



UNIVERSITY OF NAIROBI

**DIAGNOSTIC ANALYSIS AND MODELING OF TRACE BIOMETALS AND
SPECIATION ALTERATIONS IN URINARY TRACT CANCERS BY MACHINE
LEARNING BASED X-RAY FLUORESCENCE AND SCATTER (XRFS)
SPECTROMETRY**

By

OKONDA J. JUSTUS

I80/51695/2017

**A Thesis Submitted in Fulfilment of the Requirements for Award of the Degree of
Doctor of Philosophy in Physics of the University of Nairobi.**

© August, 2023

DECLARATION

This thesis is my original work and has not been submitted elsewhere for examination, award of a degree or publication. Where other people's work or my own work has been used, this has properly been acknowledged and referenced in accordance with the University of Nairobi's requirements.

Signature 

Date: 24th August 2023

Okonda J. Justus

I80/51695/2017

Department of Physics

Faculty of Science & Technology

University of Nairobi

This thesis is submitted for examination with our approval as research supervisors:

Prof. Angeyo Hudson Kalambuka

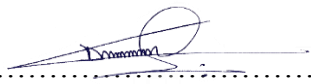
Department of Physics, University of Nairobi, Kenya

Signature 

Date: 24th August 2023

Dr. Alix Dehayem Massop

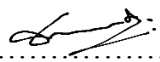
Alan Alda Center for Communicating Science, Stony Brook University, New York (USA).

Signature 

Date: 24th August 2023

Prof. Rogena Emily Adhiambo

Department Human Pathology, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Kenya.

Signature 

Date: 24th August 2023

ACKNOWLEDGEMENTS

I am indebted to the University of Nairobi for financial and material support towards the completion of this PhD research. My gratitude goes to my supervisors; Prof. Hudson A. Kalambuka Department of Physics, University of Nairobi, Dr. Alix Dehayem of Alan Alda Center for Communicating Science, Stony Brook University, New York and Prof. Emily Rogena, Department of Human Pathology, SOMED, JKUAT for their encouragement, support and guidance throughout this study. Their supervisory roles, suggestions and comments made it possible for me to successfully complete this research work in time. My gratitude also goes to Dr. Ian Kaniu Department of Physics, University of Nairobi for his unwavering support and encouragement during this research. I am also grateful to Mr. Muchiri of CRV laboratory at Kenya Medical Research Institute (KEMRI), for his substantial help during cell culturing and sample preparation. Further, I am grateful to Kenya National Research Fund (NRF) for funding the research and also thankful to the University of Nairobi for the tuition fee for my studies.

DEDICATION

I dedicate this work to my beloved wife (Saisi Mical Upendo) and children (Gloria Asuko, Moses Okonda and Carolyne Andenyi).

ABSTRACT

X-ray fluorescence spectroscopy has the capability to determine the levels, chemical speciation, and distribution of trace biometals in a biological sample for disease diagnostics. Direct biometals analysis in soft tissues and fluids by X-ray fluorescence for early cancer diagnosis has hardly been explored due to dark matrix challenges that results to weak analyte signals as well as intricate multivariate relationships. This study was aimed to develop a machine learning-aided X-ray fluorescence and scatter method for early diagnosis of urinary tract cancer (prostate and urinary bladder) based on the concentration, speciation and 2D imaging of trace biometals in cell cultures, human tissue and urine. The XRF variants enabled simultaneous analysis of biometals' levels, speciation and 2D distribution in; simulate tissue and urine, human tissue and urine, and cultured samples. The levels of the biometals (Mn, Cr, Cu, Fe, Zn and Se) were determined by multivariate calibration model (ANN) using EDXRF selected fluorescence and scatter regions from simulate urine and tissues. The levels of Fe, Cu and Zn in human tissue biopsies were in the range of; 150.0 ± 4.5 - 181.2 ± 9.5 ppm, 16.4 ± 5.4 - 25.9 ± 2.6 ppm and 60.5 ± 12.4 - 90.2 ± 3.8 ppm respectively with alteration in levels of Cr, Mn and Se. High ($p < 0.001$) concentrations of Fe, Cu and Zn were noted in cultured cancer tissues compared to normal human tissues and urine. The KNN distinguished the chemical speciation of Cu and Fe in cancerous and normal urine at 90% classification accuracy. The human cancerous urine samples were found to be rich in *Fe* and *Cu* occurring mostly as Fe^{2+} and Cu^+ possibly due to their oxidative role in Fenton reactions that accelerates carcinogenesis. SR- μ XRF enabled 2D spatial distribution of trace biometals (Mn, Fe, Cu and Se) patterns where significant ($p < 0.05$) differences in the accumulation of the analytes in cancerous and healthy cell cultures. The 2-D spatial distribution maps of the trace biometals revealed high spatial correlation between Cu and Fe (0.9406) and (0.92252) in DU_D3 and DU_D4 respectively in cancerous compared to corresponding stages in normal cell cultures for cancer diagnosis. Artificial Neural networks (ANN) distinctively classified cell cultures into cancerous and healthy groups by PC1 and PC2 scores of Cu and Fe L-lines spatial-spectral profiles. Further utility of selected fluorescence and scatter spectral profiles enabled the classification of cancerous cultured cells into early, intermediate and advanced stages of cancer development. The study demonstrated the utility of machine learning-aided XRFS analysis of hardly discernible fluorescence peaks and Compton scatter in tissues, cell cultures and urine samples to realize rapid non-invasive cancer diagnostic model.

TABLE OF CONTENTS

DECLARATION	I
ACKNOWLEDGEMENTS	II
DEDICATION	III
ABSTRACT	IV
TABLE OF CONTENTS	V
LIST OF TABLES	IX
LIST OF FIGURES	XI
LIST OF ABBREVIATIONS AND SYMBOLS	XIV
CHAPTER ONE: INTRODUCTION	1
1.1 Background to the study	1
1.2 X-ray fluorescence spectroscopy	1
1.3 Cancer survival and mortality	2
1.4 Trace Biometals in human body	3
1.5 XRF analysis of trace biometals for cancer diagnosis	6
1.6 Multivariate machine learning for XRF analysis of trace biometals	7
1.7 Statement of the Problem	8
1.8 Research objectives	9
1.8.1 General objective	9
1.8.2 Specific Objectives	9
1.9 Study Hypothesis	10
1.10 Justification and significance of the Study	10
CHAPTER TWO: LITERATURE REVIEW	12
2.1 Chapter Overview	12
2.2 Conventional cancer diagnostic methods	12
2.3 X-Ray Fluorescence (XRF) Spectroscopy	12
2.3.1 Energy dispersive X-ray fluorescence spectroscopy	13
2.3.2 Total reflection X-ray fluorescence spectroscopy	14
2.3.3 Synchrotron radiation XRF spectroscopy	15
2.4 Other spectroscopic techniques for trace biometal analysis	17
2.5 Machine Learning Techniques	18
CHAPTER THREE: THEORETICAL FRAMEWORK	20

3.1 Chapter Overview	20
3.2 X-ray fluorescence spectroscopic analysis.....	20
3.2.1 Energy dispersive X-ray fluorescence (EDXRF)	21
3.2.2 Total reflection X-ray fluorescence (TXRF).....	22
3.3 Synchrotron radiation based XRF analysis	23
3.3.1 X-ray absorption spectroscopy (XAS)	24
3.3.2 Synchrotron based micro XRF	27
3.3.3 X-ray absorption near edge spectroscopy (XANES).....	27
3.3.4 Extended X-ray absorption fine structure (EXAFS)	28
3.4 Quantification methods in XRF	28
3.4.1 Fundamental parameter method (FPM).....	28
3.4.2 Influence coefficients method	29
3.5 Machine learning multivariate analysis.....	29
3.5.1 Pattern recognition techniques.....	30
3.5.2 Multivariate calibration strategy for quantitative analysis of trace biometals.....	32
CHAPTER FOUR: MATERIALS AND METHODS.....	36
4.1 Chapter Overview	36
4.2 Reagents and materials.....	36
4.3 Sampling.....	36
4.3.1 Simulate sample preparation	37
4.3.2 Soft body tissue simulation.....	38
4.4 Urine Simulate.....	39
4.5 Cell Culturing.....	40
4.6 Urinary tract tissues and urine sampling	42
4.6.1 Ethnical approval.....	42
4.6.2 Urine sampling	42
4.6.3 Study population.....	42
4.6.4 Urinary tract tissue sampling.....	44
4.7 Sample preparation.....	44
4.7.1 Tissue simulates.....	44
4.7.2 Cell cultures	44
4.7.3 Urinary tract tissues	45
4.7.4 Urine samples	45
4.8 Quality assurance	45

4.9 X-ray fluorescence analysis	46
4.9.1 EDXRF analysis of simulate and urine samples	47
4.9.2 TXRF analysis of urine samples.....	48
4.9.3 SR-XRF analysis	49
4.9.4 Microscopy imaging of cell cultures	53
4.10 XRF data analysis.....	54
4.11 Multivariate machine learning analysis.....	55
4.11.1 Pattern recognition.....	55
4.11.2 Multivariate calibration	56
4.12 Cancer diagnostic machine learning aided XRF method development	57
CHAPTER FIVE: RESULTS AND DISCUSSIONS.....	59
5.1 Chapter Overview	59
5.2 XRF spectra of biomedical samples.....	59
5.2.1 EDXRF spectra of simulate and human urine	59
5.2.2 TXRF spectra of urine samples	61
5.2.3 SR-XRF spectra of cultured samples.....	63
5.3 Multivariate exploratory analysis of spectra data	65
5.3.1 PCA of urine samples using EDXRF spectra.....	66
5.3.2 PCA of normal and cancerous urine using TXRF data	68
5.4 Multivariate analysis of micro XRF images of the cell cultures.....	70
5.4.1 PCA of Micro XRF spatial image data of cell cultures.....	70
5.4.2 Multivariate exploratory analysis of cell cultures	71
5.4.3 Hierarchical cluster analysis of urine samples	74
5.5 Speciation analysis of cancerous and healthy urine samples	77
5.5.1 Prediction of Fe speciation in urine test simulates by KNN model.....	81
5.5.2 Prediction of chemical oxidation states of Fe and Cu in human urine	81
5.6 Direct chemical speciation of trace elements by XANES.....	84
5.7 Quantitative analysis of trace elements in human urine.....	89
5.7.1 EDXRF analysis of trace biometals in urine samples	89
5.7.2 TXRF analysis of trace biometal levels in urine samples	90
5.8 Multivariate regression models for trace biometals in urine samples.....	93
5.8.1 ANN and SVM regression models for prediction of biometals in urine samples	93
5.8.2 Detection and quantification limits of trace biometals in simulate urine samples	99
5.8.3 Predicted concentrations in urine samples.....	99

5.8.4 Correlations of trace biometal levels	104
5.9 Multivariate regression models for trace biometals in tissue simulates.....	105
5.9.1 Detection and quantification limits of trace biometals in simulate tissue samples	110
5.9.2 Predicted concentration in tissue biopsy samples.....	110
5.9.3 Correlation of trace biometals in tissue biopsy samples.....	113
5.10 Synchrotron XRF quantitative analysis of trace biometals in cell cultures	115
5.11 Correlation of trace biometals	119
5.11.1 Correlations of trace biometals distributions.....	119
5.12 Imaging absorption of trace biometals in cell cultures	123
5.12.1 Micro XRF 2D mapping of trace biometals in cell cultures.....	124
5.12.2 The quantitative distribution of trace biometals in cell cultures	128
5.12.3 PC-ANN classification of cell cultures	129
5.12.4 PC-ANN classification of tissue biopsies.....	132
5.13 Trace biometal concentrations in prostate cancer stages	134
5.13.1 Relative values of trace biometal concentrations in cancerous human urine	135
5.14 Cancer diagnostic model	136
CHAPTER SIX: CONCLUSIONS AND RECOMMENDATIONS	139
6.1 Conclusions	139
6.2 Recommendations	140
REFERENCES	141
APPENDICES	157
Appendix I: Informed consent explanation.....	157
Appendix II: Urine collection and infection control procedure	160
Appendix III: Calibration set design for simulate tissue samples.....	166
Appendix IV: Calibration set design for simulate urine samples	167
Appendix V: TXRF analysis of concentration data of 20µl urine of prostate sample.....	168
Appendix VI: EDXRF concentrations of trace biometals in normal and cancerous urine samples	169

LIST OF TABLES

Table 4.1 Source of trace biometals.....	37
Table 4.2 Elemental concentration of the stock solution	38
Table 4.3 Spiking scheme of trace elements in tissue base matrix	39
Table 4.4 Source of trace elements of interest used in simulation.....	39
Table 4.5 Spiking scheme of trace elements in urine base matrix (urea solution).....	40
Table 4.6 Photon energy of K shell emission lines.....	46
Table 4.7 Spectral feature selection for 2D spectral data.....	47
Table 5.1:Validation of speciation prediction model of Fe.....	77
Table 5.2 Predicted Fe speciation in the test set of simulate samples.....	81
Table 5.3 Predicted chemical speciation of Fe in cancerous and non-cancerous urine	82
Table 5.4 Predicted chemical speciation of Cu in cancerous and healthy urine	82
Table 5.5 Trace biometals levels (ppm) in healthy urine.....	89
Table 5.6 Trace biometals levels (ppm) in cancerous prostate urine	89
Table 5.7 Concentration of cancerous urine by TXRF	90
Table 5.8 Concentration of trace biometals in 10µl normal urine by TXRF	91
Table 5.9 ANN and SVM urine calibration model performance	93
Table 5.10 Detection limits (ppm) of elements in simulate urine.....	99
Table 5.11 ANN Predicted concentrations of trace biometals in normal urine	100
Table 5.12 ANN predicted concentrations of trace elements in cancerous urine	101
Table 5.13 ANN cancerous urine of trace elemental concentration in the urinary bladder	101
Table 5.14 Trace elements descriptive statistics in healthy urine	102
Table 5.15 Trace elements descriptive statistics in prostate cancer urine.....	102
Table 5.16 Trace elements descriptive statistics in urinary bladder cancer urine.....	102
Table 5.17 Correlation coefficients of trace biometals in normal urine.....	104
Table 5.18 Correlation coefficients of trace biometals in prostate cancer urine.....	104
Table 5.19 ANN and SVM tissue simulate model performance.....	105
Table 5.20 Detection limits (ppm) of trace elements in simulate tissue	110
Table 5.21 SVM prediction of trace biometal concentrations in normal tissues	111
Table 5.22 ANN predicted Prediction of trace biometal concentrations in normal tissues	111
Table 5.23 ANN predicted Prediction of trace biometal concentrations in cancerous tissues.....	112
Table 5.24 ANN descriptive values of trace biometal concentrations in normal tissues.....	112
Table 5.25 ANN descriptive values of trace biometal concentrations in cancerous tissues	113
Table 5.26 Trace biometal correlation coefficients in healthy tissues	113
Table 5.27 Trace biometal correlation coefficients in cancerous tissues	114
Table 5.28 Concentration of trace biometals in healthy normal cell cultures.....	115
Table 5.29 Concentration of trace biometals in cancerous cell cultures.....	115

Table 5.30 Correlations of trace biometals in cancerous cell cultures.....	121
Table 5.31 Correlations of trace biometals in healthy cell cultures.....	121
Table 5.32 Correlation coefficients of biomarkers in cancerous cultured cells.....	121
Table 5.33 Semi-quantitative distributions of trace elements in healthy Vero cell cultures.....	128
Table 5.34 Semi-quantitative distributions of trace elements in cancerous cell cultures.....	128
Table 5.35 Concentrations of trace elements in cancerous prostate urine.....	134
Table 5.36 Ratios of trace biometal levels in cancerous samples.....	135

LIST OF FIGURES

Figure 1.1 Cancer mortality national Ranking at Ages <70 Years in 2019.	3
Figure 3.1 Schematic representation of X-ray incident beam interaction with atomic electron.	20
Figure 3.2 Schematic mechanism of characteristic X-ray generation.	21
Figure 3.3: Energy dispersive X-ray fluorescence instrumentation.	22
Figure 3.4 Synchrotron radiation beamline arrangement.	23
Figure 3.5 Synchrotron radiation X-ray absorption and fluorescence spectroscopy	24
Figure 3.6 Synchrotron radiation X-ray absorption spectroscopy	26
Figure 3.7 KNN for pattern classification (Sartoros and Salin 1997)	31
Figure 3.8 Artificial neural network architecture (Hernández-Caraballo and Marco-Parra 2003).	34
Figure 4.1 Visible light microscopy images (10x) for normal and cancerous cell cultures.	41
Figure 4.2 Energy-dispersive X-ray fluorescence configuration	47
Figure 4.3 Synchrotron radiation-based X-ray spectroscopy.	49
Figure 4.4 XANES experimental setup (Harfouche et al., 2022)	50
Figure 4.5 Schematic of an X-ray fluorescence microscopy (Zhang et al., 2018).	51
Figure 4.6 Schematic photo of TwinMic micro XRF experimental setup(Gianoncelli et al., 2016).	52
Figure 4.7 Visible light microscopy images (10x) for normal cell cultures	53
Figure 4.8: Visible light microscopy images (10x) for cancerous cell cultures.	53
Figure 4.9 A conceptual framework for a machine learning based XRFS model development.	58
Figure 5.1: Typical EDXRF spectra of cancerous and simulate urine samples.	60
Figure 5.2 EDXRF spectra of cancerous and non-cancerous urine	61
Figure 5.3 Typical TXRF spectra of cancerous urine samples	62
Figure 5.4 TXRF relative sensitivity graph of biometals.	63
Figure 5.5 Micro X-ray fluorescence spectra of DU_D1 overlapped with V_D1 for day 1	64
Figure 5.6 Synchrotron X-ray fluorescence spectra of DU_D1 cancerous cell culture.	65
Figure 5.7 PC score and loadings plot utilizing fluorescence region of urinary samples	66
Figure 5.8 PC score and loadings plot of fluorescence and scatter of human urine	67
Figure 5.9 PC score and loadings plots of urine samples using the TXRF fluorescence region	68
Figure 5.10 PC score plot of cancerous urine samples using the TXRF fluorescence region.	69
Figure 5.11: PCA score and loadings plot for cell cultures using the fitted fluorescence	70
Figure 5.12 PCA Score plot for DU cell cultures of selected Fe and Cu spectral intensities	71
Figure 5.13: PCA Score plot for cell cultures using fitted spectral data.	72
Figure 5.14: PCA score and loadings plot for DU cell cultures using fluorescence and scatter.	73
Figure 5.15 HCA classification of urine samples using fluorescence region	74
Figure 5.16 HCA analysis of urine samples using fluorescence and scatter region	75
Figure 5.17 HCA characterization of cancerous prostate urine samples using fluorescence region	76

Figure 5.18 Predictive scatter plot of Fe speciation of simulate urine by KNN	77
Figure 5.19 SVM classification of Fe speciation	78
Figure 5.20 KNN speciation analysis of Cu in simulate samples	79
Figure 5.21 SVM speciation analysis of Cu in urine simulates	80
Figure 5.22 Quasar orange model for SVM and KNN speciation of Fe and Cu.	83
Figure 5.23 Normalized XANES of Fe calibration sample foil.....	84
Figure 5.24 Typical EXAFS spectra in energy, k-space, R-space and q-space.	85
Figure 5.25 Normalized XANES of Fe in human urine.....	86
Figure 5.26 Overlapped spectra of reference Fe foil and urine sample	86
Figure 5.27 Normalized and overlapped XANES spectra of Fe in simulate and reference sample	87
Figure 5.28 Overlapped spectra of reference Fe foil and cancerous human urine samples	88
Figure 5.30 Simulate SVM and ANN model validation plots of Cu	94
Figure 5.31 Simulate SVM and ANN model validation plots of Zn.....	95
Figure 5.32 Simulate SVM and ANN model validation plots of Zn.....	96
Figure 5.33 Simulate SVM and ANN model validation plots of Mn	97
Figure 5.34 Simulate SVM and ANN model validation plots of Cr	98
Figure 5.35 Simulate SVM and ANN model validation plots of Se	98
Figure 5.36 SVM and ANN multivariate calibration model.....	103
Figure 5.37 SVM and ANN regression plots of Cu	106
Figure 5.38 SVM and ANN regression plots of Zn	107
Figure 5.38 SVM and ANN regression plots of Fe.....	108
Figure 5.38 SVM and ANN regression plots of Mn	108
Figure 5.38 SVM and ANN regression plots of Se.....	109
Figure 5.39 SVM and ANN regression plots of Cr.....	109
Figure 5.40 Bar graphs of Fe concentration in cancerous and normal cell cultures	116
Figure 5.41 Bar graphs of Cu concentration in cancerous and normal cell cultures.....	116
Figure 5.42 Bar graphs of Zn concentration in cancerous and normal cell cultures.....	117
Figure 5.43 Bar graphs of Mn concentration in cancerous and normal cell cultures.....	117
Figure 5.44 The correlation of Cu and Fe intensities in DU_D1 culture sample.....	119
Figure 5.45 The correlation of Cu and Fe intensities in V_D1 culture sample.....	120
Figure 5.46 Scan region for V_D3 healthy cell culture	123
Figure 5.47 Scan region for DU_D1 cancer cell culture.....	123
Figure 5.48 Micro XRF maps of Mn, Fe, Cu and Se in the growth media	124
Figure 5.49 Micro XRF 2D maps of Mn, Fe, Cu and Se in DU145 cancer cell cultures.....	125
Figure 5.50 Micro XRF 2D maps of Mn, Fe, Cu and Se in Vero healthy cell cultures	126
Figure 5.51 Spatial distribution of Cu and Fe at different stages of development in cancer cells.	127
Figure 5.52 PC-ANN analysis of cancerous cells using the entire fluorescence region	129
Figure 5.53 ANN fluorescence (Fe and Cu) scatter plot of cultured cells	130

