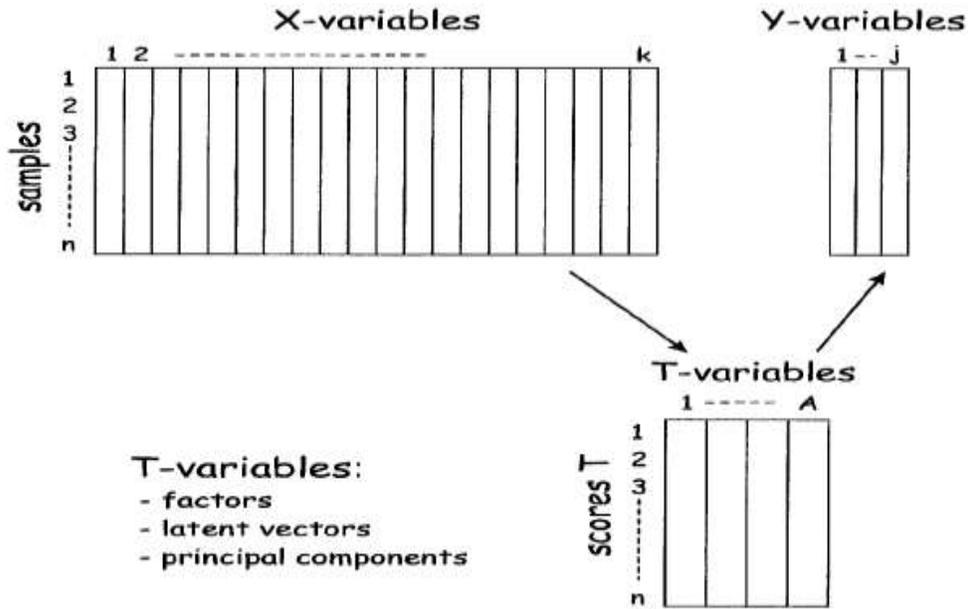


Figure 3.4.3.1: Schematic diagram of Principal Component Regression.

PCR on the centered X and Y data is thus formulated through calibration where:

$$\hat{T} = X \hat{V} \quad (12)$$

where \hat{T} is the score obtained from the linear combinations of X ; \hat{V} , loadings of the X - variable.

$$X = \hat{T} P + E \quad (13)$$

During decomposition of the X -variables in PCR, the loadings P are equal to \hat{V} and E , the residuals

$$Y = \hat{T} Q + F \quad (14)$$

The loadings Q is then compared to regression coefficient (β) in MLR; F , residuals

Finally, the prediction of concentration in the unknown sample is given by

$$\hat{y}_i = \hat{t}_i \hat{Q} \quad (15)$$

where \hat{y}_i , the predicted concentration of sample i and

$$\hat{t} = x_i \hat{V} \quad (16)$$

With x_i as the measured values (e.g. absorption) of prediction sample i .

CHAPTER 4: MATERIALS AND METHODS

4.1 Introduction

In this chapter, we present the procedures used in the preparation of standard samples for calibration training of spectral data for subsequent sample spectral data analysis for trace elemental content determination. The utility of chemometric methods using X-ray fluorescence spectra for quantification of trace elements has been investigated using calibration models developed from prepared standard simulate samples. X-ray fluorescence spectral data were obtained after samples were irradiated for 50 s using EDXRF spectrometer. These methods were validated for accuracy of measurements by analysis of Oyster tissue as a certified reference material (CRM). All chemicals used in the sample preparation in this study were of analytical grade.

4.2 EDXRF Spectrometric Instrumentation

In this study, the energy dispersive X-ray fluorescence spectrometer Shimadzu series EDX-800HS, model CE (212-23701-36), was used for spectral data measurements. The spectrometer consists of the following;

- a) X-ray generator unit consists of an air cooled of X-ray tube with Rh target operated at 5-50 kV, tube current 0.001-1mA; primary X-ray filters, automatically switched with an exposure area of 10 mm in diameter – four settings 1, 3, 5 and 10 diameters are available for automatic switching.
- b) Semi-conductor Si (Li) detector 10 mm² mounted at 45° take-off angle geometry with respect to the sample. It is cooled by liquid nitrogen only during measurement with approximate consumption of 1 litre per day. The detector is biased with a high voltage supply at 50 kV and is connected to a pre-amplifier. Detector measured resolution of about 140 eV for Mn K_{α} at 5.9 KeV.

- c) Large sample chamber which automatically opens and closes and can accommodate samples up to 300 mm wide and 150 mm high. It has optional sample turrets for continuous automatic sample measurements.
- d) Data processing unit which consists of the following modules: IBM PC/AT compatible, compiler HDD-20GB, Operating system-windows XP and pre-installed factory produced EDX software for quantitative measurements and incorporated algorithm for matrix correction FP method analysis and calibration curve method of analyses.

Figure 4.2.1 shows a typical schematic EDXRF (Shimadzu EDX-800HS) spectrometer operated at applied voltage of 50 kV in the X-ray tube, current of 0.001-1mA. The spectral data measurements were obtained in air at dead time less than 25% which was automatic as current and voltage were adjusted during acquisition of the spectrum.

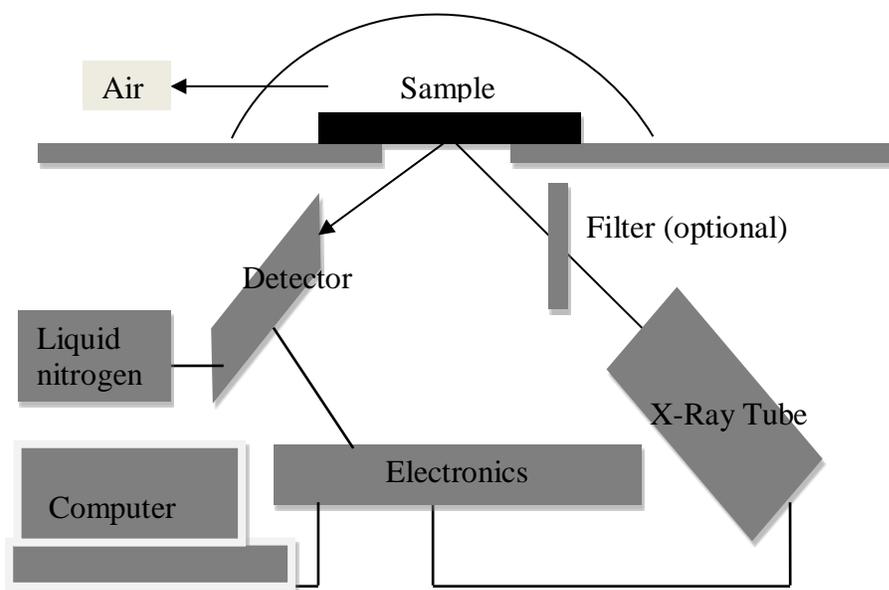


Figure 4.2.1: Schematic presentation of EDXRF instrumentation.

The elements of interest; Mn, Fe, Cu, Zn and Se were analyzed following sample irradiation with Rh-tube excitation source. The characteristic fluorescence radiations of the elements were

detected with Si (Li) detector. Spectrum deconvolution and peak identification was done using AXIL (Beckhoff *et al.*, 2006; Van *et al.*, 1989).

4.3 Sampling

Three sets of samples were considered in this study; simulates, tissue cultures and selected domestic dog cancerous tissue (prostate/mammary) samples for determination of trace elements (Cu, Fe, Zn, Mn and Se) together with their (Fe, Cu and Mn) speciation. During the entire process of sampling, standard protocols pertaining to tissue culturing and histopathological sampling were adhered to in order to minimize contaminations and any possible infection through sample handling (Appendix 1 and 2).

4.3.1 Simulate samples

Tissue simulate block samples (dimensions 3 cm × 2 cm × 1 cm) and approximately weighing 5 g, were prepared in 2 sets from highly purified paraffin wax as the base matrix, appropriate amounts of ethanol, and acetone spiked with multi-elemental stock-solution of elements of interest (Mn, Fe, Zn, Cu and Se) within the range in which they occur in body tissues; Cu (5-33 ppm), Zn (5.0-195 ppm), Fe (12-120 ppm), Mn (1.0-9.0 ppm) and Se (1-6 ppm), (Banas *et al.*, 2001; Kwiatek *et al.*, 2004). The first set of simulate samples comprises of 10 samples spiked with Fe^{3+} , Cu^{2+} , Zn^{2+} , Se^{4+} and Mn^{4+} while the second set comprised of 10 samples spiked with Fe^{2+} , Cu^{+} , Zn^{2+} , Se^{4+} and Mn^{2+} all with similar concentrations except differentiated by speciation and were labeled S1H-S10H and S1L-S10L, respectively (Table 4.4.3a and 4.4.3b). The 2 sets of samples were used to develop quantitative (for Mn, Fe, Cu, Zn and Se) and speciation (for Fe, Cu and Mn) models for elemental analysis of unknown samples.

4.3.2 Cultured tissue samples

Two sets of cultured tissue samples each of volume 5 µl in T-culture flasks; 10 cancerous and 10 non-cancerous were labeled SV1-SV10 and SL1-SL10, respectively. The samples were obtained for the study after culturing process in the incubator (model no. IF-3B) at 37 °C in humidified incubator atmosphere of 97 % and 5 % CO₂. This was done at Centre for Clinical Research Immunology Laboratory, Kenya Medical Research Institute (KEMRI). The bio-safety cabinet (for culturing processes) was sterilized using 70% ethanol to minimize contamination prior to preparation of the sterile complete media (growth media) and incomplete media (wash media). For the sterile growth media, 500 ml Roswell Park Memorial Institute (RPMI)-1640 culture medium (SIGMA-Aldrich) was supplemented with 10% foetal bovine serum (SIGMA-Aldrich), 5 ml L-glutamine (SIGMA Aldrich), 5 ml penicillin–streptomycin solution (SIGMA-Aldrich) as an anti-biotic and 10 ml HEPES (SIGMA-Aldrich). The wash media constituted of 500 ml RPMI-1640 medium (SIGMA-Aldrich) with no supplements.

4.3.3 Domestic dog tissue samples

Eighteen domestic dogs that had prostate cancer or mammary tumors were sampled for this study. Prior to sampling, the consent forms (Appendix 3) were signed by the owners of the dogs in order for them to allow their dogs to be used in the study. The animals were aged between 1.5 years to 15 years and diagnosed with cancer at the Small Animal Clinic at the Faculty of Veterinary Medicine, University of Nairobi, in 2013 following in-patient admission and were later euthanized.

Approximately 50 g of the tissue samples were surgically removed from prostate / mammary organs labeled as SP for preparation prior to analysis. The selected tissue samples were fixed in

formalin for preservation of cellular details (Su *et al.*, 2004) and later freeze-dried at -80°C for 48 hours (time optimized for removal of water and to stop cellular activities) in petri dishes.

4.4 Sample preparation

4.4.1 Simulate samples

Two sets of tissue simulates were prepared using highly purified paraffin wax as a ‘base matrix’ spiked with a multi-element stock solutions of selected elements of interest; Mn, Fe, Zn, Cu and Se. This was done at different concentrations as shown in Table 4.4.3a and 4.4.3b from carefully selected analar grade compounds as shown in Table 4.4.1.

Table 4.4.1: Chemical compounds used in the preparation of simulate samples

Source	Chemical Formula	Molecular mass (g/mol)	Analyte of interest
Cuprous chloride	CuCl	99.00	Cu^+
Copper(II) nitrate trihydrate	$\text{Cu}(\text{NO}_3)_2 \cdot 3\text{H}_2\text{O}$	241.6	Cu^{2+}
Zinc nitrate	$\text{ZnNO}_3 \cdot 6\text{H}_2\text{O}$	297.48	Zn^{2+}
Manganese (II) chloride tetra-hydrate	$\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$	197.91	Mn^{2+}
Manganese (IV) chloride	MnCl_4	197.92	Mn^{4+}
Ferrous oxalate Anhydrous	FeC_2O_4	143.91	Fe^{2+}
Iron (III) nitrate	$\text{Fe}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}$	403.99	Fe^{3+}
Selenium (IV) oxide	SeO_2	110.96	Se^{4+}

The embedding medium, paraffin wax was chosen as the base matrix as it mimics the typical soft tissue properties that must/should have a matrix that matches the tissue samples (both chemically and physically) and be homogeneous in nature (White, 1975).

Stock solutions of the elements of interest were prepared after weighing the appropriate amount of the salts determined from the following simple expression;

$$m = \frac{C_i \times V \times rmm}{1000 \times ram} \quad (18)$$

where C_i is the concentration of the stock solution prepared in $\mu\text{g/ml}$, V the volume of the stock solution pipette in ml, rmm is the relative molecular mass of the salt in grams (g) and ram is the relative atomic mass of the trace element of interest in grams (g). For each element of interest, the corresponding compound salt of determined mass m in grams (g), was dissolved in 15 ml of ethanol to obtain a stock solution of pre-determined concentrations as shown in Table 4.4.2. Ethanol was used as a solvent as it easily dissolved all the compounds with elements of interest and it is miscible with acetone.

Table 4.4.2: Elemental concentration of the stock solution

Analyte	Calculated mass m of the compound salt (mg)	15 ml Stock solution (ppm)
Zn^{2+}	101.3	1500
Se^{4+}	82	259
Cu^{2+}	46	800
Cu^+	18.5	777
Fe^{2+}	58	1200
Fe^{3+}	114	1050
Mn^{2+}	21	405
Mn^{4+}	16	400

Appropriate volumes of stock solutions corresponding to the concentrations selected for preparation of simulate samples as shown in Table 4.4.3a and 4.4.3b were determined using the following simple expression;

$$C_2 \times V_2 = C_1 \times V_1 \quad (19)$$

where C_1 and C_2 are the concentrations of stock solution and simulate sample respectively, V_1 and V_2 are the volumes of the stock solution used and the volume used for simulate sample preparation respectively.

The following experimental design (Tables 4.4.3a and 4.4.3b) was used for spiking the trace elements at various concentrations in simulate samples.

Table 4.4.3a: Calibration set design for simulate samples for Cu^{2+} , Zn^{2+} , Fe^{3+} , Mn^{4+} and Se^{4+}

Analyte	Simulate sample concentration (ppm)									
	S1H ^p	S2H	S3H	S4H	S5H	S6H ^p	S7H	S8H	S9H	S10H
Cu^{2+}	25.0	12.5	7.5	15.0	5.0	9.0	30.0	32.5	6.0	12.5
Zn^{2+}	40.0	10.0	5.0	35.0	10.0	17.5	67.5	37.5	25.0	195.0
Fe^{3+}	46.0	11.5	42.5	37.5	11.5	22.5	120.0	20.0	27.5	30.0
Mn^{4+}	2.5	1.5	3.8	3.0	1.3	9.0	4.0	6.5	3.5	0.8
Se^{4+}	2.0	0.5	2.5	6.0	1.0	1.5	5.5	3.0	2.5	1.5

p-sample used for prediction

Table 4.4.3b: Calibration set design for simulate samples for Cu^+ , Zn^{2+} , Fe^{2+} , Mn^{2+} and Se^{4+}

Analyte	Sample concentration (ppm)									
	S1L	S2L	S3L ^p	S4L	S5L	S6L	S7L ^p	S8L	S9L ^p	S10L
Cu^+	25.0	12.5	7.5	15.0	5.0	9.0	30.0	32.5	6.0	12.5
Zn^{2+}	40.0	10.0	5.0	35.0	10.0	17.5	67.5	37.5	25.0	195.0
Fe^{2+}	46.0	11.5	42.5	37.5	11.5	22.5	120.0	20.0	27.5	30.0
Mn^{2+}	2.5	1.5	3.8	3.0	1.3	9.0	4.0	6.5	3.5	0.8
Se^{4+}	2.0	0.5	2.5	6.0	1.0	1.5	5.5	3.0	2.5	1.5

p- sample used for prediction

The multi-element mixtures (S1H-S10H, and S1L-S10L) were thoroughly homogenized by stirring and topped to 1.8 ml vial using ethanol resulting to a dilution factor of 2 of the initial concentrations. The vials were tightly sealed to reduce evaporation of ethanol.

For preparation of solid simulate samples, 500 μ l of the multi-element mixtures in solutions were pipetted on a steel mould and 1ml of acetone; a universal secondary fixative with low toxicity and freely miscible with water and organic solvents was added together with 0.5 g of silicone gel to simulate the organ cells in body tissues. The mixtures were thoroughly stirred for

homogeneity prior to adding about 2 ml of molten paraffin wax (paraplast extra) heated to 56°C in a tissue cassette (serial No. 47151038) enabling acetone to evaporate thereby leaving the elements of interest embedded in paraffin wax matrix. The mixture was further stirred to increase homogeneity for about 3-5 minutes and cooled on cold plate of the tissue cassette.

The prepared solid simulate samples were dried at room temperature until solidified. The solidified tissue sample blocks were then kept in sealed plastic containers and labeled, from which $5\ \mu\text{m}$ thick section samples of radius 5 mm were cut with the aid of the microtome equipment and mounted on a $2\ \mu\text{m}$ mylar foil for EDXRF analysis.

4.4.2 Incubated cell cultures

The two sets of cancerous and non-cancerous samples used in this study were prepared in the sterilized bio-safety cabinet following the culturing protocol of sampling and sample preparations available at KEMRI (Appendix 1).

Approximately 5 ml of Peripheral Blood Mononuclear cells (PBMCs), specifically lymphocytes, were obtained from sediments of centrifuged fresh peripheral mono-nuclear blood mixed with lymphoprep and were spanned at a relative centrifugal force 1000 for 15 minutes in a centrifuge. Ten samples of Vero P_{33} cells, 0.5 ml each of cancerous cell lines and healthy human lymphocytes cells enriched in 3 ml growth media were separately cultured in $25\ \text{cm}^2$ culture flasks. For maximum cell/medium contact, the culture flasks were horizontally placed in the incubator at 37°C , in 97 % humidified atmosphere of 5 % CO_2 (Podgorczyk *et al.*, 2009).

The lymphocyte cultures in suspension were microscopically checked on daily basis to confirm for any contaminations which may lead to confluence and periodically supplied with 1 ml fresh sterilized growth medium at an interval of 3-4 days. The Vero P_{33} cancer cells adherent to the

surface were supplied with sterile growth media by pouring off the media and detaching them from the surface of the primary culture vessel through trypsinization using trypsin /EDTA before adding 3 ml fresh growth media. The essence of adding fresh sterile growth media was to ensure optimum growth conditions for both cancerous and healthy cells.

Substantial amounts of PBMCs cell suspension about 3-5 ml were centrifuged to obtain about 100 μ l pellet while Vero P_{33} cells were trypsinized (removed from the surface using trypsin/EDTA) and incubated for 5-10 minutes to detach the cells from the surface of the culture flask before centrifuged to obtain about 100 μ l pellet. The Vero P_{33} and the PBMC pellets were then labeled and stored at -80°C to minimize morphological damage to the samples prior to analysis.

Subsequent sampling was done at an interval of 24 hours for both cancerous and healthy cells in order to monitor the various stages of cancer development. Hence, SV1 and SL1 are samples cultured in 24 hours while SV10 and SL10 are samples cultured in 240 hours. Two sets (representing cancerous and non-cancerous) each consisting of 10 samples of 10 μ l of each sample was pipetted onto a 2 μ m mylar foil and dried at room temperature in the bio-safety cabinet to form a thin sample substrate for EDXRF analysis. Precaution was taken to minimize contamination during all sample preparation.

4.4.3 Domestic dog tissues

The freeze dried samples weighing approximately 30 g were thawed at room temperature prior to sample preparation. The tissues were then embedded in highly purified paraffin wax (Su *et al.*, 2004) to provide the supportive and protective aid during tissue sectioning. 5 μ m thick tissue sample sections were cut with the aid of a microtome and mounted on a 2 μ m mylar foil for

analysis with another thin section preserved for histopathological analysis. Precaution was taken to minimize sample contamination (Appendix 2).

4.5 Energy dispersive X-ray fluorescence (EDXRF) analysis

All samples analyzed in this study, were irradiated for 50 s under same operating conditions (applied voltage of 50 kV in the X-ray tube and current of 0.001-1 mA). The spectral data were recorded in the range 0 – 40 keV, with a spectral resolution of 140 eV for $Mn K_{\alpha}$ at 5.9 keV and data acquisition dead time of less than 25%.

Spectral background were corrected for by subtraction of the spectra of the sterile growth media and blank purified paraffin wax block with silicone gel mounted on a 2 μ m mylar film. To evaluate the accuracy of energy dispersive X-ray fluorescence and scattering (EDXRFS) spectroscopy procedures, standard reference material (NIST Oyster tissue1566b) were similarly prepared and analyzed in duplicate. The spectral data were converted and stored in CSV format using the EDX software ready for data analysis.

4.6 Chemometric analysis

Multivariate chemometric methods were used to analyze the EDXRFS spectral data all obtained under the same irradiation conditions, namely; 50 s irradiation time, 5-50 kV and 1-1000 μ A tube current. The multivariate chemometric methods do not rely on predefined peak shape thus peak overlaps and matrix effects are resolved through spectral pre-processing to remove spectral noises. Matrices of independent variables were constructed in such a way that the columns referred to the spectral intensities corresponding to 2047 channels of spectral data and the rows correspond to each sample. Based on the fact that full spectral data require a large memory, spectral data compression was done by selecting the appropriate spectral regions corresponding to each selected elements of interest as shown in Table 4.6.1.

Table 4.6.1: Selected spectral regions of interest (feature selection) for Mn, Fe, Cu, Zn and Se

Element		Energy (KeV)	Channel numbers
Mn	K_{α}	5.89	287-296
	K_{β}	6.49	327-335
Fe	K_{α}	6.40	318-326
	K_{β}	7.06	348-356
Cu	K_{α}	8.04	398-406
	K_{β}	8.90	442-448
Zn	K_{α}	8.63	431-438
	K_{β}	9.57	475-482
Se	K_{α}	11.21	561-568
	K_{β}	12.49	618-627

Additionally, K_{α} Compton scatter region was selected with channel numbers (950-1000) intensities corresponding to energy (19.15 keV) of Rh K_{α} . There were no peak shifts throughout the measurement periods, an indication of the stable electronic of the spectrometer.

4.6.1 Principal component analysis (PCA)

In this study, principal component analysis (PCA) was carried out using the Unscrambler software, version 9.2, from Camo[®]. This was aimed to reduce the spectral data dimensions for sample exploratory analysis. Prior to PCA analysis, the selected spectral set in Table 4.6.1 above together with the Compton scatter intensities were pre-processed by mean centering and smoothed by 5 point moving averages. PCA results were then validated by full cross validation method. The scores indicate hidden profile in the individual spectra which contain quantitative information such as concentration of trace elements while the loadings reveal the variables such as trace elements responsible for the patterns.

4.6.2 Artificial neural networks (ANN)

The ANNs analysis of spectral data was done using Matlab version 2010 software which facilitated development of training, testing and validation of the analytical method. In general, the training process involved four stages; (i) assembling training data, (ii) creating of the network, (iii) training the network, and (iv) simulating the network response to new inputs (Kaniu *et al.*, 2012). Back propagation (BP-ANN) training function with Levenberg-Marquardt (LM) algorithm was used to train the feed forward networks for function approximation (non-linear regression) as it results to lower mean square errors when compared to other algorithms for its convergence at fast rate thus was considered appropriate (Howard *et al.*, 2009).

The ANNs analysis of spectral data for trace elemental concentration determination uses special data of preferred simulate tissue samples to develop a calibration model. The spectral data sets were treated with a two-layer (input and output with a hidden layer) neural network in-built with back propagation training function. The number of neurons in the hidden layer was chosen based on the mean square error (MSE) and the training performance of each network where training was repeated several times until the best results of low MSE and satisfactory performance were achieved. In general, as a tool for quantification, ANN performance largely depended on the chosen parameters such as the transfer function employed, the number of layers and neurons used besides the linearity/non-linearity of the spectral data. The performance of a trained neural network was measured by the errors in training, validation and test sets. Training process stopped automatically when the validation error increased beyond optimal levels thereby avoiding over-fitting of the model. After developing the ANN model, results were tested for the output of the neural network with the input of the measured data and results compared. A correlation coefficient, R-value, is then used to explain the correlation between the targets and the outputs of the network. R-value is a quantitative measure of how well the variation in the output is

explained by the targets. If its numerical value is close to 1, then there is good correlation between targets and outputs.

4.6.3 Principal Component Regression (PCR)

PCR analysis was performed using Matlab version 2010 software where the spectral data utilized in development of ANN model were used to also develop the PCR model for comparison.

In order to explain the variation of the X-variables with the dependent Y-variable(s), the regression of Y (concentration) was based on selected principal components of PCA analysis (PC1 and PC2) instead of the original X-variables. The 2 PCs were then regressed against the given concentrations to obtain the regression coefficients which were used on the test data to predict the concentrations of the sample constituents.

4.6.4 K-nearest neighbors (KNN) technique

This chemometric tool enabled pattern recognition in the spectral data based on a nearest-neighbor approach where the classification method describes the similarity of the new sample by all the other known spectral outputs (Brereton, 2002).

The spectral input data, K_α and K_β mean intensities of Cu, Fe and Mn were selected as indicated in Table 4.3 from the entire spectrum of each sample in the training sets (simulate samples), cultured tissue sets and dog tissue samples for the application. Qualitative classification (speciation of trace Cu, Fe, and Mn) were based on discrimination and/or similarities between the sample (cancerous and non-cancerous tissue) and the training set of all the standards (simulate samples).

Fig. 4.6.1 shows the steps for generating an analysis model for determination of speciation of trace Fe, Cu and Mn in which the unknown sample and the training set in KNN technique were

exposed to classification rules (nearest neighbors to the unknown sample) for identification. The unknown sample to be classified was assigned the same class as the class of the set of neighboring training samples so found based on the Euclidean distances of the nearest training set samples to the unknown sample.

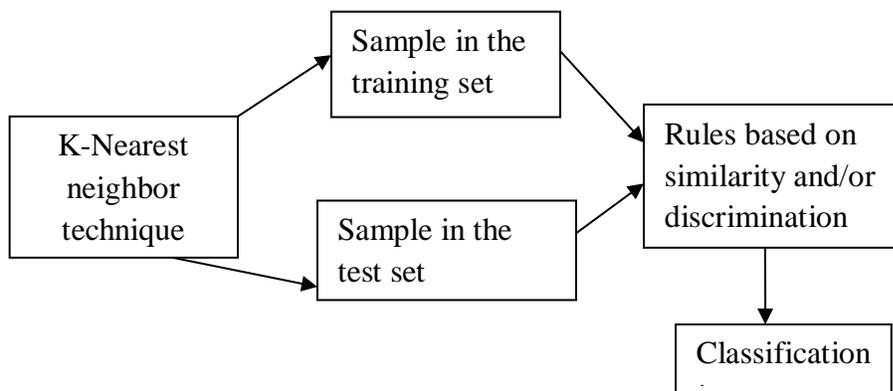


Figure 4.6.1: Flow chart for KNN technique (After: Sartoros and Eric, 1997).

The schematic diagram, Fig. 4.6.2 shows the steps involved in the development of our cancer diagnostic model based on quantitative and/or qualitative profiles of Mn, Cu, Fe, Zn and Se, as hypothesized cancer biomarkers. The EDXRFS spectra were pre-processed by selecting appropriate region of interest of the trace elements prior to exploratory analysis by PCA in which various classes in spectral data are discernible. The simulate samples were used to develop and validate the diagnostic model comprising of quantitative (PCR or ANN) and qualitative (KNN) for determination of concentration and speciation of trace elements. The model was then applied to unknown samples (cultured and selected domestic dog tissue samples) for diagnosis of cancer based on the concentration and/or speciation of trace elemental signatures and their multivariate alterations/correlations in the soft tissues.