Figure 5.54 ANN scatter plot of cancerous cultured cells by fluorescence and Compton scatter	131
Figure 5.55 ANN scatter plot of tissue biopsies by fluorescence and scatter	132
Figure 5.56 ANN classification of cancerous prostate urine by selected fluorescence and scatter	133
Figure 5.58: Machine learning aided XRFS cancer diagnostic model.....	137

LIST OF ABBREVIATIONS AND SYMBOLS

ANNs	Artificial Neural Networks
BPH	Benign Prostate Hyperplasia
DA	Discriminant Analysis
DNA	Deoxyribonucleic acid
EDXRF	Energy dispersive X-ray fluorescence
EDXRFS	Energy dispersive X-ray fluorescence and scattering
KNN	K-Nearest Neighbor
MAA	Molecular Activation Analysis
MCA	Multichannel analyzer
ML	Machine learning
LDA	Linear discriminant analysis
NAA	Neutron Activation Analysis
PCA	Principal Component Analysis
PIXE	Proton Induced X-ray Emission
PR	Pattern Recognition
ppm	parts per Million
ppb	parts per billion
SR	Synchrotron radiation
SR-XRF	Synchrotron Radiation X-ray Fluorescence
TXRF	Total Reflection X-ray Fluorescence
XANES	X-ray Absorption Near-Edge Structure
XRF	X-ray fluorescence
XRFS	X-ray fluorescence and scatter
SRIXE	Synchrotron Radiation Induced X-ray Emission

CHAPTER ONE

INTRODUCTION

1.1 Background to the study

Noninvasive XRF spectral diagnosis of disease is preferable because it is, together with treatment the two most important factors in medical care. The multivariate machine learning based X-ray fluorescence and scatter (XRFS) has potential to determine concentration, speciation and spatial distribution profiles of trace biometals for early disease (cancer) diagnosis. XRF analysis of biomedical samples is thus limited by the complexity (dark matrices due to low Z elements) of the samples which result in weak spectral signatures. In this study, XRFS is aimed at exploiting the fluorescence analyte signals masked in the pronounced Bremsstrahlung background and Compton Scatter peaks in conjunction with multivariate machine learning techniques.

1.2 X-ray fluorescence spectroscopy

The XRF analysis is a spectroscopic analytical method based on principle of radiation-interaction with matter through photoelectric effect (absorption and emission) and scattering (Compton and Rayleigh). Spectrometric X-ray fluorescence (XRF) enable rapid, direct and non-invasive method that enables simultaneous analysis of elements in variety of samples (Beckhoff *et al.*, 2007). In a typical sample XRF spectrum, XRF fluorescence and scatter peaks (due to low elements ($Z < 20$)) are predominant and are linked to enhanced background and increased Compton and Rayleigh scatter. The characteristic fluorescence lines are influenced by total analyte content but also the oxidation states of the analytes (Markowicz *et al.*, 1993, Van Espen 2002). However, the classical XRF analysis of biomedical samples is faced with challenges due to the dark matrix which gives rise to elevated background and prominent scatter region (Kaniu *et al.*, 2012). The classical XRF spectra in this context contain vital information but are limited by spectral overlaps, weak analyte signals and extreme matrix effects due to low-Z elements with complex interactions (relationship between trace elements and cancer) among various trace biometals. XRF is based on photoelectric phenomenon that generates energy spectra composed of various characteristic energies of the sample elements thus enhances simultaneous analysis of elements from Na to U (Carvalho *et al.*, 1998, Morona *et al.*, 2017). The conventional classical XRF spectrometry uses one fluorescent peak (normally the most intense with highest signal to noise ratio and

interference-free) and assumes a linear correlation between the characteristic intensities and concentrations of analytes. In practice, XRF spectra from dark matrix-dominated samples such as biological specimens consist of both fluorescence lines and scatter peaks as well as pronounced Bremsstrahlung which masks the weak analyte fluorescence signals. This poses a challenge in direct rapid XRF analysis of trace elements in body tissues and fluids.

X-ray fluorescence and scatter (XRFS) method is attractive in this regard as it has potential for utility of multiple X-ray fluorescence signals (regardless of interference) in conjunction with the scatter peaks to furnish accurate trace biometal profiles. Valuable additional information on morphology and sample density can be extracted from the Compton and Rayleigh scatter spectral region. However, the complexity and organic nature of biomedical tissue samples remains a major limitation in the application of XRF spectrometry. These constitute a multivariate problem that when tackled with powerful machine learning techniques can yield important information (concentrations and speciation) of trace elements in low-Z matrices (Angeyo *et al.*, 2012, Kaniu *et al.*, 2012).

1.3 Cancer survival and mortality

Cancer is a multi-factorial complex non-communicable disease as a results of cells growing out of control and dividing to form numerous abnormal cells (Kainth *et al.*, 2020). Further, cancer impairs the immune system and lead to chemical and structural changes in the body that alter the trace biometal spectral characteristic information obtained by X-ray spectrometry of body tissues and fluids.

Worldwide, about 19 million new cases of cancer were recorded and about 10.0 million reported deaths due to cancer occurred in 2020 (Sung *et al.*, 2021). Cancer is the third cause of death after infectious and cardiovascular diseases and accounted for 7.9 million deaths in 2009 and has increasingly had an upward trend in developing countries. In Kenya, cancer cases have been on the rise with resultant loss of lives and resources during treatment. Most of the cancer patients seen in Kenya are screened at an advanced stage when management and treatment are difficult and expensive (Nairobi Cancer Registry). Based on the prognostic features of the tumor, the available unreliable method (excision/incision-based histopathology) for genital urinary tract cancers involves optical microscopic examination of samples obtained through biopsy and fine needle aspiration. The burden of cancer in the world is projected to rise to 28.4 million new cases in the year 2040 (47% rise from 2020) as

a result of demographic changes which will also be further exacerbated by increasing cancer statistics (Sung *et al.*, 2021) as shown in Figure 1.1.

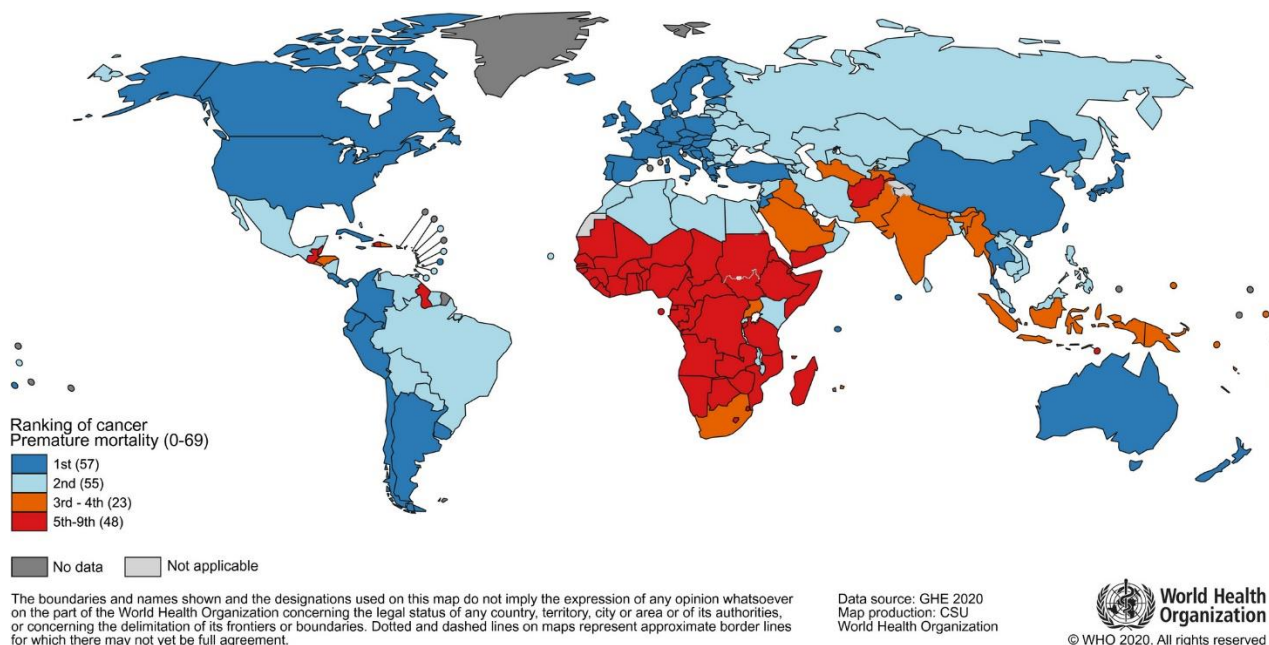


Figure 1.1 Cancer mortality national Ranking at Ages <70 Years in 2019 (Source: World Health Organization).

In Figure 1.1, bladder cancer recorded high cancer incidence which was four times higher in men as compared to women in developed countries. Prostate cancer ranks among the most common tumor worldwide where about 20% tends to metastasize (Navdaev and Eble 2011) and the second most prevalent malignant (21%) cancer for men while Benign Prostate Hyperplasia (BPH) prevails in men above the age of 50 (Leitão *et al.*, 2014). Although at average, 60% of world cancer cases were recorded with half of these cases reported in developing countries (Wagner and Brath 2012). The effective remedy for reduction of cancer mortality is through early diagnosis of cancer at the local stage of development. Urine and tissue trace biometal dynamics in onset and development of cancer is still challenging. There exists potential in developing a machine learning based XRFS for rapid, less invasive and objective method for diagnosis of cancer especially at local stage of development. It is not only imperative to correctly diagnose cancer, but also to determine its stage of development.

1.4 Trace Biometals in human body

The trace biometals are normally present in trace concentrations and when in abnormal concentrations, they reflect the various pathologies in human body (Guidotti *et al.*, 2008). The trace biometals are important due to their role in several physiological functions in

biochemical processes for the well-being of human beings and can be either beneficial or detrimental based on their concentration, speciation and distribution in human body. The trace biometals are either essential or non-essential with essential trace biometals vital for maintenance of normal physiological functions (Zaichick *et al.*, 1996). Essential biometals (Skalnaya and Skalny 2018) occur at low levels in human tissue (Koch *et al.*, 1956) for normal regulatory physiological processes (Prasad 2013, Rose 2016, Zaichick and Zaichick 2016). Further, elevated or deficiency of trace biometals may result to production of free radicals that is a precursor of cancer (Kainth *et al.*, 2020). It is thus reasonable to presume that elevated or depressed levels of trace biometals results to carcinogenesis (Imtiaz *et al.*, 2010). Essential trace elements are components to human health thus deficient and excess levels of trace elements may result in onset and development of diseases (Ates *et al.*, 2008). In normal concentrations, trace biometals stabilize cellular structures while in deficiency they stimulate pathways that cause diseases through activation or inhibition of certain biochemical reactions thus alter the cell membrane permeability.

Trace biometals are important in normal physiological function due to their potential role via activation, inhibition or promotion of enzymatic reactions (Silva *et al.*, 2009). They play vital roles by regulating the immune system, mitochondrial activity and membrane potential (Schubert *et al.*, 1998). The biochemical changes of trace biometals remains to be complex and that higher levels may be toxic whereas depletion may result to metabolic instabilities (Rao and Sciences 2005). As a result of biological alterations generated by the diseases (cancer), trace biometals levels and chemical speciation are altered. (Geraki *et al.*, 2002). Trace element changes may result in the development of cancer. There exist correlations between the level of trace biometals in human soft tissues and fluids, and the presence of cancer (Kwiatek *et al.*, 2002). The root cause of cancer in humans remains unknown but is possibly linked to genetic and epigenetic factors such as age, race and eating habits (Geraki *et al.*, 2002). These can all be linked to changes (or alterations) in biochemical profiles of trace biometals (Tan and Chen 2011, Aslam and Neubauer 2013). Timely diagnostics of impaired trace biometals in human body would aid in personalized treatment of cancer (Skalnaya and Skalny 2018). The spectrometric analysis of concentrations of trace biometal levels is of great importance due to their known roles (activators or inhibitors of enzymes) in biochemical activities in human tissues (Carvalho *et al.*, 1998).

Each fluid and tissue have unique characteristic biochemical composition that changes in response to pathological stimuli (for instance cancer). Therefore, tissue and body fluid

chemical compositions, the trace biometal speciation and distributions have the potential to determine the status of health. There is a need therefore to rapidly and non-invasively obtain information on the concentration, speciation and multivariate alterations of trace biometals and their distributions in tissues for early cancer diagnosis.

The utility of oxygen in human body can be associated many toxic effects associated with cancer in human body. Trace biometals (Fe, Cu, Cr and Mn) have the ability to catalyze Fenton reactions resulting to production of highly reactive hydroxyl radical (Kwiatek *et al.*, 2005). High concentrations of essential trace biometals (Cu, Fe, Mn, Cr and V) are greatly involved in Fenton reactions (Halliwell and Aruoma 1991) based on their speciation levels. Chemical oxidation states of trace biometals in body tissues and fluids occur in two forms (INAGAKI *et al.*, 2000); Protein-associated chemical species and protein non-associated chemical species that aid in excretion of the elements. The biological effects can be explained based on their potential to catalyze the production of free radicals as a result of decomposition of peroxides (Tariq *et al.*, 2016) which are linked to pathogenesis (Fraga and Oteiza 2002). A free-radicals with one or more unpaired electrons increases the concentration of the hydroxyl radical (*OH) which are active oxidizing agents that induce mutation of DNA (Aruoma *et al.*, 1989). The ability of increased free radicals to cause DNA mutation and damage of other cellular components is well documented (Teebor *et al.*, 1988). Reactive oxygen species are then generated (Kong *et al.*, 2022) and causes cellular damage when their levels increase and thus anti-oxidant defense mechanisms are overwhelmed.

In human body, Fe is always attached to ferritin proteins that stores about 4500g of Fe, while transferrin is responsible for Fe transport in human blood (Silva *et al.*, 2013). Fenton reaction for Fe consist of two stages (Halliwell and Gutteridge 1984);



The *OH free radical results to oxidative stress where the free radical generation exceeds the ability to eliminate and neutralize their effects.



Fe loosen weakly bound electron thus catalyzes free radicals which reacts with Deoxyribonucleic Acid (DNA) thus breaking of DNA chains for carcinogenesis(Silva *et al.*, 2013). Determination of the present of Fe (II) and Fe (III) may help to determine the extent of damage of genetic materials by free radicals.

Copper is a cofactor of Fe metabolism and also aid in free radical production and synthesis of melanin (Silva *et al.*, 2013) where Cu^{2+} aids in the formation of free radicals. It's clear that free radicals play critical role in the initiation, invasion and metastasis of diverse human pathologies such as cancer pathogenesis (Frenkel 1992) and (Girotti and Thomas 1984). Zinc is imperative in metabolic process in cell division and regulation of proteins that are actively involved in replication of the DNA (Silva *et al.*, 2013).

Their changes due to biochemical processes in cancer is directly associated with structural and metabolic changes (Silva *et al.*, 2013) thus 2D spatial mapping the trace elements can be correlated to carcinogenesis. The concentrations and chemical oxidation alterations of trace biometals and their correlations in human body tissues and fluids can thus be potential markers for cancer diagnostics.

1.5 XRF analysis of trace biometals for cancer diagnosis

The XRF is non-destructive method for analysis of liquids, solids and powder samples and relies on detection of the emitted characteristic secondary X-rays of atoms as a result of excitation (Pushie *et al.*, 2014). The characteristic fluorescence lines are influenced by total analyte content but also the oxidation states of the analytes (Markowicz *et al.*, 1993, Van Espen 2002).

Micro XRF can non-destructively provide both full spectral elemental maps and single-point spectra. Thus, micro XRF can explore the sub-cellular topography and speciation of biometals thus enhances the study of the roles of trace biometals in their structural context. Synchrotron radiation spectroscopic and imaging methods are the prime techniques for early cancer diagnosis as they are sensitive to changes in structure and composition of pathologically altered fluids and tissue. They provide automated, fast, non-intrusive and allow simultaneous sensitive detection and analysis of different biomarkers for characterization of tissues and fluids based on trace biometal profiles. Considering the sensitivity, spatial resolution and robustness required to rapidly and non-invasively elucidate the role of trace metals and biochemical in disease bioprocesses diagnostics at cellular level (as most diseases start at the level of single cells), novel analytical approaches are required which can provide additional diagnostic information on the chemical states of elements in microsystems of interest, as knowledge of bulk concentrations is insufficient. Microanalysis methods, combining spatial and spectral resolutions are the most attractive in this regard as disease results to structural changes in body tissue and/or fluid spectral images. Trace

biometals levels and their patterns/correlations are associated with cancer, and as the cancer progresses (severity increases) alterations can be detected in the levels and chemical speciation. The focus is to identify, quantify and map the trace elements and their speciation to determine their role in biopathological activity in cancer, including modelling their evolution with cancer (dynamic analysis and imaging) in body fluids and tissues. These will enhance non-invasive simultaneous analysis of trace biometals for cancer diagnostics at localised stage of development.

1.6 Multivariate machine learning for XRF analysis of trace biometals

XRF analysis of biomedical samples is faced with challenges due to the dark matrix which gives rise to elevated background and prominent scatter region (Kaniu *et al.*, 2012). Rapidly and directly acquired XRF spectra contain vital information but are limited by spectral overlaps, weak analyte signal and enhanced background due to low-Z elements with complex interactions (relationship between trace elements and cancer) among various trace biometals (Okonda *et al.*, 2022). The patterns are complex and multivariate thus are discernible by use of machine learning techniques for information mining. The goal is to develop machine learning enabled X-ray fluorescence and scatter methodology for simultaneous analysis of trace biometals' levels, distribution and speciation (as cancer biomarkers) for early cancer diagnostics.

Machine learning (in the domain of XRF spectra) involves application of supervised and unsupervised univariate and multivariate techniques for simultaneous extraction of important information inherent (Eriksson *et al.*, 2014). A novel machine learning enabled X-ray fluorescence and scatter (XRFS) methodology has the potential for identification and quantification of trace biometals and their speciation (Kaniu *et al.*, 2012). This provides greater sensitivity, versatility, multivariate modeling and exploratory analysis capability, speed, and possibility to obtain real time information for clinical purposes. Multivariate analytical tools have the ability to reveal hidden relationships (correlation of trace elements to cancer in tissues) thus can enhance utilization of characteristic (fluorescence peaks) and non-characteristic information (Compton and Rayleigh scatter) of the spectrum by maximizing extraction of relevant chemical information (concentration and speciation). Multivariate image analysis enables extraction of comprehensive information from the spectroscopic datasets in complex multivariate space (Brereton 2003) and has the ability to reduce the

(spectral) data complexity while maintaining the analytical information contents (underlying biomedical phenomena).