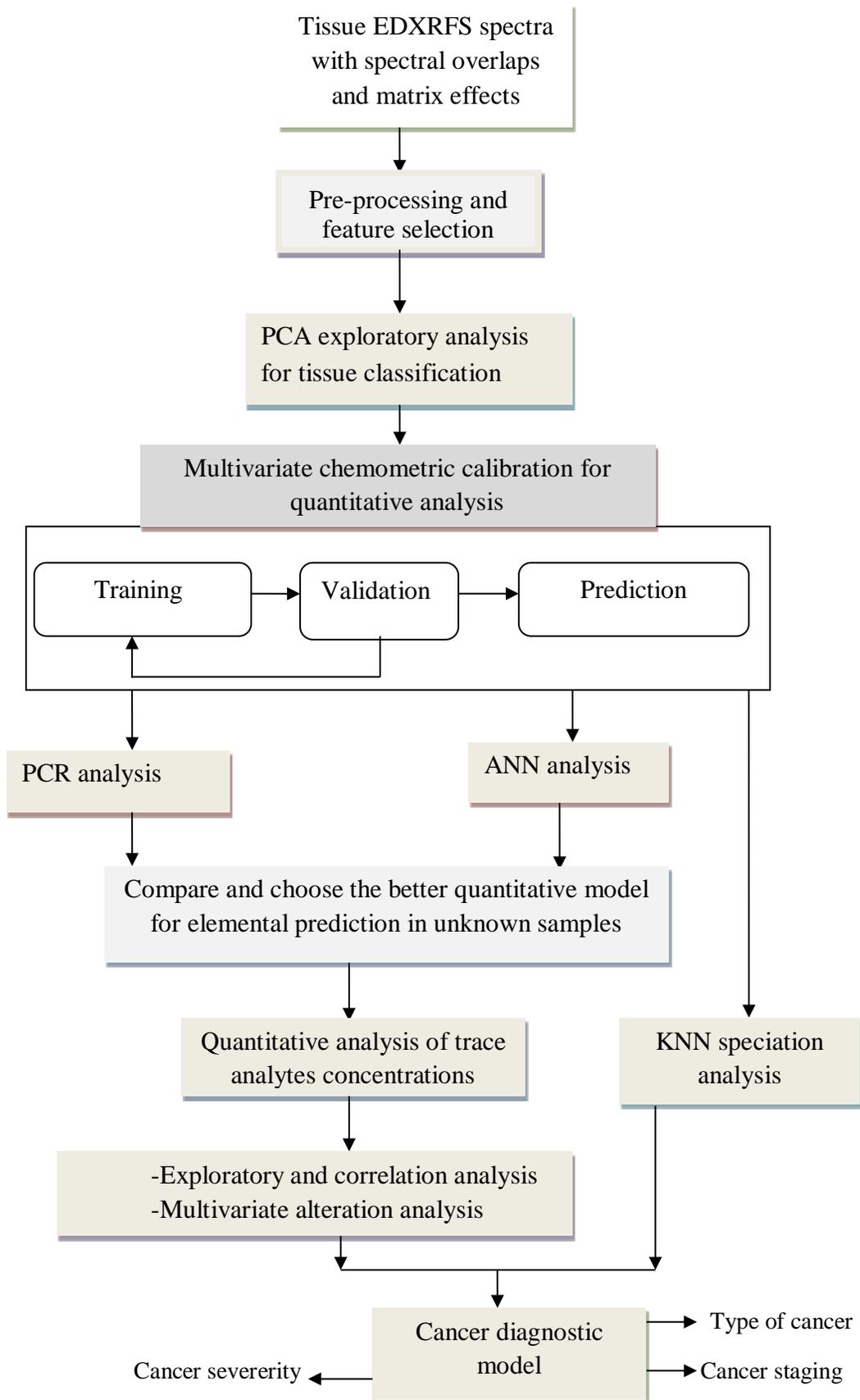


Figure 4.6.2 Block diagram showing the steps involved in model development.

CHAPTER 5: RESULTS AND DISCUSSION

5.1 Introduction

In this chapter, we present results of EDXRFS spectroscopy of tissue samples (simulate tissues, incubated cultures and selected domestic dog tissues) using for trace concentration and exploratory analysis multivariate chemometric methods. Included, are results of EDXRFS spectral data analyzed with PCA and KNN of simulate tissue samples for method development. Quantitative multivariate calibration (PCR and ANN) of EDXRFS spectral data for elemental speciation content of simulate samples are presented for validation of the quantitative methods. Finally, the application and performance evaluation of the developed chemometrics- EDXRFS method on typical tissue samples is reported.

5.2 EDXRFS spectra of simulate samples

Fig. 5.2.1 shows a typical EDXRFS spectrum with various macro-elements; Ca, K, P, S and Cl; and fluorescence peaks of the spiked trace elements namely Mn, Fe, Cu, Zn and Se in simulate tissues. The tissue major elements constituents; K, Ca and Cl can be attributed to tissue structure but had no utility in cancer differentiation for purpose of this study. The trace elements of interest; Mn, Fe, Cu, Zn and Se are linked to their important roles in physiological and metabolic processes (Rose, 1983). The enhanced Compton scatter region is very informative on the overall sample composition in terms of low Z and trace element ionic composition (Kaniu *et al*, 2011). The enhanced Se peak can be attributed to the enhanced and broadened peak of Br while the indiscernible fluorescence peak of Mn is due to low spiked concentration in the simulate sample. This complicates the classical XRF quantitative analysis which calls for an alternative multivariate approach.

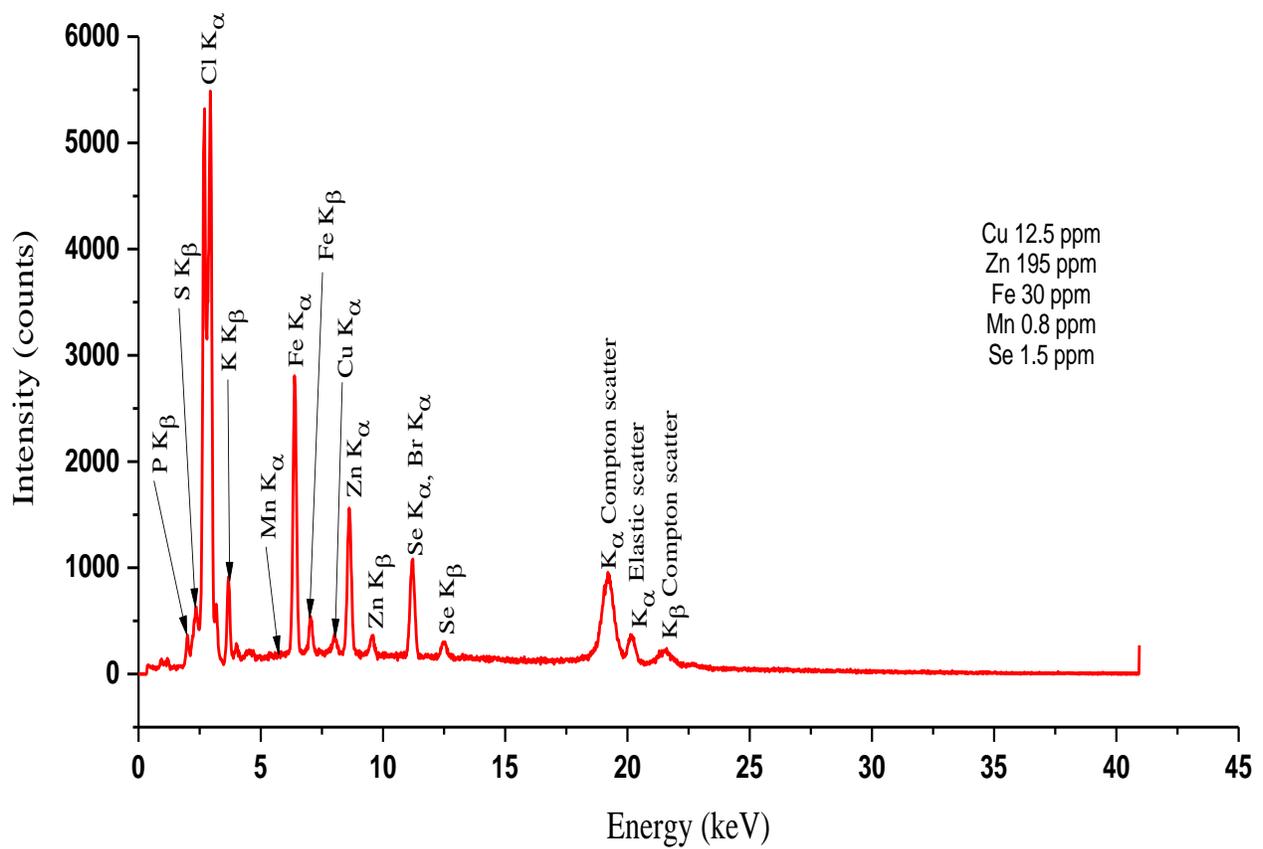


Figure 5.2.1: Typical EDXRFS spectrum of S10H simulate tissue.

Fig. 5.2.2 shows 2 spectral overlaps of simulate sample, S10H obtained at different irradiation times; 50 s and 100 s under the same irradiation conditions (50 kV and 0.001-1 mA). It can be deduced that increase of irradiation time does not result in significant increase in peak intensities which can be attributed to stable electronics of the 800HS spectrometer.

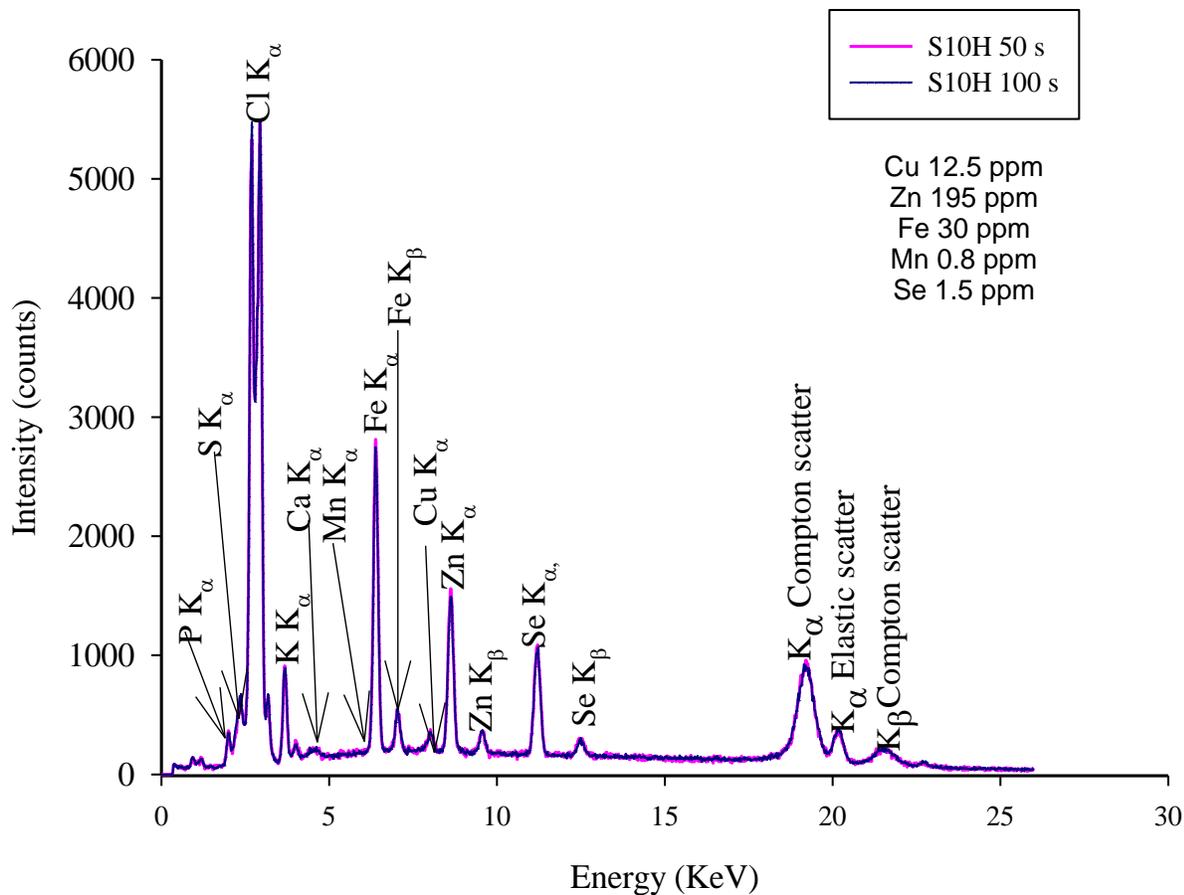


Figure 5.2.2: Typical EDXRFS profiles of simulate tissue acquired at 50 and 100 s.

A rapid chemometric aided EDXRFS analysis of tissue samples was therefore based on the spectra acquired at 50 s irradiation time for development of cancer diagnostic model.

Fig. 5.2.3 shows spectral overlaps of simulate sample (S10H) and the base matrix spectra analyzed at similar irradiation conditions (50 kV and 0.001-1 mA). The base matrix spectrum was used for spectral background subtraction as the complex organic nature of simulate sample has high background and more enhanced scatter region due to the low-Z elemental profiles in the base matrix.

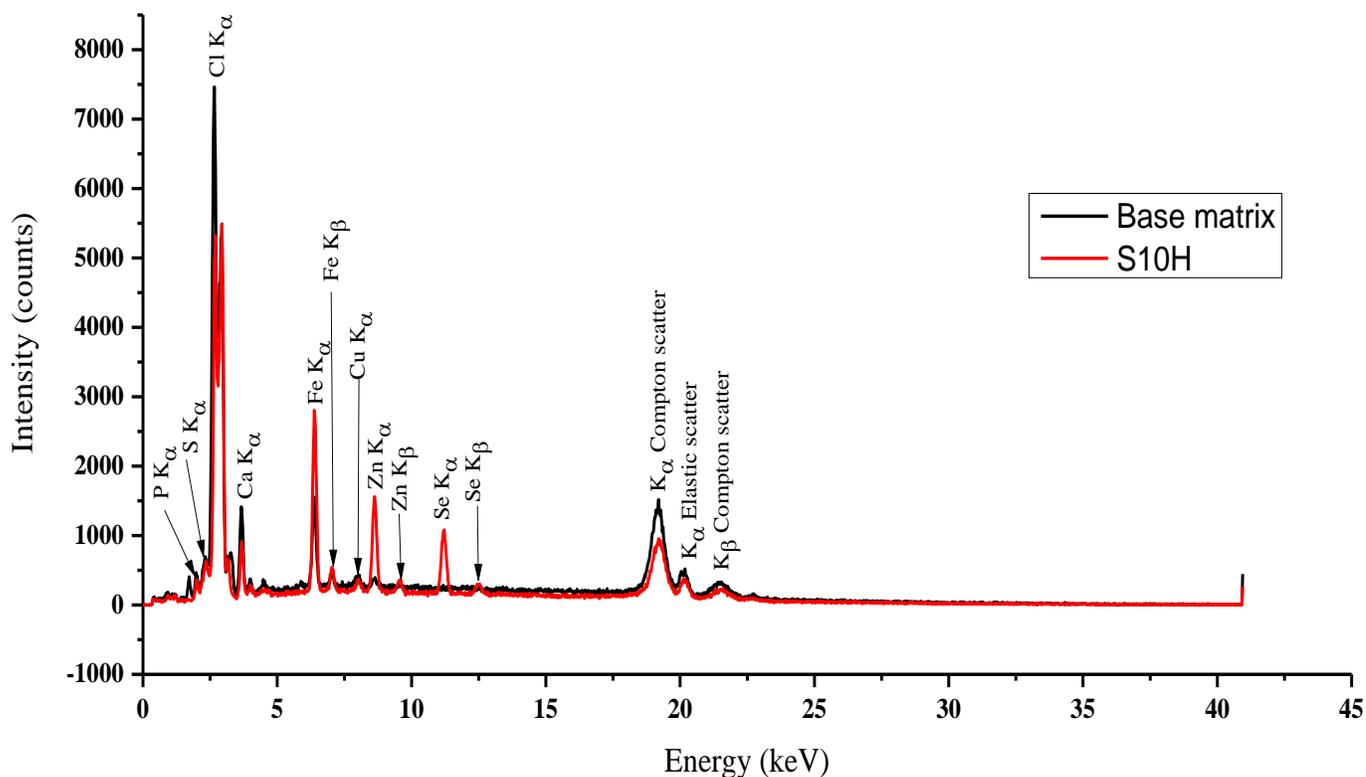


Figure 5.2.3: Typical EDXRFS spectral overlaps for S10H and paraffin wax (base matrix).

The presence of Ca in the base matrix can be attributed to its structural function in tissues while Fe and Cu can be linked to contaminations during sample preparation.

Fig. 5.2.4 shows typical EDXRFS spectrum of incubated Vero SV9 cancerous tissue sample obtained following irradiation for 50 s. The fluorescence peaks of elements of interest are not clearly discernible. However, multivariate techniques have the potential to mine information of the analytes from such spectral data (Kaniu *et al.*, 2012). This could be useful for development of trace element concentration and speciation model for cancer diagnosis at the local stage of development. This could be done using various chemometric techniques; PCA, for exploratory analysis (Oliveira *et al.*, 2010) which could enhance classification of tissue samples into cancerous and non-cancerous groups, KNN for determination of chemical speciation of Fe, Cu

and Mn, ANN and PCR for quantitative determination of concentration of Mn, Fe, Cu, Zn and Se in soft tissues.

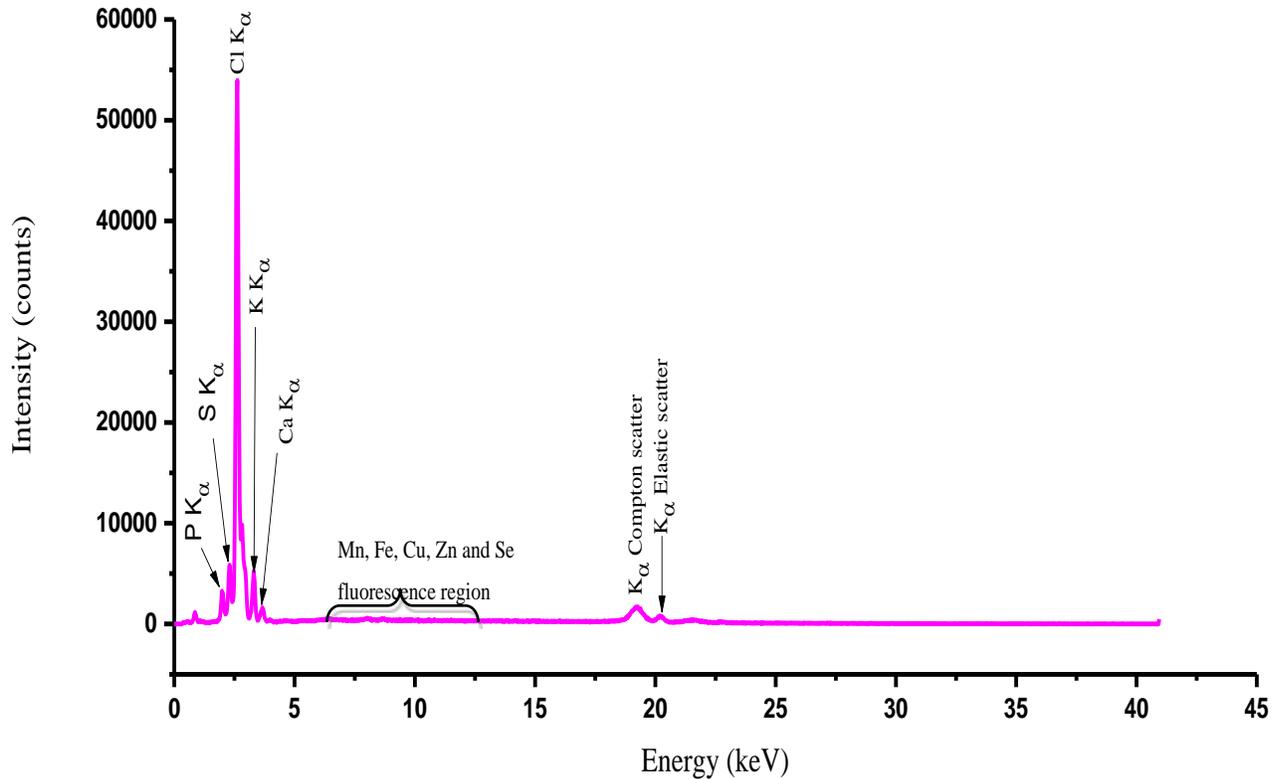


Figure 5.2.4: Typical EDXRFS spectrum of incubated Vero SV9 cancerous tissue sample.

5.3.1 Cultured tissue photomicrographs

Figures 5.3.1 and 5.3.2 show the photo micrographs of healthy and cancerous incubated tissue cultures in which there is great inhomogeneity with no clear histopathological differences in tissue contents for the diseased (cancerous) and non-diseased (healthy) cultured tissues. An alternative technique that has the potential to mine diagnostic information about cancer tissues is therefore necessary.



Figure 5.3.1: Photomicrograph of cultured lymphocyte cells harvested at day 4 (300×).

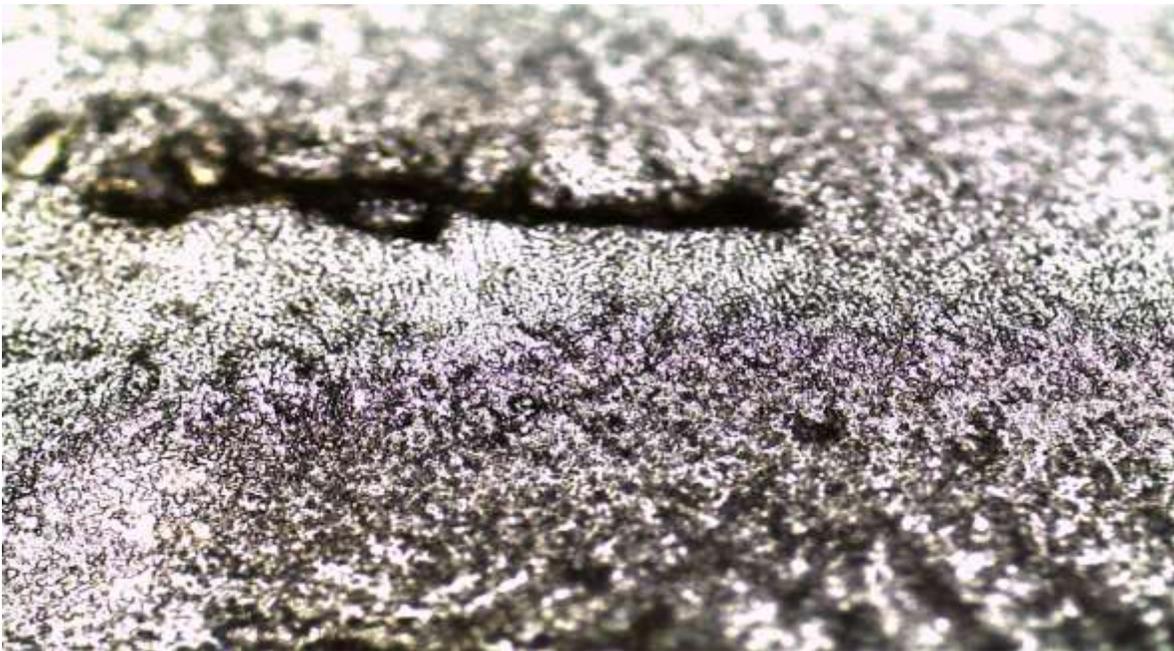


Figure 5.3.2: Photomicrograph of cultured Vero cancer cells harvested at day 4 (300 ×).

5.3.2 Detection limits of elements in simulate samples

In order to ascertain the sensitivity of the method, the detection limits (*DL*) of the standards (simulate samples) in Table 5.3.1 were calculated for the five trace elements of interest (Mn, Fe, Cu, Zn and Se) using equation 16 (Debertin and Helmer, 1988).

$$DL = 3 \times C \sqrt{\frac{Bkg}{Pa}} \quad (16)$$

where Bkg is the background intensity, C the concentration of element and Pa the analyte fluorescence intensity.

Table 5.3.1: Detection limits (ppm) for direct univariate analysis of elements in simulate tissue

Element	K_{α} Energy (keV)	Atomic number, Z	Detection limit \pm standard deviation (ppm)
Mn	5.89	25	3.2 \pm 1.4
Fe	6.40	26	13.0 \pm 5.7
Cu	8.04	29	8.5 \pm 2.8
Zn	8.63	30	10.6 \pm 3.5
Se	11.21	34	1.6 \pm 0.7

The concentrations of trace elements below or close to the detection limits in Table 5.3.1 are hardly achieved using routine EDXRF analysis for modeling towards disease diagnostics in soft tissues.

5.4 Principal component analysis (PCA) exploration of simulate samples

In the present study, principal component analysis (PCA) was used to reduce the spectral data dimensions for sample exploration in order to identify possible discriminating features between the 2 sets of simulate samples (higher and lower speciation) at irradiation time of 50 s. The obtained PCA results are based on the analyte fluorescence peaks of the elements of interest (Mn, Fe, Cu, Zn and Se) and the K_{α} Compton scatter as inputs.

Fig. 5.4.1 shows PCA score plot in which the simulate samples were classified into 2 clusters based on their matrix composition; explained by variance of 93 % (77 % and 16 % for PC1 and PC2 respectively) which clusters samples containing Fe^{3+} , Cu^{2+} and Mn^{4+} and Fe^{2+} , Cu^{+} and Mn^{2+} , respectively with same concentration of Zn and Se. Samples S5L, S7L and S10L were

identified as outliers and constituted the remaining 7 % variation which can be attributed to noise.

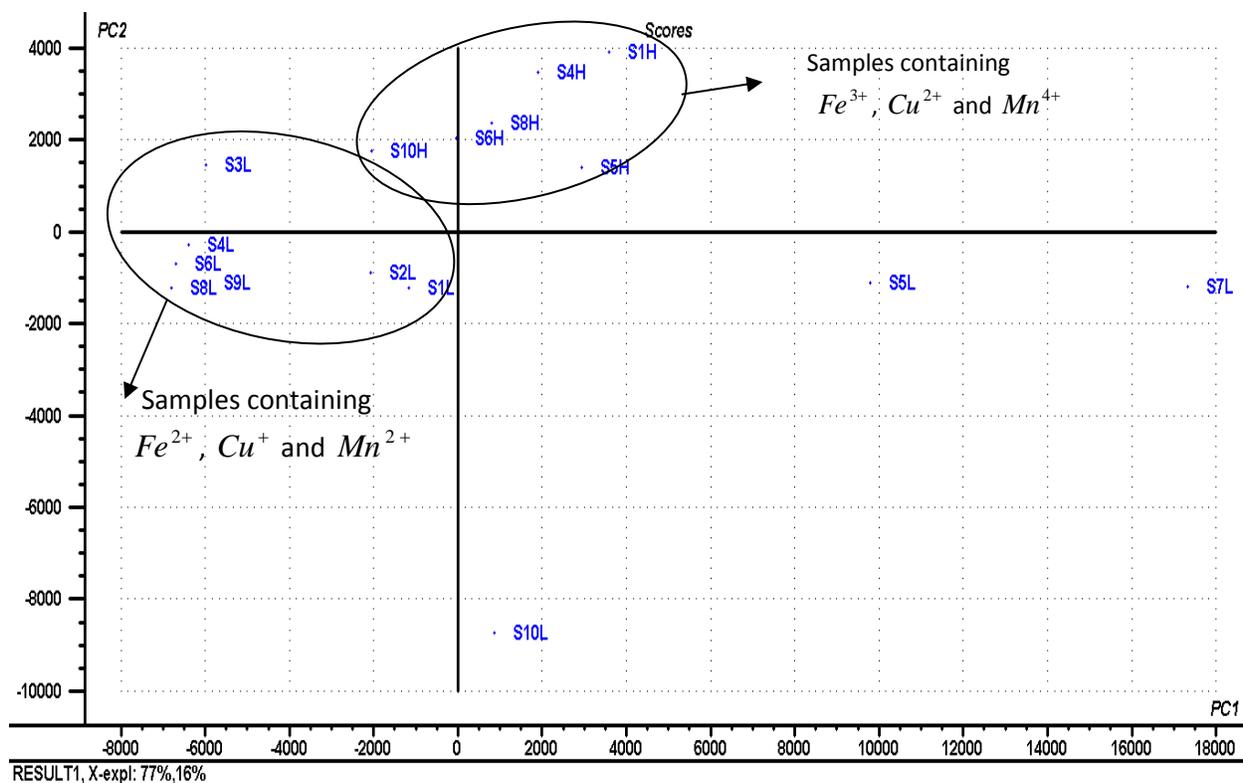


Figure 5.4.1: PC1 (77%) × PC2 (16%) score plot of simulate samples.

The two clusters in Fig. 5.4.1 are influenced by both K_{α} and K_{β} intensity signals of Ca, Cl, Mn, Fe, Cu, Zn and Se as shown on the PCA loadings plot in Fig. 5.4.2. The presence of Cl in the loadings plot was as a result of $CuCl$, $MnCl_2 \cdot 4H_2O$ and $MnCl_4$ used as sources of Cu^{+} , Mn^{2+} and Mn^{4+} respectively for simulate sample preparation. The inclusion of Ca in the loadings plot can be attributed to its structural function in body tissues.

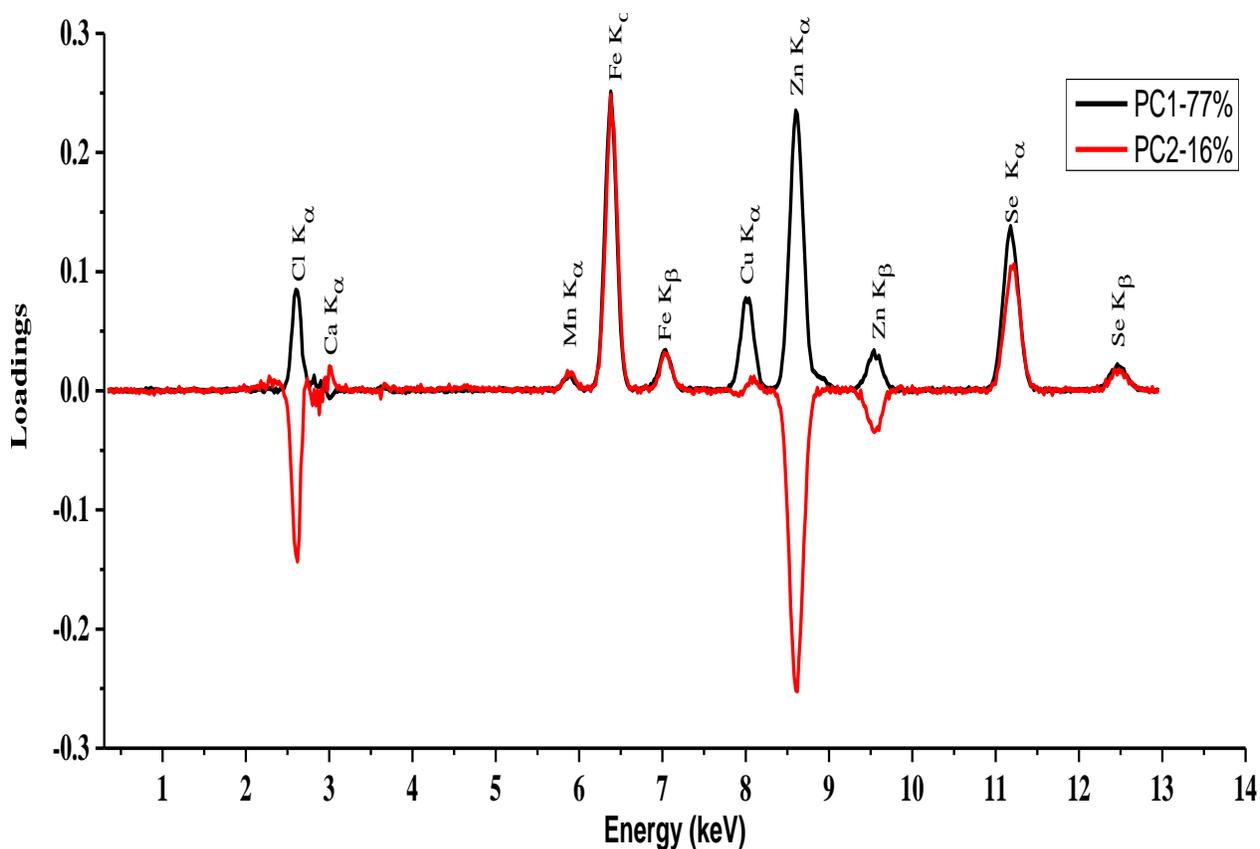


Figure 5.4.2: PCA loadings plot of simulate samples at 50s live time for a total fluorescence.

The positive loadings of the elements of interest (Mn, Fe, Cu, Zn and Se) in PC1 significantly (77%) contribute to the classification of the simulate samples into 2 clusters. The manifestation of PC2 loadings in Figure 5.4.2 is due to its orthogonal graphical trend to PC1. Zinc (Zn) loadings have a positive influence on PC2 clusters while Mn, Fe, Cu and Se have a negative influence on PC2 clustering.

Fig. 5.4.3 shows analysis on data acquired at live time 100 s, explained by 79 % and 15 % variance for Principal Components (PC1 and PC2) respectively.

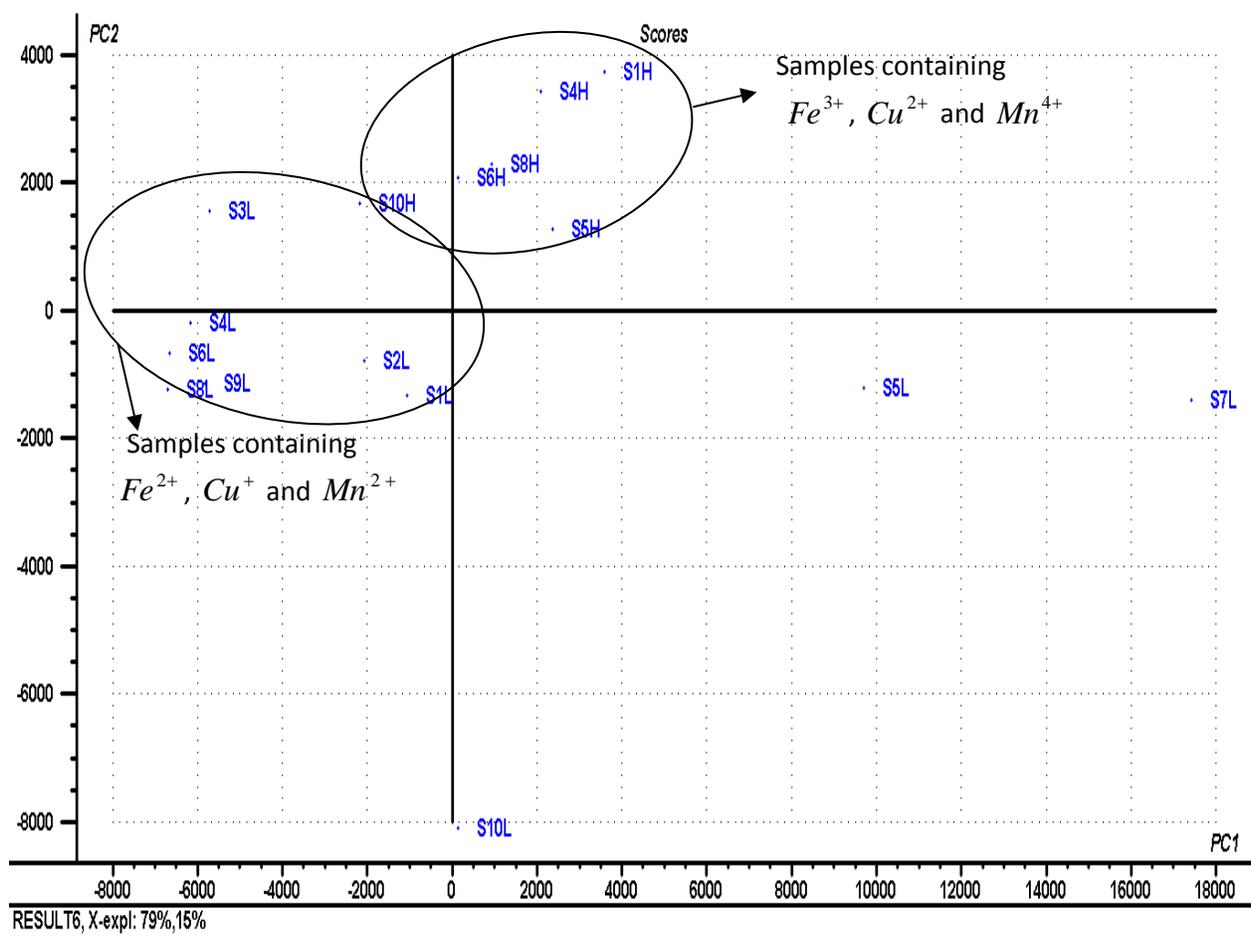


Figure 5.4.3: PC1 (79%) × PC2 (15%) score plot of simulate samples for 100s live time.

It is clear that despite the increase in irradiation time, the 2 clusters are still discernible with explained variance of 94% which can be attributed to stable electronics of the 800HS spectrometer.

Figure 5.4.4 shows loadings of signal intensities of K, Ca, Cl, Mn, Fe, Cu, Zn, and Se that accounts to the PCA score plot classification in Fig. 5.4.3.

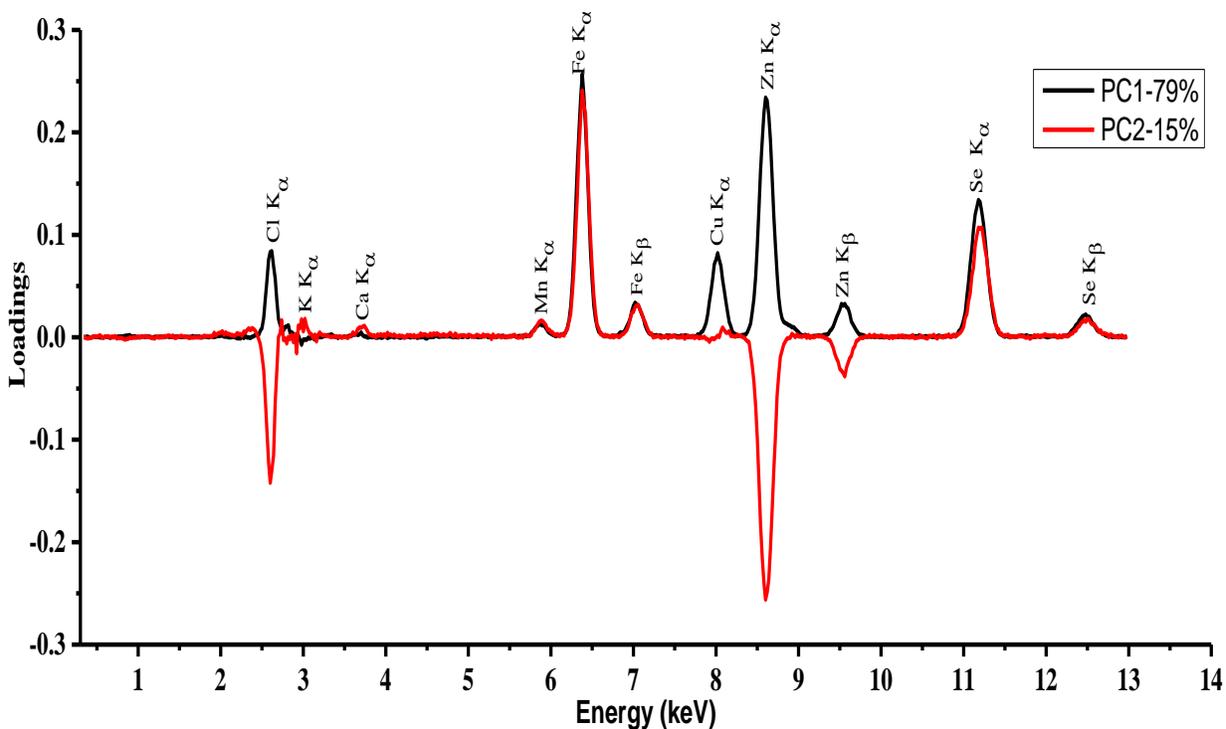


Figure 5.4.4: PCA loadings plot of simulate samples at 100s live time for total fluorescence

In general, there is no significant variation in PCA score plot results. Therefore, increase in irradiation time does not result in significant increase in explained variance of the PCA score plots. In this study, analytical results reported are consequently based on the irradiation time of 50s.

Fig. 5.4.5 shows analysis of simulate spectral data acquired at live time 50 s but using the fluorescence peaks plus Compton scatter peaks. Almost the entire information at 95 % variance was explained by 75 % and 20 % for Principal Components; PC1 and PC2 respectively.

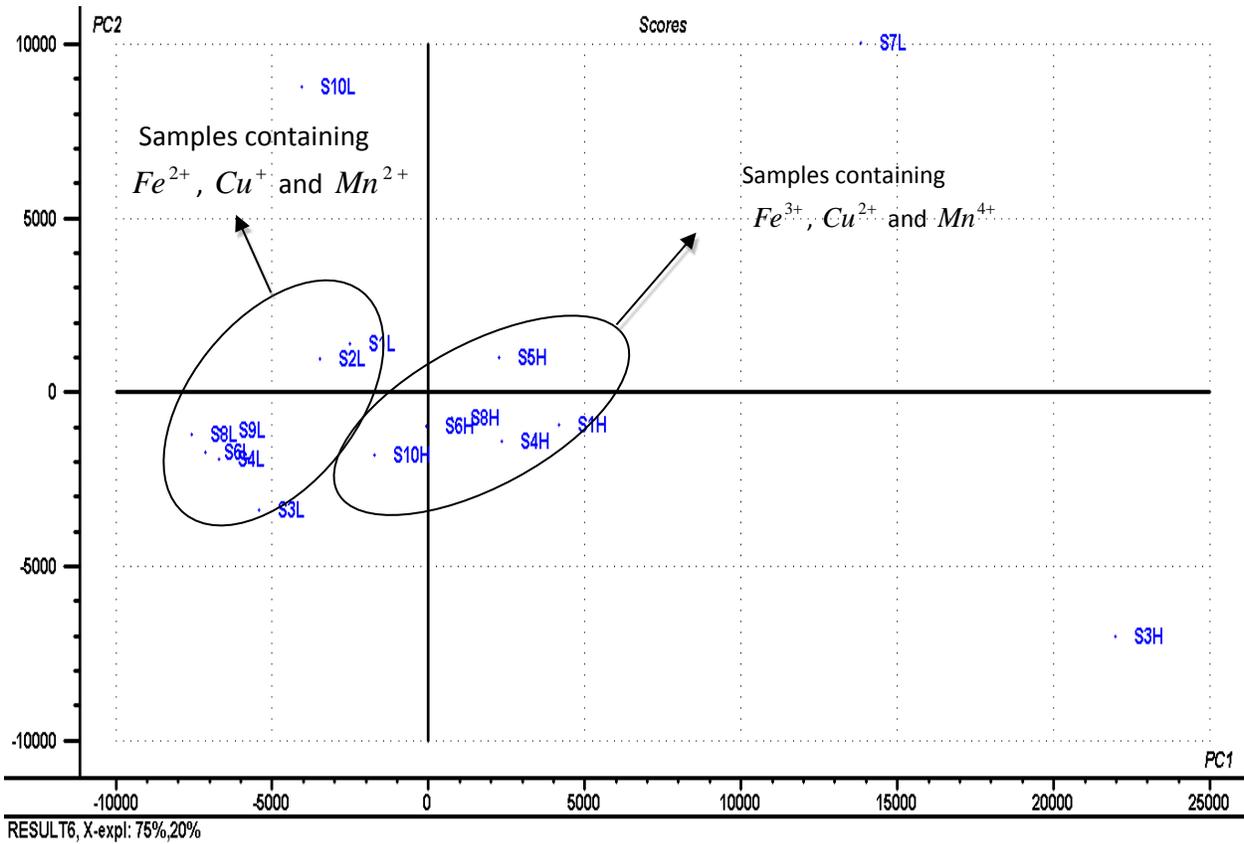


Figure 5.4.5: PC1 (75 %) × PC2 (20 %) score plot of simulate samples for 50s live time for the fluorescence peaks plus Compton scatter.

There is a decrease of 2 % in PC1 and significant increase of 4 % in PC2 explained variance with reference to Figure 5.4.1 which can be attributed to the additional information from the Compton scatter that should not be ignored in the development of cancer diagnostic model.

As seen from Figure 5.4.6, the 2 clusters in Fig.5.4.5 are still based on the speciation of Mn, Fe, and Cu in the samples with a modulation according to the entire matrix composition but with no significant contribution from the Compton scatter.

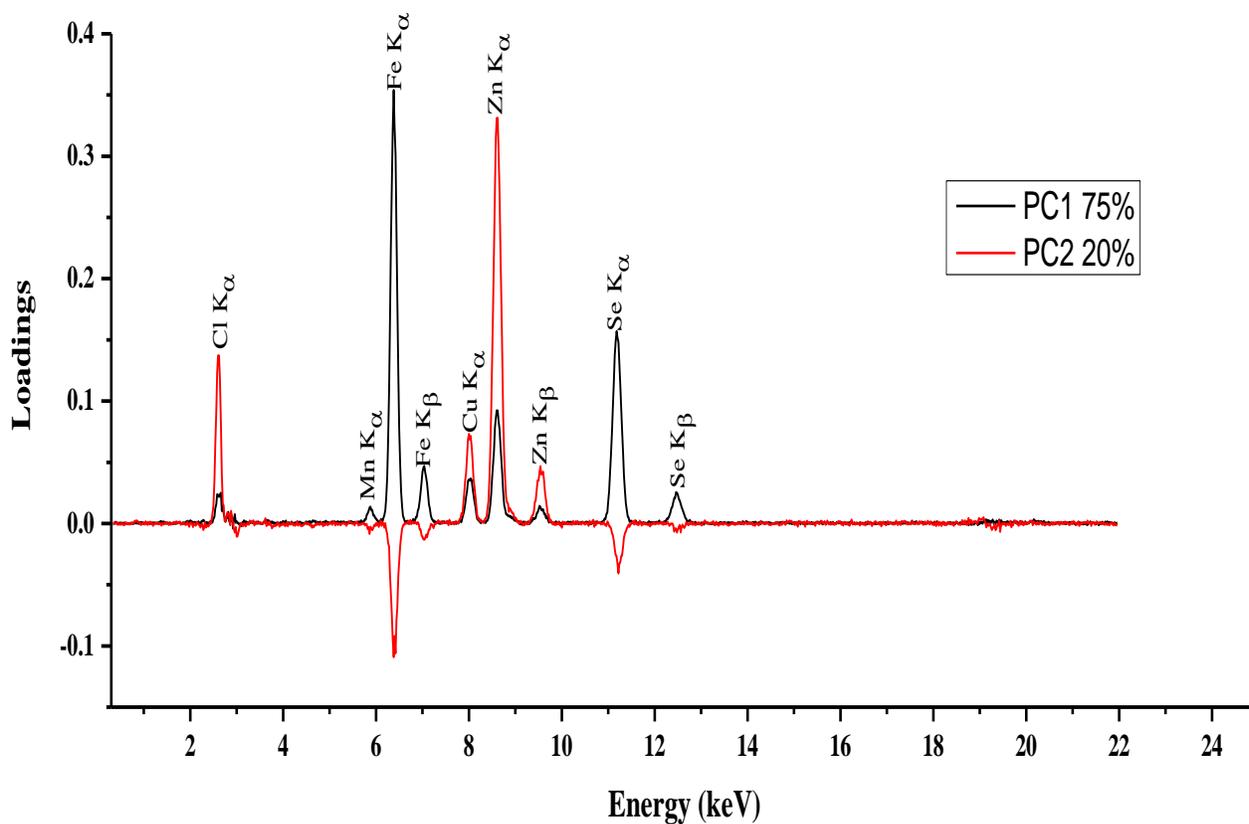


Figure 5.4.6: PCA loadings plot of simulate samples for 50s live time using both fluorescence plus Compton scatter.

As envisaged from the loadings plot, all the trace elements of interest (Mn, Fe, Cu, Zn and Se) positively influence PC1. However, PC2 is positively influenced by Fe and Se but with negative influence of Zn. This clearly illustrates the significance of Compton scatter which alters the role of Zn in the loadings plot as compared to its contribution to PC2 loadings in Figure 5.4.1.

In order to explore the role of speciation contributions to the PCA classification model, PCA analysis of the simulate tissue samples was thereafter carried out using the selected fluorescence spectral regions of K_{α} and K_{β} identified for the trace elements; Fe, Cu and Mn as input data in the PCA (Table 4.3) as shown in Fig. 5.4.7.

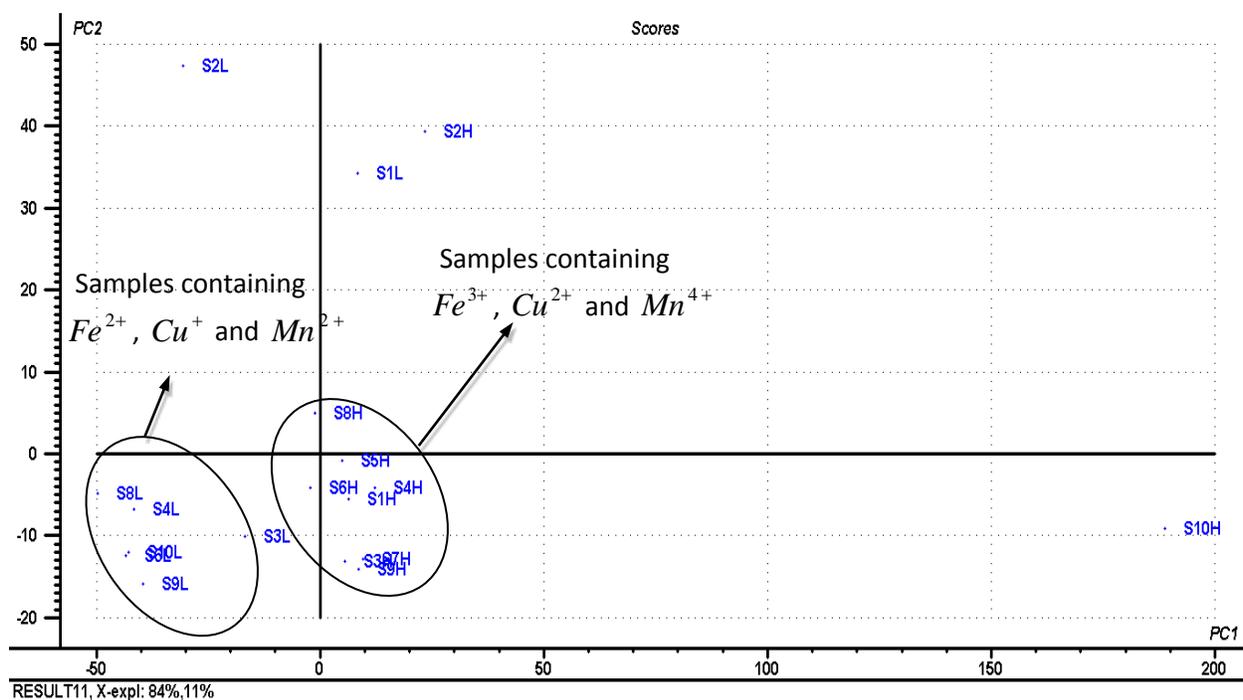


Figure 5.4.7: PC1 (84%) × PC2 (11%) score plot of simulate samples using selected fluorescence region of Cu, Fe and Mn for 50s live time.

The 2 clusters in Fig. 5.4.7 are majorly influenced by K_{α} and K_{β} signals of Fe and Cu as shown on the loading plot in Fig. 5.4.8. This therefore lays a foundation for further speciation analysis of Fe and Cu to ascertain their possible role in carcinogenesis of cancer for model development.

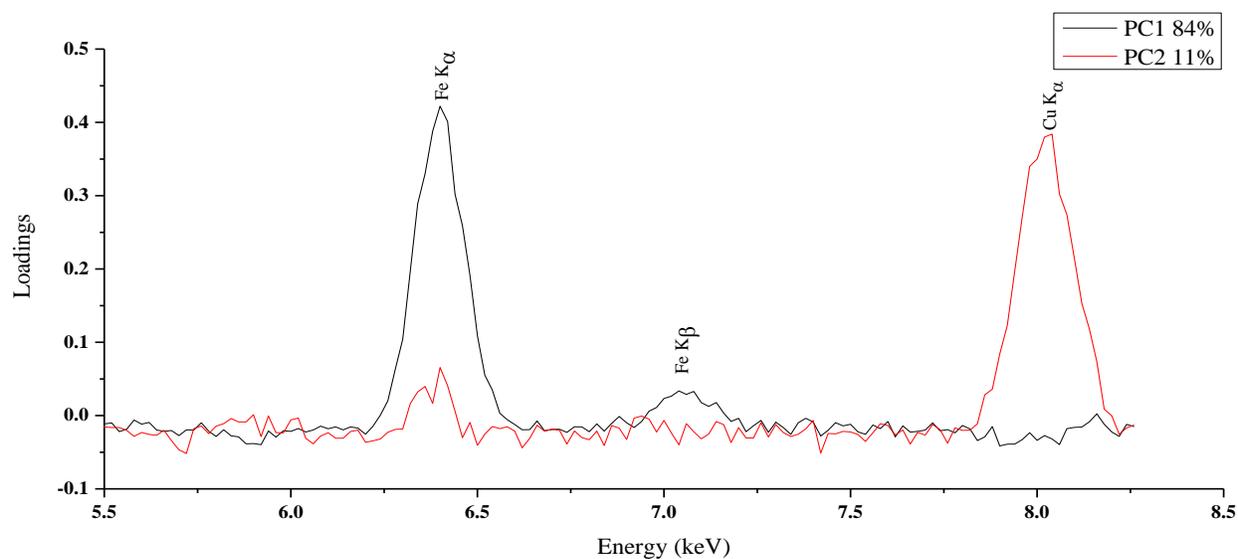


Figure 5.4.8: PCA loading plot of simulate samples for 50s live time using the fluorescence region of Cu, Fe and Mn.

In general, the loading plots in Figures 5.4.2, 5.4.4, 5.4.6 and 5.4.8 indicate that PC1 and PC2, based on 2 major clusters were able to distinguish between simulate samples containing either the lower speciation (Fe^{2+} , Cu^+ and Mn^{2+}) or higher speciation (Fe^{3+} , Cu^{2+} and Mn^{4+}). The 2 clusters were influenced by K_{α} and K_{β} intensities of Fe^{2+} , Cu^+ and Mn^{2+} , and Fe^{3+} , Cu^{2+} and Mn^{4+} together with total elemental content of spiked Mn, Fe, Cu, Zn and Se but with no significant influence from the Compton scatter, based on the positive contribution of PC1 loadings. This generally validates PCA as an exploratory multivariate tool for analysis of cultured and selected dog tissues towards cancer diagnosis based on trace elemental content.

5.5 Quantitative multivariate chemometric techniques

The calibration models for method development were based on building a relationship between the desired chemical and biological property (concentration of elements) in a sample and its corresponding spectral data. Multivariate calibration techniques, ANN and PCR were best utilized to exploit the relationship (linear or non-linear) between several responses (sample spectral data matrix) and vector of properties (concentrations of analytes). This was aimed for prediction of the levels of trace elements (Fe, Mn, Cu, Zn and Se) in future tissue samples based on observed characteristics (concentration and speciation).

5.5.1 Artificial neural networks (ANN) for quantitative analysis of trace elements

The predictive ability of multivariate calibration by ANN was based on selection of 4 neurons in the hidden layer which was obtained as a result of optimization of the neural network architecture. This was achieved by trial starting from one neuron and increasing the number by one in the subsequent trials as the models were trained, tested and validated. This was aimed at achieving the smallest training error as a function of the addition of neurons to the intermediate layer.

Sample training performance plots shown in Fig 5.5.1.1 indicate the variation of mean square error (MSE) decreasing to lower values for convergence of the network.

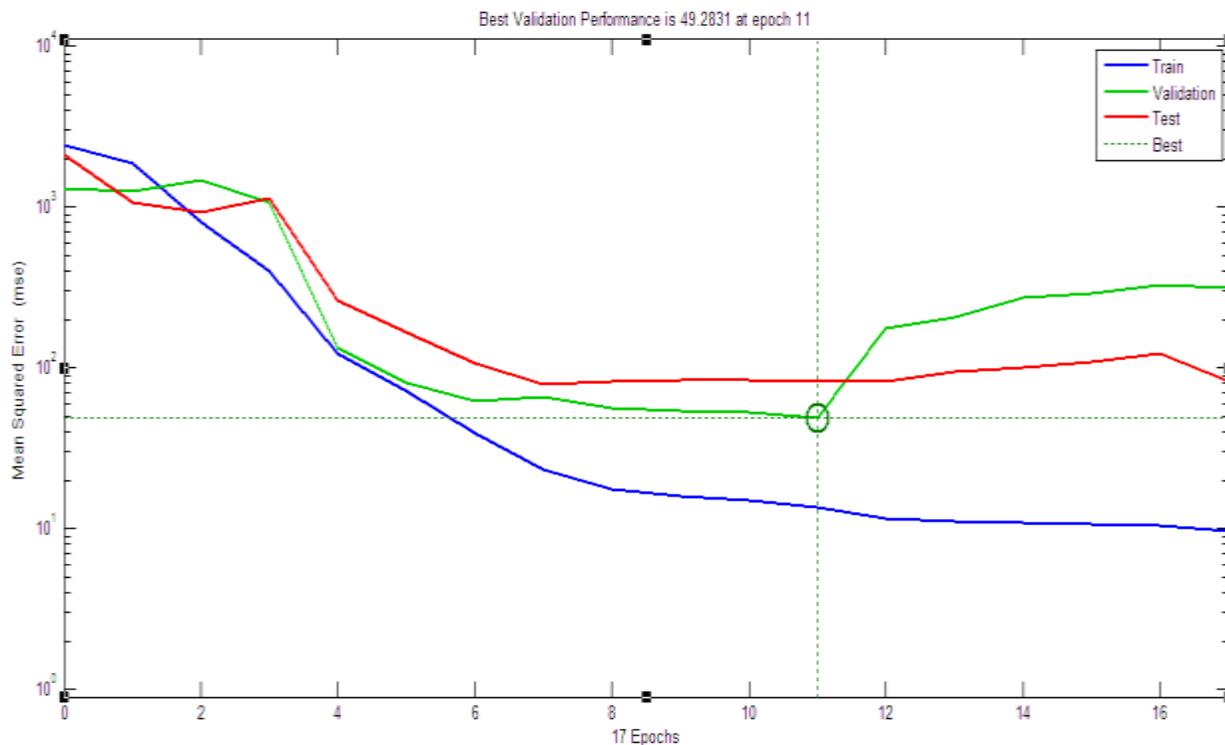


Figure 5.5.1.1: Performance plots for ANNs training errors for 4 neurons exploiting selected (Fe, Cu, Zn, Mn and Se) fluorescence regions plus Compton scatter.

The 3 lines in the plot are linked to the training (blue), validation (green) and testing (red) analysis of the model and are as a result of the inputs fed into the network being divided into three sets; 60 % for training the network, 20 % for validating how well the network generalizes and 20 % to test how the network will perform on new set of data.

Table 5.5.1 shows the results obtained following training the neural network at various neurons in the hidden layer. The efficiency in predicting is based on the R value which is a measure of how well the variation in the output is explained by the targets. If its numerical value is close to 1, then there is good correlation between targets and the outputs. The best model was initialized as that corresponding to high R value, with low mean-squared error and with fewer neurons in

the hidden layer. A high number of neurons in the hidden layer would cause over-fitting while too few would not be able to classify accurately. As such, the model with 4 neurons was selected because it had lower validation mean-squared error with a consistent R- value that was close to 1.