It is apparent that the complexity of the biomedical samples and data interpretation (the identity and distribution of the trace biomarkers does not appear in a straightforward way as their spectral characteristics overlap (Rosenberg 2002) constitutes is a multivariate problem. Machine learning may be used to address this problem as it has potential to reduce the dimensions of multispectral images and to extract analytical information. This may be realized via a variety of unsupervised algorithms such as cluster analysis, independent component analysis or supervised algorithms such as support vector machines, partial least-squares and artificial neural networks. Machine learning broaden the applicability and sensitivity of the proposed analytical methodology by enabling mining thus facilitating multivariate modeling of the trace biometal distribution. Further advantages gained include ability to reduce spectral noise, and interferences, and outliers.

The ultimate goal is to exploit the full potential of machine learning-assisted X-ray fluorescence and scatter (XRFS) spectrometry towards early cancer diagnosis based on the correlative speciation, concentration and multivariate alterations of trace biometals in human body tissues and fluid (urine). This is due to the increasing realization that successful treatment of cancer depends on accurate diagnosis of cancer at the local stage of development based on trace biometals as potential biomarkers. The method is aimed to provide needed cancer diagnostic information, its type and stage of development. The envisaged method has the potential for the combination of trace biomarkers of cancer that would have the ability to accurately diagnose and characterize cancer at cellular level as well as distinguish between normal, malignant and benign types.

1.7 Statement of the Problem

The current cancer diagnostic methods are inapplicable at an early stage of cancer development, due to their invasiveness besides being subjective thus results to challenging pathological interpretations and examination of tumor morphology. The limitation in early diagnostics of cancer is the lack of integrated approach that combines chemical with spatial imaging to perform non-invasive trace quantitative analysis in human body tissue and fluid microenvironments. Although the XRF trace spectroanalytical and imaging tools have high versatility, they are limited by the sample complexity and extreme matrix effects which result in weak spectral signatures as well as intricate interpretations of the analyte multivariate

spectral data sets. This generates a multivariate problem that requires adequate interpretation using powerful multivariate chemometric techniques. The proposed machine learning based XRFS has the potential to mine trace biometal concentrations, speciation and 2D mapping in complex human body tissue and fluids for cancer diagnosis.

1.8 Research Objectives

1.8.1 General Objective

To exploit machine learning for integrated X-ray fluorescence and scatter spectrometric analytical strategy for simultaneous identification, mappings, quantification of trace biometals and their speciation in urinary tract cancer pathogenesis.

1.8.2 Specific Objectives

The specific objectives of the study are;

- (i) To develop XRFS machine learning enabled multivariate calibration strategies (ANN and SVM) based on simulates for analysis of trace biometals concentrations in human tissues (urinary ladder and prostate) and urine.
- (ii) To quantify trace elements levels in selected urinary tract tissues (urinary ladder and prostate) and urine samples using the above developed calibration strategies.
- (iii) To determine the speciation levels of Fe, Cu, Cr and Mn in urinary tract tissues and urine samples utilizing the multivariate calibration and SR-XANES.
- (iv) To spatially map Fe, Cu, Mn and Se using machine learning enabled micro-XRF spectrometry in selected model body tissue microenvironments for cancer diagnosis.
- (v) To develop an integrated machine learning based diagnostic model for prediction of the status of urinary tract cancer (identify the presence and severity) in clinical samples utilizing the identified trace biomarkers.

1.9 Study Hypothesis

This study is hereby founded on the premises that;

(a) The development and metastasis of genital urinary tract cancer in human body tissue (urinary bladder and prostate) and urine is influenced by;

- (i) Concentration and alterations of trace elements.
- (ii) Chemical oxidation states of trace biometals.
- (iii) Correlative spatial distribution of Fe, Cu, Mn and Se.

(b) The age of the sampled patients does not affect the biomarkers levels in this study.

(c) Determination and modeling of the biomarkers' occurrence and multivariate relationships in human urine is associated with genital urinary tract cancers thus possible to track the type of cancer in soft body tissues and fluids.

1.10 Justification and significance of the Study

Histopathology as the “gold standard” method of assessing abnormal changes in tissues for cancer diagnostics is less rapid, invasive and also prone to subjective interpretations. Early cancer diagnosis thus remains a challenge due to the limitations of the above conventional diagnostic methods. Early and precise diagnosis of cancer enhances effective intervention. The envisaged method based on trace biomarkers of cancer has the potential to accurately prognose, diagnose and characterize cancer at cellular level at early stage of development. Successful diagnosis of cancer at local stage can thus be based on the combined analysis of the concentration of trace elements (Fe, Cu, Mn, Zn, Cr and Se) and their speciation (Cu and Fe) together with their correlations and multivariate alterations as cancer markers in soft body tissues and fluids (urine).

The XRFS analytical method is based on the determination of the trace analyte levels in the sample despite the matrix effects and relies on fluorescence profiles of heavy elements. The highly sensitive XRF in micro XRF and XANES modes using the more intense synchrotron radiation beam has the potential for utility in this study for determination of 2D maps and chemical oxidation states of trace biometals. XRF can simultaneously map out biometals without any use of dyes at high resolution with subsequent XANES analysis to provide rapid and non-destructive analysis of chemical speciation of biometals (Fe, Cu, Mn and Cr) at high resolution. Further, the EDXRFS spectrometry also exploits scatter peaks in addition to the

analyte fluorescence peaks for determination of cancer biomarkers. The XRFS method is not suited and applicable for direct rapid analysis of trace elements, their distribution and their speciation in low-Z matrices. These challenges are complex but they are discernible by use of machine learning aided techniques since they are subtle and the relationships multivariate. The adoption of machine learning (ML) based techniques for greater sensitivity, versatility, multivariate modeling and exploratory analysis enhance real time information for clinical cancer diagnostic. This proposed approach has the ability to provide necessary genital urinary tract cancer diagnostic information, type and stage of cancer development. The real quest for early diagnosis is to enhance early intervention to improve survival rate and successful treatment of the cancer at an early stage of development before invasion and metastasis to adjacent tissues.

CHAPTER TWO

LITERATURE REVIEW

2.1 Chapter Overview

This chapter contains five sections in which section 2.2 highlighting on the current conventional cancer diagnostics methods. Section 2.3 gives an overview of X-ray fluorescence (XRF) Spectroscopy with sub-sections emphasizing on XRF variants of trace analytical spectroscopic analysis. Subsequent 2.4 introduces other analytical for trace biometal analysis while section 2.5 presents the applications of multivariate machine learning techniques in XRF analysis.

2.2 Conventional cancer diagnostic methods

The current histopathological methods for urinary tract cancer diagnosis (biopsy and FNAB) are faced with challenges. Core biopsy and fine-needle-aspiration biopsy (FNAB) are the routine techniques for cytology and histopathology for urinary tract cancer diagnosis. These methods do not give definitive diagnosis of cancer besides not being cost effective (Griffiths *et al.*, 2007). The methods are painful, time consuming (as they involve special sample preparation) and expensive yet they do not give definitive diagnosis of cancer especially at early stage of development to enable effective treatment. Cancer is challenging not only in its early diagnosis and successful treatment but also in its complexity and increased mortality rate. The conventional cancer diagnostic methods are insensitive, laborious due to the considerable sample preparation involved thus unreliable for detecting the stage of cancer. It is important to not only diagnose cancer correctly, but also to determine its stage using subtle trace signatures analyzed at cellular levels.

2.3 X-Ray Fluorescence (XRF) Spectroscopy

X-ray fluorescence (XRF) spectroscopy enhances qualitative and quantitative analysis of elements in body fluids and tissues potential for disease (cancer) diagnosis. XRF is simple and rapid method with high sensitivity and relatively low detection limits of elemental concentrations. However, XRF is limited by matrix effects and poor detection limits for low Z elements ($Z < 19$) (Brouwer, 2003). XRF spectroscopy and its applications to samples of various kinds have attracted considerable interest due to its ability to directly detect and quantify simultaneously and non-destructively the elements in samples. Trace biometal

analysis by XRF spectroscopic methods enhance characterization of tissues and fluids for disease diagnosis. XRF spectroscopic and spectral imaging of trace biometals are sensitive to alteration in the structure and chemical composition of pathologically altered fluids and tissue hence can provide rapid and non-intrusive diagnostic characterization of tissues and fluids.

2.3.1 Energy dispersive X-ray fluorescence spectroscopy

EDXRF is based on photoelectric phenomenon that generates energy spectra composed of various characteristic energies of the elements thus enhances simultaneous analysis of analytes (Carvalho *et al.*, 1998, Morona *et al.*, 2017). The conventional EDXRF spectrometry uses one fluorescent peak (normally the most intense with highest signal-to-noise ratio and which should be interference-free) to perform quantification. However, important information cannot only be obtained by evaluation of characteristic emission lines of analytes (Henrich *et al.*, 2000) but can be extracted from the scatter spectral region. Further, EDXRF and X-ray diffraction enabled determination of the concentrations of Zn, Cu and Fe in cancer tissues for characterization of cancer (Geraki *et al.*, 2003). The potential utility of chemometric based Energy dispersive X-ray fluorescence (EDXRF) analysis for direct rapid analysis of Fe, Cu, Mn, Se and Zn in soft tissues as cancer biomarkers has been established (Okonda *et al.*, 2022).

2.3.1.1 Energy dispersive X-Ray fluorescence and scatter spectroscopy

Additional exploitation of the scatter (Compton and Rayleigh) in EDXRFS spectrometry coupled with chemometrics has the potential for cancer diagnosis. The EDXRFS aided multivariate modeling of cancer diagnostics utilizing selected fluorescence and Compton scatter profiles in model human tissues and cultures successfully classified cancer staging (Okonda *et al.*, 2022) using artificial neural network. In this study, low levels of Fe (101 ± 28 ppm) and Cu (21 ± 1 ppm) were found in SV10 cancer sample due to the lag phase stage of cancer development. Additionally, EDXRFS method utilized multiple X-ray fluorescence signals combined with the scatter peaks for quantitative analysis (Sichangi *et al.*, 2019) of trace elements in model tissue. EDXRFS thus exploits fluorescence analyte signals masked in the pronounced Bremsstrahlung background and Compton scatter peaks. An EDXRFS spectrum of a diseased tissue biopsy or fluid is mathematically a complex multivariate dataset that requires application of multivariate machine learning techniques. However, the scatter regions of similar sample matrices can be very informative on the overall sample composition when combined with a multivariate analytical tool.

2.3.2 Total reflection X-ray fluorescence spectroscopy

Total reflection X-ray fluorescence involves totally reflective mode where the X-ray beam strikes the sample at a grazing angle of 0.1° or less. TXRF is well-established spectroanalytical method elemental analysis in liquids and powdered microsamples (Beltrán *et al.*, 2020). TXRF has increasingly been utilized in trace biometal analysis in biomedical samples due to its capability for multi elemental determination (Klockenkämper and von Bohlen 2014). In this regard, it enables qualitative and quantitative analysis of elements by measurement of their spectral emission energy and intensity respectively (Van Grieken and Markowicz 2001, Silva *et al.*, 2012). The calibration by internal standard enhances its ability for simultaneous multi-element quantitative analysis in minute sample amounts (Gruber *et al.*, 2020).

TXRF has been utilized to determine trace biometals in serum specimen from healthy and cancerous human tissues utilizing Compton peaks as internal standard (Hernández-Caraballo and Marco-Parra 2003). No significant difference was observed in p values of Se levels in both healthy and cancer group. The concentration of Cu in cancerous and normal human serum by TXRF spectroscopy (Kubala-Kukuś *et al.*, 2014). The Cu levels in chemotherapy cancerous human serum was found to be 1.78 ± 0.91 ppm with a corresponding 108 ± 0.55 ppm in non-cancerous human serum. This results thus indicate the potential utility of levels of Cu in characterization of cancerous and non-cancerous serum applicable to cancer diagnosis. TXRF spectrometry has been employed to determine the concentrations of Ca, Ti, Fe, Cu and Zn in digested human breast tissues (normal and malignant) where multivariate discriminant analysis (MDA) enabled the tissue classifications into cancerous and normal groups (da Silva *et al.*, 2009). TXRF is therefore a non-invasive sensitive analytical method for multi-elemental analysis with lower detection limits for determination of trace element concentrations in human body tissues and fluids (Khuder *et al.*, 2007).

SR-TXRF has been used to determine Fe, Cu and Zn in biological human sample (Leitão *et al.*, 2014) where the results indicated suppressed concentrations of Fe and Zn in BPH compared to normal prostate tissue samples. Additionally, low levels of Fe, Zn and Rb were found in prostate cancer tissues suggesting their association with carcinogenesis. TXRF has been applied in analysis of chemical composition in human kidney stones (Kubala- Kukuś *et al.*, 2017) in which Fe, Cu and Se levels were determined and the elemental correlations aided in classification of kidney stones. The levels of trace biometals in healthy and cancerous human tissues (Benninghoff *et al.*, 1997) have been determined by TXRF

spectrometry. The statistical analysis demonstrated the possibility of using trace biometals in cancerous and normal human tissues for cancer diagnosis. The ratio of Zn to Cu and the levels of Cu were further determined as useful profiles for cancer diagnosis (Jiménez *et al.*, 2001).

2.3.3 Synchrotron radiation XRF spectroscopy

Synchrotron radiation is well suited to the quantitative as well as spatial analysis of biological samples at micrometer to millimeter scale (Szalóki *et al.*, 2021). Synchrotron's (Gianoncelli *et al.*, 2009) high brilliance, linear polarization and beam collimation (Castillo *et al.*, 2017) makes it highly suited for trace elemental analysis at cellular level. For these reasons, synchrotron radiation can extensively unravel the distribution and chemical oxidation profiles of elements at trace levels in human specimens (Gherase and Fleming 2019). Synchrotron beamline light sources generate a micro-focused monochromatic X-ray beam that has potential for speciation, quantification and imaging at high spatial and sensitivity (Khan *et al.*, 2019). The concentrations of trace elements in prostate tissues have been analyzed using more sensitive method such as synchrotron radiation (SR)-XRF where the concentrations of Fe and Zn were found to be lower in prostate cancerous compared to healthy tissues where the observed high levels of Mn and Fe were correlated to free radicals as precursors to carcinogenesis (Halliwell and Gutteridge 1984).

Synchrotron based X-ray fluorescence (A Castillo-Michel *et al.*, 2016) is best suited for analysis of trace elements at cellular and subcellular level. The highly sensitive μ -XRF and XANES variants of synchrotron radiation (SR) are suitable techniques for investigating the 2D mapping of trace elements and their speciation. X-ray Absorption spectroscopy and micro-spectroscopy SR methods with high resolution for trace biometals are most attractive in this context. This allows combination of micro X-ray fluorescence and XANES (A Castillo-Michel *et al.*, 2016) as the most powerful tools for mapping local structure and oxidation states of trace biometals. The highly sensitive XANES (Ogunlewe and Osegbe 1989) enhances direct determination of the chemical oxidation states (speciation) of the trace biometals. X-ray fluorescence in microscopy domain enabled simultaneous detection of multiple elements for 2D mapping of trace biometals in specimens (Kourousias *et al.*, 2015).

2.3.3.1 Synchrotron radiation based micro-XRF spectroscopy

Synchrotron radiation (SR)-XRF microimaging utilizes the flexibility of variable incident X-ray energy and advanced sample manipulator stages to realize rapid analysis of elements. Micro-spectroscopy combined with spatial distributions (for trace elements analysis) are attractive in this regard to disease diagnostics as it enables simultaneous analysis of trace elements (Kourousias *et al.*, 2015) in biomedical samples. Synchrotron X-ray fluorescence microscopy enabled 2D mapping of trace biometals at cellular level and also for quantification of their semi-quantitative analysis of their levels with submicron resolution (Fahrni 2007, Marmorato *et al.*, 2011). SR-XRF micro-spectroscopy can therefore be used for 2D and 3D spatial mapping of elements in trace biometals in sample specimen (Gherase and Fleming 2020). Synchrotron radiation micro-XRF enabled inhomogeneous distributions of Cu and Fe in human prostate spheroid cells using DU145 (cancerous) and RWPE-1 (normal) cell lines supplemented with zinc chloride (Rocha *et al.*, 2019).

Synchrotron radiation (SR)-based XRF has been used to determine the spatial mapping of Zn, Fe and Cu in incubated prostate cancer cells (Kawakami *et al.*, 2003) using Monte Carlo simulations in conjunction with X-ray diffraction. In this study, the levels of Fe and Cu were found to be elevated in tumor specimens compared to normal specimens. SR-XRF enabled determination of trace biometals concentrations for cancer diagnostics in breast tissues (Silva *et al.*, 2006). The levels of Fe and Cu were also higher in cancer tissues than in normal tissue thus a possible breast cancer diagnostic tool. The concentration of Zn was low in cancerous tissues as compared to non-cancerous prostate tissues. Similarly, the concentrations of Fe and Se were found to be in high levels in cancerous tissues (Banaś *et al.*, 2001). The increased levels of Mn and Fe are correlated to free radicals production which is a precursor to carcinogenesis (Halliwell and Gutteridge 1984). SR-XRF enabled determination of trace biometals in breast cancer patients' serum (Naidu *et al.*, 2019) where elevated levels of Fe, Cu and Pb were observed in cancerous with respect to normal subjects with suppressed levels of Mn, Zn and Se. Further, multivariate analyses on trace elements' data by discriminant analysis classified healthy and malignant tissues at 98.8% accuracy where the correlation analysis revealed significant different correlations Zn and Se in the two studied groups. Scanning X-ray fluorescence microscopic analysis enabled 2D mapping of Fe, Cu and Zn in cell cultures (Matsuyama *et al.*, 2009) where Zn and Cu were colocalized and Fe highly distributed in the nucleus of the biological samples. SRIXE has been used to analysis trace metals in cancerous tissues (Kwiatek *et al.*, 2004) where a negative correlation of Zn and Cu

was established in cancerous soft tissues. Similarly levels of Zn and Se were suppressed and not correlated with high concentration of prostate specific antigen (PSA) in prostate cancer tissues which tend to be anti-carcinogenic to prostate cancer (Platz and Helzlsouer 2001, Gray *et al.*, 2005). Micro XRF imaging enabled segmentation of elemental distribution images in stromal and neoplastic ovarian tumour and healthy tissue (Wróbel *et al.*, 2019). In this study, the non-negative matrix factorization and K-means clustering and were used for image segmentation for characterization of stromal, tumour or unclassified areas utilizing elemental (P, Mg and Zn) distribution maps as potential biomarkers for cancer diagnosis.

2.3.3.2 Synchrotron based XANES spectroscopy

X-ray absorption spectroscopy (XAS) technique enhances determination of chemical oxidation states by the highly sensitive synchrotron radiation XANES (Castillo-Michel *et al.*, 2017) due to utility of more intense synchrotron beams. X-ray absorption near edge structure (XANES) enhanced rapid determination of chemical speciation at high resolution and provides non-destructive analysis to provide chemical information on the elements present in the sample (A Castillo-Michel *et al.*, 2016, Aquilanti *et al.*, 2017). XANES quantitatively determined oxidation state (Jones *et al.*, 2020) of Fe where Fe K-edge absorption edge slightly shifted towards higher energies for Fe (III). Iron (Fe) as an abundant element occurs as either Fe^{2+} (ferrous iron) or Fe^{3+} (ferric iron) where Fe K-edge absorption edge slightly shifts towards higher energies when the chemical oxidation state increases (Jones *et al.*, 2020). Synchrotron-based XANES has the ability to determine the chemical oxidation states of biometals for disease diagnostics (Christensen *et al.*, 2004).

2.4 Other spectroscopic techniques for trace biometal analysis

Neutron activation analysis (NAA) (Zaichick and Zaichick 2016) has been utilized in an exploratory analysis of trace biometal concentration variations in cancerous prostate adenocarcinoma and healthy prostate tissues. The levels of Cr, Mn and Fe were high with suppressed levels of Zn and Se in cancer tissues compared to normal tissue sections. In this study, it is evident that malignancy significantly altered the levels of trace biometals in prostate tissue. Atomic absorption spectrophotometry has also been utilized in analysis of Cd, Ni, Co, Mn and Zn in urinary bladder cancer samples (Gecit *et al.*, 2011). High concentrations of Co, Ni and Cd were evident ($p < 0.05$) with decreased concentrations of Mn and Zn ($p < 0.05$). In this study, it was established that there was a correlation between the biometals' concentrations and bladder cancer occurrence. Graphite furnace atomic absorption

spectrophotometer (GFAAS) has been used to determine the levels of trace biometals (Zn, Cu, Se, Pb and As) in human urine (Lin *et al.*, 2009). The concentrations of Zn and Se were significantly higher in urinary bladder cancer patients. This study indicated that elevated Zn and Se can be linked to the proliferation of cells in urinary bladder cancer patients as they are excreted in urine.