Table 5.5.1: Analytical performance indices for different ANN models

Model	Number of neurons in the hidden layer	Best validation performance mean squared error	No. of epochs/ iterations	Correlation coefficients			
				Training R-value	Validation R-value	Test R-value	Overall R-value
A	1	105.2114	2	0.8803	0.9021	0.7690	0.8745
B	2	81.6165	7	0.9814	0.8155	0.8808	0.9632
C	3	74.4968	8	0.9674	0.7208	0.9749	0.9619
D	4	49.2831	11	0.9981	0.9687	0.9421	0.9888
E	7	7.2299	2	0.9892	0.9564	0.9895	0.9795

By 11th iteration, no significant over-fitting had been reported; the test and validation sets behaved in a similar manner. Over-fitting was automatically avoided by stopping the training when validation error increases beyond the expected low limit (Howard *et al.*, 2009) as shown in Table 5.5.1.

In addition, the network response was investigated by performing a regression analysis between the network response and the corresponding targets as shown in Fig. 5.5.1.2 and 5.5.1.3.

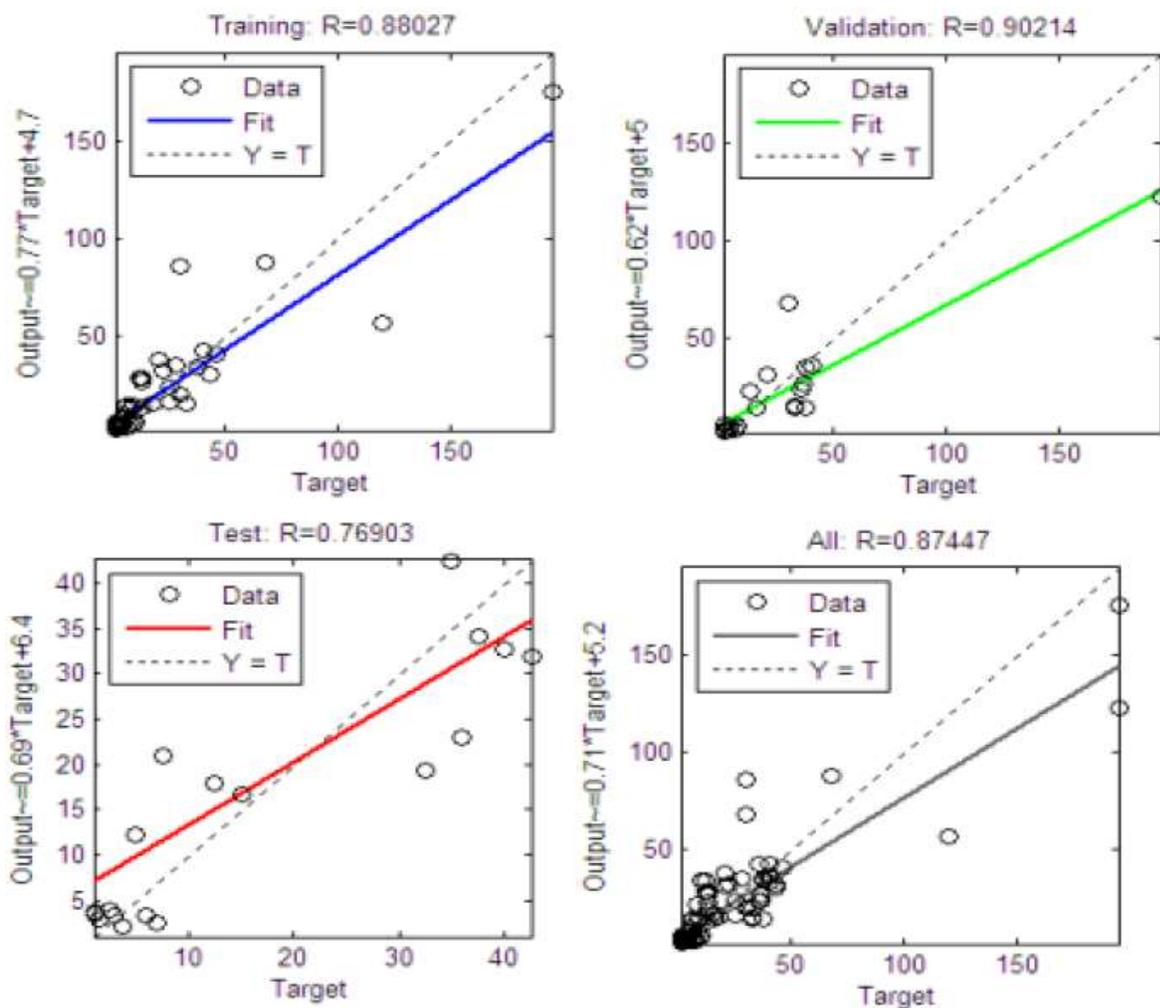


Figure 5.5.1.2: Linear regression of ANN outputs and corresponding targets utilizing the fluorescence regions plus Compton scatter regions for simulates.

ANN was able to simultaneously train the model for the elements of interest (Mn, Fe, Cu, Zn and Se) with R value close to 1 (0.88). This was subsequently followed by good validation and prediction with R values of 0.90 and 0.77 respectively which resulted to an overall R value of 0.87.

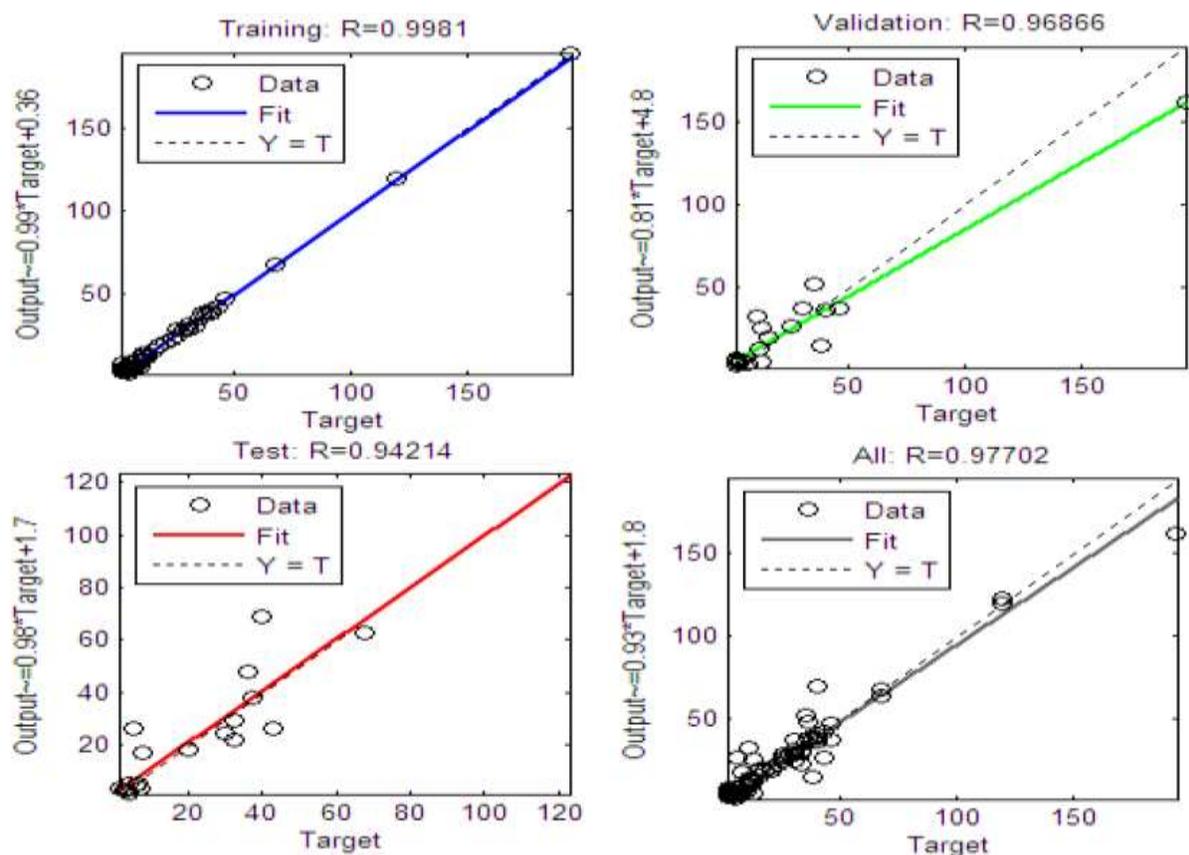
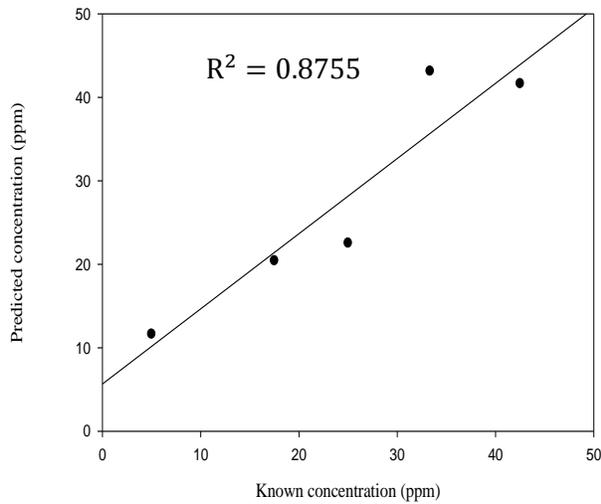


Figure 5.5.1.3: Linear regression of ANN outputs and corresponding targets utilizing fluorescence regions.

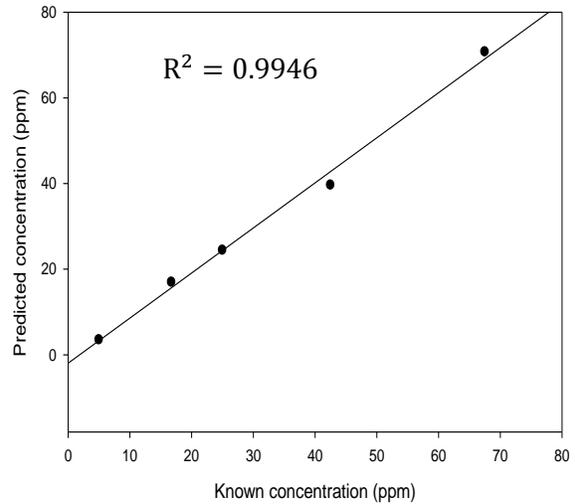
It can be deduced that selected fluorescence regions give a better model with R values so close to 1; 0.998, 0.97 and 0.94 for training, validation and prediction respectively. This resulted to a good overall R value of 0.98 which is over 95 % for the total response.

In general, the addition of Compton scatter to the selected fluorescence region has potential utility in development of quantitative model for trace elemental determination. This can be viewed from the fact that the output satisfactorily tracked the targets (Fig. 5.5.1.2) for an overall R value of 0.87 for the total response.

The predicted verses measured (known) regression plots, Fig. 5.5.1.4-8, show how the ANN model was able to predict Zn, Cu, Mn, Fe and Se in simulate tissue samples and how well the model is likely to perform for future tissue samples of similar matrix composition.

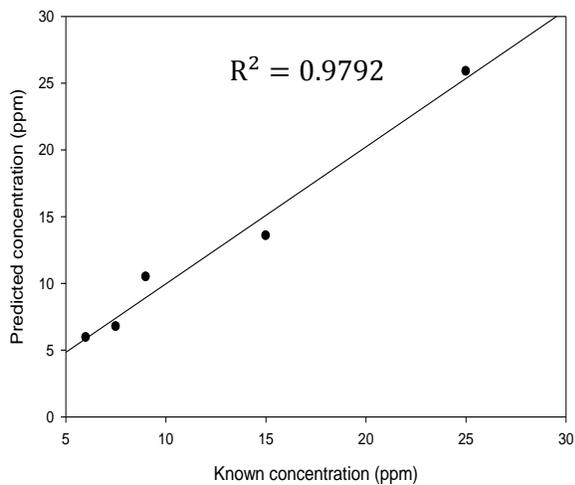


(i)

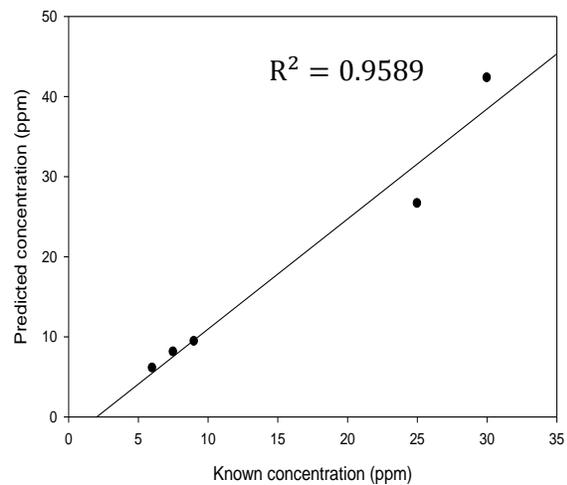


(ii)

Figure 5.5.1.4: ANN regression plots for predicted versus known Zn utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.

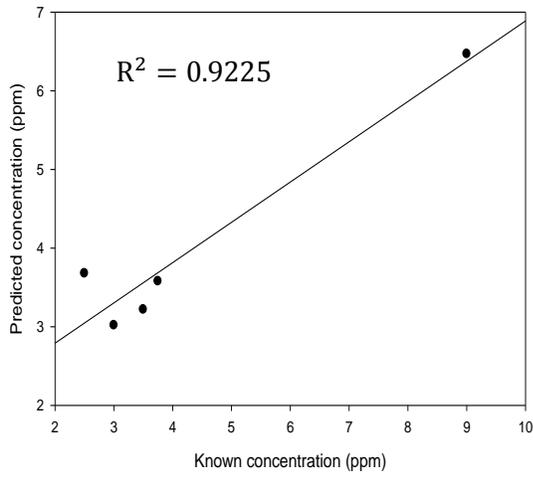


(i)

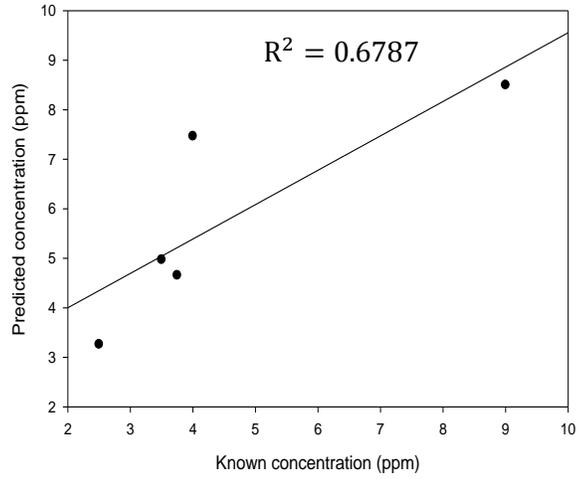


(ii)

Figure 5.5.1.5: ANN regression plots for predicted versus known Cu utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.

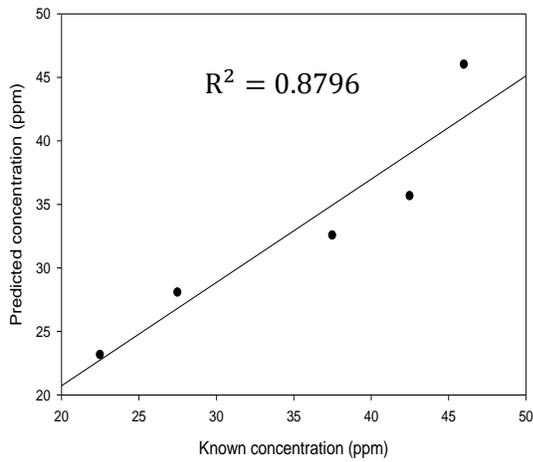


(i)

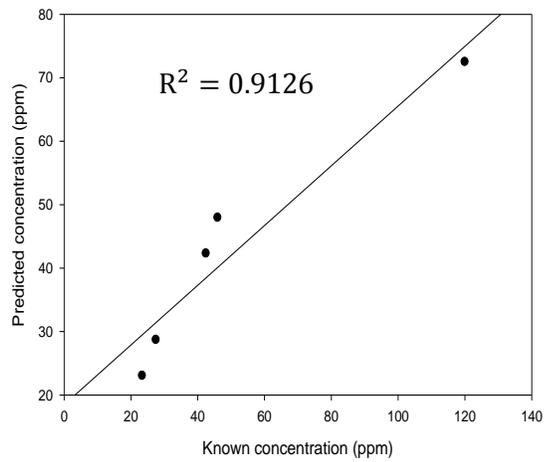


(ii)

Figure 5.5.1.6: ANN regression plots for predicted versus known Mn utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.



(i)



(ii)

Figure 5.5.1.7: ANN regression plots for predicted versus known Fe utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.

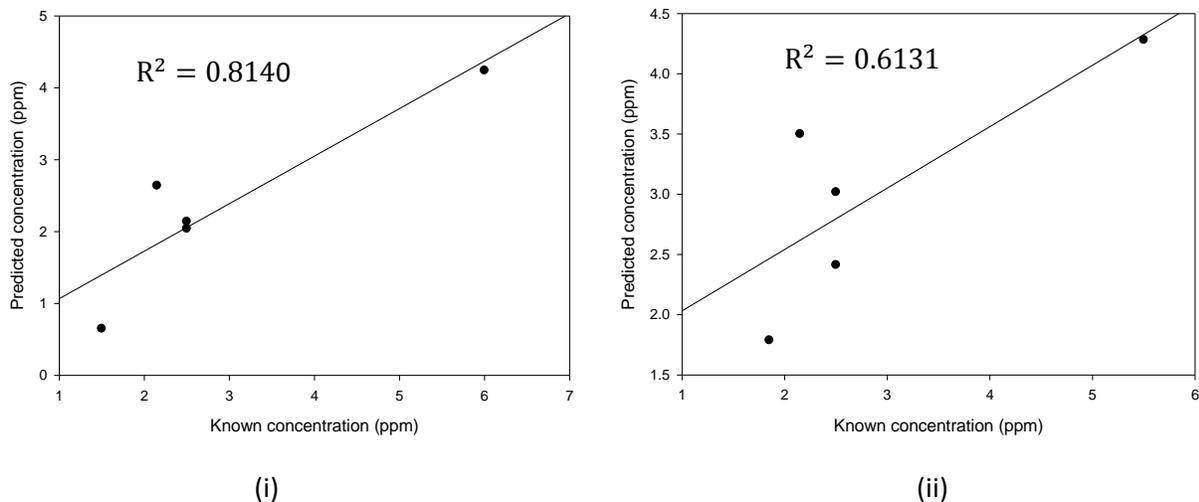


Figure 5.5.1.8: ANN regression plots for predicted versus known Se utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.

The low R^2 values for Mn and Se can be attributed to less enhanced structural background, low spiked concentration levels in simulate tissues and the influence of noise. The ANN prediction results were closer to the true values of known (spiked) concentration for most trace elements of interest.

5.5.2 Principal Component Regression (PCR) of trace elements

PCR was performed using selected fluorescence peaks/plus Compton scatter peak regions of the trace elements (Mn, Fe, Cu, Zn and Se) where the regression of Y (concentration) was based on selected principal components of X instead of the original X-variables. The calibration matrix was decomposed by PCA to obtain the first 2 principal components (PC1 and PC2) with the largest variances to explain as much of the total variation of the X-variables as possible for good correlation with the dependent Y-variable(s). The 2 PCs were then regressed against the given concentrations to get the regression coefficients which were used on the test data to predict the concentrations of the constituent elements.

The predicted vs. measured (known) regression plots (Fig. 5.5.2.1-5) shows how the PCR model was able to predict the trace elements; Zn, Cu, Mn, Fe and Se for simulate tissue prediction samples.

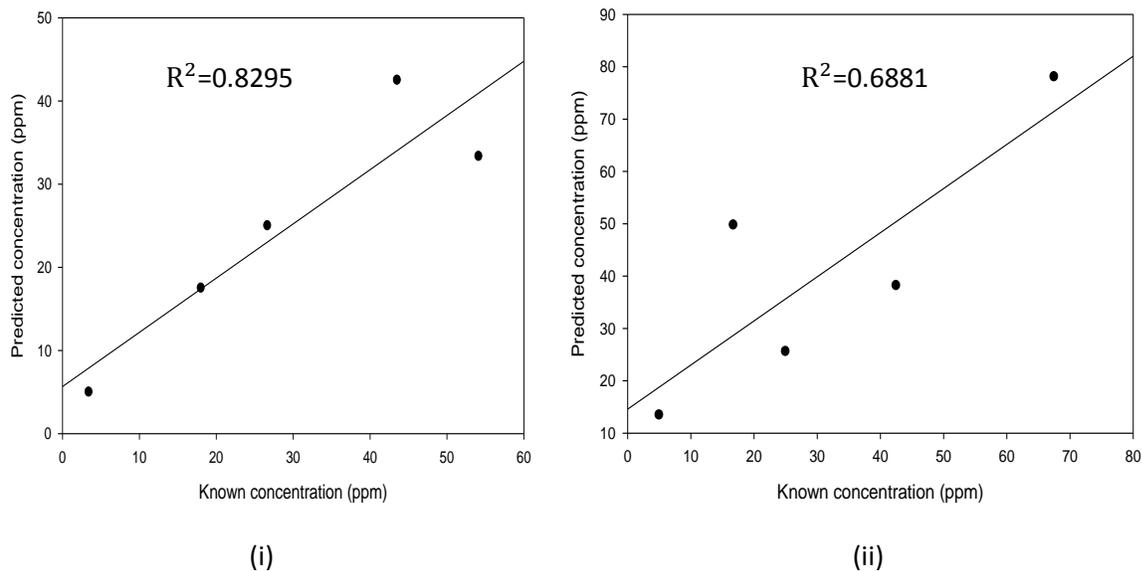


Figure 5.5.2.1: PCR regression plots for predicted verses known Zn utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.

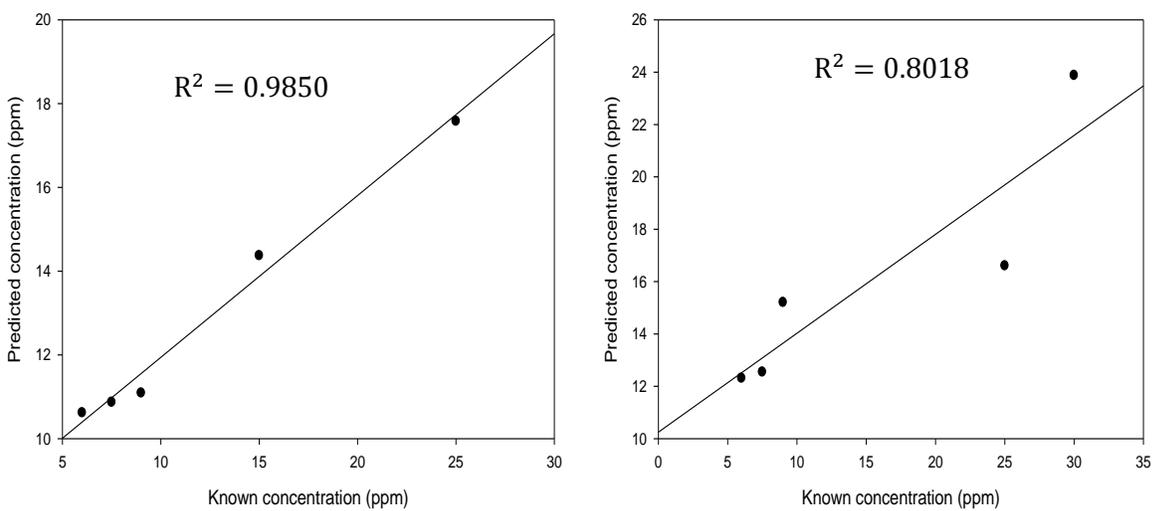
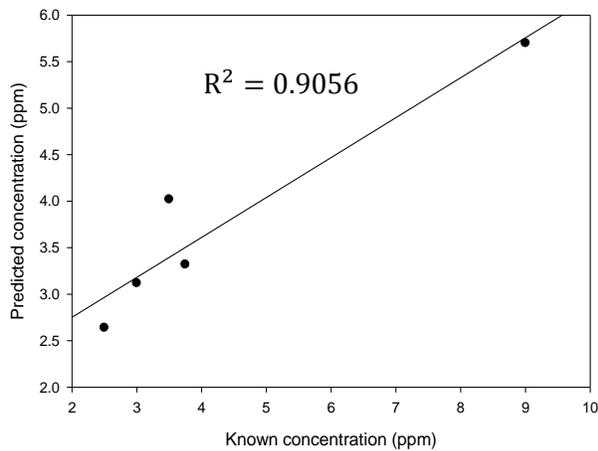
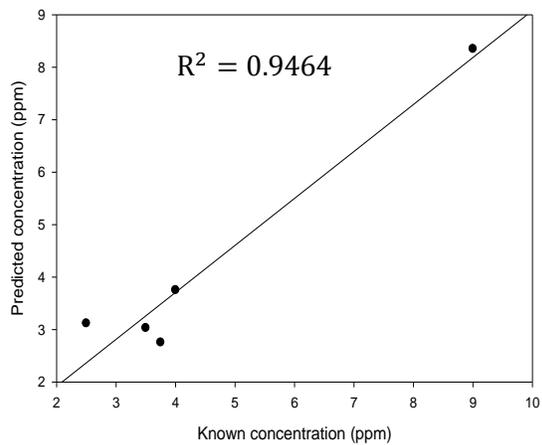


Figure 5.5.2.2: PCR regression plots for predicted verses known Cu utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.

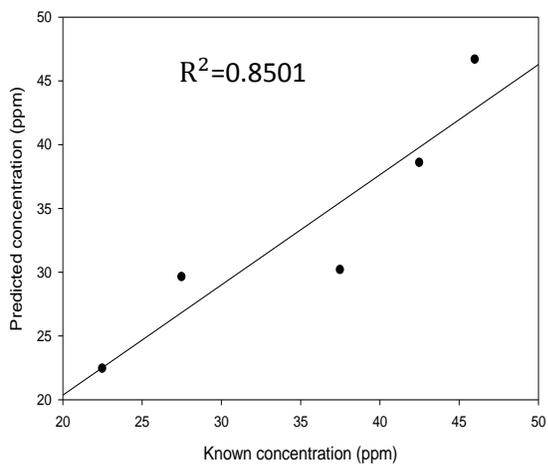


(i)

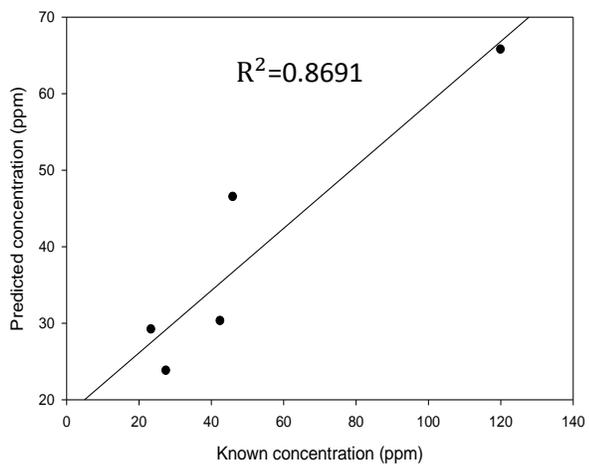


(ii)

Figure 5.5.2.3: PCR regression plots for predicted versus known Mn utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.



(i)



(ii)

Figure 5.5.2.4: PCR regression plots for predicted versus known Fe utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.

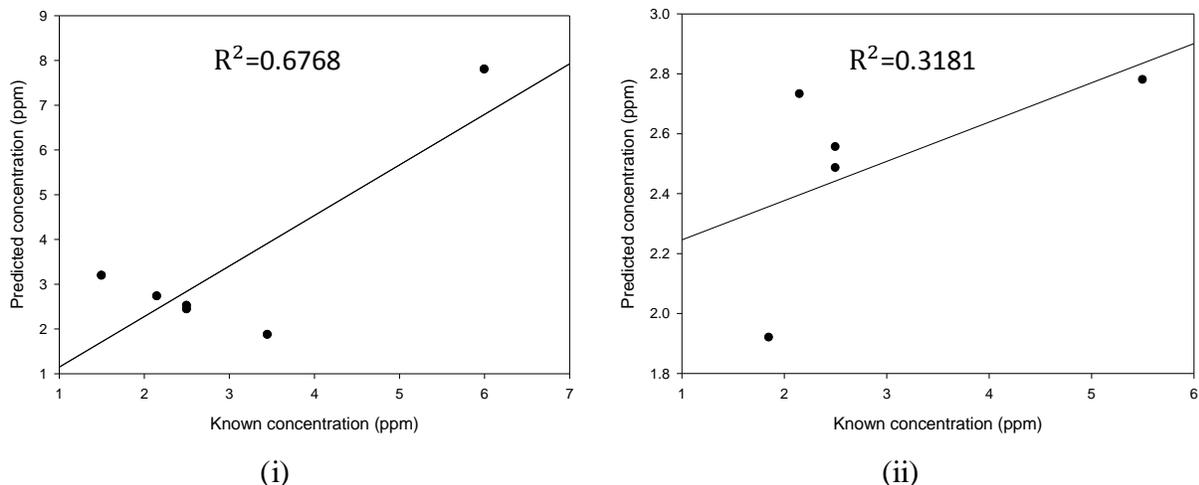


Figure 5.5.2.5: PCR regression plots for predicted versus known Se utilizing (i) the fluorescence peaks and (ii) the fluorescence plus Compton scatter region.

The PCR model gave low R^2 values (as low as 0.3181 for Se) which could be linked to low spiked concentrations of Se and also due to the non-linearity nature of the model as it could not deal with non linear spectral relations to analyze concentrations in the standard samples.

The ANN and PCR models were compared to determine the most appropriate multivariate method for quantitative analysis of trace elements of interest; Fe, Cu, Mn, Zn and Se in cultured tissues and selected dog tissue samples in this study. Table 5.5.2 shows the comparison of regression coefficients for both ANN and PCR. This can give guidance on the better quantitative model for a specified trace element.

Table 5.5.2: Regression coefficients for ANN and PCR

Element	Regression coefficients (R^2 values)			
	ANN (i)	ANN (ii)	PCR (i)	PCR (ii)
Mn	0.9225	0.6787	0.9056	0.9464
Fe	0.8796	0.9126	0.8501	0.8691
Cu	0.9792	0.9589	0.9850	0.8018
Zn	0.8755	0.9946	0.8295	0.6881
Se	0.8140	0.6131	0.6768	0.3181

In general, the overall ANN model predictions of elemental concentrations of Mn, Fe, Cu, Zn and Se were better than the corresponding values obtained through PCR model as shown in Table 5.5.2. This observation shows that prediction of elemental concentration in unknown samples by ANN model is better as compared to PCR model. This can be attributed to the fact that ANN model can learn both linear and non-linear mappings in samples thus considered more reliable (Goodacre, 2003). After training, the neural network model was challenged for analysis of new unknown samples of similar matrices for determination of trace elements content (concentration).

5.6 Speciation analysis of simulates using K-nearest neighbors (KNN)

In this method, the spectral subset features; K_α and K_β mean intensities of Mn, Fe and Cu were utilized to develop the speciation model. For a successful training, 2 sets each with 10 samples were considered; the first set comprised 10 samples with lower ionic speciation of the trace elements; Fe^{2+} , Mn^{2+} and Cu^+ , while the second set comprised of 10 samples with higher ionic speciation of trace elements; Fe^{3+} , Mn^{4+} and Cu^{2+} . The procedure was adopted to determine the 10 nearest neighbors to the unclassified sample defined by the Euclidean distance of the training set from the unclassified sample. The Euclidean distance between the unclassified sample and every classified sample in the training set was calculated, and the majority of smallest distances from the training sets, i.e. the nearest neighbors, determined the set in which the unknown sample was assigned.

Figure 5.6.1 indicates how the KNN technique differentiated simulate samples, S6H and S6L based on their speciation of the 10 nearest neighbors in the training set.

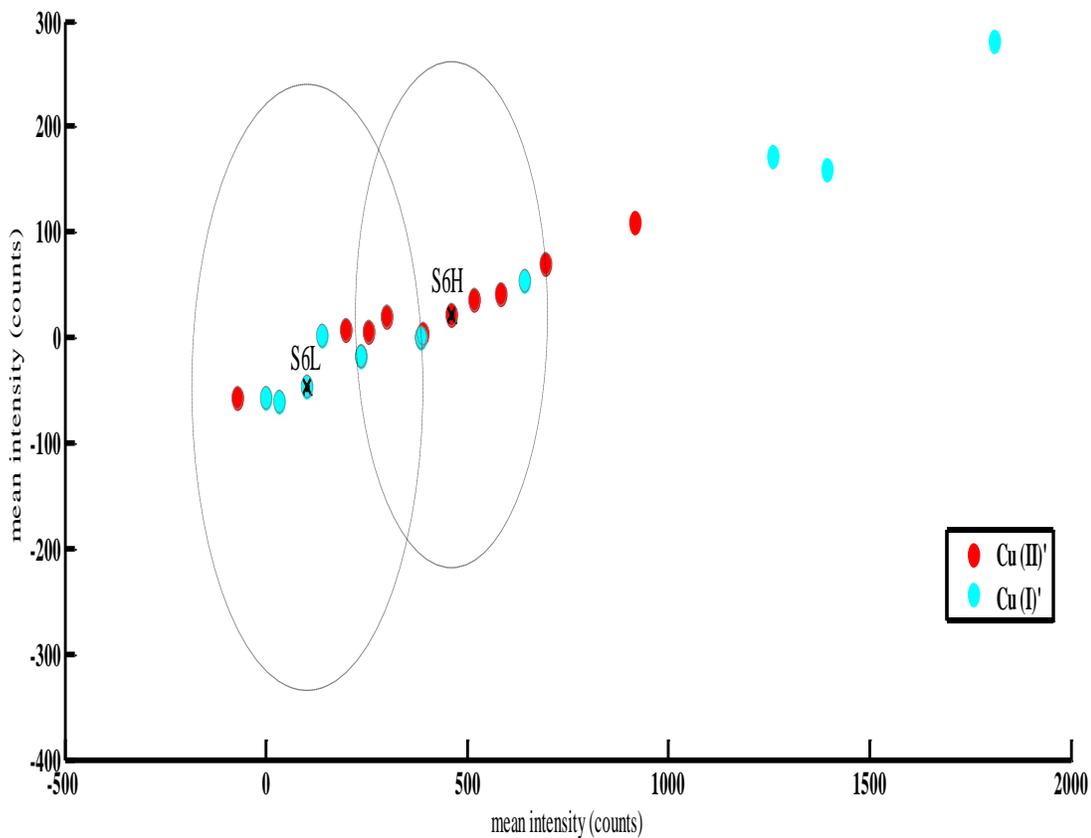


Figure 5.6.1: Speciation of Cu for S6L and S6H samples using KNN technique.

Sample S6H was sensed to have 30 % Cu^+ (3 neighbors) and 70% Cu^{2+} (7 neighbors) thus correctly predicted to have higher ionic speciation of Cu (Cu^{2+}). On the other hand, sample S6L had 60 % Cu^+ (6 neighbors) and 40 % Cu^{2+} (4 neighbors) hence predicted to have a lower ionic speciation of Cu (Cu^+).

Further, speciation analysis of Fe in simulate samples S3H and S3L were carried out using KNN with 10 neighbors as shown in Figures 5.6.2.

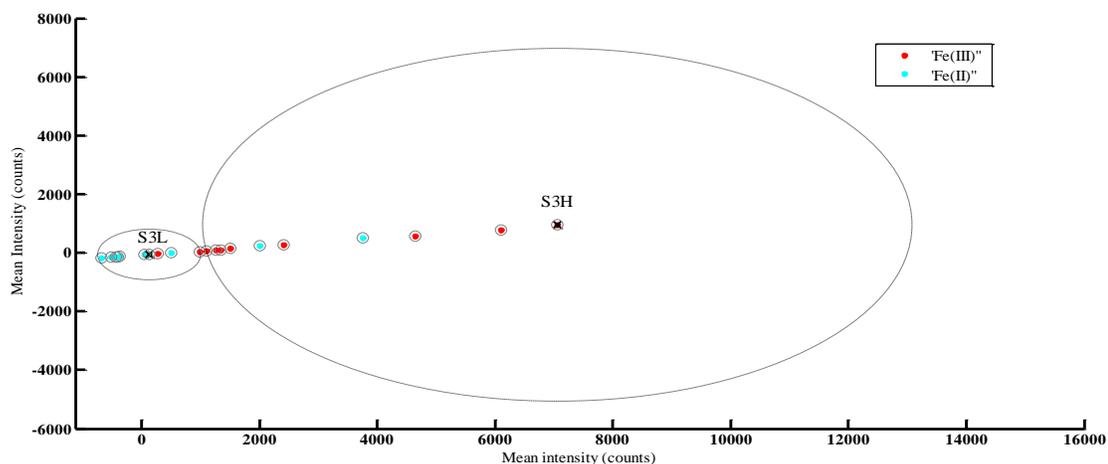


Figure 5.6.2: Speciation of Fe for S3L and S3H samples using KNN technique.

Sample S3L was found to have 80 % of Fe^{2+} while S3H had 80 % of Fe^{3+} which confirmed known chemical speciation of Fe in these two samples.

Fig. 5.6.3 indicates the speciation of Mn in simulate samples S1L and S1H in which S1L was found to have 70 % of Mn^{2+} while S1H had 60 % of Mn^{4+} which was in agreement with the known speciation of Mn in samples S1H and S1L.

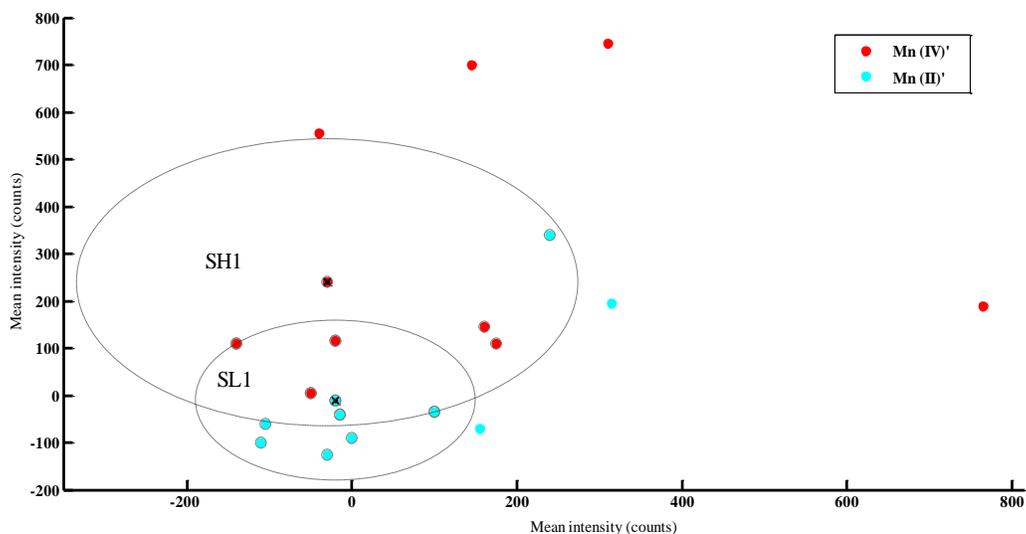


Figure 5.6.3: Speciation of Mn for S1L and S1H samples using KNN technique.

In general, the above results are in agreement with the known speciation of Cu, Mn and Fe in the simulate samples thus indicates the potential of KNN to determine the speciation of Cu, Mn and Fe in unknown tissue samples of the same matrix. The ability of the method to perform speciation analysis has enabled for tissue trace metal and speciation profiles in biological samples for possibilities to develop a cancer diagnostic tool based on trace elemental levels and speciation alterations.

5.7 Principal Component Analysis (PCA) of cultured tissues

In the present study, principal component analysis (PCA) was used to characterize the cancerous tissue samples under study. Figure 5.7.1, PCA score plot shows the transition of cancer from the early stage to late stage of development where spectra have been clustered into early and late stage cancer based on the elemental signatures of Mn, Fe, Cu, Zn and Se and speciation profiles of Fe, Cu and Mn.

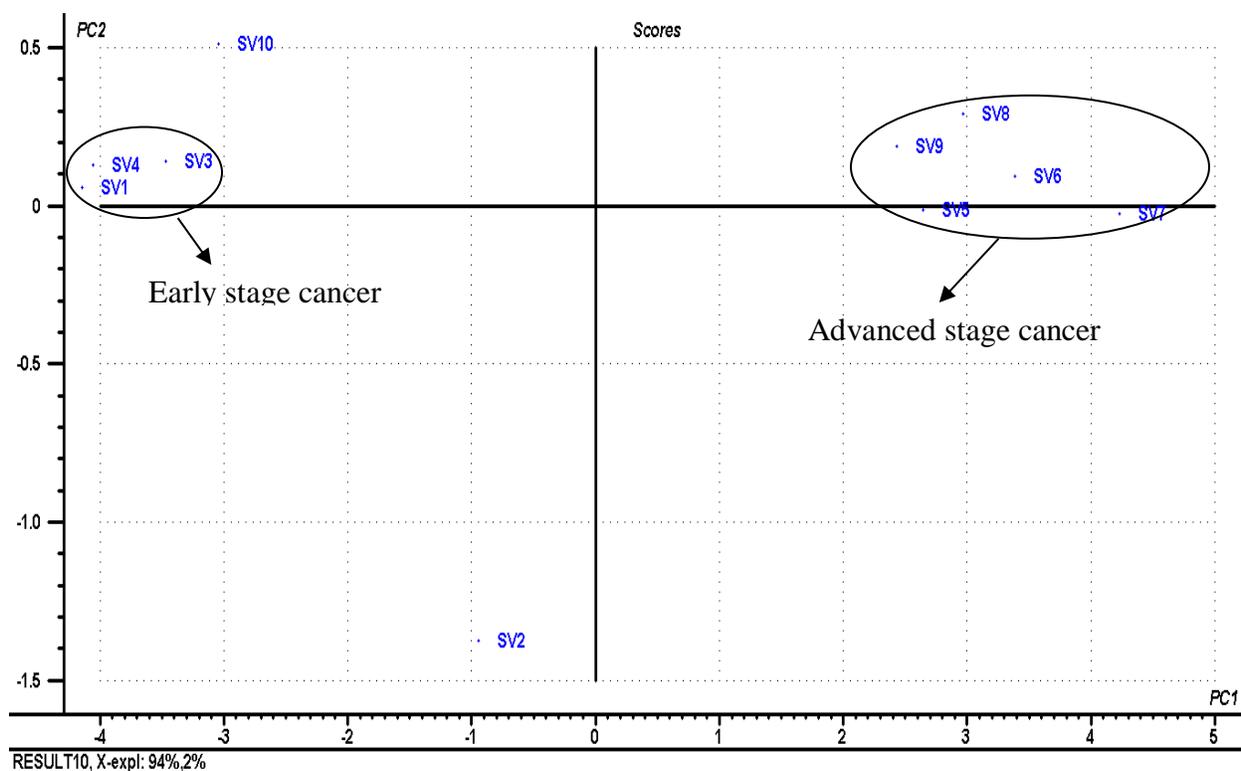


Figure 5.7.1: PCA score plot for staging of cancer from in cancer tissue cultures using selected fluorescence peaks of Mn, Fe, Cu, Zn and Se.

The entire information (variance of 96 %) was explained (94 % and 2% for Principal Components, PC1 and PC2 respectively) and the remaining 4% variation could be attributed to presence of outliers (SV2 and SV10). SV2 can be explained as a result of sample contamination during periodic feeding of the cultures in the incubator and SV10 as a result of confluence of the tissue culture where the cells clump together hence interfering with cell division.

Fig. 5.7.2 shows the variables in the loadings plot responsible for the above clusters in Fig. 5.7.1 where the trace elements especially Fe, Cu, Mn, and Zn are part of the valuable tracers for the classification of cultured cancerous tissue samples.

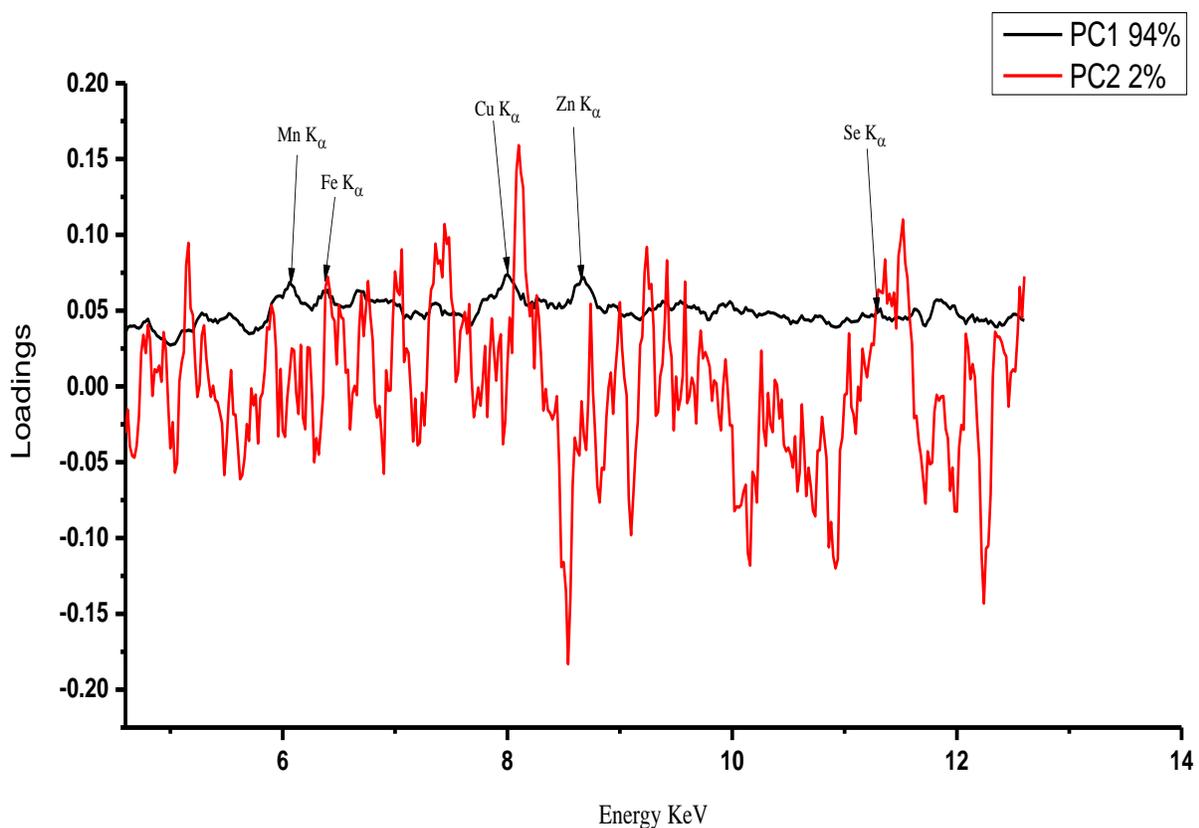


Figure 5.7.2: PCA loadings plot of cultured samples for 50s live time using partial fluorescence.

The two clusters in Fig. 5.7.1 are positively influenced by K_{α} intensity signals of Mn, Fe, Cu and Zn as shown on the PCA loadings plot in Fig. 5.7.2.

Figure 5.7.3 shows the PCA score plot of the same incubated tissue cultures using fluorescence plus Compton scatter peaks in which similar classification was evident but with reduced explained variance of 90% (PC1 84% and PC2 6% respectively). The reduced explained variance can possibly be attributed to real biochemical and structural changes in the Compton scatter which in this case could be having a negative contribution to staging of cancer.

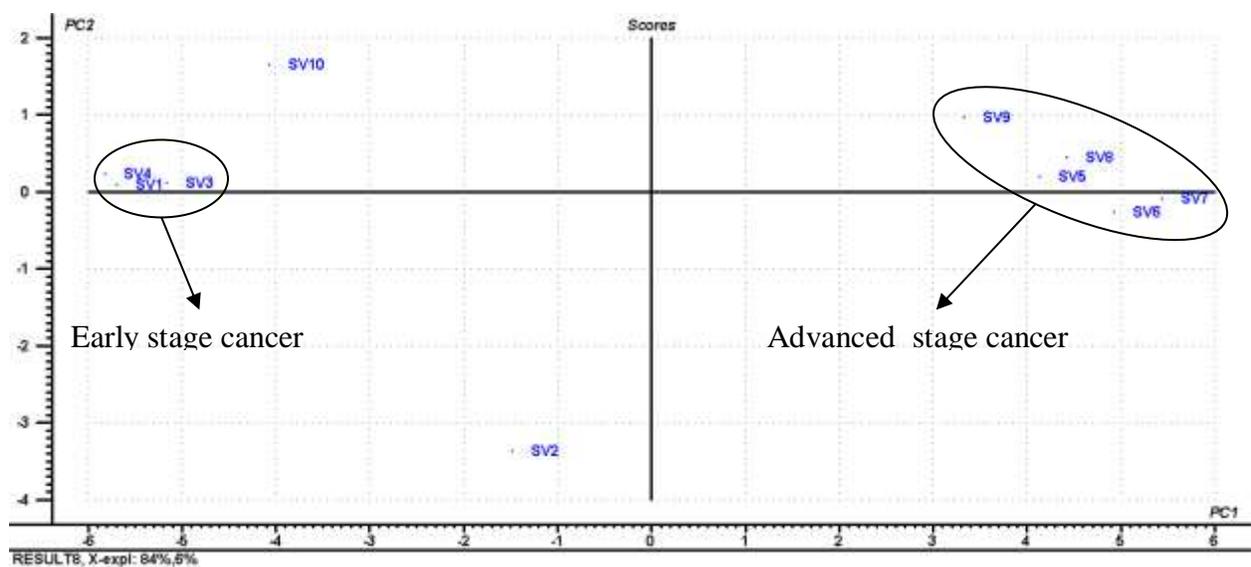


Figure 5.7.3: PCA score plot analysis for staging of cancer in cancer tissue cultures using both fluorescence plus Compton scatter.

Fig. 5.7.4 shows the variables in the loadings plot responsible for the above clusters in Fig. 5.7.3 where the trace elements (Fe, Cu, Mn, and Zn) can be traced to be responsible for the classification patterns of cultured cancerous tissue samples but with strong positive influence from the Compton scatter for PC1 clustering.

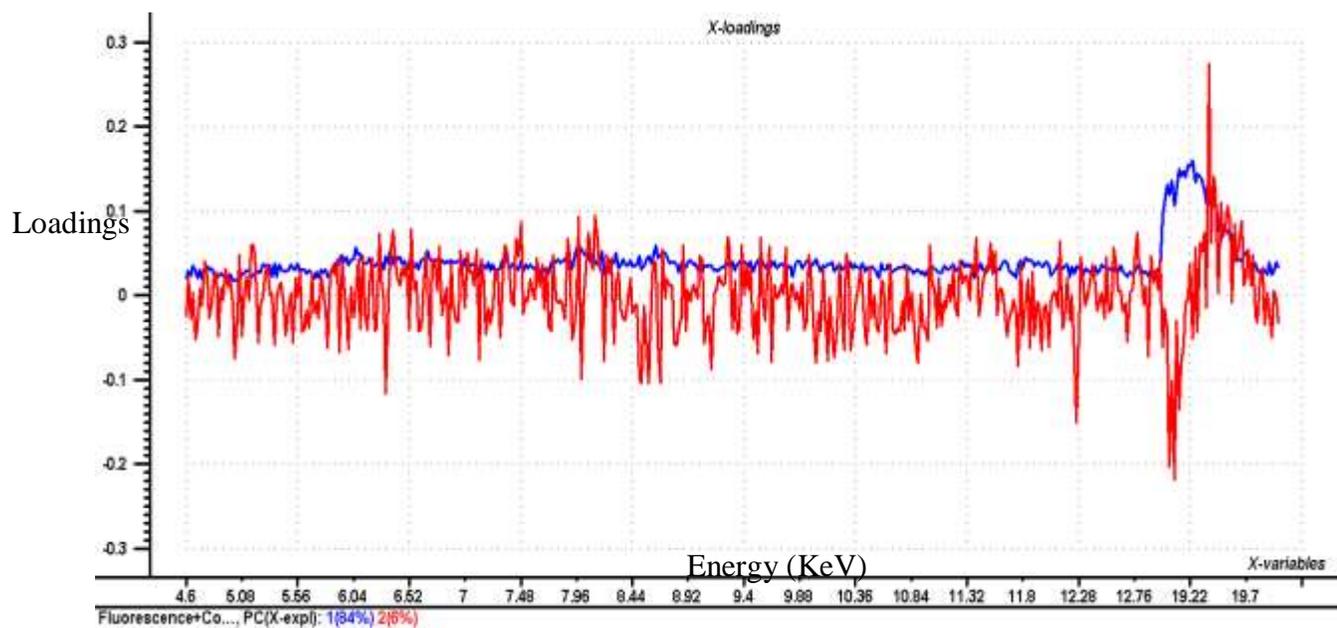


Figure 5.7.4: PCA loadings plot of cultured samples for 50s live time using both fluorescence plus Compton scatter.

5.8 Analysis of NIST Oyster tissue using ANNs

The accuracy of the developed method was evaluated through analysis of standard reference material (Oyster tissue NIST 1566b), prepared under the same conditions of the simulate tissue samples. The concentrations for the trace elements obtained in Table 5.8.1 are in good agreement ($\leq 10\%$ or less for most elements) with the certified values except for Se which can be attributed to probably low spiked concentrations that are close to the detection limit in this study.

Table 5.8.1: Evaluation of elemental concentrations in Oyster tissue (NIST 1566b)

Element	EDXRFS Measured value \pm Standard deviation (ppm)	Certified value \pm Standard deviation (ppm)	% deviation from certified values
Mn	16.9 \pm 1.2	18.5 \pm 0.2	8.6
Fe	238 \pm 15.6	205.8 \pm 6.8	4.9
Cu	66.8 \pm 4.6	71.6 \pm 1.6	6.7
Zn	1280.5 \pm 29.0	1 424 \pm 46	6.8
Se	3.6 \pm 1.1	2.06 \pm 0.15	31.1

5.9 Prediction of elemental concentrations in cultured tissue samples by ANN

Tables 5.9.1 and 5.9.2 show the predicted elemental concentration in the ANNs model based on the mean-centered and smoothed data for fluorescence and fluorescence plus Compton scatter regions.

The quantitative analysis revealed significant differences in the concentration of trace elements between normal and cancerous tissues. Cultured cancerous tissues (SV5-SV10) had elevated levels of trace elements (Fe, Cu Zn and Se) concentrations as compared to cultured healthy tissues (SL5-SL10) probably due to increased cellular activity and active enzymatic systems in malignant tissues. Based on the above models, high concentration of Fe (102.7 \pm 1.5-169.7 \pm 1.1 ppm) in cultured cancerous tissues compared to 40.3 \pm 11.4-82.9 \pm 2.6 pmm in healthy cultured tissues can be attributed to generation of free radical ions which attack cellular membranes

leading to breakage of DNA strands (Ames, 2001) thus resulting to carcinogenesis (Stephen, 2013). High concentrations of Cu (22.2 ± 1.8 - 24.9 ± 1.7 ppm) in cultured cancerous samples compared to 14.6 ± 4.7 - 19.2 ± 0.4 ppm in healthy cultured tissues could have enhanced cancer by damaging DNA with toxic free hydroxyl radicals (Theophanides and Anastassopoulou, 2002) which could lead to cell growth proliferation. Conflicting results were observed in cancerous tissues (VonCzarnowski *et al.*, 1997) in which no difference in concentration of Cu in cancerous and healthy body tissues. The observed elevated levels of Fe in cancerous tissues are consistent with previous studies (Kwiatek *et al.*, 2004; Banas *et al.*, 2001) as a result of increased cellular activity in cancerous tissues. The increased concentration of Fe in cultured cancerous tissues compared to cultured healthy tissues can be attributed to increased blood supply to the malignant tissues which can be linked to rapid cell division in a growing tumor.