It's apparent that the utility of the trace biometals' levels, their speciation and distribution profiles were not simultaneously determined thus the stage and types of cancers could not be identified. The above XRF and other spectroscopic methods are limited by weak spatial and spectral signals besides difficult analytical interpretation of highly-dimensional trace biometals' data for disease diagnostics. Effective cancer management calls for early detection and diagnosis preferably at the cellular level. Cellular microenvironments are complex with subtle spatial distribution of the trace biometals. Cancer diagnostics at the local stage of development based on trace biometals as biomarkers may be achieved by SR-XRF quantitative imaging at sub-micrometer resolution and low detection limits at cellular levels.

2.5 Machine Learning Techniques

The concentrations of cellular trace biometals are subtle thus the quantitative analysis requires to be ultra-sensitive analysis followed by robust multivariate analysis to process and extract the relevant analytical information that informs comprehensive diagnostic interpretation. Further, the identity and distribution of trace biometals in soft body tissues in their spectra or image domain is not straightforward despite the ability of combining trace analytical spectral sampling and imaging. The complexity of tissue structure makes the interpretation of spectral data difficult. The method is further limited by the weak spectral signals, spectral characteristics overlaps, analytical and interpretative challenges of the high-dimensional data. These constitute a multivariate problem which requires application of multivariate machine learning techniques for rapid mining and extraction of vital information from the complex data for greater sensitivity and versatility (Mitchell 1999).

Machine learning through self-adaptive learning has the ability to preprocess the data and can deal with peak overlaps and matrix effects hence its potential for indirect function relationships from robust results of large complex data sets (Luo 2006). Machine learning approaches deal with broad problem formulations (Kowalik and Einax 2006); reduction of large data sets into fewer dimensions for exploratory analysis, elimination of redundancy

information, visualization of multidimensional data and extraction of relevant chemical information.

Multivariate statistical tools have been used for characterization of cancerous and healthy lung samples (Ren *et al.*, 1997). Four multivariate chemometric tools namely; discriminant analysis (DA), PCA, cluster analysis and PR correctly distinguished the two groups based on the levels of Zn, Cu and Se in serum and hair samples. The discriminant multivariate analysis and decision tree induction (Cunha *et al.*, 2006) have been applied to the scattering profiles on XRF data in which the univariate discriminant variable was used to identify the differences between cancerous and normal breast tissues. Exploratory analysis enabled characterization and classification of breast cancer tissues (97%) which suggests an effective scattering biomarker for cancer diagnostics. XRF coupled with PLSR and PCA has been successfully applied to chromium speciation (De Oliveira *et al.*, 2010). Artificial neural network enabled characterization of human tissues into normal and cancerous where 3 layers were used with hyperbolic tangent and sigmoid functions hidden and output layers respectively which resulted to a prediction of 94% (Hernández-Caraballo and Marco-Parra 2003).

The conventional methods of trace biometals analysis are expensive and involve tedious sample preparation with a limitation to simple comparisons of concentration levels in biomedical samples. Hardly are any attempts directed in the exploitation of trace biometal levels, speciation and imaging profiles together with their multivariate alterations in relation to the biochemical composition of diseased tissues and fluids. These limit the applicability of the conventional methods to direct and rapid analysis of trace elements levels and their speciation alterations in human body tissues and fluids (urine) for reliable cancer diagnostics in multivariate domain. Additionally, less efforts in incorporation of multivariate or other advanced data analysis techniques such as machine learning to elucidate the role of the analytes in carcinogenesis. The complexity of the samples and data interpretation (the identity and distribution of the trace biomarkers does not appear in a straightforward way as their spectral characteristics overlap that constitutes is a multivariate problem. These patterns are complex but they are discernible by use of machine learning techniques since they are subtle and the relationships multivariate. In this study, machine learning techniques has the potential to address this problem as it has ability to reduce the dimensionality of multispectral images and to extract the subtle analytical information. These can be realized via a variety of unsupervised exploratory algorithms, artificial neural networks and support vector machines.

CHAPTER THREE

THEORETICAL FRAMEWORK

3.1 Chapter Overview

In this chapter, the basic principles of X-ray fluorescence spectroscopy and the classical quantitative methods used in XRF quantitative analysis of elements are presented. In addition, the principles of multivariate machine learning technique principles are discussed as well as their utility in multivariate calibration modeling of XRFS spectra.

3.2 X-ray fluorescence spectroscopic analysis

In 1895, Wilhelm Conrad Roentgen did observe radiation during his study of cathode rays in high-voltage where he discovered X-rays. XRF spectroscopy is thus based on the interactions between EM and matter. Later, Henry Moseley in 1913 did discover a mathematical relationship between the atom number (Z) and wavelength of emitted X-ray characteristics of different atoms. The X-ray fluorescence is based on the principle of radiation-interaction with matter through photoelectric effect (absorption and emission) and scattering (Compton or Rayleigh). The energy source (X-ray tubes or X-ray emitting radioisotopes) emits photons which excite and eject electrons in the sample (Bueno *et al.*, 2005). When an X-ray beam interacts with an atomic electron as shown in Figure 3.1, it is either absorbed or scattered where the absorption gives rise to the characteristic X-rays.

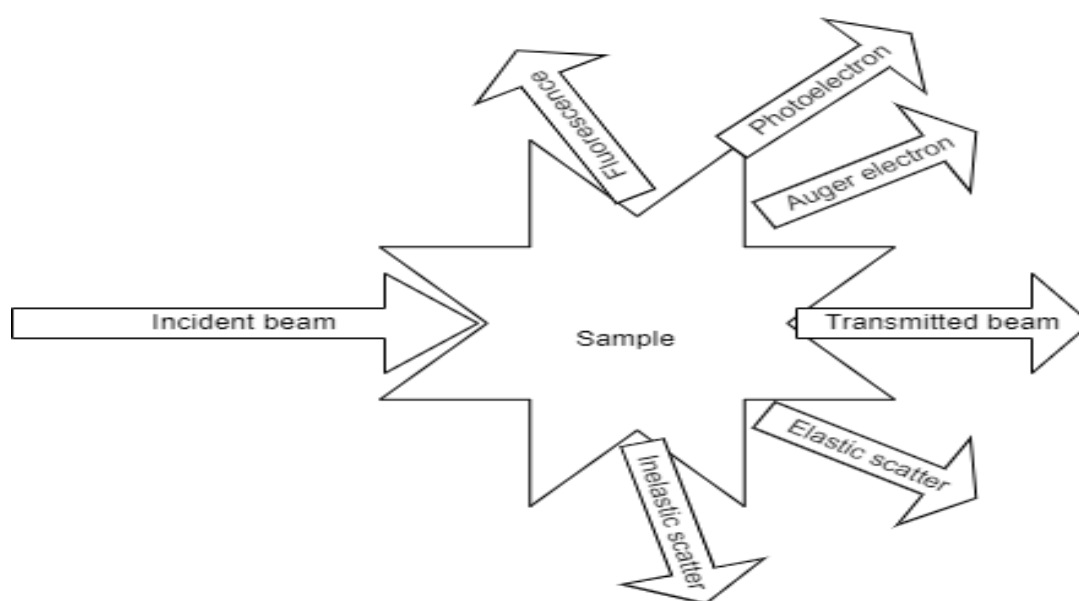


Figure 3.1 Schematic representation of X-ray incident beam interaction with atomic electron.

XRF is normally based on the absorption of incident X-ray photon by atoms that gives rise to photo-electron emissions that leave vacancies in the inner shell as shown in Figure 3.2.

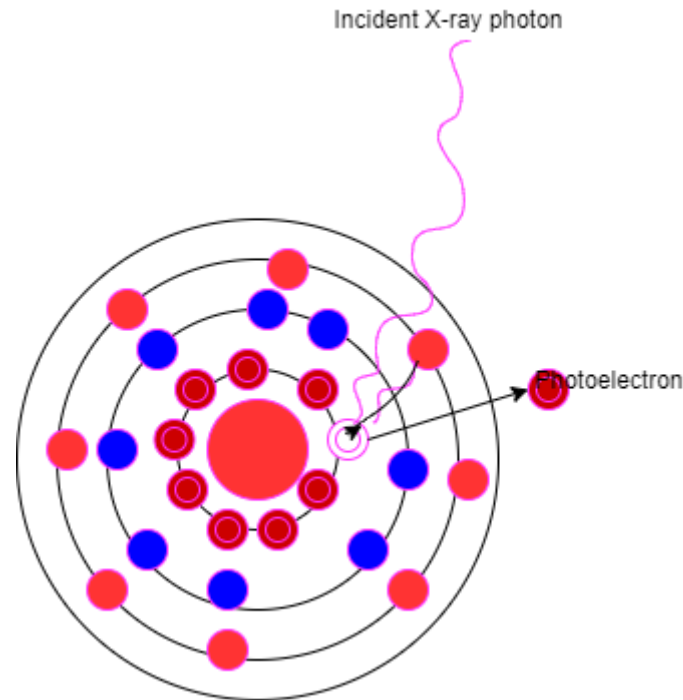


Figure 3.2 Schematic mechanism of characteristic X-ray generation

Higher shell electrons fill the vacancies through electronic transitions hence the characteristic X-rays are emitted. The characteristic energy and intensity indicate the element and its levels respectively (Markowicz 2011, Garg *et al.*, 2021). The transition results to an X-ray of fixed characteristic energy that is detected by a Si(Li) drifted detector resulting to K_{α} and K_{β} transitions. The qualitative and quantitative elemental information on sample specimens can be obtained through analysis of characteristic energy and intensity of emitted photons.

3.2.1 Energy dispersive X-ray fluorescence (EDXRF)

EDXRF instrumentation represented in Figure 3.3 is well-established approach for simultaneous and non-destructive analysis of elements in sample specimens (Chen *et al.*, 2008).

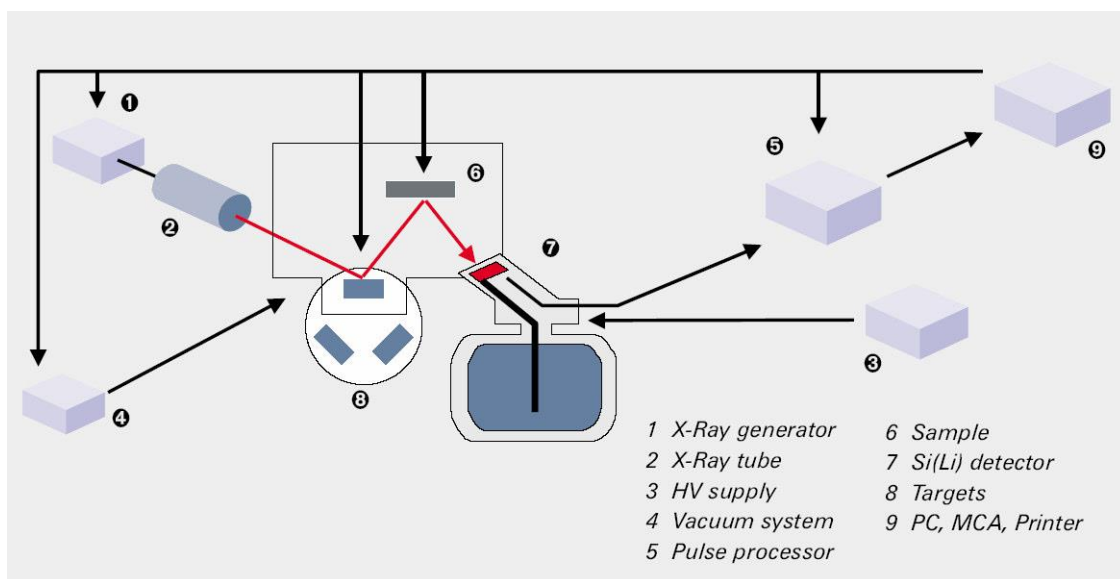


Figure 3.3: Energy dispersive X-ray fluorescence instrumentation

EDXRF enhance rapid, direct and non-invasive simultaneous elemental analysis in a variety of samples (Beckhoff *et al.*, 2007) where the fluorescence and scatter peaks (due to light elements $Z < 20$) are predominant. The detector measures the characteristic energies and intensities of the emitted X-rays which allows the identification of the element present in the sample and the elemental quantitative levels.

3.2.2 Total reflection X-ray fluorescence (TXRF)

TXRF is a surface sensitive non-destructive analytical technique (Schmeling 2019) that involves irradiation of a near optically flat sample with a beam of X-rays for microanalysis. In TXRF mode, the sample is placed on a flat substrate and scanned at small angles of incidence lower than 0.1° . TXRF quantification of elements is basically by internal standardization where addition of internal standard to determine the relationship between net count rate and analyte concentration (Greaves *et al.*, 2000). Additionally, the Compton scatter can be used as an internal standard (Yap *et al.*, 1988) to determine the levels of the analytes.

Excitation in total reflection geometry gives a significant reduction of the background and doubles the fluorescence that has the potential for analysis of trace elements (Nagata *et al.*, 2006). The large spectral background in conventional XRF as a major challenge is greatly reduced in TXRF spectroscopy. TXRF is thus a cost-effective spectrometric technique with penetration depth due to primary X-rays a few nanometers into the sample hence significantly

reduce the scatter background intensity in XRF spectra and enhanced sensitivity of analytes detected.

3.3 Synchrotron radiation based XRF analysis

A synchrotron is an electron-accelerating machine of a high radio frequency (RF) electromagnetic radiations attained by multi-MeV electrons moving in a magnetic field (Gherase and Fleming 2020). The synchrotron light source has a linear accelerator, a booster ring, a storage ring and beamlines where the X-ray beam interacts with the sample as shown in Figure 3.4.

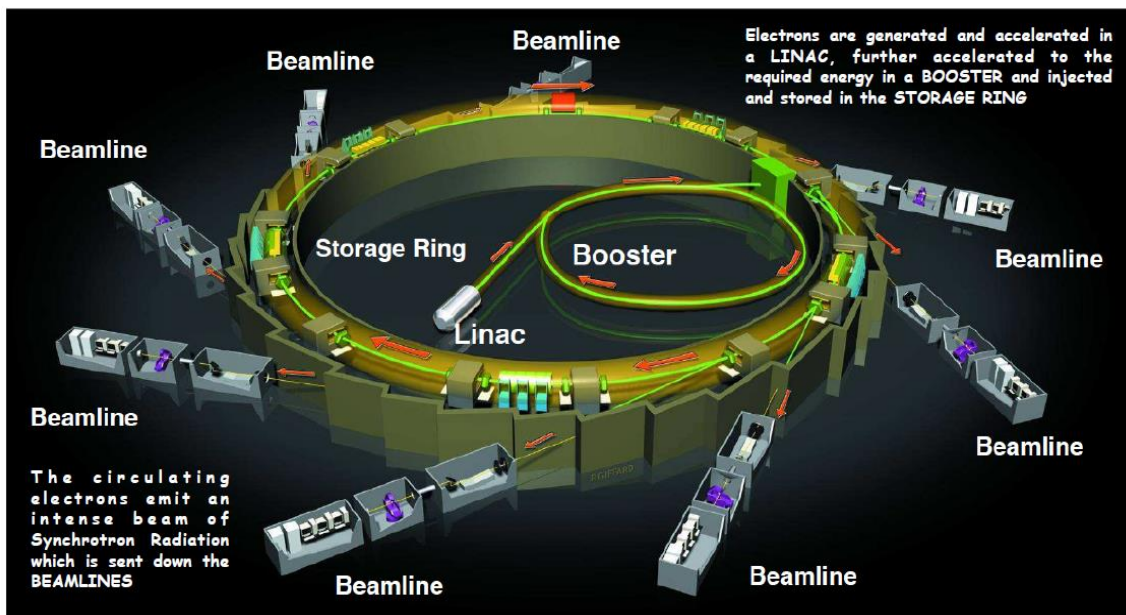


Figure 3.4 Synchrotron radiation beamline arrangement

In the linear accelerator, an electron gun fires electron that move relativistic as it enter a booster ring for further acceleration where it circulates with an optimized energy to accelerate the electrons (Bilderback *et al.*, 2005). The bending magnet together with the insertion devices (wigglers and undulators) in the storage ring bend (accelerate) the electron that results in production of synchrotron light. Beamline run off tangent to the storage ring to transport SR from the source to the sample. Beamlines are designed to utilize chromatic optics to focus beam onto the sample (Northrup *et al.*, 2016). The photons then travel at a tangent to the storage ring into experimental hutch where it is used for XRF analysis (Garg *et al.*, 2021).

The SR source is normally polarized with low emittance, high collimation, reliability in energy due to monochromatic emission (Kainth *et al.*, 2020). The synchrotron radiation

(SR)-XRF combines the flexibility of variable incident X-ray energy and advanced sample manipulator stages for sample analysis with detection limits below 10 ag and lateral resolution better than 100 nm (Silversmit *et al.*, 2009) thus can solve some of the shortcomings classical laboratory XRF approaches due to remarkable features (tunability, high flux, and linear polarization in the ring plane).

3.3.1 X-ray absorption spectroscopy (XAS)

A synchrotron-based XAS approach involves modulation of X-ray absorption coefficient over energies near and above the X-ray absorption edge. XAS measures the absorption coefficient of X-rays as a function of energy above the absorption edge of a given element (Aquilanti *et al.*, 2017, Aquilanti *et al.*, 2017) where an atom absorbing an X-ray of energy E which destroys a core electron at energy E_0 and creates a photo-electron with energy $(E-E_0)$. These results to the photo-electric effect of an electron from the absorption of an X-ray photon. XAS thus leads to change in linear absorption coefficient (μ) of an element in a sample (Liotti *et al.*, 2015, Terzano *et al.*, 2019).

XAS enables rapid data acquisition at high signal-to-noise ratio and measurements can be acquired at room temperature. The analysis is done at a synchrotron radiation facility with intense and tunable monochromatic beams of X-ray. Figure 3.5 shows how the XAS measurements are typically carried out based on the incoming beam photon intensity that is measured before the sample (I_0) and the transmitted intensity is measured after the sample (I_t) for different monochromatic photon energies besides the fluorescence emissions.

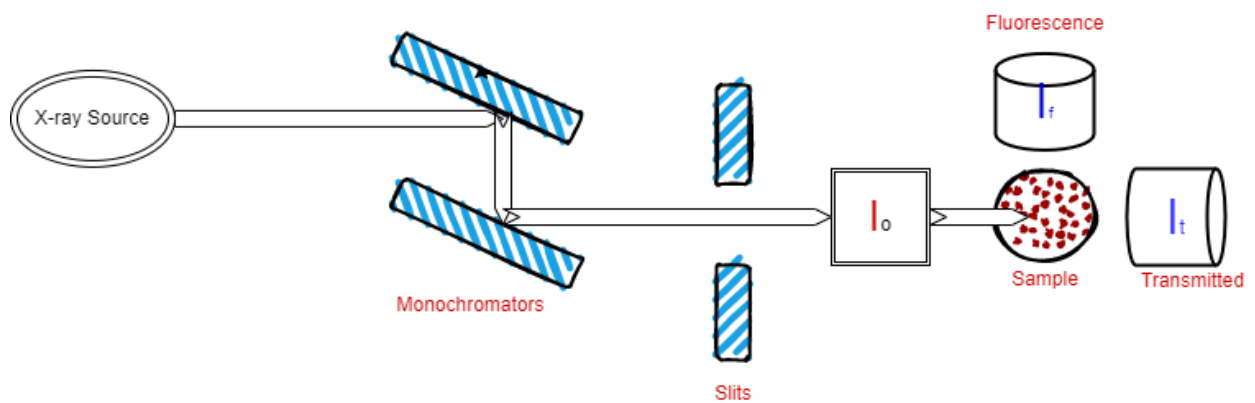


Figure 3.5 Synchrotron radiation X-ray absorption and fluorescence spectroscopy

For a sample of length (d), the absorption coefficient $\mu(E)$ is the probability of absorbed X-rays as per the Lambert-Beer's law; (Terzano *et al.*, 2019).

$$I = I_0 e^{-\mu(E)d}$$

3.1

Transmission measurement directly quantifies the attenuated beam through the sample as incidence beam energy is normally scanned across an elemental absorption edge.

Thus,

$$\mu(E)d = \ln\left[\frac{I_o(E)}{I_t(E)}\right] \quad 3.2$$

For X-ray fluorescence mode a solid-state multi-detector is placed perpendicular^o with respect to the incidence X-ray beam where the absorption coefficient is measured as a function of X-ray fluorescence yield I_f (Porcaro *et al.*, 2018) normalized by I_o is given by

$$\mu(E) \propto \frac{I_f}{I_o} \quad 3.3$$

In transmission mode, $\mu(E)$ is measured as a portion of the signal that is being absorbed while in fluorescence mode the intensity of the secondary process of photon re-radiation after absorption is implied (A Castillo-Michel *et al.*, 2016). Fluorescence intensity measurement quantifies the characteristic X-ray fluorescence emitted as a consequence of absorption (Northrup *et al.*, 2016). An edge results from a core electron absorbing energy equal to or greater than its binding energy. Synchrotron-radiation based XAS offers a direct method for chemical speciation analysis in a local chemical environment and occurs in two regimes: XANES) and EXAFS as shown in Figure 3.6 (Porcaro *et al.*, 2018).

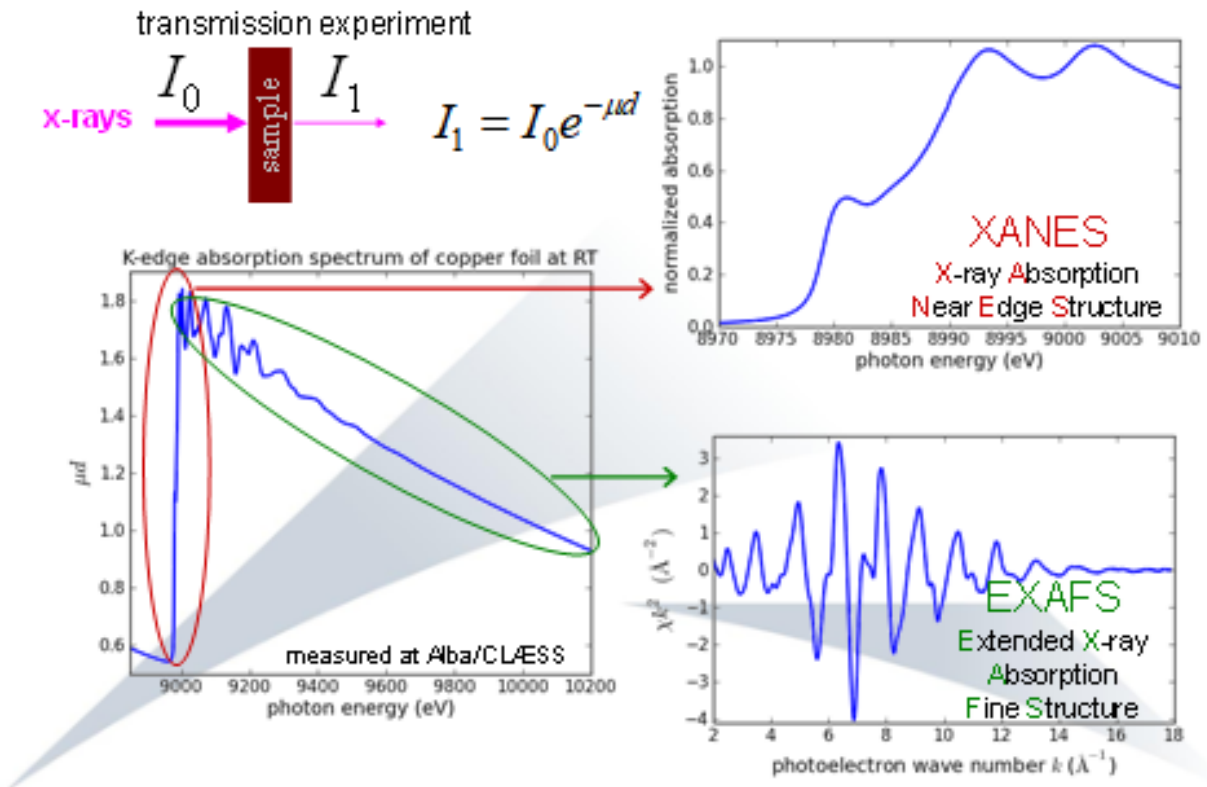


Figure 3.6 Synchrotron radiation X-ray absorption spectroscopy

XANES is sensitive to chemical oxidation states and coordination of the absorbing atom. Spectral structure closer to the absorption edge is dominated a small shifts in the absorption edge position depending on chemical oxidation state (Terzano *et al.*, 2019).

In EXAFS, the wave function due to excited photoelectron is modulated by interference of the outgoing wave and the backscattered wave on surrounding neighboring atoms. EXAFS results from backscattering of the photoelectron by neighboring atoms, and provides information about neighbor identities, the coordination numbers and interatomic distances (Terzano *et al.*, 2019).

Normalized modulation of absorption coefficient (Mastelaro and Zanotto 2018) is given by;

$$\chi(E) = \frac{\mu(E) - \mu_o(E)}{\Delta\mu_o(E)} \quad 3.4$$

Where $\mu(E)$ is the measured absorption coefficient, $\mu_o(E)$ is the absorption of a smooth function due to the background of isolated atom and $\Delta\mu_o(E)$ is the measured absorption jump at threshold energy E_o .

The corresponding wave vector of the photoelectron ejected due to the absorption of X-ray photons is described by (Mastelaro and Zanotto 2018);

$$k = \frac{\sqrt{2m(E - E_o)}}{\hbar} \quad 3.5$$

where E is the incident photon energy and E_o is particular absorption edge energy.

The EXAFS is basically described by dampened harmonic oscillation through the following equation (Porcaro *et al.*, 2018);

$$\chi(k) = S_0^2 \sum_j N_j \frac{|f_j(k)|}{kR_j^2} \sin(2kR_j + \phi_j(k)) e^{-2R_j/\lambda(k)} e^{-2\sigma_j^2 k^2} \quad 3.6$$

For each coordination shell j : R_j , N_j , s_j^2 are the sought distance, coordination number and variance of distance $f_j(k) = |f_j(k)| e^{i\phi_j(k)}$ is the scatter amplitude, λ is the electron free path (calculated), S_0^2 accounts for many-electron excitations.

3.3.2 Synchrotron based micro XRF

In order to probe the sub-cellular complexity of trace elements in relation to disease, sub-micrometer resolution at low detection limits is necessary. X-ray fluorescence (XRF) analysis with a submicron probe μ XRF spectroscopy is the most direct and sensitive method for such a task at the cellular level. Micro XRF (μ XRF) in this context with a spatial resolution significantly smaller than conventional XRF enables micron size sample analysis for elemental mapping elements. In this study, X-ray Fluorescence Microscopy (XFM) is aimed at quantitatively determining and mapping of trace biometals by specialized microprobes at synchrotron facilities for high spatial resolution. The XRF microscopy set-up can resolve spectral spatial information at high spatial resolution the biometal present at low energy X-ray regime.

3.3.3 X-ray absorption near edge spectroscopy (XANES)

X-ray fluorescence lines shift can be used to determine chemical oxidation state (Jalilehvand 2006). XANES is useful for probing the oxidation state and local electronic structure to identify and quantify chemical species and gives a direct measurement of chemical state

(valence state) of an element. Spectral structure close to the absorption edge is dominated by a combination of a small shifts in the absorption edge position depending on electronic (oxidation) state (Northrup *et al.*, 2016).

3.3.4 Extended X-ray absorption fine structure (EXAFS)

EXAFS measures the energy dependence of the absorption spectrum and enhances excitation of deep core electrons into higher unoccupied or continuum states. It enhances measurement of the energy dependence of the absorption spectrum coordination and elemental association hence the radial distribution of atoms around the photo absorber (Northrup *et al.*, 2016).

3.4 Quantification methods in XRF

This involves conversion of characteristic intensities measured from spectral analytes to the concentrations of elements based on excitation source intensity, overall composition and absorption properties (Markowicz 2011).

The conventional matrix correction methods are either by internal standard addition and the dilution method. Additionally, matrix effects are mathematically eliminated by means of linear regression method (fundamental parameters method and the influence coefficient method) and machine learning techniques (Qi *et al.*, 2015). The linear regressions-based models (fundamental parameter methods) and the influence coefficient method basically use the element measured to determine a regression equation without considering the roles of other major elements. The methods determines a regression relation from the element to be measured with little attention to the roles of other elements, especially major elements that have a great influence on analytical determination of trace elements for instance Fe causes significant increase in fluorescence yield of Co (Lu *et al.*, 2022).

3.4.1 Fundamental parameter method (FPM)

The fundamental parameters method is based on theoretical mathematical expressions for quantification of fluorescence emissions. The fundamental parameters include; mass absorption coefficients, fluorescence yields and characteristic X-ray energies (Beckhoff *et al.*, 2007). The method is built based on the Sherman equation 9 with consideration of both primary and secondary fluorescence of calibration and analytical steps and describes the relationship between the emitted intensities of the sample and its elemental content.

$$I_j = f(C_j, C_k, \dots, C_m) \quad 3.7$$

Where I_j is element's j net intensity with C_j as the concentration of j and $C_k \dots C_m$ as other elements' concentrations.

In the calibration step, elements in the pure sample are measured for determination of the relative intensities of analytes. In the second step, iteration compares the experimental spectrum with the theoretical result for determination of the concentrations of the analytes. This method is limited by the requirement to accurately know the spectral distribution of the incident beam.

3.4.2 Influence coefficients method

Influence coefficients quantify matrix effects individually, the matrix effect of element j , the matrix effect of element k , ... on analyte i . The elemental concentration of an element i can then be written as a function of its characteristic line intensity, corrected for the concentrations of other elements by influence coefficients.

$$C_i = R_i \left[1 + \sum_j \alpha_{ij} C_j \right] \quad 3.8$$

In which C_i is the concentration of the analyte, C_j expresses the concentration of other elements in the sample and α_{ij} is 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.

3.5 Machine learning multivariate analysis

Machine learning (in the domain of XRF spectra) involves application of supervised and unsupervised univariate and multivariate analytical methods for simultaneous analysis of data (Eriksson *et al.*, 2014). This enables extraction of comprehensive chemical information from the spectroscopic analysis in complex multivariate spectra (Brereton 2003). ML enables the representation of multivariate data into few dimensions to extract important information from complex multivariate. This is achieved by reducing the data complexity while increasing the information gained even in the presence of large uncorrelated spectral variations or noise. Biomedical sample analysis results to a multivariate analytical problem that requires application of multivariate techniques. The combination of spectro-analytical techniques with machine learning broadens the applicability for mining and extraction of useful information

from spectral data for disease diagnostics. The method surpasses current practice in that a set of advanced machine learning techniques to extract chemical information from the integrated spectroscopic images. In the domain of XRF, the application of supervised and unsupervised multivariate analytical methods has the ability to reduce the dimensions of multispectral images and extract the subtle analytical information (Mitchell 1997).

3.5.1 Pattern recognition techniques

Pattern recognition (PR) assigns objects or samples to a class based on similarity or dissimilarity in matrix composition (Marini 2010). PR is based on the fact that samples of similar chemical compositions tend to produce similar spectral shape when excited by X-rays. Pattern recognition is either pure classification or class-modeling technique (Marini 2010). In this case, each category is modeled separately where samples that fit the model belong to the class while those which don't fit do not belong to the class.

The XRF spectral data of materials with similar chemical composition tend to produce spectral pulse-heights of identical shape which is characteristic of the sample's chemical composition which can be recognized through pattern recognition (Yin *et al.*, 1989). The classification can either be by discrimination between classes or similarities within a class (supervised and unsupervised approaches). A significant portion of machine learning applications in analytical chemistry falls within the general framework of pattern recognition (Kowalski and Bender 1972).

3.5.1.1 Principal Component analysis

PCA is an exploratory method for extraction of information (variance) from large data sets thus creates a "window" in a multidimensional space where the matrix with several variables correlated to each other (Freitas *et al.*, 2010). It is the most utilized multivariate chemometric tool for exploratory data analysis and predictive modeling (De Leeuw 2011) to extract and visualize vital hidden characteristic information. The multivariate PCA model is developed by orthogonal (independent) basis vectors (eigenvectors) called principal components (PCs) which reduce the highly dimensional datasets to a lower dimensional space. Each PC summarizes the data sets by generating scores for the observations with corresponding loadings for the variables (Bortoleto *et al.*, 2005). These removes redundant information while retaining the most important information and differentiate samples by groupings (Romanenko and Stromberg 2007) for characterization of samples for similar or different

spectral characteristics.

PCA transforms uncorrelated variables that represent the structure of the original data set (Deming *et al.*, 1988) where the matrix X_c is decomposed to T and also loadings P by orthogonal computation until the variation correlated (Brereton 2003):

$$X_c = TP^T + E \quad 3.9$$

3.5.1.2 K-Nearest Neighbors (KNN)

KNN assigns a training set to known class by calculating the distance (Euclidean) of an unknown sample to the training set and ranks them in order. It then picks the k smallest distance and project the classes in which the unknown is closest to through the ‘majority vote’ and use this for classification (Brereton 2003). The Euclidean distance for the closest neighbours is given by;

$$d(a,b) = \sqrt{\sum_{i=1}^k (a_i - b_i)^2} \quad 3.10$$

Where a_i denotes the training set for the i^{th} sample and b_i is the test set and k indicate the number of neighboring points to consider when classifying a data point where the Euclidean distance between them is used to classify the sample. The test sample is normally assigned to a class with majority of the training set.

Popular other chemometrics techniques that will be used include pattern recognition include k-nearest neighbours (KNN) which predict samples based on the nearest training instances as shown in Figure 3.7.

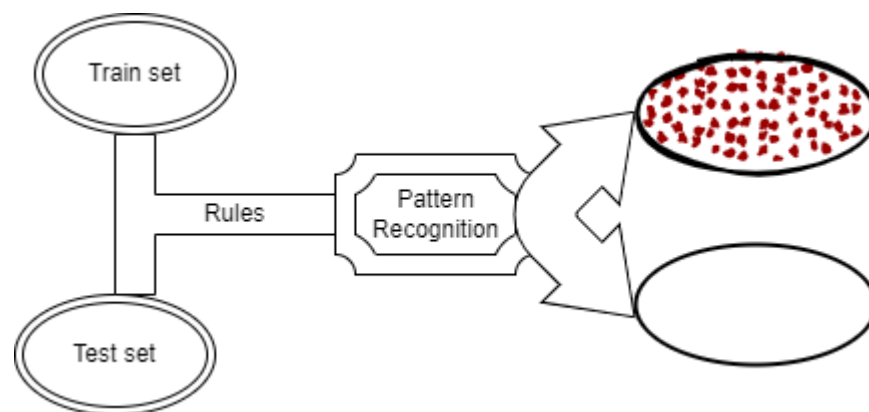


Figure 3.7 KNN for pattern classification (Sartoros and Salin 1997)

In this case, it will search for k nearest training examples in feature space and uses their mean matrix for prediction for instance average Euclidean distance between 2 samples (Wilson and

Martinez 2000) as follows. The k nearest neighbours are first selected and their frequencies determined and the class to which majority of training set belongs is then assigned to the test sample (Sartoros and Salin 1997). If a tie occurs, then the class with close neighbors is automatically selected.

3.5.1.3 Hierarchical Cluster Analysis

Hierarchical Clustering algorithms are data analysis techniques that allows partitioning of the data by a rooted (binary) tree (dendrogram) (Bueno *et al.*, 2005). Each cluster (starting with the whole dataset) is divided into two, then divided again, and so on. HCA partitioned the samples in clusters which in turn can be partitioned in sub-clusters and so on to a certain level to detect the hierarchical structure of the system and decomposes the data objects into tree of clusters.

3.5.2 Multivariate calibration strategy for quantitative analysis of trace biometals

Multivariate calibration involves mathematical modelling that reproduce concentration matrix Y from a matrix X for detection and modeling optimization for prediction of unknown sample of similar matrices (Nagata *et al.*, 2006). Multivariate calibration models enhance accurate linear and nonlinear prediction with good robustness. Neural-network or kernel based (Support Vector machines) are more reliable and flexible than linear models (PCR) for fitting the response calibration curve for quantitative analysis as they utilize non-linear approximations to model non-linear relations in the data (Olivieri *et al.*, 2011).

3.5.2.1 Support Vector Machines

Support vector machine (SVM) method that uses a kernel matrix to transform a non-linear separation to produce boundaries between classes (Xu *et al.*, 2006). SVM uses kernel matrix (radial basis, sigmoidal and polynomial functions) for data transformation to classify and quantify the attribute space by a hyperplane that maximizes on the margin between different classes (Xu *et al.*, 2006). They are used for both as a characterization and calibration technique for both classification and regression technique (Thissen *et al.*, 2004). This technique often yields a good predictive model where it shows superior performance compared to traditional classical methods applicable to non-linear problems for complex data (Thissen *et al.*, 2004). Support vector machine regression (SVMR) has the potential for optimization to map the data X to a higher-dimension space through non-linear mapping and regression (Xu *et al.*, 2006) given by;

$$y = \sum_{i=1}^l w_i \phi_i(x) + b$$

3.11

Where $\phi_i(x)$ is a set that maps input features, w_i and b are coefficients.

3.5.2.2 Artificial neural networks (ANNs)

Artificial Neural Network (ANN) involves mathematical modeling aimed at training, visualizing, and validating models (Fahlman 1988, Baffes 1989) which does not rely on assumption (Chang and Islam 2000). This involves combination of computer technology and biology to solve many problems such as process control, structure analysis and matrix calibration. The artificial neural networks (ANNs) contain artificial neurons inspired in the natural neurons in human brain. ANN algorithms (utilizing backward error propagation algorithm) for regression enhances prediction of the relationship (biometal concentrations) between its inputs and the desired output in model simulates (Luo *et al.*, 1997). Backpropagation algorithm in feed-forward enables the nodes to propagate their signals “forward” with errors propagated backwards (Rumelhart *et al.*, 1986). The nodes are computational units which receive inputs, and transform them into output by a complex process (single node might be in more than one network).

A natural neuron usually receives a signal via a synapses at the dendrites and when the signal is strong beyond a certain threshold, the neuron is then activated and emits a signal (Gershenson 2003). An artificial neural network has layered arrangement as shown in Figure 3.8 where the input layer supplies the data to hidden layer which intern propagates to the output layer.

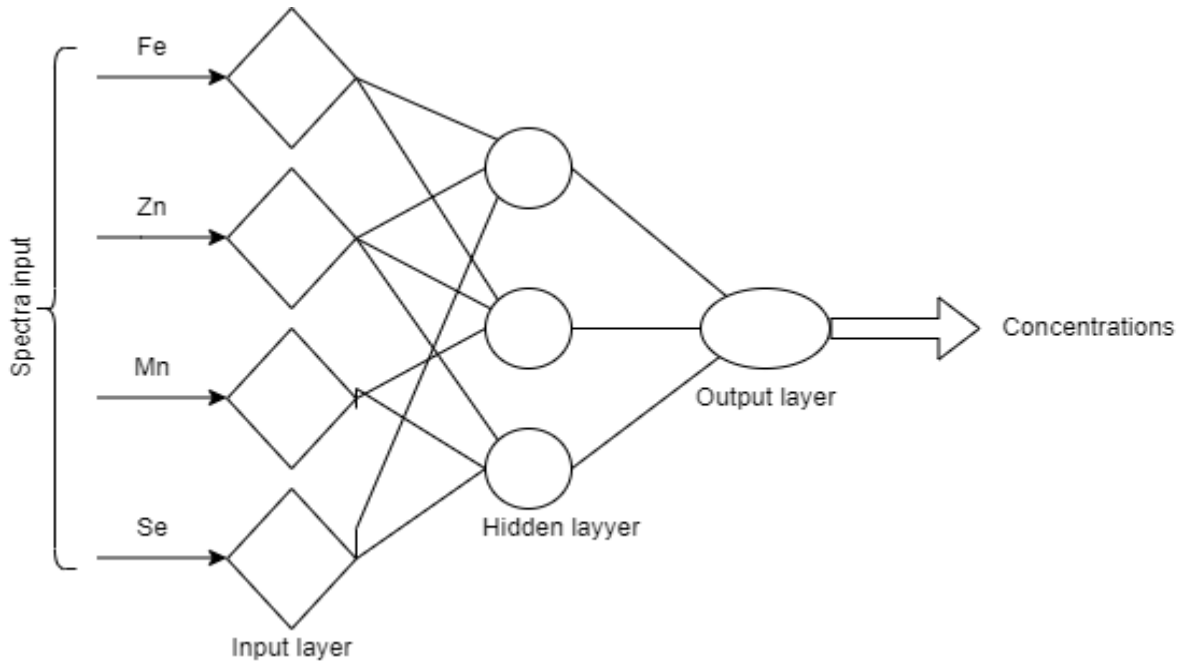


Figure 3.8 Artificial neural network architecture (Hernández-Caraballo and Marco-Parra 2003)

The input data (matrix) contain one row for each spectral image and one column for each measured independent variable upon which the calibration is based. The dataset is split into three subsets; say 65% training set, 25% for testing and 10% validation of the calibration models. The selection of the samples into subsets is made according to the principal component analysis (PCA) approach (Gershenson 2003). This algorithm utilizes supervised learning thus it's provided with examples of the inputs variables together with the outputs to be computed and the difference between expected and actual results (error) calculated (Heermann and Khazenie 1992). The tan-sigmoid and linear transfer functions are then used to determine the output (Pao 1989).