Table 5.9.1: ANN predicted concentration for cultured samples using fluorescence peaks

Stage	Sample	Mean elemental concentrations \pm standard deviation (ppm)				
		Mn	Fe	Cu	Zn	Se
I	SV1	4.8 \pm 0.1	44.3 \pm 4.6	19.4 \pm 1.1	43.3 \pm 2.2	1.4 \pm 0.1
	SL1	4.5 \pm 0.0	55.3 \pm 6.5	19.2 \pm 0.4	49.8 \pm 0.6	1.7 \pm 0.1
II	SV2	4.3 \pm 0.1	130. \pm 15.2	22.6 \pm 1.3	81.9 \pm 4.8	3.2 \pm 0.5
	SL2	4.5 \pm 0.0	75.4 \pm 4.9	20.1 \pm 1.1	57.8 \pm 4.1	2.4 \pm 0.1
III	SV3	4.9 \pm 0.4	75.1 \pm 3.4	22.6 \pm 2.9	56.0 \pm 1.9	2.3 \pm 0.3
	SL3	4.0 \pm 0.4	85.3 \pm 24.9	18.9 \pm 0.3	66.4 \pm 16.2	2.2 \pm 0.6
IV	SV4	4.6 \pm 0.1	69.8 \pm 2.1	19.5 \pm 0.5	50.8 \pm 0.2	1.8 \pm 0.1
	SL4	4.3 \pm 0.2	81.9 \pm 0.8	18.6 \pm 0.3	57.1 \pm 2.5	1.9 \pm 0.1
V	SV5	4.1 \pm 0.1	169.7 \pm 1.1	24.3 \pm 0.7	94.8 \pm 6.4	4.0 \pm 0.2
	SL5	4.4 \pm 0.0	52.9 \pm 14.8	18.5 \pm 0.8	47.3 \pm 5.1	1.6 \pm 0.3
VI	SV6	3.9 \pm 0.1	160.9 \pm 7.9	22.7 \pm 0.9	92.1 \pm 4.6	3.8 \pm 0.2
	SL6	4.4 \pm 0.1	67.3 \pm 6.7	19.2 \pm 0.4	57.4 \pm 1.1	1.9 \pm 0.1
VII	SV7	3.8 \pm 0.2	146.9 \pm 3.9	22.2 \pm 1.8	82.1 \pm 1.5	3.8 \pm 0.2
	SL7	4.2 \pm 0.3	76.2 \pm 10.1	18.1 \pm 0.8	57.7 \pm 6.5	2.2 \pm 0.1
VIII	SV8	3.9 \pm 0.1	151.6 \pm 5.4	22.5 \pm 0.1	89.3 \pm 8.1	3.8 \pm 0.1
	SL8	4.4 \pm 0.1	61.2 \pm 4.3	19.3 \pm 0.5	54.6 \pm 2.7	1.9 \pm 0.1
IX	SV9	4.4 \pm 0.1	165.1 \pm 7.5	24.9 \pm 1.7	102.6 \pm 1.3	4.1 \pm 0.3
	SL9	3.6 \pm 0.0	35.5 \pm 3.3	14.6 \pm 4.7	40.2 \pm 21.3	BDL
X	SV10	4.4 \pm 0.2	102.7 \pm 17.5	20.7 \pm 0.1	59.0 \pm 4.3	2.4 \pm 0.1
	SL10	4.5 \pm 0.2	49.1 \pm 20.47	17.9 \pm 1.0	40.1 \pm 12.6	1.4 \pm 0.5

Table 5.9.2: ANN predicted concentration for cultured samples using both fluorescence and Compton scatter

Stage	Sample	Mean elemental concentrations \pm standard deviation (ppm)				
		Mn	Fe	Cu	Zn	Se
I	SV1	4.4 \pm 0.1	41.8 \pm 8.7	20.3 \pm 1.5	43.8 \pm 3.8	2.4 \pm 0.2
	SL1	4.4 \pm 0.2	53.8 \pm 11.3	18.7 \pm 0.1	55.0 \pm 8.3	1.5 \pm 0.2
II	SV2	3.8 \pm 0.7	125.9 \pm 14.4	21.1 \pm 0.3	81.9 \pm 6.8	3.2 \pm 0.7
	SL2	4.2 \pm 0.5	75.9 \pm 7.7	20.6 \pm 2.3	57.8 \pm 5.8	2.5 \pm 0.2
III	SV3	4.5 \pm 0.2	98.8 \pm 28.7	21.7 \pm 21.7	56.0 \pm 2.7	2.5 \pm 0.6
	SL3	3.9 \pm 0.7	83.8 \pm 33.3	19.2 \pm 0.8	66.3 \pm 22.8	1.9 \pm 0.5
IV	SV4	4.5 \pm 0.1	72.3 \pm 0.6	24.3 \pm 7.5	50.8 \pm 0.3	1.7 \pm 0.1
	SL4	4.4 \pm 0.2	82.9 \pm 2.6	19.1 \pm 0.3	52.1 \pm 3.5	2.0 \pm 0.1
V	SV5	3.6 \pm 0.4	159.7 \pm 15.8	24.8 \pm 1.7	93.3 \pm 6.9	3.5 \pm 0.5
	SL5	3.9 \pm 0.6	56.5 \pm 15.8	19.9 \pm 1.1	47.4 \pm 7.1	1.7 \pm 0.2
VI	SV6	3.8 \pm 0.3	160.8 \pm 11.2	23.7 \pm 0.1	87.1 \pm 13.5	3.9 \pm 0.3
	SL6	4.3 \pm 0.2	63.8 \pm 4.4	18.7 \pm 1.3	57.4 \pm 1.7	1.6 \pm 0.4
VII	SV7	3.6 \pm 0.7	148.2 \pm 7.2	22.2 \pm 2.6	86.1 \pm 3.6	3.8 \pm 0.4
	SL7	3.9 \pm 0.8	75.2 \pm 12.9	18.1 \pm 1.1	52.7 \pm 2.1	2.6 \pm 0.5
VIII	SV8	3.9 \pm 0.1	156.3 \pm 14.3	23.9 \pm 2.2	87.3 \pm 8.6	4.3 \pm 0.8
	SL8	4.6 \pm 0.3	59.2 \pm 9.0	19.2 \pm 0.5	54.6 \pm 3.8	1.9 \pm 0.1
IX	SV9	4.2 \pm 0.1	167.6 \pm 14.2	24.9 \pm 2.5	107.6 \pm 5.3	3.9 \pm 0.2
	SL9	3.4 \pm 0.2	40.3 \pm 11.4	18.1 \pm 1.7	40.2 \pm 30.2	1.7 \pm 0.3
X	SV10	4.3 \pm 0.1	100.7 \pm 27.5	24.2 \pm 5.1	59.0 \pm 6.1	2.2 \pm 0.5
	SL10	4.3 \pm 0.1	51.5 \pm 25.5	16.23 \pm 3.8	40.1 \pm 17.9	1.5 \pm 0.6

Cancerous cultured samples (SV5, SV6, SV7, SV8, SV9 and SV10) at late stage of development had elevated concentration of Zn (82.1 \pm 1.5-102.6 \pm 1.3 ppm) compared to corresponding non-cancerous cultured samples (40.1 \pm 12.6-57.7 \pm 6.5 ppm). This may be attributed to demands for excessive cell division in rapidly growing tumors (Ng *et al.*, 1997). Sample SV10 had low concentrations of Fe (102.70 \pm 17.48 ppm), Zn (59.01 \pm 4.30 ppm) and Cu (20.69 \pm 0.12 ppm) with alterations in Mn and Se levels. This can be explained based on the lag phase stage of cancer development which does not require a lot of Fe, Cu and Zn due to reduced cell division and proliferations in mature cancer tissues. The elevated concentration levels of Cu, Fe and Zn in cancerous tissues can be attributed to the physiological process and increased enzymatic activities in tumors. The non-cancerous cultured samples had lower concentration levels of Fe, Cu and Zn which can be attributed to normal biochemical processes in the tissues.

Mn and Se concentrations were almost at the same concentration levels in both cancerous and non-cancerous cultured samples although showing alterations from one stage of development to another.

Figure 5.9.1 shows how the concentration of Cu, Fe, Se, Zn and Mn are varying with staging of cancer and also for normal growth of healthy tissues. Fe, Cu and Zn show similar pattern for early staging of cancer (SV1-SV4) and also for advanced stage cancer (SV5-SV9) with alteration in concentration of Mn and Se for all the stages of cancer. This may possibly explain the observed clustering in PCA results early and late stage of cancer development.

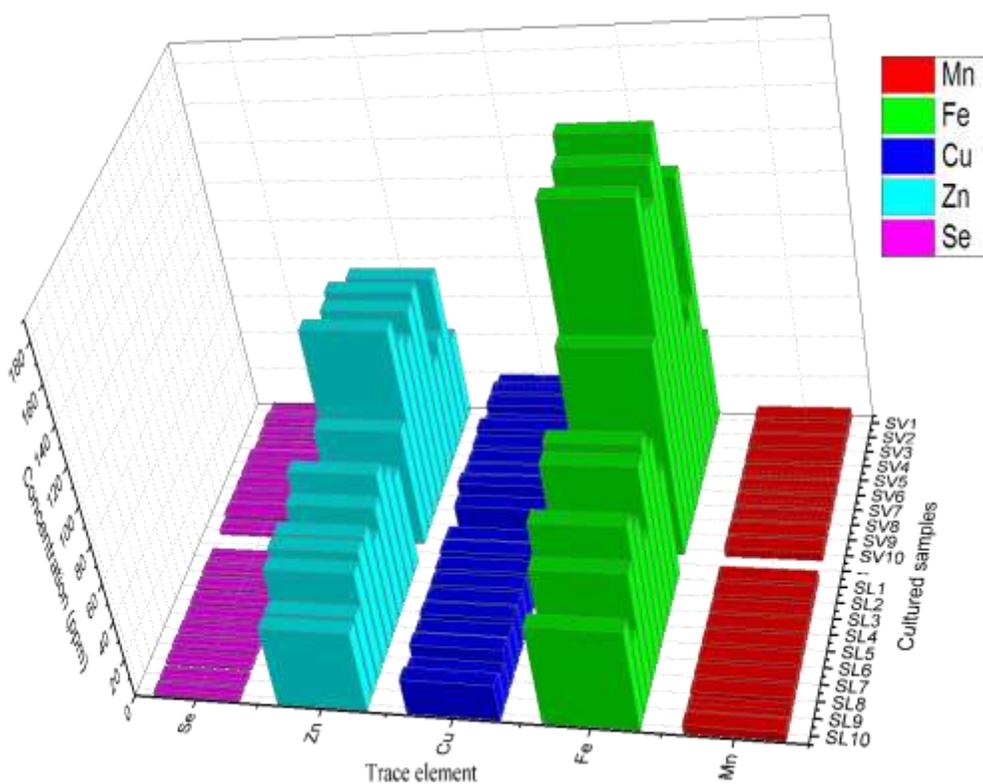


Figure 5.9.1: 3-D Grouped bar graph showing the variation in concentration of trace elements in cultured cancerous and healthy tissues.

In general, the concentration of Fe, Cu and Zn are higher in cancerous tissues as compared to the corresponding normal cultured tissues. This may be attributed to increased cell division and proliferations in cancerous tissues as compared to normal healthy tissues.

5.9.1 Interrelationships between trace elements and biological processes

In order to better understand the associations between trace elements and biological processes related to tumor prognosis (tumor growth and metastasis) and differences between trace elemental concentrations in malignant tissue, correlation tests were carried out on the concentration levels of Mn, Fe, Cu, Zn and Se in cultured tissue samples. Pearson correlation coefficients between the determined elemental concentrations in cancerous tissues are as presented in Table 5.9.1.1 together with the corresponding P-value which shows how the trace elements vary significantly with respect to a property being analyzed (concentrations of trace elements).

Table 5.9.1.1: Correlation coefficients of elemental concentrations in cultured cancer tissues

Element	Pearson correlation coefficients between elemental concentrations (and their corresponding P values)				
	Mn	Fe	Cu	Zn	Se
Mn	1				
Fe	-0.8470 (P<0.01)	1			
Cu	-0.425 (0.22)	0.827 (P<0.01)	1		
Zn	-0.764 (0.01)	0.979 (P<0.01)	0.879(P<0.01)	1	
Se	-0.813 (P<0.01)	0.988 (P<0.01)	0.868(P<0.01)	0.981(P<0.01)	1

Strong positive correlations between Fe and other trace elements were evident, but a negative correlation discerned between Mn and all other elements noted. The observed strong positive Pearson correlation between Fe and Cu (0.827) in cultured neoplastic tissues is consistent with the fact that these elements collectively contribute to angiogenesis process in which new blood vessels are formed thus leading to tumor growth and metastasis (Nasulewicz *et al.*, 2004). It further emphasizes the function of Cu as an important cofactor for metabolism of Fe in tissues.

Furthermore, the observed correlation between Fe and Cu and their significant positive association support the hypothesis that there is a connection with both increased cellular activity and blood supply in the formation of neoplasies in tissues (Richardson, 2002). The even stronger positive correlation between Fe and Zn (0.979) can be linked to increase in metabolic activities and angiogenic process in malignant tissues (Kubala-Kukus *et al.*, 2004). Additionally, Se showed strong positive correlations with Fe (0.988), Cu (0.868) and Zn (0.981) in cultured cancerous samples which can be attributed to its essential role in carcinogenesis (Banas *et al.*, 2001).

The results in Table 5.9.1.1 above indicate statistically significant positive association ($p < 0.05$) between Fe and Cu, Fe and Zn, and, Fe and Se in neoplastic cultured tissues. These clearly indicate the importance of these elements in biochemical processes associated with development of cancer as a result of increased cellular activity in cancerous tissues (Kwiatek *et al.*, 2004).

Table 5.9.1.2 shows correlation of elemental concentrations in healthy cultured tissues in which positive correlations between trace elements were evident. This can be attributed to normal physiological body tissue functions (Rose, 1983; Prasad, 1976). The positive correlation between Fe and Cu (0.64) can be linked to normal biochemical processes in healthy tissues as plays an important role in angiogenesis as the main essential enzyme regulator and an important co-factor for metabolism of Fe (Nasulewicz *et al.*, 2004).

Table 5.9.1.2: Correlation coefficients of elemental concentrations in healthy cultured tissues

Element	Pearson correlation coefficients between elemental concentrations (and their corresponding P values)				
	Mn	Fe	Cu	Zn	Se
Mn	1				
Fe	0.251 (0.485)	1			
Cu	0.824 (P<0.01)	0.640 (0.046)	1		
Zn	0.126 (0.730)	0.932 (P<0.01)	0.631 (0.051)	1	
Se	0.293 (0.411)	0.925 (P<0.01)	0.706 (0.022)	0.913(P<0.01)	1

The positive correlation between Mn and other elements may possibly indicate the anti-carcinogenic nature of Mn which needs further speciation analysis to ascertain the exact role of Mn in carcinogenesis.

5.10 Prediction of concentration of trace elements in domestic dog tissues by ANN

The elemental concentrations in domestic dog tissues were obtained by using the validated ANN model. As clearly indicated in Tables 5.10.1, it was possible to detect and quantify the following trace elements: Fe, Cu, Zn, Mn and Se in which statistically significant differences and alterations in concentration in cancerous tissues were observed.

Table 5.10.1: Concentrations of trace elements in histopathologically classified dog tissue samples analyzed using fluorescence peaks

Cancerous Samples	Sample name	Mean elemental concentrations \pm standard deviation (ppm)				
		Mn	Fe	Cu	Zn	Se
Early stage prostate cancer	SP1B	3.9 \pm 0.3	167.5 \pm 3.6	21.9 \pm 0.7	94.4 \pm 2.6	3.7 \pm 0.1
	SP43	3.6 \pm 0.3	129.7 \pm 35.9	17.4 \pm 5.4	74.5 \pm 26.4	2.9 \pm 1.2
	SP45	3.7 \pm 0.3	154.0 \pm 4.5	21.1 \pm 1.1	87.7 \pm 6.2	3.6 \pm 0.1
	SP20	3.7 \pm 0.4	168.9 \pm 4.7	20.9 \pm 3.5	94.2 \pm 5.5	3.8 \pm 0.3
	SP13	3.3 \pm 0.4	158.7 \pm 12.1	21.2 \pm 0.8	95.2 \pm 2.9	3.8 \pm 0.1
	SP24	3.9 \pm 0.1	170.1 \pm 3.7	23.0 \pm 0.5	93.6 \pm 4.6	4.0 \pm 0.1
	SP66	4.1 \pm 0.2	166.3 \pm 3.3	23.2 \pm 3.9	98.9 \pm 10.1	4.0 \pm 0.5
Late stage prostate cancer	SP35	3.8 \pm 0.1	165.4 \pm 0.2	21.9 \pm 0.3	96.9 \pm 5.2	3.9 \pm 0.2
	SP5B	3.6 \pm 0.6	161.1 \pm 5.9	19.8 \pm 3.6	89.8 \pm 3.8	3.5 \pm 0.1
	SP7B	3.9 \pm 0.5	161.9 \pm 8.9	23.4 \pm 3.2	97.3 \pm 6.8	3.9 \pm 0.5
	SP3B	4.1 \pm 0.3	167.6 \pm 6.5	23.7 \pm 3.1	99.5 \pm 4.7	4.3 \pm 0.1
Advanced mammary cancer	SP4B	3.3 \pm 0.2	171.1 \pm 3.9	21.3 \pm 2.2	92.6 \pm 6.1	4.4 \pm 0.3
	SP29	3.8 \pm 0.2	191.2 \pm 9.5	23.7 \pm 3.1	99.5 \pm 4.7	4.9 \pm 0.1
	SP6B	4.7 \pm 0.5	184.9 \pm 6.8	19.8 \pm 3.6	89.8 \pm 3.8	5.1 \pm 0.7
	SP42	4.1 \pm 0.4	175.1 \pm 3.9	23.2 \pm 3.9	98.9 \pm 10.1	4.5 \pm 0.1
	SP2B	3.3 \pm 0.3	152.9 \pm 10.3	21.9 \pm 0.3	96.9 \pm 5.2	3.4 \pm 0.9

The results in Table 5.10.1 above indicate elevated concentrations of trace elements (Fe, Cu and Zn) in neoplastic tissues that could strongly be affected by the severity of cancer. The above levels of Fe, Cu and Zn in cancerous tissues are consistent with previous studies (Carvalho *et al.*, 2007) as a result of increased cellular activity in cancerous tissues. Neoplastic tissues at an early

stage of development were found to have elevated concentration of Fe due to increased blood supply which results to rapid cell growth and proliferation that calls for increased requirement for DNA synthesis that is accompanied by increased expression of the Fe-dependent enzyme ribonucleotide reductase (RR) (Elford *et al.*, 1970; Jessica *et al.*, 2013) for tumor angiogenesis.

This study indicates accumulation of Zn in cancerous samples which is important to cell proliferation which supports tumor growth (Sukumar *et al.*, 1983; Lee *et al.*, 2003). The uncontrolled growth of cells in the tumor tissues tends to accumulate more zinc to cope with the demands of excessive dividing of the cells.

High concentrations of Cu in cancerous samples were evident. The mechanism of Cu elevation in cancerous tissues can be explained by increased generation of free radicals in Fenton reaction and/or by interference with DNA repair processes (Theophanides and Anastassopoulou, 2002). Conflicting results regarding concentrations of Cu have been observed in cancerous and normal tissues (VonCzarnowski *et al.*, 1997) where there were no significant differences in concentrations of Cu between cancerous and normal stomach tissue samples. The deficiency of Cu as cofactors of Fe in enzymes could severely impair the host's resistance against carcinogenic stress (Schrauzer, 1980).

In general, most of the above trace elements are very important for various biological and enzymatic processes thus the changes in concentrations of these elements (Mn, Fe, Cu, Zn and Se) in cancerous tissue are realistic. This strongly supports the fact that the trace elements can be used as biomarkers for diagnosis of cancer.

The associations between trace elements and biological processes related to tumor growth and metastasis in selected dog tissues were determined using correlation tests between trace elements (Mn, Fe, Cu, Zn and Se) concentrations in malignant tissues. Pearson correlation coefficients

between the pre-determined elemental concentrations in cancerous tissues are as presented in Tables 5.10.2-4.

Table 5.10.2: Pearson correlation coefficients of elemental concentrations in prostate tissue samples for early stage cancer

Element	Pearson correlation coefficients between elemental concentrations and their corresponding P values				
	Mn	Fe	Cu	Zn	Se
Mn	1				
Fe	0.488 (0.220)	1			
Cu	0.612 (0.107)	0.904 (P<0.01)	1		
Zn	0.317 (0.366)	0.928 (P<0.01)	0.903 (P<0.01)	1	
Se	0.469 (0.242)	0.927 (P<0.01)	0.974 (P<0.01)	0.946 (P<0.01)	1

Table 5.10.3: Pearson correlation coefficients of elemental concentrations in prostate tissue samples for advanced stage cancer

Element	Pearson correlation coefficients between elemental concentrations and (their corresponding P values)				
	Mn	Fe	Cu	Zn	Se
Mn	1				
Fe	-0.283 (0.717)	1			
Cu	0.832 (0.168)	0.174 (0.826)	1		
Zn	0.874 (0.126)	0.174 (0.826)	0.989 (0.011)	1	
Se	0.005 (0.995)	0.0927(0.073)	0.507 (0.493)	0.482 (0.518)	1

Table 5.10.4: Pearson correlation coefficients of elemental concentrations in mammary tissue samples for advanced stage cancer

Element	Pearson correlation coefficients between elemental concentrations and (their corresponding P values)				
	Mn	Fe	Cu	Zn	Se
Mn	1				
Fe	0.675 (0.3425)	1			
Cu	-0.535 (0.465)	0.085 (0.915)	1		
Zn	-0.667 (0.334)	-0.108 (0.892)	0.981 (0.019)	1	
Se	0.824 (0.176)	0.0974 (0.026)	-0.109 (0.891)	-0.296 (0.704)	1

At an early stage of development, the Pearson correlation coefficient between Fe and Cu was found to be 0.904 with a corresponding P value (<0.01) which can be attributed to the

exponential phase of cell division and proliferations in tissues as a result of increased cellular activity (Raju *et al.*, 2006). As cancer progressed to an advanced stage, the Pearson correlation coefficient between Zn and Cu increased to 0.981, which can be linked to the lag phase stage of cell growth in the tissue as Zn and Cu tends to lower important physiological functions.

The stronger positive correlation coefficient between Mn and Se was evident (0.824) in cancerous mammary tissues as compared to prostate cancerous tissues (0.005). These may be probably attributed to Mn as a co-factor of Se in cancerous mammary tissues thus resulting to increased metabolic activities for carcinogenesis of cancer.

The most significant differences in elemental associations between prostate and mammary advanced cancer may possibly be based on the strong positive correlation between Mn and Cu (0.832) and, Mn and Zn (0.874) in prostate tumors with corresponding negative correlations in mammary tumors; Mn and Cu (-0.535), Mn and Zn (-0.667). However, there seems to exist strong positive correlations between Fe and Se, Cu and Zn in both prostate and mammary advanced cancer tissues. This can be linked to their co-factor roles to advanced stage of development of cancer.

5.11 Speciation analysis of cancerous tissues using KNN

The determination of the oxidation state of Fe, Cu and Mn was aimed to give more insight in the mechanism of cancer development. Simulate validated KNN multivariate technique was utilized for determination of chemical speciation of trace elements (Fe, Cu and Mn) in cancerous tissues (cultured and dog tissues) to elucidate their role in progression of cancer.

5.11.1 Results for cultured cancerous tissues

Figure 5.11.1.1 indicates how the KNN technique was utilized to determine the chemical speciation of Fe in sample SV1 based on the 10 nearest neighbors in the training set. SV1 was found to have 80 % Fe^{2+} (8 neighbors) and 20% Fe^{3+} (2 neighbors) thus predicted to have lower ionic speciation of Fe (Fe^{2+}).

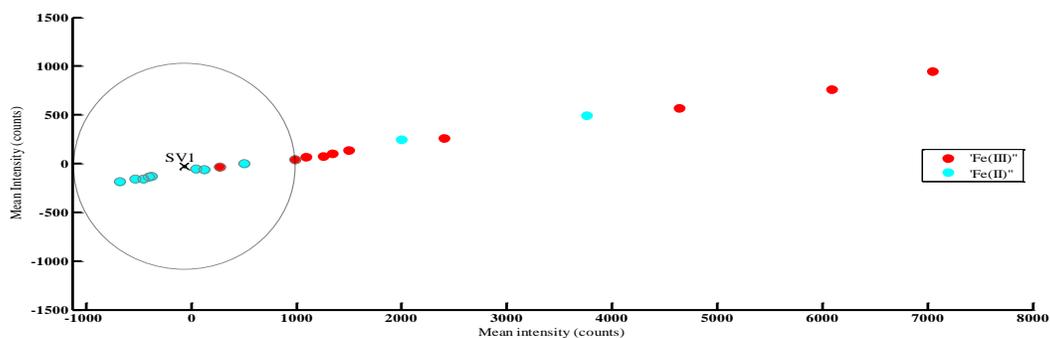


Figure 5.11.1.1: Speciation of Fe in typical sample SV1 using KNN technique.

Fig. 5.11.1.2 indicates that SV7 was found to have 70 % of Mn^{4+} and 30 % of Mn^{2+} thus predicted to have higher speciation of Mn.

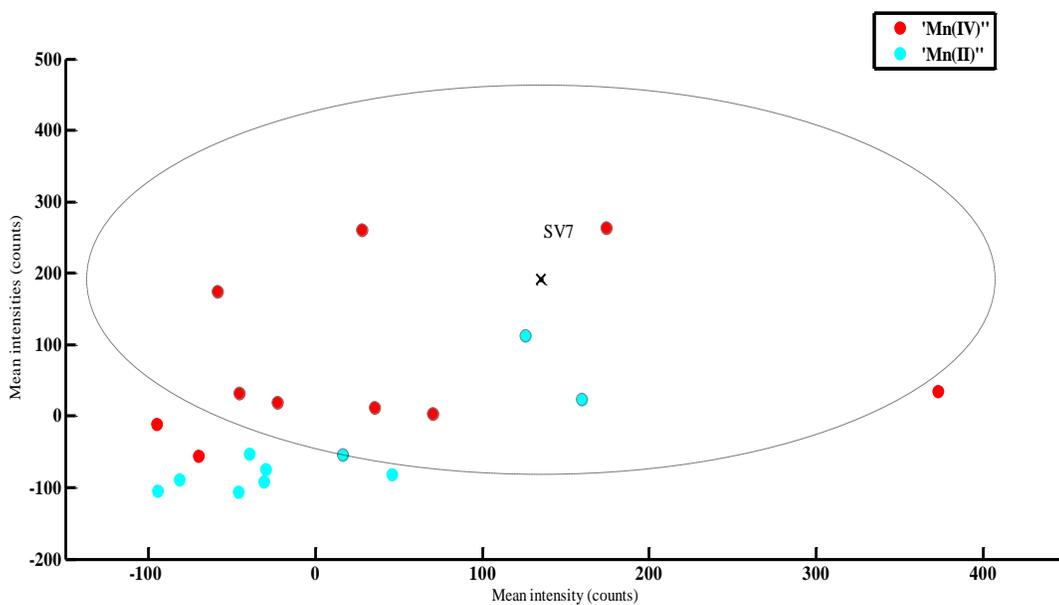


Figure 5.11.1.2: Speciation of Mn in sample SV7 using KNN technique.

Fig. 5.11.1.3 shows simultaneous determination of speciation of Cu and Fe in sample SV1 which has 40 % Cu^+ (4 neighbors) and 60% Cu^{2+} (6 neighbors), 80 % Fe^{2+} (8 neighbors) and 20% Fe^{3+} (2 neighbors) thus predicted to have higher and lower ionic speciation of Cu and Fe respectively.

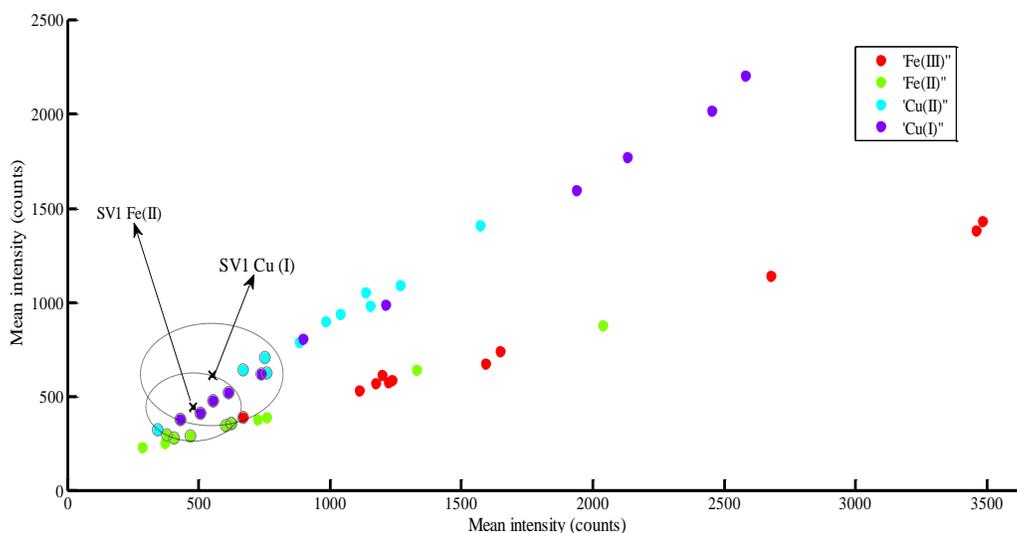


Figure 5.11.1.3: Speciation of Fe and Cu in sample SV1 using KNN technique.

The results obtained from KNN analysis of cultured samples (cancerous) in Tables 5.11.1 indicates direct determination of chemical speciation of Fe, Mn and Cu in cancerous tissues.

Table 5.11.1: KNN results for speciation of Fe, Cu and Mn in cultured cancer samples

Sample	Element					
	Fe		Cu		Mn	
	Fe^{2+} %	Fe^{3+} %	Cu^+ %	Cu^{2+} %	Mn^{2+} %	Mn^{4+} %
SV1	80	20	60	40	70	30
SV2	80	20	60	40	40	60
SV3	80	20	60	40	40	60
SV4	80	20	60	40	50	50
SV5	70	30	40	60	20	80
SV6	70	30	40	60	20	80
SV7	70	30	40	60	30	70
SV8	70	30	50	50	30	70
SV9	70	30	40	60	30	70
SV10	80	20	60	40	50	50

The cultured cancer tissues were rich in Fe, Cu and Mn occurring mostly in the following oxidation states; Fe^{2+} , Cu^{2+} and Mn^{4+} which demonstrated the possible production of free radicals in cancerous tissues in Fenton reaction based on the oxidation states of Cu and Mn (Rojas *et al.*, 1999). Due to the ability of Cu to change between its two oxidation states; Cu^+ and Cu^{2+} , Cu has been found to cause generation of Reactive Oxygen Species (ROS) which produce hydroxyl radicals that modify proteins, lipids, and nucleic acids thus leading to DNA damage (Armendariz and Vulpe, 2003). The oxidation states Cu and Mn clearly elucidate the mechanism of damage to genetic material in cells by the chemically active oxygen-containing molecules (free radicals). The results conflicts with an earlier study using XANES (Kwiatek *et al.*, 2004), in which the cancerous tissue were found to be dominated by Fe^{3+} .

The KNN results of cultured normal tissue samples in Tables 5.11.2 show direct determination of speciation of Fe, Mn and Cu in normal cultured tissues.

Table 5.11.2: KNN results for speciation of Fe, Cu and Mn in cultured healthy tissues

Sample	Element					
	Fe		Cu		Mn	
	Fe^{2+} %	Fe^{3+} %	Cu^+ %	Cu^{2+} %	Mn^{2+} %	Mn^{4+} %
SL1	80	20	60	40	50	50
SL2	80	20	60	40	50	50
SL3	80	20	60	40	50	50
SL4	80	20	60	40	50	50
SL5	80	20	60	40	50	50
SL6	80	20	60	40	50	50
SL7	80	20	60	40	50	50
SL8	80	20	60	40	50	50
SL9	40	60	60	40	60	40
SL10	70	30	60	40	50	50

The incubated normal tissue samples were rich in Fe, Cu and Mn occurring mostly in the following oxidation states; Fe^{2+} and Cu^+ with Mn occurring in the ratio 1:1 for Mn^{2+} : Mn^{4+} respectively. In general, there is no change in speciation of Fe, Cu and Mn for all stages of

growth of a normal cultured tissue which may be attributed to normal biochemical processes in healthy tissues.

The proposed method was able to distinguish samples containing just one of these species using spectral region of interest of Mn, Fe and Cu. Based on the above speciation results in Table 5.11.1, the elemental concentrations of Mn, Fe and Cu in their respective oxidation states in cancerous samples were obtained as shown in Table 5.11.2 based on the fluorescence peak concentrations of Mn, Fe and Cu reported in Table 5.9.1.

Table 5.11.3: Speciation of Fe, Cu and Mn in cultured cancer tissue

Sample	Mean elemental concentrations \pm standard deviation (ppm)					
	Mn		Fe		Cu	
	Mn^{2+}	Mn^{4+}	Fe^{2+}	Fe^{3+}	Cu^{+}	Cu^{2+}
SV1	3.4 \pm 0.08	1.44 \pm 0.04	35.42 \pm 3.70	8.86 \pm 1.00	11.62 \pm 0.66	7.75 \pm 0.44
SV2	1.73 \pm 0.01	2.60 \pm 0.01	104.76 \pm 12.18	26.19 \pm 3.04	13.55 \pm 0.77	9.04 \pm 0.52
SV3	1.98 \pm 0.14	2.97 \pm 0.22	60.06 \pm 2.74	15.01 \pm 0.69	13.58 \pm 1.71	9.05 \pm 1.14
SV4	2.28 \pm 0.04	2.28 \pm 0.04	55.83 \pm 1.67	13.96 \pm 0.42	11.67 \pm 0.32	7.78 \pm 0.21
SV5	0.81 \pm 0.03	3.25 \pm 0.11	118.79 \pm 0.80	50.91 \pm 0.34	9.73 \pm 0.29	14.59 \pm 0.43
SV6	0.79 \pm 0.01	3.17 \pm 0.02	112.60 \pm 5.53	48.26 \pm 2.37	9.08 \pm 0.38	13.61 \pm 0.57
SV7	1.17 \pm 0.05	2.72 \pm 0.12	102.87 \pm 2.72	44.09 \pm 1.17	8.88 \pm 0.73	13.32 \pm 1.09
SV8	1.18 \pm 0.02	2.75 \pm 0.04	106.09 \pm 3.77	45.47 \pm 1.61	11.25 \pm 0.04	11.25 \pm 0.04
SV9	1.31 \pm 0.04	3.05 \pm 0.08	115.57 \pm 5.25	49.53 \pm 2.25	10.00 \pm 0.69	14.99 \pm 1.04
SV10	2.22 \pm 0.10	2.22 \pm 0.10	82.16 \pm 13.98	20.54 \pm 3.50	12.41 \pm 0.07	8.28 \pm 0.05

5.11.2 Results for domestic dog tissues

Figure 5.11.2.1 indicates how the KNN technique was utilized to determine the chemical speciation of Mn in sample SP4B based on the 10 nearest neighbors in the training set. SP4B was found to have 50 % Mn^{2+} (5 neighbors) and 50% Mn^{4+} (5 neighbors) thus predicted to have both lower and higher ionic speciation of Mn.

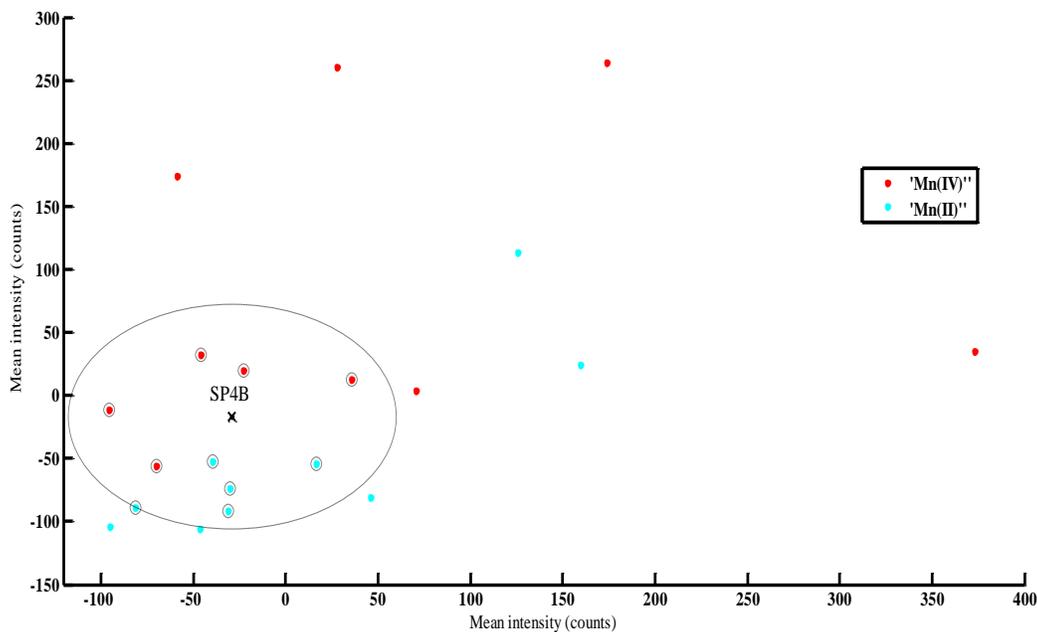


Figure 5.11.2.1: Speciation of Mn in sample SP4B using KNN technique.

The results of speciation analysis of Fe, Cu and Mn obtained using KNN analysis of domestic dog tissue samples are presented in Table 5.11.2.1.

Table 5.11.2.1: KNN results for speciation of Fe, Cu and Mn in selected dog tissues

Cancerous Samples	Element						
	Fe		Cu		Mn		
	Fe^{2+} %	Fe^{3+} %	Cu^{+} %	Cu^{2+} %	Mn^{2+} %	Mn^{4+} %	
Early prostate cancer	SP1B	80	20	60	40	40	60
	SP43	80	20	60	40	50	50
	SP45	80	20	60	40	40	60
	SP20	80	20	40	60	30	70
	SP13	80	20	60	40	30	70
	SP24	80	20	60	40	30	70
	SP66	80	20	60	40	50	50
SP35	80	20	60	40	30	70	
Advanced prostate cancer	SP5B	80	20	60	40	30	70
	SP7B	80	20	60	40	60	40
	SP3B	80	20	60	40	40	60
	SP4B	80	20	60	40	50	50
Advanced mammary cancer	SP29	80	20	30	70	30	70
	SP6B	80	20	30	70	20	80
	SP42	80	20	20	80	20	80
	SP2B	80	20	20	80	20	80

The elemental concentrations of Mn, Fe and Cu in their respective oxidation states in dog tissue samples were obtained as shown in Table 5.11.2.2.

Table 5.11.2.2: Speciation concentrations of Fe, Cu and Mn in selected dog tissue samples

Sample	Element concentration \pm standard deviation (ppm)					
	Fe		Cu		Mn	
	Fe^{2+}	Fe^{3+}	Cu^+	Cu^{2+}	Mn^{2+}	Mn^{4+}
SP75	136.29 \pm 0.40	34.07 \pm 0.10	7.29 \pm 0.48	17.12 \pm 1.13	0.74 \pm 0.03	2.97 \pm 0.14
SP66	133.06 \pm 2.67	33.26 \pm 0.67	13.91 \pm 2.33	9.28 \pm 1.55	2.04 \pm 0.09	2.04 \pm 0.09
SP45	123.23 \pm 3.62	30.81 \pm 0.91	12.66 \pm 0.66	8.44 \pm 0.44	1.49 \pm 0.11	2.24 \pm 0.17
SP43	103.78 \pm 28.69	25.94 \pm 7.17	10.42 \pm 3.22	6.95 \pm 2.15	1.80 \pm 0.17	1.80 \pm 0.17
SP42	140.10 \pm 3.12	35.03 \pm 0.78	4.64 \pm 0.78	18.55 \pm 3.10	0.82 \pm 0.07	3.30 \pm 0.28
SP35	132.36 \pm 0.18	33.09 \pm 0.04	13.16 \pm 0.16	9.59 \pm 0.10	1.15 \pm 0.02	2.67 \pm 0.04
SP32	122.00 \pm 10.72	30.50 \pm 2.68	14.38 \pm 1.18	9.59 \pm 0.79	2.10 \pm 0.12	2.10 \pm 0.12
SP29	152.94 \pm 7.57	38.23 \pm 1.89	7.12 \pm 0.93	16.62 \pm 2.17	1.15 \pm 0.05	2.67 \pm 0.13
SP24	136.10 \pm 2.95	34.02 \pm 0.74	13.82 \pm 0.28	9.21 \pm 0.19	1.19 \pm 0.02	2.79 \pm 0.06
SP20	135.11 \pm 3.78	33.78 \pm 0.94	8.35 \pm 1.40	12.52 \pm 2.09	1.10 \pm 0.12	2.57 \pm 0.27
SP13	126.92 \pm 9.66	31.73 \pm 2.41	12.71 \pm 0.50	8.48 \pm 0.33	1.00 \pm 0.13	2.33 \pm 0.31
SP7B	129.56 \pm 7.18	33.19 \pm 1.80	14.03 \pm 1.90	9.36 \pm 1.26	2.38 \pm 0.27	1.58 \pm 0.18
SP6B	147.90 \pm 5.43	36.97 \pm 1.36	5.94 \pm 1.07	13.86 \pm 2.51	0.95 \pm 0.10	3.77 \pm 0.42
SP5B	128.85 \pm 4.74	32.21 \pm 1.19	11.88 \pm 2.15	7.92 \pm 1.32	1.06 \pm 0.12	2.49 \pm 0.43
SP4B	136.90 \pm 3.14	34.22 \pm 0.78	12.79 \pm 1.30	8.52 \pm 0.86	1.68 \pm 0.12	1.68 \pm 0.12
SP3B	134.08 \pm 5.18	33.52 \pm 1.29	14.24 \pm 1.86	9.50 \pm 1.24	1.64 \pm 0.12	2.45 \pm 0.18
SP2B	122.32 \pm 8.26	30.58 \pm 2.06	4.39 \pm 0.05	17.54 \pm 0.21	0.65 \pm 0.06	2.62 \pm 0.24
SP1B	134.00 \pm 2.84	33.5 \pm 0.71	13.16 \pm 0.40	8.77 \pm 0.26	1.60 \pm 0.14	2.39 \pm 0.20

The ability of the method to perform speciation analysis has enabled us to develop an approach for spectral diagnosis of cancer utilizing tissue trace metal and speciation profiles (and their multivariate alterations) as cancer biomarkers. From these results it is clear that concentration levels and speciation alterations, as well as correlations of trace chemicals in body tissues can be used as parameters for cancer diagnostic at the local stage of development.

5.12 Principal Component Analysis (PCA) of selected dog tissue samples

Fig. 5.12.1 shows how the selected dog tissue samples have been classified based on the tissue pathological status, 2 classes are identified in the PC1×PC2 score plot with explained variance of 92% (86 % and 6% for PC1 and PC2 respectively). The outliers constituted 8% which can be explained based on the age and medical history of the dogs.

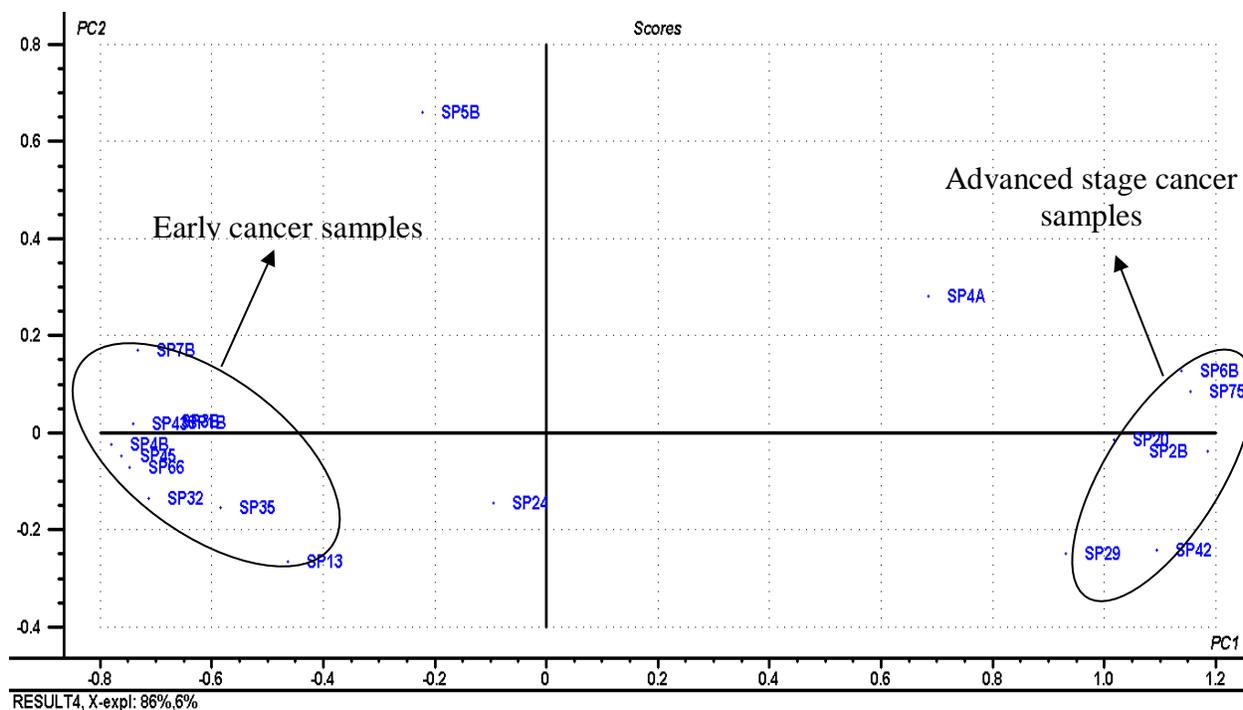


Figure 5.12.1: PCA score plot analysis for domestic dog tissue samples using partial fluorescence region.

Fig. 5.12.2 shows the variables (trace elements) responsible for the above classification of selected dog tissue samples. Based on the PC1 loadings plot, trace elements; Fe, Cu and Zn greatly influence development of cancer from one stage to another depending on the concentration levels. The results clearly confirm the fact that trace elements especially; Fe, Cu, Mn, Se and Zn are valuable biomarkers for the identification of pathological disorders in tissues.

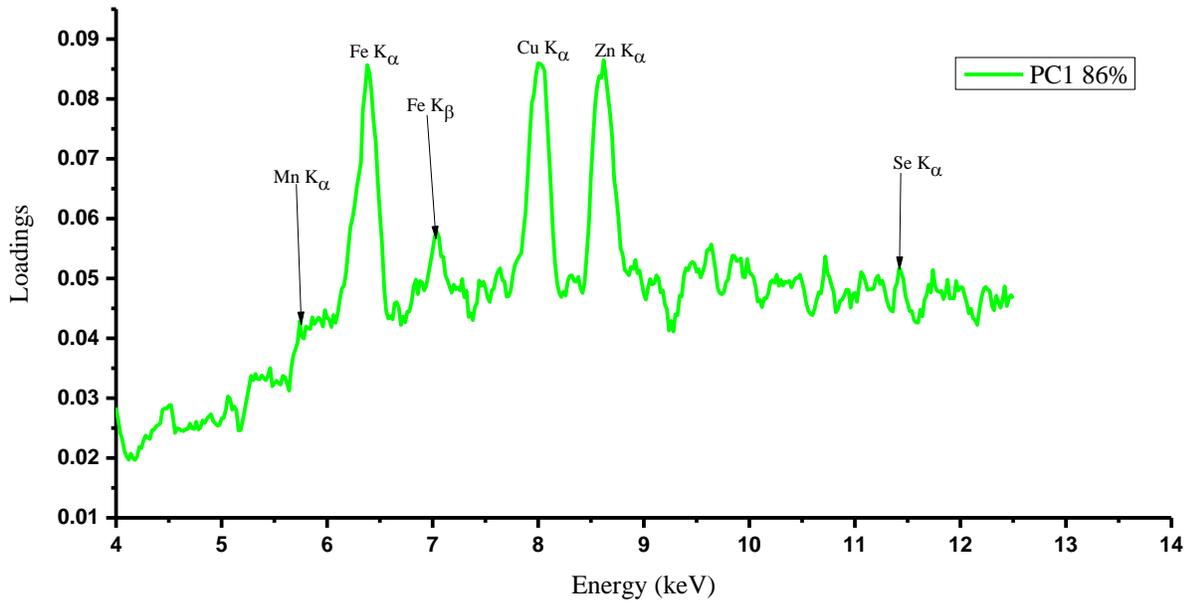


Figure 5.12.2: PCA loadings plot of domestic dog tissue samples using selected fluorescence regions.

Fig. 5.12.3 shows PCA score plot analysis for domestic dog tissue samples using partial spectral regions in which 3 clusters are now evident.

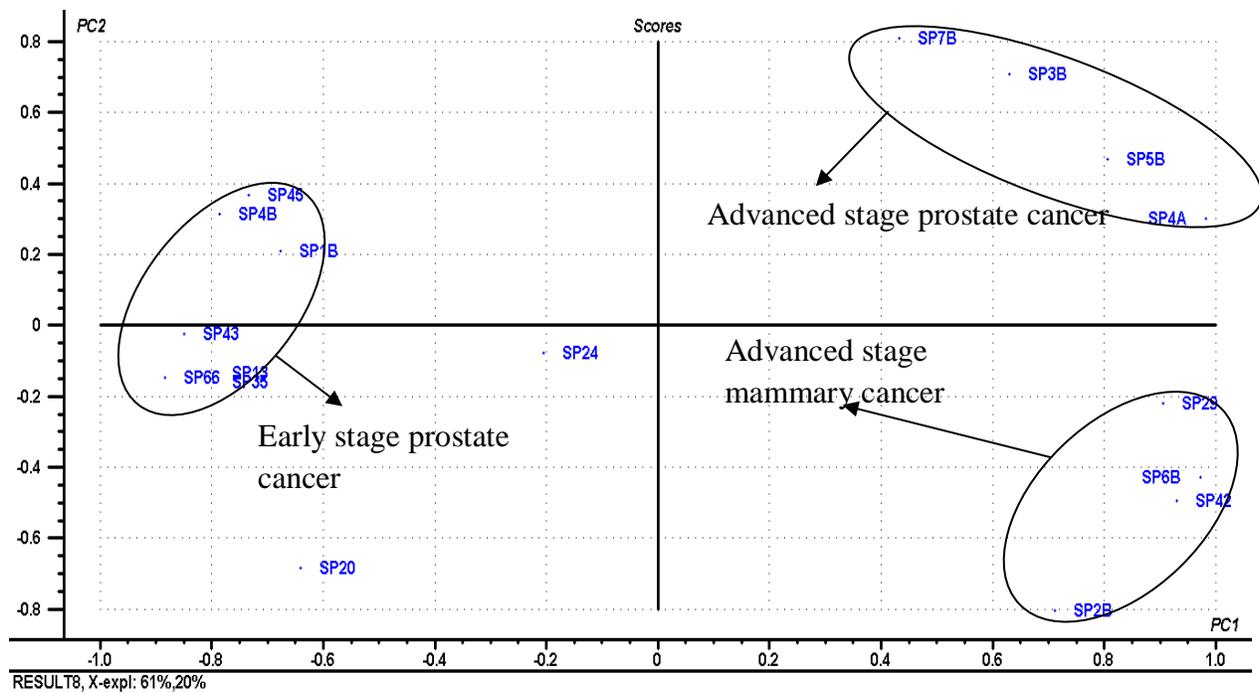


Figure 5.12.3: PCA score plot analysis of domestic dog tissue samples using selected spectral regions.

Fig. 5.12.4 shows the variables (trace elements) responsible for the above classification of selected dog tissue samples based on selected spectral region. From the PC1 loadings plot, it is clear that trace elements; Fe, Cu and Zn are cancer biomarkers.

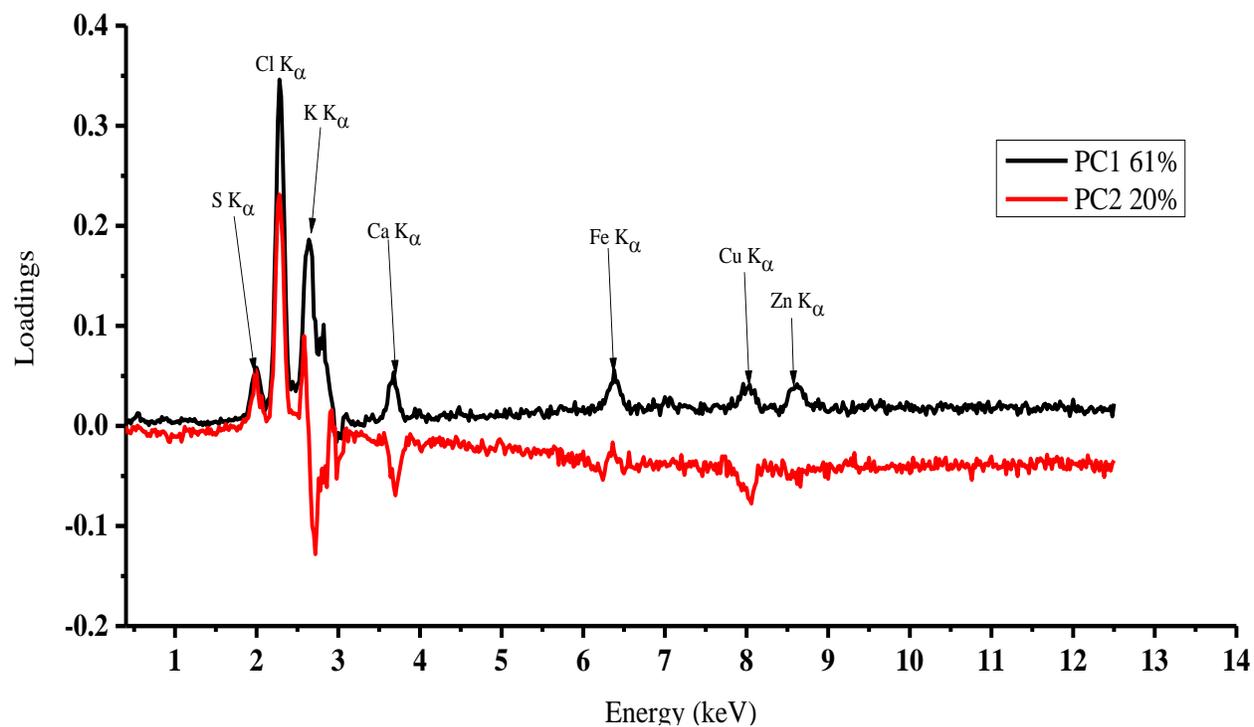


Figure 5.12.4: PCA loadings plot in domestic dog tissue samples.

In general, these results indicate that PCA exploratory analysis can be used to classify body tissue samples based on their healthy status thus be used as parameters for development of cancer diagnostic at the local stage of development.

Based on the exploratory, quantitative and speciation analysis of trace elements (Mn, Fe, Cu, Zn and Se) in tissue (simulate, cultured and dog) samples using chemometric techniques, a potential cancer diagnostic model (Figure 5.12.5) has been developed. The model has the ability to not only characterize cancer but also determine staging of cancer (early and late stage) using chemometric aided EDXRFS technique.

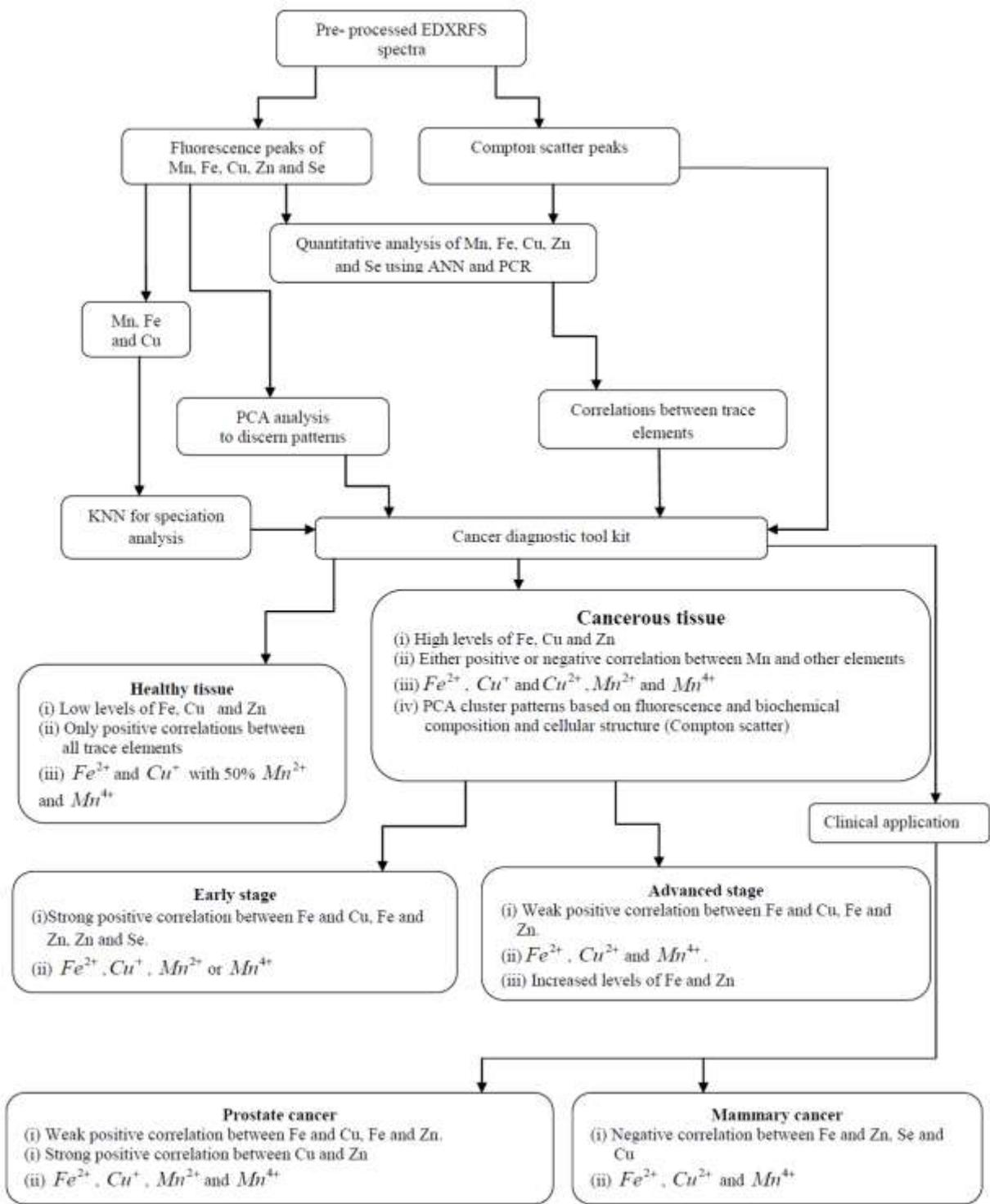


Figure 5.12.5: Chemometric aided EDXRFS diagnostic model for cancer in human body tissues.

CHAPTER 6: CONCLUSION AND RECOMMENDATION

This study was undertaken to develop and evaluate the potential of energy dispersive X-ray fluorescence and scattering (EDXRFS) spectroscopy technique in conjunction with multivariate chemometrics namely PCA, ANNs, PCR and KNN for analysis of trace elements concentration and speciation in cancer tissues. The reported methodology and results demonstrate analysis of biomedical samples by EDXRF fluorescence and Compton scatter peaks for Mn, Fe Cu, Zn and Se can be used as a tool to not only quantify trace elements but also determine speciation of the trace elements (Mn, Fe and Cu) in diagnosis of cancer at local and advanced stages of development.

The results obtained from the analysis of cultured samples using PCA analysis indicate that it is possible to characterize cancer at early and late stage of development based on concentration of trace elements (Fe, Cu, Zn, Mn and Se) and speciation of Cu and Mn. Energy Dispersive X-Ray Fluorescence and Scattering (EDXRFS) combined with chemometrics has been demonstrated as a method for diagnosis of cancer as it is efficient for determining the spectral signatures in presence of complex matrices in biomedical samples. The method is non-destructive, highly efficient and rapid in the analysis of biological samples as it had the capability of determining low elemental concentrations (ppm) of Mn and Se. The methodology used to develop quantitative calibration models proved to be equally efficient as it was possible to determine the concentration levels of the above trace elements in cultured and domestic dog tissue samples.

Furthermore, it has been demonstrated that chemometric aided EDXRFS cancer diagnostic model is a superior alternative for the traditional quantitative methods used in XRF such as FP technique. Multivariate calibration strategy, ANN dealt with spectrum evaluation and quantification in one single step which is vital for EDXRFS studies of low-Z-element matrices

when a large number of samples are to be analyzed. KNN speciation analysis enabled determination of speciation of Cu, Mn and Fe in cancerous and non-cancerous tissue samples. Based on the already developed model, cancer can be classified into 2 stages (early and late stage) based on levels of trace biomarkers (Mn, Fe, Cu, Zn and Se) and their speciation (Fe, Mn and Cu) but with no provision for intermediate staging of cancer. The results confirmed that the trace elements (Mn, Fe, Cu, Zn and Se) are valuable cancer bio-makers and should be considered as tracers for the identification of pathological disorders in tissues due to their strong correlation to histopathological changes.