The spectral intensities of the training data are input into the first layer of the network. The input sum of x_i and the weights W_{ij} of threshold θ_{ij} (Luo *et al.*, 1997). The input is given by

$$Net_j = \sum (x_i w_{ij}) + \theta_{ij} \quad 3.12$$

The sigmoid transfer function used in this study is given by;

$$f(u) = \frac{1}{(1 + e^{-u})} \quad 3.13$$

The output is given by;

$$Y_j = f(Net_j) \quad 3.14$$

For node k , the output is given by the weighted sum as;

$$Net_k = \sum(Y_j W_{jk}) + \theta_{ik} \quad \mathbf{3.15}$$

The regressions are then built from the transformation by the sigmoid function given by

$$Out_k = f(Net_k) \quad \mathbf{3.16}$$

where Out_k is the concentration to be predicted from the function $f(x)$ and x the vector acquired from the spectrum. The Multivariate calibration can be obtained from $f(x)$ using the calibration simulates.

The artificial Neural Network multilayer perceptron (MLP) and the radial basis function (RBF) (Chakraborty *et al.*, 2000) with a single input layer, several hidden layers of neurons, and an output layer (Brown and Lo 1998) has the potential for optimal classifications. The known sample spectral data are then used for training and validating the ANN, which in turn uses it's learnt knowledge for predictive classification of unknown samples of similar matrices (Brown and Lo 1998).

CHAPTER FOUR

MATERIALS AND METHODS

4.1 Chapter Overview

This chapter contains ten sections that form the materials and method. Section 4.2 presents the reagents and materials used in method development while section 4.3 emphasizes on the sampling of simulate and real biomedical samples. Subsequent sections 4.4, 4.5 and 4.6 emphasizes on simulation process, cell culturing and urinary tract sampling. Section 4.7 highlight on tissues and urine sample preparation while section 4.8 gives quality assurance standards. Section 4.9 illustrates the XRF spectroscopic analysis while section 4.10 presents the data analytical procedure. Section 4.11 highlights on application of multivariate machine learning techniques for the development of cancer diagnostic kit in section 4.12.

4.2 Reagents and materials

To closely mimic the human tissue and urine, properties like absorption coefficient and scattering coefficients were taken into consideration and all the analar reagents were used in sampling, cell culture and sample preparation. Quartz glass as TXRF sample carriers were mechanically cleaned and washed in doubly distilled water and rinsed successively with doubly de-ionized water and acetone then placed in basic cleaning solution at 60 °C for 2 hours and then in 10% nitric acid for another 2 hours. The carriers were then placed in a drying chamber at 60 °C for further 2 hours to cool down to room temperature. After cooling, the quartz sample carriers were then siliconated at the centre with 10 μ L of silicone solution (Gruber *et al.*, 2020) for homogenous distribution of the sample. Nickel due to the fact that that it was part of the analytes and no detected in the samples in this study was used as TXRF internal standard for quantification prepared by diluting 1000 mg of Ni in 1 litre of doubly distilled water.

4.3 Sampling

During sampling, the integrity of the specimen was not be compromised before analysis by minimizing contamination during sample collection, preparation and storage. Human tissue and urine simulation process were intended to provide a simple and economic ways of preparing reliable multivariate calibration strategy for quantitative determination and imaging of trace biometals in human soft tissue and urine samples.

4.3.1 Simulate sample preparation

Synthetic simulates were prepared from analytical grade base reagents spiked with multi-elemental stock solution of elements of interest (Cr, Mn, Fe, Cu, Zn and Se). The two sets of multi elemental stock solutions with well-known concentrations were made using paraffin wax and urea as the base matrices to simulate soft human tissues and urine respectively as they mimic real sample properties of relevance to X-ray interaction such as absorption coefficient and scattering (Culjat *et al.*, 2010). The masses of the salt to be dissolved in were as shown in Table 4.1;

Table 4.1 Source of trace biometals

Source	Symbol	Molecular weight	% Assay	Trace element of Interest
Copper (I) Bromide	$CuBr$	143.45	98	Cu
Copper (II) Chloride	$CuCl_2H_2O$	170.48	98	Cu
Ammonium Ferrous Sulphate	$(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$	392.13	98	Fe
Iron (III) chloride	$FeCl_3$	162.2	99	Fe
Zinc Chloride	$ZnCl_2$	136.29	98	Zn
Selenium Dioxide	SeO_2	110.99	99	Se
Manganese (II) Sulphate	$MnSO_4H_2O$	169.01	98	Mn
Chromium Acetyl Acetonate	$C_{15}H_{21}CrO_6$	349	98	Cr

For 1000 ppm of trace analyte, mass m was calculated as;

$$mass = \left(\frac{1g}{Z} \times M.W \times \frac{100}{assay} \right) g / litre \quad 4.1$$

Where $M.W$ is the molecular weight of the salt and Z is the atomic number.

For 2000 ppm of trace analyte, mass m was calculated as;

$$mass = \left(\frac{2g}{Z} \times M.W \times \frac{100}{assay} \right) g / litre \quad 4.2$$

Stock solutions of the elements of interest in their respective chemical speciation were made as per Table 4.2.

Table 4.2 Elemental concentration of the stock solution

Analyte	Mass of the compound salt (g/1000 ml)	1 litre Stock solution (ppm)
Zn^{2+}	4.27	2000
Se^{4+}	2.83	2000
Cu^{2+}	2.74	1000
Cu^+	2.31	1000
Fe^{2+}	14.30	2000
Fe^{3+}	5.91	2000
Mn^{2+}	3.13	1000
Cr^{3+}	6.71	1000
Cr^{6+}	1.49	500

Simulate samples were prepared through spiking appropriate levels (Cu, Zn, Fe, Se, Cr and Mn) and speciation (Cu, Fe, Cr and Mn) of the elements of interest in which they occur in the human body tissue and urine. The volume of stock solutions corresponding to the selected levels of simulate samples was determined by;

$$C_1V_1 = C_2V_2 \quad 4.3$$

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 respectively.

4.3.2 Soft body tissue simulation

Model soft body tissue samples were prepared from analar grade paraffin wax as the tissue equivalent material (Bethesda 1989, Ferreira *et al.*, 2010). The base matrices were doped with Zn, Fe, Cu, Cr, Mn and Se from the stock solution using predetermined dilution factors in their respective chemical speciation and concentration ranges selected to span the wide occurrence of trace elemental compositions in human soft tissues (De Silva *et al.*, 2009) as shown in Table 4.3 and Appendix III for higher and lower speciation respectively.

Table 4.3 Spiking scheme of trace elements in tissue base matrix

Analyte	Simulate sample concentration (ppm)								
	S1T ^h	S2T ^h	S3T ^h	S4T ^h	S5T ^h	S6T ^h	S7T ^h	S8T ^h	S9T ^h
<i>Cu</i> ²⁺	2.0	10.0	14.0	20.0	35.0	30.0	27.0	40	50
<i>Zn</i> ²⁺	25	300	200	450	10	80	30	95	70
<i>Fe</i> ³⁺	32	24	65	80	180	200	300	350	420
<i>Mn</i> ⁴⁺	9.6	1	100	5	1.5	50	3	1.7	13
<i>Cr</i> ⁶⁺	0.5	15	10	0.6	1.5	0.8	0.3	1	2
<i>Se</i> ⁴⁺	0.5	1	2	1.6	5	10	3.5	7.5	9

Analyte	Simulate sample concentration (ppm)					
	S10T ^h	S11T ^h	S12T ^h	S13T ^h	S14T ^h	S15T ^h
<i>Cu</i> ²⁺	100	120	150	5	350	200
<i>Zn</i> ²⁺	105	150	55	250	40	130
<i>Fe</i> ³⁺	250	5	400	600	750	800
<i>Mn</i> ⁴⁺	100	90	2.5	4.5	7	3.5
<i>Cr</i> ⁶⁺	25	20	38	60	41	50
<i>Se</i> ⁴⁺	0.7	1.2	15	30	20	12

Sample size; 15

4.4 Urine Simulate

To closely mimic the human urine properties like absorption coefficient and scattering coefficients, all the analar reagents were added in about 100 ml of doubly distilled water and homogenously stirred for all the crystals to dissolve. The human urine majorly composed of water (95%) and urea (2%) was simulated (Putnam 1971) as shown in Table 4.4 below.

Table 4.4 Source of trace elements of interest used in simulation

Source	Chemical Symbol	Molecular weight	Mass (g)/litre
Urea	<i>CH₄N₂O</i>	60.06	15.000
Uric Acid	<i>C₅H₄N₄O₃</i>	168.11	0.250
Creatinine	<i>C₄H₇N₃O</i>	113.12	0.881
Potassium Chloride	<i>KCl</i>	74.55	2.308
Sodium Citrate	<i>Na₃C₆H₅O₇.H₂O</i>	294.10	0.720
Sodium Chloride	<i>NaCl</i>	58.44	1.756
Ammonium chloride	<i>NH₄Cl</i> ⁺	53.49	1.266
Calcium Chloride	<i>CaCl₂</i>	110.99	1.850
Magnesium Sulphate	<i>MgSO₄.7H₂O</i>	120.37	1.082
Sodium Sulphate	<i>Na₂SO₄</i>	142.04	1.700
Sodium dihydrogen phosphate dihydrate	<i>NaH₂PO₄.2H₂O</i>	156.01	2.912

Sodium phosphate dibasic dihydrate	$Na_2HPO_4 \cdot 2H_2O$	141.96	0.0831
Potassium hydrogen Oxalate	$COOKCOOH$	128.12	0.0035

Appropriate amounts of salts (all analar grade) were added and stirred until the solution was clear (Shmaefsky 1990). The urine base standard solutions were doped with Zn, Fe, Cu, Cr Mn and Se predetermined from the stock solution in their respective concentration ranges (Suzuki *et al.*, 2004, Majewska *et al.*, 2014) and chemical speciation in well selected for trace elemental compositions in human urine as shown in Table 4.5 and Appendix IV for higher and lower speciation respectively.

Table 4.5 Spiking scheme of trace elements in urine base matrix (urea solution).

Analyte	Simulate sample concentration (ppm)								
	S1U ^h	S2U ^h	S3U ^h	S4U ^h	S5U ^h	S6U ^h	S7U ^h	S8U ^h	S9U ^h
Cu^{2+}	0.5	2	1	2.5	0.8	5	0.7	10	7
Zn^{2+}	11	70	100	10	90	200	300	500	32
Fe^{3+}	2	1.5	5	20	1.8	2.3	3.3	2.5	30
Mn^{4+}	2	5	1	0.5	3	2.4	0.8	15	20
Cr^{6+}	2	3	7.5	4.7	2.5	6.0	5	8	10
Se^{4+}	1.8	2.7	3.5	3	5	1	2	6	1.2

Analyte	Simulate sample concentration (ppm)					
	S10U ^h	S11U ^h	S12U ^h	S13U ^h	S14U ^h	S15U ^h
Cu^{2+}	15	20	3	5.5	12	25
Zn^{2+}	600	40	30	15	9	5
Fe^{3+}	12	45	80	35	100	60
Mn^{4+}	1.5	3.5	2.5	1.8	4.5	25
Cr^{6+}	15	12	1.5	3.5	20	30
Se^{4+}	2.5	10	15	35	18	20

Sample size; 15

4.5 Cell Culturing

Prior to culturing process, the incubator was sterilized with 70% ethanol to minimize contamination. The cell culture substrate support robust cell growth as it is optically transparent for cell growth (Finney and Jin 2015). Silicon nitride membrane windows were used as substrate due to their low fluorescence background that supports robust cell attachment and proliferation (Carter *et al.*, 2010).

Three sets of cancer cell lines; DU 145, Hela 229 and Vero (control) each of aliquots 1 μ L cell suspension in culture growth media at density 1×10^6 cells/mL were seeded on 100 nm thick silicon nitrite windows as shown in Figure 4.1.

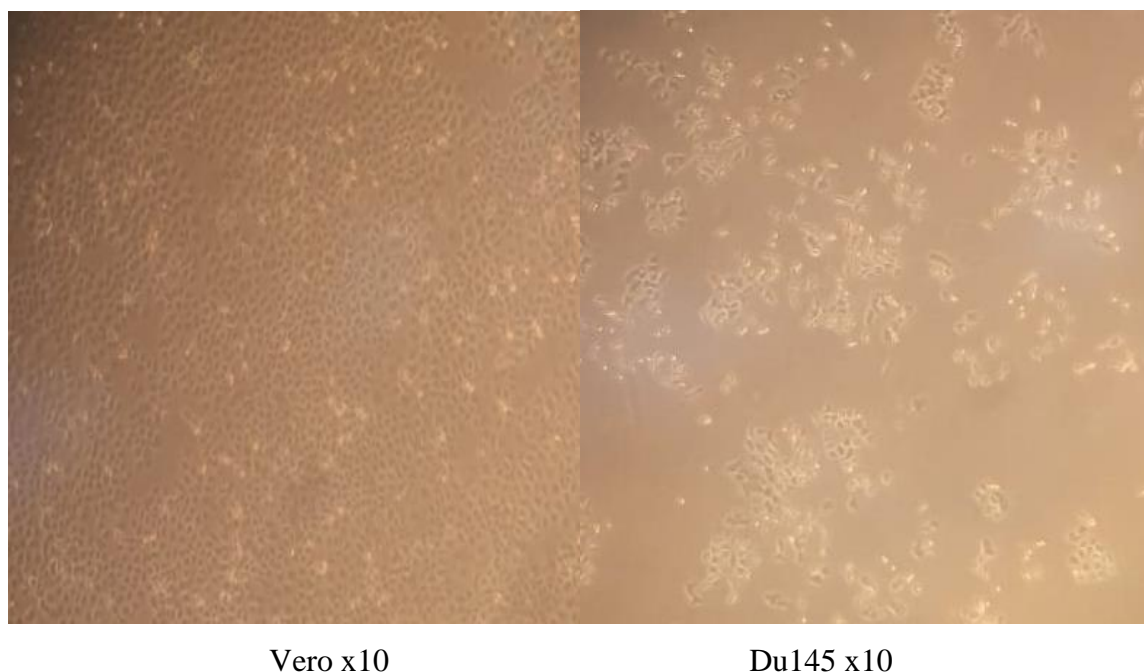


Figure 4.1 Visible light microscopy images (10x) for normal and cancerous cell cultures

Cells were cultured as a monolayer in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum and 1% of an antibiotic-antimycotic mixture (1%; Sigma), and trypsinized before culturing (Podgorczyk *et al.*, 2009). After 24 hours, the culture medium was replaced with fresh DMEM growth medium or complete essential (growth) medium supplemented with 10% foetal bovine serum (SIGMA-Aldrich), 100 U/ml penicillin-streptomycin solution (SIGMA-Aldrich), 10-mM zwitterionic sulfonic acid buffer (HEPES) and 1-mM sodium bicarbonate (SIGMA-Aldrich) (Pascolo *et al.*, 2014). The cultures were fed periodically by complete essential (growth) at 37°C and later harvested at interval of 48 hours for 10 days at Cell culture lab, Kenya Medical Research Institute (KEMRI). The cells were then sub-cultured through trypsinization (using Tripsin). During the entire culture period, the non-adherent cells were removed and adherent cells washed twice with phosphate buffered saline (PBS), and the medium was replaced. The cultures were harvested at intervals of 48 hours for 10 days, fixed in Paraformaldehyde (PFA), centrifuged at 1200 rpm for 6 min and the supernatant discarded before being secured on a 4 μ m Mylar foils in a spectral cup. The monolayer cultures were grown be grown directly a borosilicate microscopic cover slips suspended by 4 μ m Mylar foil spread on a spectral cup under same

culturing conditions. After the culture, the cells were fixed with 4% paraformaldehyde at room temperature for about 20 minutes and then washed with PBS for micro XRF imaging.

4.6 Urinary tract tissues and urine sampling

4.6.1 Ethical approval

This study was approved by Kenyatta National hospital/University of Nairobi Ethics Research Committee (P604/10/2017). The study conformed to Helsinki Declaration where participation was voluntary and informed consent (Appendix I) was given by participants. Patients diagnosed histopathologically with prostate and urinary bladder cancers but with no prior chemotherapeutic treatment were included in this study for urine sampling. The ultimate aim of the study was explained to the patients or their relatives before they gave an informed consent by signing the consent forms. The consented patients also gave additional demographic information (age, medical history, weight, and sex). For all the participants in the study, their confidentiality was strictly observed by ethically releasing all the results to participant individual file.

4.6.2 Urine sampling

Urine as a significant biomarker in disease diagnostics was used in this study due to its simplicity in collection, storage and preparation. As an attractive non-invasive sampling compared to the biopsy specimen, urine biomarkers enabled non-invasive approach for cancer diagnostics (Mahugija *et al.*, 2018). The participants were instructed on the standard way of voiding and sampling urine into a labelled biochemically clean universal tube. All participants or their relatives (in severely ill cases) gave an informed written consent (Appendix I).

4.6.3 Study population

The participants were recruited from Kenyatta National Hospital Centre at adult medical wards level seven and eight where the patients had been admitted for urinary tract cancer and those attending the comprehensive care centre (CCC-KNH). The patients who volunteered after a health talk about the study were included in the study after giving an informed written consent in the language (English or Kiswahili) they best understood. The urine specimens were taken from patients from January 2018 to December 2021. A total of 100 participants were recruited in the study. The histopathologically classified tissue biopsies were:

hyperplasia cases, well differentiated cases, moderately differentiated cases and poorly differentiated cases.

The target group for the study participants were those willing to participate in the study aged 18 years and above. With 200 cases expected to be reported in a year, the sample size for the study population will be calculated using the Fishers formulae method(Cochran 2007);

$$N = \frac{Z^2 pq}{e^2} \quad 4.4$$

Where p is the incidence, $q = (1-p)$, Z is the standard deviation of 1.97 at 95%, N is the sample size, p the estimated prevalence of an attribute present in the population at 6%, e the desired width of confidence interval and Z^2 the square of the normal standard deviation corresponding to 95% confidence interval.

$$N = \frac{1.96^2 \times 0.06 \times (1 - 0.06)}{0.05^2}$$

$$N = 87$$

4.6.3.1 Inclusion criteria

1. Patients diagnosed with no other significant diseases (diabetes mellitus, arthritis and liver malfunction) or malignancies but cervical, prostate and bladder cancers urinary tract cancer and aged 18 years and above.
2. Those who consented to participate in the study

4.6.3.2 Exclusion criteria

1. Patients who objected to be included in the study due to their own reasons.
2. Those diagnosed with other significant diseases (diabetes mellitus, arthritis and liver malfunction) or malignancies and aged below 18 years.
3. Patients that had already began cancer treatment (for instance chemotherapy).

Patients were given biochemically clean universal polyethylene bottles for urine sampling. (Yoshinaga *et al.*, 2000) to minimize contamination and absorption effects. The 5 mL of urine samples were filtered using a 5 mm membrane filter to remove debris and stored at -80 °C (Lin *et al.*, 2009).