For future result reliability, correct assessment of the role of each trace elemental concentration and/or speciation in regard to carcinogenesis needs to be closely monitored for a large set of samples. The predetermined cancer diagnostic model can possibly be refined by incorporating macro elements and increasing the number of trace elements to be analyzed in tissue samples for determination of early, intermediate and advanced stages of cancer development. For correct assessment of the functions of macro and trace elements in regard to carcinogenesis, there is need for acquisition of more data from several body tissue investigations.

REFERENCES

- American Cancer Society. Cancer Facts and Figures, (2012) Atlanta, Ga: American Cancer Society.
- Ames, B. (2001) DNA damage from micronutrient deficiencies is likely to be a major cause of cancer. *Mutation Res* **475**, 7-20.
- Angeyo, K.H., Gari, S., Mustapha A.O. and Mangala, J.M. (2012) Feasibility for direct rapid energy dispersive X-ray fluorescence (EDXRF) and scattering analysis of complex matrix liquids by partial least squares. *Applied Radiation and Isotopes* **70**, 2596-2601.
- Armendariz, D. and Vulpe, A. D. (2003) 11th International Symposium on Trace Elements in Man and Animals Abstracts. *The Journal of Nutrition* **133(5)**, 203E–282E.
- Aruoma, I., Halliwell, B., Gajewski, E. and Dizdaroglu, M., (1989), Damage to the Bases in DNA Induced by Hydrogen Peroxide and Ferric Ion Chelates. *J. Biol. Chem* **264**, 20509-20512.
- Banas, A., Kwiatek, W. M. and Zajac, W. (2001) Trace element analysis of tissue section by means of synchrotron radiation: the use of GNUPLLOT for SRIXE spectra analysis. *Journal of Alloys and Compounds* **328**, 135–138.
- Beckhoff, B., Kanngierfer, B., Weddel, R., Wolff, H. and Langhoff, N. (2006) Handbook of Practical X-Ray Fluorescence Analysis. Springer, Berlin.
- Brereton, R.G. (2002), Chemometrics: Data Analysis for the Laboratory and Chemical Plant. *John Wiley & sons Ltd*, West Sussex, England.
- Brereton, R.G. (2003), Chemometrics: Data Analysis for the Laboratory and Chemical Plant. *John Wiley & sons*, Bristol.
- Brouwer, P.N., (2003) Theory of XRF, Getting Acquainted with the Principles, Panalytical BV: Almelo, 23 – 30.
- Bertin, P.B., (1991) Principles and Practice of X - Ray Spectrometric Analysis, Plenum Press, New York; 1079-1101.
- Bueno, M. I. M. S. T., Martha T.P.O. , Aline M., Erica B. S. O. and Alete P. T. (2005) X-ray scattering processes and chemometrics for differentiating complex samples using conventional EDXRF equipment. *Chemometrics and Intelligent Laboratory Systems* **78**, 96– 102.
- Carvalho, M., Brito, J. and Barreiros, M. A. (1998) Study of trace element concentrations in human tissues by EDXRF spectrometry, X-ray spectrometry **27**, 198-204.

- Compton, A. H. and Allison, S.K. (1967) X-Rays in Theory and Experiment. *Van Nostrand*: New York; Chapt. III.
- Criss, J.W. and Birks, L.S. (1968) Calculation methods for fluorescent X-ray spectrometry empirical coefficients vs. fundamental parameters. *Anal Chem* **40**, 1080–1086.
- Caruso, J. A., Klaue, B., Michalke, B. and Rocke, D.M. (2003) Ecotoxicology and Environmental Safety **56**, 32.
- Daniel, B. T. and Anusuya C. (2011) Chemometric analysis of EDXRF measurements from fossil bone. *X-Ray Spectrom* **40**, 441–445.
- Debertin, K. and Helmer, R. G. (1988) Gamma and X-Ray Spectrometry with Semiconductor Detectors. *Elsevier*, Amsterdam.
- Durak, R., Kurucu, Y., Erzeneoglu, S., Ertugrul, M. and Sahin, Y. (1997) Scattering Contribution of Fluorescent Radiation to XRF Intensity in Ge(Li) Detector. *Physica Scripta* **55**, 547-549.
- Elford, H. L., Freese, M., Passamani, E. and Morris, H. P. (1970) Ribonucleotide reductase and cell proliferation. I. Variations of ribonucleotide reductase activity with tumor growth rate in a series of rat hepatomas. *J. Biol.Chem* **245**, 5228–5233.
- Everitt, B. S. (1981) Cluster Analysis. Heineman, London.
- Fabíola, M., Verbi, Edenir, R., Pereira-Filho, Maria Izabel, M.S., Bueno (2005) Use of X-ray scattering for studies with organic compounds: a case study using paints. *Micro chimica Acta* **150**, 131-136.
- Feather, C.E. and Willis, J.P. (1976) A Simple method for background and matrix correction of spectral peaks in trace element determination by X-ray fluorescence spectrometry, *X-ray Spectrometry* **5**, 41.
- Fausett, L. (1994) Fundamentals of Neural Networks: Architectures, Algorithms and Applications. Prentice-Hall, Upper Saddle River.
- Feinendegen, E. L. and Kasperek, K. (1980) Medical aspects of trace element research, in Trace Element Analytical Chemistry in Medicine and Biology, P. Bratter and P. Schramel, Eds., pp. 1–36, Walter de Gruyter, Berlin, Germany.
- Geraki, K. and Farquharson, M.J. (2001) An X-ray fluorescence system for measuring trace element concentrations in breast tissue. *Radiation Physics and Chemistry* **61**, 603–605.
- Geraki, K., Farquharson, M. J. and Bradley, D.A. (2002) Concentrations of Fe, Cu and Zn in breast tissue, a synchrotron XRF study. *Phys. Med. Biol.* **47**, 2327–39.

- Geraki, T. K., Farquharson, M. J. and Bradley, D. A. (2004) X-Ray Fluorescence and Energy Dispersive X-Ray Diffraction for the quantification of elemental concentrations in breast tissue. *Phys. Med. Biol.* **49**, 99–110.
- Goodacre, R. (2003) Explanatory Analysis of Spectroscopic data using machine learning of simple, interpretable rules. *Vibrational Spectroscopy* **32**, 33-35.
- Green, P.E. (1976) *Mathematical Tools for Applied Multivariate Analysis*, New York.
- Grodner, M., Anderson, S. and Deyoung, S. (2000) *Foundation and clinical applications of nutrition*, 2nd ed., Mosby Inc., St. Louis, Missouri; 45-107.
- Guardiner, P. H. E., (1987) *Topics in Current Chemistry* 141:147.
- Halliwell, B. and Gutteridge, J. M. C. (1999) *Free Radicals in Biology and Medicine*, third ed., Oxford University Press, New York.
- Halliwell, B. and Gutteridge J. M. C. (1984) Oxygen toxicity, oxygen radicals, transition metals and disease. *Biochem J.* **219**, 1-14.
- Hambidge, K.M., Casey, C.E. and Krebs, N.F. (1986) Zinc in: Mertz, W. ED. *Trace elements in human and animals nutrition*, 5th ed., Orlando, F.A, Academic press Inc.; 1-137.
- Hartwig, A., Asmuss, M., Ehleben, I., Herzer, U., Kostelac, D. and Pelzer, A. (2002) Interference by toxic metal ions with DNA repair processes and cell cycle control: molecular mechanisms. *Environ Health Perspect* **5**, 797-9.
- Henry, J., Kocri, J. K., Eerr, M. A., Smith, B. S., Neil F., Shimp, M. and Jane C. M. S. (1995) Analysis of trace elements in human tissues. *Cancer* **9**, 499-511.
- Howard, D., Mark, B. and Hagan, M. (2009) *Neural Network Toolbox: User's Guide*. The MathWorks, Inc, US.
- Huang, Y., Sheu J. and Lin T., (1999) Association between oxidative stress and changes of trace elements in patients with breast cancer. *Clin Biochem* **32(2)**, 131-136.
- Janssens, K. (2004) X-ray based methods of analysis. *Non-Destructive Micro Analysis of Cultural Heritage Materials* **42**, 129–226.
- Jessica, L. H., Joshua, M. W., Catherine, P. L. and Daniel, S. W. (2013) Iron Deprivation in Cancer–Potential Therapeutic Implications. *Nutrients* **5**, 2836-2859.
- Jose, M. A. G. and Neil, W. B. (2009) *Basic Chemometric Techniques in Analytical Spectroscopy*. *RSC Analytical Spectroscopy Monographs* **10**, 244-263.
- Jolliffe, I. T. (1982) A note on the use of principal components in regression. *Applied Statistics* **31**, 300–303.

- Kallithrakas-Kontos, N. and Moshohoritou, R. (1998) Vanadium Speciation by Energy Dispersive X-Ray Fluorescence (EDXRF). *X-ray spectrometry* **27**, 173-176.
- Kaniu, M. I., Angeyo, K.H., Mwala, A. K. and Mangala, M. J. (2012) Direct Rapid Analysis of Trace Bioavailable Soil Macronutrients by Chemometrics-assisted Energy Dispersive X-Ray Fluorescence and Scattering spectrometry. *Analytica Chimica Acta* **729**, 21– 25.
- Kaniu, M. I., Angeyo, K.H.,Mangala, M.J.,Mwala, A.K., Bartilol, S.K. (2011). Feasibility for Chemometric Energy Dispersive X-Ray Fluorescence and Scattering (EDXRFS) Spectroscopy Method for Rapid Soil Quality Assessment. *X-Ray Spectrom* **40 (6)**, 432–440.
- Kaniu, M. I. (2011) Development of a Chemometric Energy Dispersive X-Ray Fluorescence and Scattering Spectroscopy (EDXRFS) Method for Rapid Soil Quality Assessment. *MSc Thesis. University of Nairobi, Kenya.*
- Kessler, T., Hoffmann, P., Greve, T. and Ortner, H. M. (2002) Optimization of the identification of chemical compounds by energy dispersive X-ray fluorescence spectrometry and subsequent multivariate analysis, *X-Ray Spectrom* **31**, 383-390.
- Konemann, S., Bolling, T., Matzkies F., Willich, N., Kisters, K. and Micke, O. (2005) Iron and iron-related parameters in oncology. *Trace Elem Electrolytes* **22(2)** 142-149.
- Kowalik, C. and Einax J. W. (2006) Modern chemometric data analysis-methods for the objective evaluation of load in river systems. *Acta hydrochim, hydrobiol* **34**, 425-436.
- Kryston, T. B., Georgiev, A. B., Pissis, P., and Georgakilas, A. G. (2011) Role of oxidative stress and DNA damage in human carcinogenesis, Mutation Research—Fundamental and Molecular. *Mutation Research* **711 (2011)**, 193–201.
- Kuehnelt, D., Kienzl, N., Traar, P., Hoang, L. N., Francesconi, K.A. and Ochi, T. (2005) *Analytical and Bioanalytical Chemistry* **383(2)**, 235.
- Kubala-Kukus, A., Kuternoga, E., Braziewicz, J., Pajek, M., (2004), “Log-stable concentration distributions of trace elements in biomedical samples”, *Spectrochim. Acta Part B* **59**, 1711–1716.
- Kurt, V. and Peter, F. (2008) Introduction to Multivariate Statistical Analysis in Chemometrics, CRC: New York, Chapter 2–4.
- Kwiatek, W.M., Hanson, A. L., Paluszkiewicz, C., Galka, M., Gajda, M. and Cichocki, T. (2004) Application of SRIXE and XANES to the determination of the oxidation state of iron in prostate tissue sections. *Journal of Alloys and Compounds* **362**, 83–87.

- Kwiatek, W.M., Drewniak, T., Gajda, M., Galka, M., Hanson, A.L. and Cichocki, T. (2002) Preliminary study on the distribution of selected elements in cancerous and non-cancerous kidney tissues. *J. Trace Elem. Med. Biol* **16**, 155-160.
- Kwiatek, W. M., Banas, A., Banas, K., Gajda, M., Gałka, M., Falkenberg, G. and Cichocki, T. (2005) Iron and other elements studies in cancerous and non-cancerous prostate tissues. *Journal of Alloys and Compounds* **40**, 178–183.
- Lachance, G. R. (1999) Demystification of algorithms and influence coefficients in quantitative XRF analysis, *JCPDS-ICDD*, 718–731.
- Lee, R., Woo, W., Wu, W.B., Kummer, A., Duminy, H. and Xu, Z. (2003) Zinc accumulation in N- methyl-N-nitrosourea-induced rat mammary tumors is accompanied by an altered expression of ZnT-1 and metallothionein. *Exp Biol Med* **228**, 689-696.
- Lennart, E., Johan, T. and Svante, W. (2014) A chemometrics toolbox based on projections and latent variables. *J. Chemometrics* **28**, 332–346.
- Liqiang and Luo, (2006) Chemometrics and its applications to X-Ray Spectrometry. *X-Ray Spectrom* **35**, 215–225.
- Liu, Y., Upadhyaya, B. R. and Masoud, N. (1993) Chemometric data analysis using artificial neural networks, *J. Appl. Spectroscopy* **47**, 12-23.
- Luo, L., Changlin, G., Guangzu, M. and Ang, J. (1997) Choice of Optimum Model Parameters in Artificial Neural Networks and Application to X-ray Fluorescence Analysis. *X-Ray Spectrometry* **26**, 15-22.
- Magos, L. (1991) Epidemiological and experimental aspects of metal carcinogenesis, physicochemical properties, kinetics and the active species, *Environment health perspect.* **95**, 157-189.
- Mehmet, Y., Gokce, K. and Hayrettin, Y. (2007) Distribution of trace metal concentrations in paired cancerous and non-cancerous human stomach tissues. *World J Gastroenterol* **13** (4), 612-618.
- Massart, D.L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y. and Kaufman, L. (1988) Chemometrics. A Textbook, Elsevire, Amsterdam.
- Mathers, C. D. and Loncar, D. (2006) Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine*, **3**, 2011–2030.
- Mathias, O. (1999) Chemometrics: Statistics and Computation in Analytical Chemistry, Wiley-VCH, Weinheim.

- Marini, F., Bucci, R., Magrì, L. and Magrì D. (2008) Artificial neural networks in chemometrics: History examples and perspectives; *Microchemical Journal* **88**, 178–185.
- Markowicz, A. (2011) An overview of quantification methods in energy-dispersive X-ray fluorescence analysis. *Pramana-J. Phys* **76** (2), 321-329.
- Martens, H. and Nas, T. (1991) *Multivariate calibration*. Chichester: John Wiley & Sons.
- Nasulewicz, A., Mazur, A. and Opolski, A. (2004) Role of copper in tumor angiogenesis-clinical implications, *Trace Elem. Med. Biol.* **18** (1), 1–8.
- Ng, K. H., Bradley D. A. and Looi, L. M. (1997) Elevated trace element concentrations in malignant breast tissues. *British Journal of Radiology* **70**, 375–382.
- Nielson, K. K. (1977) Matrix corrections for energy dispersive X-ray fluorescence analysis of environmental samples with coherent/incoherent scattered X-rays. *Anal. Chem.* **49**(4), 641-648.
- Oliveira, L., Antunes, A. M. and Bueno, M. I. M. S. (2010) Direct Chromium Speciation using X-Ray Spectrometry and Chemometrics. *X-Ray Spectrom* **39**, 279–284.
- Prasad Ed., A. S. O. D. (1976) *Trace Elements in Human Health and Disease*. Academic Press, New York.
- Pires, J., Martins, F., Sousa, S., Alvim-Ferraz, M. and Pereira, M. (2008) Selection and Validation of Parameters in Multiple Linear and Principal Component Regressions. *Environmental Modelling & Software* **23**(1), 50-55.
- Podgorczyk, M., Kwiatek, W. M., Zaja, W. M., Dulinska-Litewka, J., Welter, E. and Grolimund, D. (2009) Zinc in Native Tissues and Cultured Cell lines of Human Prostate studied by SR-XRF and XANES. *X-Ray Spectrom* **38**, 557–562.
- Potts, P. J., Ellis, A. T., Kregsamer, P., Strelly, C., Vanhoof, C., West, M. and Wobrauschek, P. (2006) Atomic spectrometry update-X-ray fluorescence spectrometry. *J Anal At Spectrom*, **21**(10), 1076-1107.
- Poletti, M. E., Gonçalves, O. D., Pérez, C. A. and Magalhães, S. D. (2004) A preliminary study of the distribution of trace elements in healthy and neoplastic breast tissues with synchrotron radiation X-ray fluorescence. *Radiat. Phys. Chem.* **71**, 975–976.
- Raju, G. J. N., Sarita, P., Kumar, M. R., Murty, G., Reddy, B. S., Lakshminarayana, S., Vijayan, V., Lakshmi, P., Gavarasana, S. and Reddy, S. B. (2006) Trace elemental correlation study in malignant and normal breast tissue by PIXE technique. *Nuclear*

Instruments & Methods in Physics Research Section B-Beam Interactions with Materials and Atoms, **247(2)**, 361-367.

- Richardson Le, N. T. V. (2002) The role of iron in cell cycle progression and the proliferation. *Biochim Biophys Acta-Rev Cancer* **1603(1)**, 31-46.
- Rizk, S. L. and Skypeck, H. H. (1984) Comparison between Concentrations of Trace-Elements in Normal and Neoplastic Human-Breast Tissue. *Cancer Res* **44(11)**, 5390–5394.
- Rojas, E., Herrera, L. A., Poirier, L. A. and Ostrosky-Wegman, P. (1999) Are metals dietary carcinogens. *Mutat Res* **443**, 157-181
- Rose, J. (1983) Trace Elements in Health, Butterworth-Heinemann: London.
- Rousseau, R. M. (1984) Fundamental algorithm between concentration and intensity in XRF analysis – theory and Practical application. *X-Ray Spectrom.*, **13**, 115–125.
- Schrauzer, G. N. (1980) The Role of Trace Elements in the Etiology of Cancer. *Trace Elem Anal Chem Med Biol*, 183-195.
- Schramm, R. (2000) Chemometric Methods in Energy Dispersive X-Ray Fluorescence. *Analytica Chimica Acta* **420**, 197–203
- Sherman, J. (1955) The Theoretical Derivation of Fluorescent X-Ray Intensities from Mixtures. *Spectrochim. Acta* **7**,283-306.
- Sharaf, M. A., Illman, D. L. and Kowalski, B. R. (1986) Chemometrics, John Wiley, New York.
- Sherman, J. (1958) A Theoretical Derivation of the Composition of Mixable Specimens from Fluorescent X-ray Intensities. *Adv. X-ray Anal* **1**, 231.
- Shiraiwa, T. and Fujino, N. (1966) Theoretical Calculation of Fluorescent X-ray Intensities in Fluorescent X-ray Spectrochemical Analysis. *Japan Journal of Applied Physics* **5**, 886–899.
- Siddiqui, M. K., Jyoti, Singh, S., Mehrotra P. K., Singh, K. and Sarangi, R. (2006) Comparison of some Trace Elements Concentration in Blood, Tumor free Breast and Tumor Tissues of Women with Benign and Malignant Breast Lesions, an Indian Study. *Environ Int* **32 (5)**, 630-7.
- Silva, M. P., Tomal, A., Perez, C. A., Ribeiro-Silva, A. and Poletti, M.E. (2009) Determination of Ca, Fe, Cu and Zn and their Correlations in Breast Cancer and Normal adjacent Tissues. *X-Ray Spectrom* **38 (2)**, 103–111.

- Silva, Marina, P., Danilo, F., Soave, Alfredo, Ribeiro-Silva, Martin, E. and Poletti, (2012) Trace Elements as Tumor Biomarkers and Prognostic Factors in Breast cancer: a study through Energy Dispersive X-Ray Fluorescence. *Biomed central* **5**, 194
- Sigel, H. (1980) Metal Ions in Biological Systems: Carcinogenicity and Metal Ions, Volume **10**. New York: Marcel Dekker.
- Sky-Peck, H. H. and Joseph, B. J. (1981) Determination of Trace Elements in Human Serum by Energy Dispersive X-Ray fluorescence. *Clin. Biochem.* **14** (3), 126–131.
- Stephen, J. M. (2013) Comparative Trace Elemental Analysis in Cancerous and Non-cancerous Human Tissues Using PIXE. *Journal of Biophysics* **2013**, 1-8.
- Sukumar, S., Notario, V., Martin-Zanca, D. and Barbacid, M. (1983) Induction of Mammary Carcinomas in Rats by Nitroso-methylurea involves Malignant Activation of H-ras-1 locus by Single Point Mutations. *Nature* **306** (5944), 658-661.
- Szupunar, J. and Lobinski, R. (1999) Species-selective Analysis for Metal Biomacromolecular Complexes using Hyphenated Techniques. *Pure and Applied Chem* **71** (5), 899-918.
- Su, J. M. F., Perlaky, L., Li, X. N., Leung, H. C. E., Antalffy, B., Armstrong, D. and Lau, C.C. (2004) Comparison of ethanol versus formalin fixation on preservation of histology and RNA in laser capture micro dissected brain tissues. *Brain Pathol* **14**, 175–18.
- Thanh, N. T., Wehrens, R. and Buydens, L. M. C. (2004) Clustering multispectral images: a tutorial. *Chemometrics and Intelligent Laboratory Systems* **77**, 3-17.
- Theophanides, T. and Anastassopoulou, J. (2002) Copper and carcinogenesis. *Crit Rev Oncol Hematol* **42**, 57-64.
- Valko, M., Morris, H. and Cronin, M. T. D. (2005) Metals, Toxicity and Oxidative Stress, *Current Medicinal Chemistry* **12**(10), 1161–1208.
- Van Espen, P. and He F. (1989) Axil Quantitative X-Ray Fluorescence Analysis, Technical Note. Wilrijk, Antwerp, Belgium.
- Vandeginste, B. G. M., Massart, D. L., Buydens, L. M. C., Jong, S. D., Lewi, P. J. and Smeyers, V. J. (1998) Supervised Pattern Recognition. In: Handbook of Chemometrics and Qualimetrics. *Part B. Elsevier*, Amsterdam, 207-241.
- Virendra, S., Anrawal, H., Joshi, C., Sudersha, M. and Sinha, K. (2011) Elements Profiles of Agricultural Soil by EDXRF Technique and Use of the Principal Component Analysis

- (PCA) method to interpret the Complex Data, *Applied Radiation and Isotopes* **36**, 181-194.
- VonCzarnowski, D., Denkhaus, E. and Lemke, K. (1997) Determination of Trace Element Distribution in Cancerous and Normal Human Tissues by Total X-Ray fluorescence Analysis. *Spectrochim Acta B* **52**, 1047-1052.
- West, M., Ellis, A. T., Kregsamer, P., Potts, P. J., Strelci, C., Vanhoof, C. and Wobrauschek, P. (2007) Atomic spectrometry update. X-ray fluorescence spectrometry, *J. Anal. At. Spectrom.*, **22**, 1304–1332.
- White, D. R. (1975) The Formulation of Tissue Substitute Materials Using Basic Interaction Data. *Physics in Medicine and Biology* **22** (5), 22-25.
- World Health Organization (1996b) Trace Elements in Human Nutrition and Health, Geneva; 72-103.
- Yaman, M. (2006) Comprehensive Comparison of Trace Metal Concentrations in Cancerous and Non-cancerous Human Tissues. *Current Medicinal Chemistry* **13**(21), 2513–2525.
- Zhai, H., Chen, X. and Zhidong, H. (2003) Study on the Relationship between Intake of Trace Elements and Breast Cancer Mortality with Chemometric Methods. *Comp. Bio. Chem.* **27**, 581-586.
- Zomer, S., Guill, C., Brereton, R. G. and Hanna, B. M. (2003) Toxicological Classification of Urine Samples using Pattern Recognition Techniques and Capillary Electrophoresis. *Anal Bioanal. Chem* **378**, 2008-2020.
- Zupan, J. and Gasteiger, J. (1999) Neural Networks in Chemistry and Drug Design. 2nd ed. Wiley VCH Ed., Weinheim ISBN: 978-3527297795.

APPENDICES

Appendix 1: Tissue culturing protocol used in the study at KEMRI

- When working with human blood, cells or infectious agents, all appropriate bio-safety practices must be followed.
- All cell incubators should be performed in a humidified 37⁰C, 5 % CO₂ incubator unless specified.
- Always wear protective lab coat and sterilized gloves when handling cells to minimize contaminations.
- During incubation process, the cells have to be examined microscopically to monitor their confluence and viability and for sub-culturing if necessary.
- The growth media have to be changed periodically until the cells have reached confluence stage (stop growing).

Appendix 2: Histopathological tissue sampling protocol used in the study at KEMRI

Tissue sampling generally follows standard protocols established by each laboratory for particular sample specimens;

- Tissue blocks thickness for processing should consistent with the purpose for which they are required for instance 1-2 mm thick for urgent specimens.
- Specimens should not be tightly packed into processing tissue cassettes but should have sufficient free space to facilitate fluid exchange where small specimens and tissue fragments are processed in fine mesh containers, wrapped in lens tissue, sandwiched between sponge biopsy pads or more safely, double embedded in agar-paraffin wax.
- Specimens are generally identified by a numbering system that is not bleached by subsequent fluid and solvent treatment. These can be (a) a numbered card label generated by computer-printer, or handwritten in soft lead pencil or waterproof ink (b) colour coded plastic cassettes, machine or manually labelled.

Appendix 3: CONSENT FORM

X-RAY FLOURESCENCE AND SCATTERING SPECTROMETRY OF TRACE

METALS AND THEIR SPECIATION IN MODEL CANCER TISSUE MEDIA

You are asked for your dog to participate in a research study conducted by Mr. Okonda Justus, for an award of master of science degree, P. O. Box 30197-00100, Nairobi, Telephone number 4447552, from the department of Physics at the University of Nairobi. Your dog has been selected to participate in this research because it (put × in appropriate box):

- Is non-cancerous (health dog).
- Has mammary cancer.
- Has Prostate cancer.

Give the following information pertaining your dog.

Age	
Weight	
Any Treatment given	
Diet	

Consent by the owner/next of kin of the dog for its participation in the research

I(Mr/Mrs/Miss/Dr/Prof).....

have read (or someone has read to me) the information provided above. I have been given an opportunity to ask questions, and all of my questions have been answered to my satisfaction. I have been given a copy of the consent explanation form. Thus I accept my dog to participate in this research. I understand that the results of these tests shall be used for research work only and be kept confidential.

BY SIGNING THIS FORM, I WILLINGLY AGREE FOR MY DOG TO PARTICIPATE IN THE RESEARCH.

SignatureDate.....

I Prof./Dr./Mr/Mrs/Miss.....

confirms that I have explained to the owner of the dog the nature of the research, tests to be done, and given him/her a copy of the consent explanation form.

SignatureDate.....

Appendix 4: ANNs Training Algorithm Used in this Work

```
load data
% Normalize the inputs and targets.
[pn,pp1] = mapstd(p);
[tn,tp] = mapstd(t);
sn=mapstd(sv);
% Create a feedforward network with 4 hidden neurons, 3 output neurons
and assign the Levenberg-Marquardt training function - TRAINLM. The
NEWFF command initialized the weights in the network.
net = newff(pn,tn,4);
% Train the network.
net = train(net,pn,tn);
net = init(net);
% Simulate the trained network.
y = sim(net,sn);
% Convert the output of the network back into the original units of
the targets. Since the targets were transformed using MAPSTD so that
the mean was 0 and the standard deviation was 1, we need to use MAPSTD
and the original mean and standard deviation (stored in tp) to
transform the network outputs back into the original units.
a = mapstd('reverse',y,tp);
```

Appendix 5: PCR Algorithm Used in this study

```
% Load a data set comprising spectral intensities
>> X = pn;
>> y = tn;
>> [n,p] = size(X);
% Perform Principal Components Analysis on X, using the princomp
function, and retaining two principal components
>> [PCALoadings,PCAScores,PCAVar] = princomp(X);
>> betaPCR = regress(y-mean(y), PCAScores(:,1:2));
% Transform to regression coefficients for the original, uncentered
variables.
>> betaPCR = PCALoadings(:,1:2)*betaPCR;
>> betaPCR = [mean(y) - mean(X)*betaPCR; betaPCR];
>> yfitPCR = [ones(n,1) X]*betaPCR;
% Plot predicted vs. known response for the PCR fits.
>> plot(y,yfitPCR,'r^');
>> xlabel('Known concentration');
>> ylabel('predicted concentration')
```

Appendix 6: KNN Training Algorithm Used in this study

```
% load KNN data
x = meas(:,1:2);
t=normc(x);
gscatter(t(:,1),t(:,2),species)
set(legend,'location','best')
% Plot the new point:
newpoint = [1340 1460];
line(newpoint(1),newpoint(2),'marker','x','color','k',...
      'markersize',10,'linewidth',2)
% Locate the 10 sample points closest to the new point:
[n,d] = knnsearch(t,newpoint,'k',10);
line(t(n,1),t(n,2),'color',[.5 .5 .5],'marker','o',...
      'linestyle','none','markersize',10)
% Determine the species of the 10 neighbors:
Tabulate (species (n))
Value      Count Percent
'Fe (II)''   8 80.00%
'Fe (III)''  2 20.00%
% Define the center and diameter of a circle, based on the location of
the new point:
ctr = newpoint - d(end);
diameter = 2*d(end);
% Draw a circle around the 10 nearest neighbors:
h = rectangle ('position',[ctr,diameter,diameter],...
              'curvature',[1 1]);
set(h,'linestyle',':')
```