4.6.4 Urinary tract tissue sampling

Tissue blocks from patients confirmed by histopathologist as having; prostate and urinary bladder cancer were included in this study. The FFPE blocks for patients diagnosed with no other significant disease or malignancies but cervical, prostate and bladder cancers without prior chemotherapeutic treatment. Human soft tissue blocks (Su *et al.*, 2004) were obtained from the KNH/UON tissue archives. The cancer tissue biopsy samples had been classified as hyperplasia and adenocarcinoma. The patients were categorized in two groups; the cancer group (70 patients with histopathologically confirmed adenocarcinoma cancer) and the control group (30 patients with benign hyperplasia). Paraffin embedded samples (non-cancerous and cancerous) were extracted from select tissue using circular biopsy punches (2 mm thick sections) with the help of microtome tissue cassette at KEMRI. The microtomed tissue section samples were fished onto 4 μm thick Mylar films stretched over plastic spectral cups and dried at room temperature. In order to avoid contamination, the tissue sections were not stained.

4.7 Sample preparation

4.7.1 Tissue simulates

Tissue sections 8 μm thick tissue sections were microtomed from the tissue blocks for XRF analysis.

4.7.2 Cell cultures

Prior to preparation, the cell culture samples were defrosted (Wei *et al.*, 2018) and about 0.5 mL pipetted into a 1.8ml vial. This solution was later spiked with appropriate amount of Ni (internal standard) in a 1.8ml vial obtained from the equation;

$$C_1V_1 = C_2V_2 \quad 0.5$$

About 10 μl of the homogenate solution was deposited at the centre of quartz sample carrier and dried on a heating plate at a temperature of 60° C analyzed as the dry residue using total reflection XRF (TXRF) analysis. In order to minimize contaminations, the plates were covered with glass dishes in a class 100 laminar flow box. Precaution was taken to minimize contamination during all sample preparation. Each sample was prepared in duplicate for TXRF and micro-XRF analysis.

4.7.3 Urinary tract tissues

Routine histopathological cancer tissue sample preparation had been carried to preserve the cell morphology of the tissues. Tissue sections (2mm thick) were microtomed from the tissue blocks for XRF analysis. Samples tissues were sectioned (5 μ m thick) from select tissue that was Paraffin embedded. The microtomed sample tissues were fished onto 4 μ m thick Mylar film stretched over a plastic holder.

The sectioned samples were fished onto 4 μ m thick Ultralene and Mylar film stretched over a plastic holder. Freeze-dried samples were kept at -80°C while sectioned unfixed samples were snap frozen in liquid isopentane.

4.7.4 Urine samples

Prior to preparation, the urine samples were thawed (Wei *et al.*, 2018) in a water bath for 15 min at 30°C . 1ml of the sample was pipetted into a 1.8ml vial and topped to 1.5ml. A calculated amount (10 μ l) of Ni standards (100 mg/L in 2% HNO_3) was added as internal standard and further mixed for 60 s for TXRF analysis. The mixture was shaken thoroughly mixed in the 1.8ml vial by shaking and 10 μ l of the homogenate solution was pipetted at the center of a quartz glass sample carrier and 10 μ l of concentrated HNO_3 added and then placed on a heating plate 60°C at a low pressure (400 mbar) to dry for TXRF analysis. About 10 μ l aliquot of urine sample solution was pipetted at the centre of the quartz-glass carrier and evaporated to dryness on a heating plate. The reflectors were then dried on a hot plate at 80°C for about 10 min. 10 μ l of each sample in Ni internal standard was pipetted on sample carriers and dried on heating plate for about 1 minute in order to remove the liquid matrix. After evaporation of the solvent, the residue was analyzed by TXRF spectrometer S2 PICOFOX for trace element determination under fixed glancing angle. The characteristic fluorescence radiations were then recorded by a Si (Li) detector as an energy spectrum. The high reflectivity of the sample substrate reduced the spectral background for low trace biometals detection limits.

4.8 Quality assurance

The retrieved FFPE tissue blocks and voided urine samples were assigned unique codes and matched for respective block number and patient's files respectively. The histopathology technologist was engaged in tissue sample sectioning and staining. The retrieved blocks were processed (Okonda *et al.*, 2022) while adhering to histological standard operating procedures

(SOPs) where the slides from best-preserved blocks containing adequate tissue were selected for sectioning. All reagents were prepared based on existing SOPs as per the manufacture's instruction.

4.9 X-ray fluorescence analysis

The XRF variant techniques that are non-invasive and reliable with minimum sample preparation (John *et al.*, 2001) were used for analysis of biomedical samples. Fully optimized XRF methodologies were utilized for simultaneous analysis of trace elements in human tissue, urine, simulates and tissue cultures. The input data matrix were intensities of the fluorescence and scatter where each row represented the spectral intensities and the column indicated the measured energy as independent variable upon which the exploratory and calibration analysis were developed. The quantification of trace elements were achieved through analysis of calibrated samples on the same sample substrate under similar irradiation conditions.

The detected characteristic X-rays fluorescence energy (Table 4.6) relates to characteristic fingerprints of the element with the corresponding intensity linked to the concentration of the element (Van Grieken and Markowicz 2001, Beckhoff *et al.*, 2007).

Table 4.6 Photon energy of K shell emission lines

Atomic number	Element	K α_1 (eV)	K α_2 (eV)	K β_1 (eV)
24	Cr	5,414.72	5,405.51	5,946.71
25	Mn	5,898.75	5,887.65	6,490.45
26	Fe	6,403.84	6,390.84	7,057.98
29	Cu	8,047.78	8,027.83	8,905.29
30	Zn	8,638.86	8,615.78	9,572.00
34	Se	11,222.40	11,181.40	12,495.90

Additionally, Synchrotron radiation X-ray fluorescence (SR-XRF) spatial data were used to analyze of trace elemental (using L lines as shown in Table 4.7) distributions in normal and cancerous human tissues and fluids for cancer diagnostics.

Table 4.7 Spectral feature selection for 2D spectral data

Biometal	Line series	Energy range (keV)
Mn	L	0.59-0.62
Fe	L	0.64-0.76
Cu	L	0.79-0.85
Se	L	1.27-1.49

4.9.1 EDXRF analysis of simulate and urine samples

The EDXRF spectrometer equipped with Cu, Al and Mo as secondary targets, 2 KW Mo tube and a Si (Li) semiconductor detector was used in this study. The samples were excited by X-ray from Rh X-ray tube where electrons were ejected from the core levels of the sample under investigation. The indirect excitation by use of Cu and Mo targets enhanced monoenergetic excitation that resulted to induced excited states for the emission of characteristic fluorescent that were detected. Figure 4.2 shows the EDXRF setup used with polarizing excitation where the radiations from the source (tube) were deflected by 90° to irradiate the sample with radiations polarized reduced spectral background.

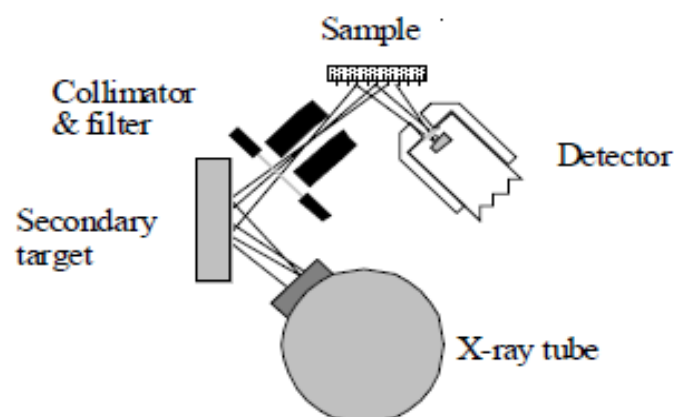


Figure 4.2 Energy-dispersive X-ray fluorescence configuration

Each soft tissue section on a Mylar film was placed directly in X-ray beam for elemental determination. The energy of the detected X-rays was characteristic of the element, while the intensity enabled determination of concentration of the characteristic element. The detector counted and sorted out intensities for all characteristic photons. A pulse height spectrum indicated the number of photons or impulses for a given energy established. The EDXRF

mode enabled simultaneous determination of elements, but is limited by insufficient sensitivity to low-Z elements and enhanced inter-element spectral effects.

4.9.2 TXRF analysis of urine samples

TXRF analysis were done at the institute of nuclear science and technology utilizing a high-efficiency S2 Picofox module (Bruker Nano GmbH, Berlin, Germany) with Mo X-ray tube. The spectrometer was first calibrated with KB multi element standard solutions before analysis. The sample residue on a quartz sample carrier was analyzed using TXRF spectrometer where the incident beam radiations will be reflected for a reduced spectral background for high fluorescence yield due to minimized absorption effects (matrix effects). An optically flat quartz containing aliquots of the digested (using nitric acid) and pre-concentrated substrate of the matrix enabled determination of concentration of trace biometals in complex sample matrices.

The sample substrates were analyzed by TRXRF through primary radiation that impinged the surface of the substrate at an angle less than 0.1° for total internal reflection mode on the surface at irradiation time of 200s. Spectral evaluation was achieved by spectra PicoFox software. The spectrometer and the associated computer were operated at a voltage (100-240V), frequency (50-60Hz) and power (150W). As a result, lower detection limits and low background were achieved and simultaneous multi-elemental information with low amount of sample acquired. The quantitative analysis using Nickel internal standard was achieved by the following relation;

$$C_A = \frac{N_A / S_A}{N_{IS} / S_{IS}} \cdot C_{IS} \quad 4.6$$

Where C_A is the concentration of the analyte, C_{IS} concentration of the internal standard, N_A net intensity of the analyte, N_{IS} net intensity of the internal standard, S_A relative sensitivity of the analyte and S_{IS} is relative sensitivity of the internal standard. The lowest limits of detection were calculated as;

$$LLD_i = \frac{3C_i \sqrt{N_{BG}}}{N_i} \quad 4.7$$

Where C_i is the concentration of the i^{th} element, N_{BG} background area under the fluorescent peak and N_i the area under the fluorescent peak intensity (Gruber *et al.*, 2020). In order to reduce the inhomogeneity in the sample preparation, each sample were analyzed in triplicate.

4.9.3 SR-XRF analysis

The highly linearly polarized Synchrotron beam was utilized in order to reduce the background and scatter signal for enhanced fluorescence peaks. Advanced synchrotron-based X-ray fluorescence (XRF) in Figure 4.3 enhanced determination of quantitative levels, 2D distribution and chemical oxidation states of trace biometals in cellular microenvironments.

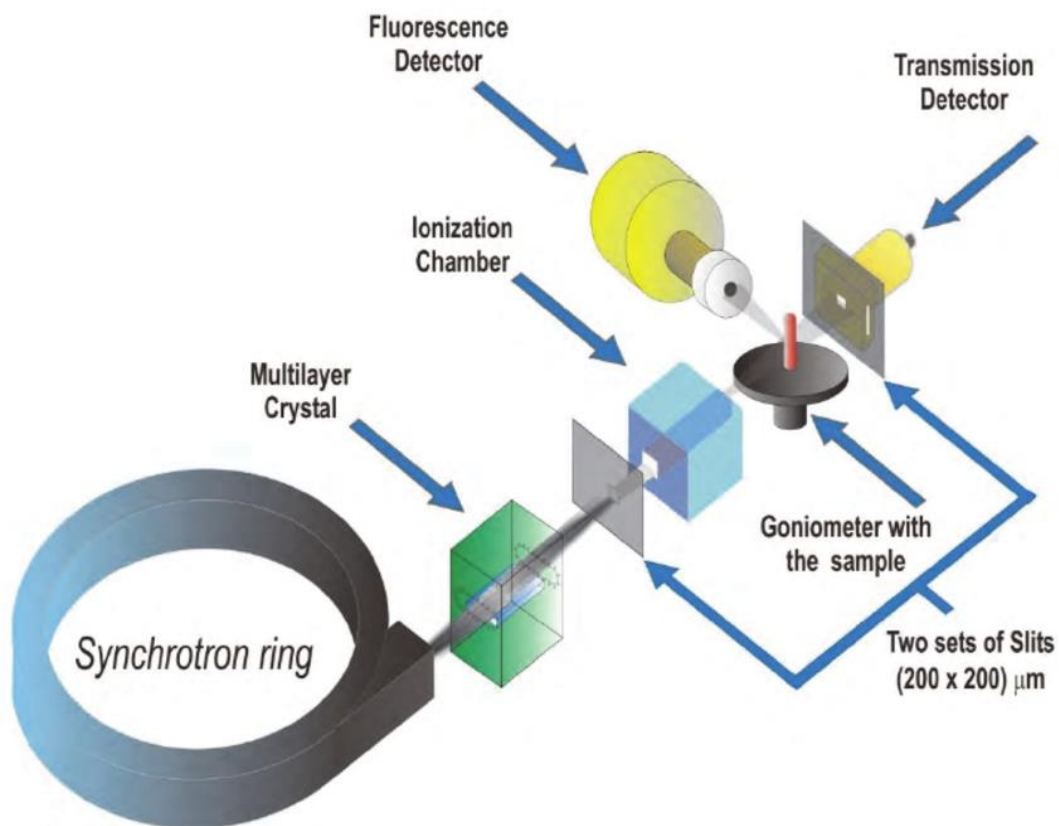


Figure 4.3 Synchrotron radiation-based X-ray spectroscopy

Due to the SR-XRF high sensitivity of trace biometals, the 2D spatial maps and chemical oxidation states of trace biometals were determined in body fluid and tissues for disease (cancer) diagnosis. The samples were placed on a goniometer translation stage for rotation as well as translating it normally to the incident beam was raster scanned across the sample at spot sizes in the range of 1 to 10 μm for XANES and micro-XRF analysis.

4.9.3.1 XANES analysis of cell cultures

The XANES analysis was enabled by tuning the X-ray beam energy at the binding energy for an absorption edge. The Fe K-edge XANES spectra were analyzed at XRF beamline at Synchrotron-light for Experimental Science and Applications in the Middle East (SESAME). X-ray absorption near edge structure (XANES) schematic arrangements at SESAME (Jordan) in Figure 4.4 at lower energies enabled acquisition of Fe oxidation state for cancer diagnosis.

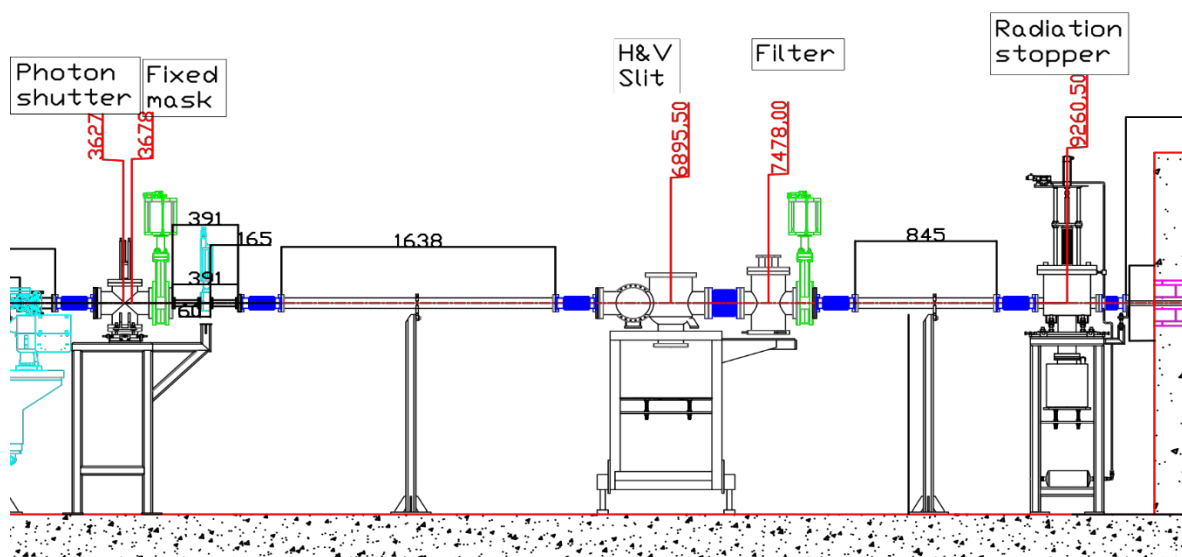


Figure 4.4 XANES experimental setup (Harfouche *et al.*, 2022)

For maximum fluorescence, the incident beam was tuned to 500 eV above the absorption edge of Fe at a step size of 0.2 eV at the XANES. The analysis was performed in fluorescent mode (due to dilute nature of the samples) with the sample at 45° to the incoming beam which resulted to improved fluorescence signal (Geraki *et al.*, 2004).

Urine Samples on microscopic cover slips were fixed onto a sample holder using Kapton tape and placed within the synchrotron beam. The analysis was carried out in air and at room temperature and the acquisition time for set to 10s for enhanced signal-to-noise ratio for three energy scanning regions in the broader energy range covering the Fe K-edge from 7100 to 7200 eV with the step of 1 eV. The XANES spectral data were preprocessed by the ATHENA part of the DEMETER software package (Di Cicco *et al.*, 2009). Fe reference calibration foil sample enabled determination of the oxidation state of Fe in the sample species. The XANES spectra of Fe reference foil and urine samples were then analyzed in fluorescence mode and the spectra normalized using ATHENA part of the DEMETER software package.

4.9.3.2 Micro XRF analysis of cell cultures

Micro XRF analysis were performed at the TwinMic beamline at ELETTRA synchrotron source in Trieste, Italy (Kaulich *et al.*, 2006) where the microscope scanning module was used and X-ray microprobe by Zone Plate focusing optics were utilized. Linearly polarized synchrotron beam reduced the scatter effects by the detector. Further, a perpendicular geometry between synchrotron beam and detector as shown in Figure 4.5 resulted to reduction of the scatter signal which improved the fluorescence signal and reduced irradiation time.

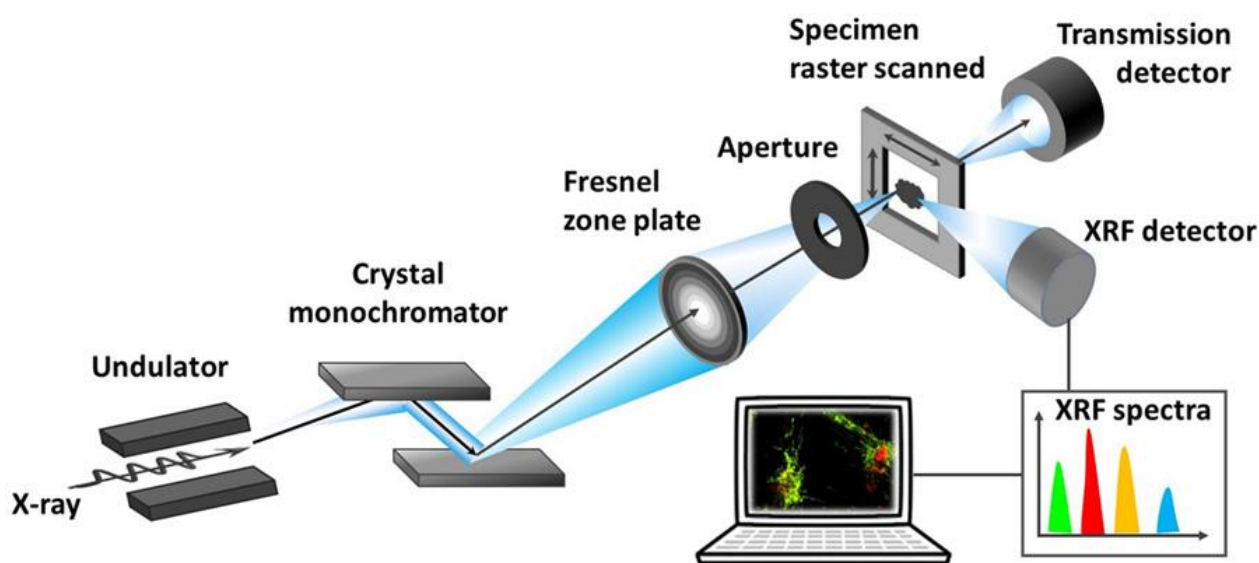


Figure 4.5 Schematic of an X-ray fluorescence microscopy (Zhang *et al.*, 2018)

The low-medium (2200 eV) photon energy range (Gianoncelli *et al.*, 2009) set-up enabled elemental analysis at high spatial resolution. Figure 4.6 shows the TwinMic micro-X-ray fluorescence setup used for rapid simultaneous acquisition of 2D maps of trace elements through refocusing of chromatic zone plate (ZP) lens setups at an energy range of 2200 eV.

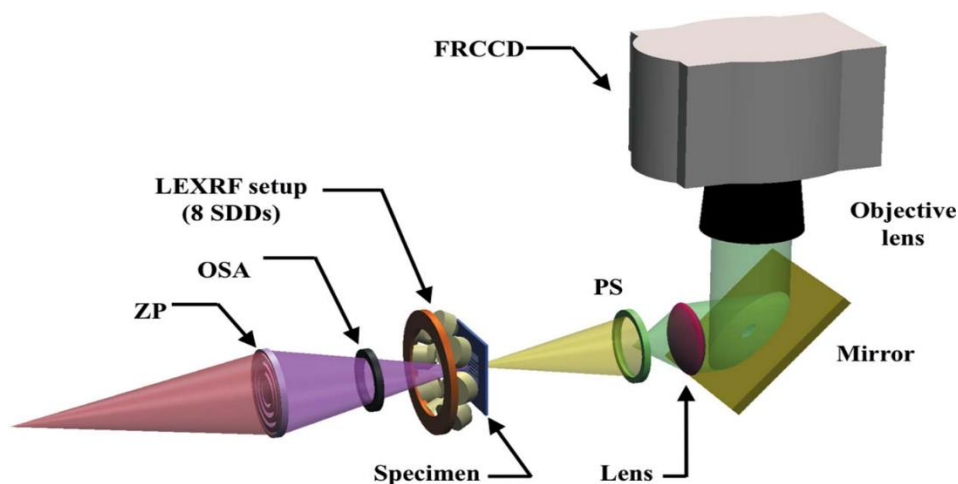


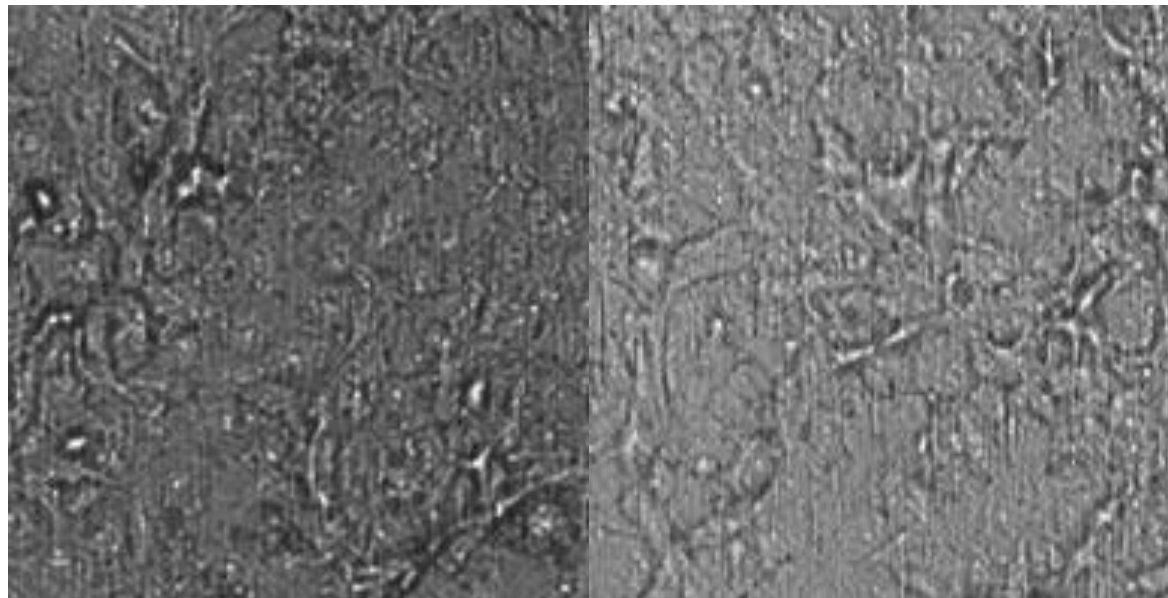
Figure 4.6 Schematic photo of TwinMic micro XRF experimental setup (Gianoncelli *et al.*, 2016).

Prior to sample analysis, the samples were sandwiched in acrylic glass frames and then mounted in a kinematic holder for direct analysis to reduce the scattering contributions of sample substrate. Ten representative cell cultures; 5 cancerous (DU145) and 5 normal (Vero cell line) were analyzed in this study. The cell culture samples were mounted on an x-y stage at 45/45 geometry (with the incident beam normal to the sample surface) facing 8 Silicon Drift Detectors. The analysis was carried out at photon energy of 1.7 KeV and spot size of 450 nm for sufficient fluorescence signal at high spatial resolution. The analytical conditions were as follows: 10 μ m step size, 10 s irradiation time per spot (real time) and vacuum pressure (<3 mbar). XRF images were acquired in vacuum where the samples were raster scanned in step sizes of 100 μ m across the incident beam to obtain elemental maps. Automatic 2D scans were performed in steps of 5 μ m in each direction (x, y) on the selected area. The XRF images were collected using an exciting beam at 9.4 KeV with an energy resolution of 1 eV, with the aid of Si (111) as monochromator crystal. The fluorescence spectra of trace elements were obtained at each pixel for 2D imaging and recorded by the detector normally to the incident x-ray beam.

The beam was focused to a few micro meters for the resolution to be in the similar range. The sample under an optical microscope with motorized zoom enabled the area to be chosen for scanning. The 2D mapping of the elements was achieved by narrowing on the window of the biometals of interest in the XRF spectra and the full XRF spectra data saved at each pixel.

4.9.4 Microscopy imaging of cell cultures

Pathological cancer advancement resulted in chemical and structural changes in the cultured cells that possibly altered the micro XRF spectra. The microscopy images in Figure 4.7 clearly demonstrated the difference in stages of cancerous (DU 145) cell cultures for day 1 and day 4.

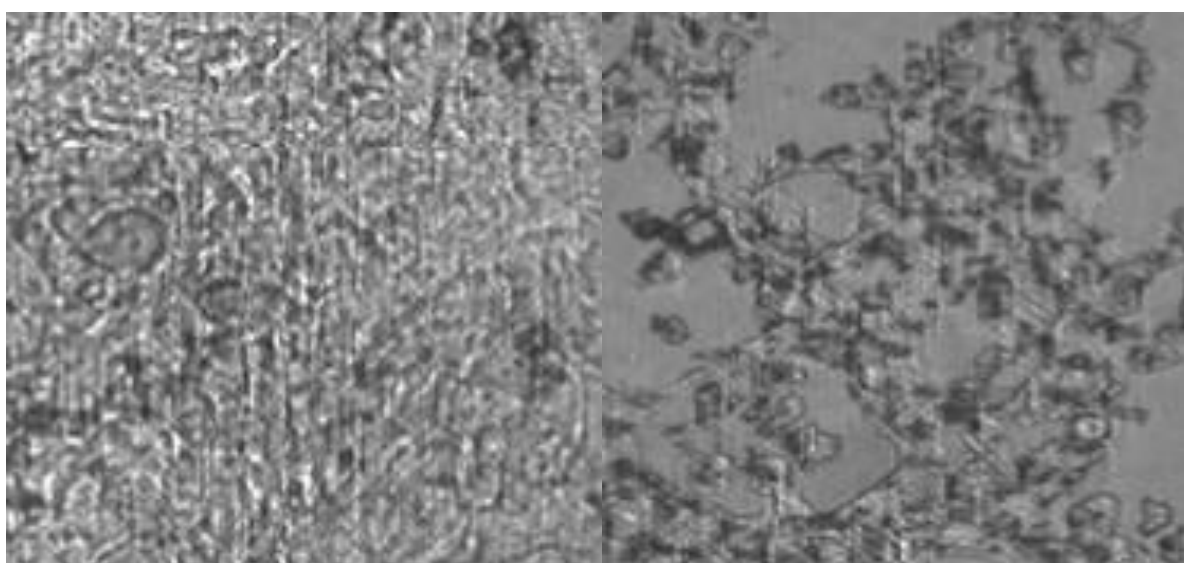


Day 1

Day 4

Figure 4.7 Visible light microscopy images (10x) for normal cell cultures

Figure 4.8 shows the corresponding light microscopy images of the normal cell culture (Vero) for day 1 and day 4.



Day 1

Day 4

Figure 4.8: Visible light microscopy images (10x) for cancerous cell cultures

From the above images, the morphological transformations and alterations (Gonçalves *et al.*, 2006) of DU145 and Vero monolayers could not differentiate cancerous and healthy cells as the cells were growing in colonies and not distinct. From these microscopy images, there is need to determine the cell morphological information from the Compton scatter region besides the selected fluorescence.

4.10 XRF data analysis

Prior to XRF data preprocessing, the channel numbers were converted to the corresponding energy and the spectra fitted for the identification of the elements. Further, equal steps of energy were determined and the counts were taken as input variable for data preprocessing. The data was preprocessed by auto scaling for similar contribution to model development and elimination of noise prior to multivariate machine learning analysis to enhance extraction of vital information from the spectra.

Statistical analysis was performed using Origin and Orange quasar software with comparison trace biometal concentrations in urine samples between cancerous and non-cancerous samples by using unpaired t-test. Orange as a user-friendly and an open-source software with machine learning and statistical tools for visual programming (Toplak *et al.*, 2017) was used for multivariate machine learning analysis. The Quasar orange-software with embedded machine learning tools enabled classification and multivariate regression for quantitative analysis achieved through workflows by widgets (basic functional units) which read, pre-processed and visualized data.

TXRF spectral data deconvolution was done by QXAS-AXIL (Analysis of X-ray spectra by Iterative Least-squares fitting) software package. For the XANES data analysis, XAS data were preprocessed by the ATHENA (DEMETER) software package (Di Cicco *et al.*, 2009) to mine useful diagnostic information. The micro XRF characteristic spectra were deconvoluted by PyMca software where the spectral overlaps and matrix effects were taken care of and background subtracted for extraction of important cancer diagnostic information. In PyMca software, the fluorescence data were presented as color maps displayed in RGB, with the brightest spots attributed to high elemental fluorescence. They were then deconvoluted using the PyMca Hypermet algorithm, where the color maps (RGB) were placed in grey to minimize differences in human color perception.

The concentrations and speciation data of the trace biometals in the tissue and urine samples (both cancerous and non-cancerous) were processed for descriptive statistics. Correlation coefficients were calculated from the distances to group the data thus the samples was considered similar or dissimilar and the results presented as dendrograms for visualization of sample relationships(Alexandre and Bueno 2006).

4.11 Multivariate machine learning analysis

The XRFS spectrometric data was analyzed by machine learning techniques in Quasar orange-software to identify the species of each trace element present. Machine learning techniques were utilized for data analysis and interpretation due to their ability to model both linear and nonlinear data sets. Machine learning techniques was used to analyze the spectra obtained from the samples to identify the species of trace elements present.

The multivariate data analysis enabled data reduction, elimination of redundant and “noise”, extraction of information and hypothesis formation (Kowalik and Einax 2006). These enabled extraction of important information (underlying biological phenomena) from complex multivariate large data sets. The multivariate data (linear and non-linear) analysis was realized via a variety of unsupervised and supervised algorithms (Luo 2006).

4.11.1 Pattern recognition

Spectral analysis, imaging and subsequent spectral processing and decomposition to retrieve diagnostic biomarker information were realized via pattern recognition-based machine learning techniques. PCA reduced the data complexity from high dimensional and complex data for rapid cancer diagnostic information mining that enabled exploratory analysis of the samples into subsets. These further enhanced mining vital diagnostic information from the large data sets in reduced dimensions where the exploration enabled identification of possible distinct discriminating features between cancerous and healthy human urine. PCA analysis was done using the Orange quasar software for exploratory analysis where transformed the original variables into new orthogonal principal components (PCs) in which the new axes were uncorrelated with each other for classification of the tissue samples (Massart *et al.*, 1988, Mark 2001). In this case, PCA aided in resolving overlapping spectral features for a rapid and direct cancer diagnostic technique based on simultaneous determination of biometals and Compton scatter profiles in complex matrices. Each principal component summarised the data set by generating scores for the observations with corresponding

loadings for the variables. Scores and loadings were compared to each other to mine cancer diagnostic information. Further, the first 5 PCs from PCA were used as input in ANN multilayer perceptron (MLP) model algorithm utilizing radial basis function (RBF) function (Brown and Lo 1998) to train, validate, and test the model for accurate classification. The classification approaches were used to learn from the available training data to develop rules to classify similar samples with same matrices.

Similarly, Hierarchical cluster analysis (HCA) based on similarity and dissimilarity distance measure among the samples (Bueno *et al.*, 2005) enabled examination of natural unsupervised classifications existing in the samples. These enabled classification of soft tissue biopsies and urine samples as either cancerous or non-cancerous and probably speciation of trace elements. K-nearest neighbors (KNN) technique enabled pattern recognition in the spectral data based on a nearest-neighbor. The classification method described the similarity of the new unknown sample using the other known spectral outputs (Brereton 2003) thus a potential tool for speciation trace biometals in human tissues and urine.

4.11.2 Multivariate calibration

The multivariate calibration necessitated determination of concentration and speciation of trace biometals in human body tissue and urine to realize accurate quantification of trace biometals' levels in human tissue and urine samples (Zhang *et al.*, 2017). Due to the complex and multivariate relationships of trace biometals in cancer development, a multivariate approach for determination of trace biometals was utilized in this study.

ANN enabled data analysis by mimicking biological neural systems (Looney 1996). The data sets were mean-centered before being cross-validated and all concentrations or speciation normalized for prediction of the concentration and speciation in new samples of the similar matrices with high robustness and accuracy (Marini *et al.*, 2008). ANN algorithms (utilizing back propagation techniques) will be constructed based on the simulate biometal levels. The dataset will be split into two subsets, one to be used in the calibration procedure and other to validate the calibration models. The input data (matrix) contained one row for each spectral image and one column for each measured independent variable upon which the calibration model was based. The selection of the samples into subsets will be made according to the principal component analysis (PCA) approach.

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 inputs fed into the network were divided into three sets; 60 % for network training, 20 % for validating how well the network generalizes and 20 % for testing how the network will perform on new set of data of similar matrices. The prediction ability of multivariate calibration by ANN was based on optimization of the neural network architecture where the models were trained, tested and validated for smaller training errors. The best model was initialized as that corresponding to high R value, with low mean-squared error and with 3 neurons in the input layer, 4 neurons in the hidden layer and 5 neurons in the output layer.

Support vector machines (SVMs) was used for classification algorithms and also to develop predictive regression models (Davis 2014). SVM regression was also applied for determination of trace biometal concentrations due to its robustness in high dimensional space analysis. For each class, a kernel function was used to distinguish the classes based on the measured parameters for probabilistic quantification and classification. The SVM kernel-based classification was robust to in high dimensional space hence took care of over-fitting. This enabled X-ray fluorescence analytical quantitative determination of trace biometals' levels using multivariate calibration strategies in human tissues and urine.

The mean square error (MSE) determined the mean of the squares of deviations (difference between predicted and spiked elemental concentrations) for judgment of the model performance. Root mean square error (RMSE) was used to determine of imperfection of the fit of the concentrations estimate. The Mean Absolute Error (MAE) was also used to determine how close the predictions were related to target outcomes. The R^2 was then used to determine the proportion of the variance in the dependent variable that was predicted from the independent variable. The ideal and robust calibration models with low RMSEC values and high R^2 values was developed for correlative analysis of the trace biometals to carcinogenesis.

4.12 Cancer diagnostic machine learning aided XRF method development

The multivariate ANN and SVM calibration strategies were developed for prediction of trace biometal concentration while exploratory analysis of cell cultures, human tissue and urine spectral data was achieved by PCA, KNN and HCA. Figure 4.9 shows a machine learning based X-ray fluorescence and scattering model for renal tract cancer diagnostics based on the

exploratory, quantitative and speciation analysis of trace biomarkers in soft human tissues and fluids (urine).

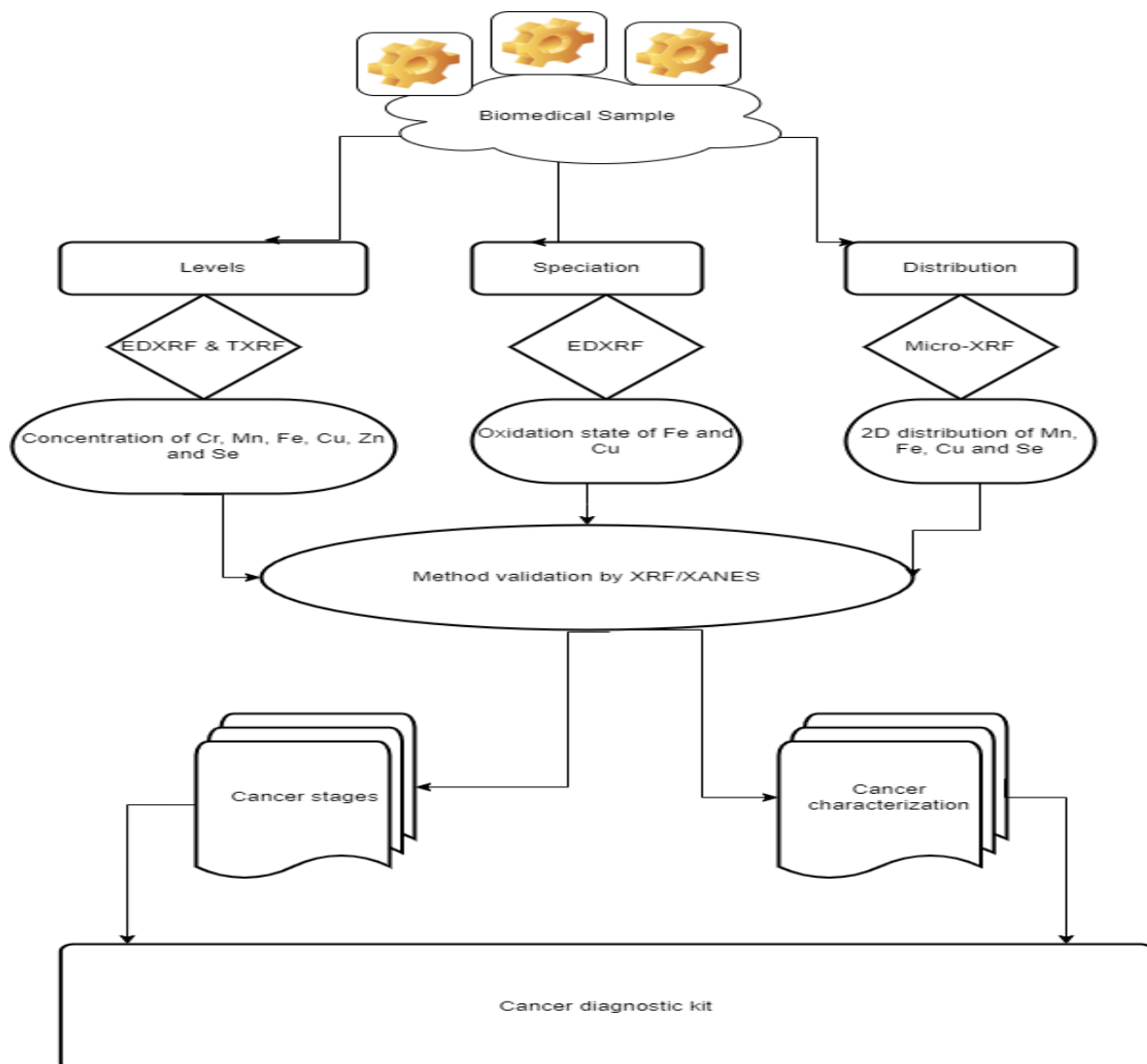


Figure 4.9 A conceptual framework for a machine learning based XRFS model development

The X-ray fluorescence peaks and scatter profiles with spectral overlaps and matrix effects with characteristic qualitative and quantitative information were analyzed by for rapid accurate prediction in both linear and non-linear biomedical data with good robustness.

CHAPTER FIVE

RESULTS AND DISCUSSIONS

5.1 Chapter Overview

This chapter contains fifteen sections that form the discussion of the results. Section 5.2 presents the XRF spectra of biomedical samples while section 5.3 report on the results of multivariate machine learning analysis of simulate samples. Sections 5.4, 5.5 and 5.6 emphasizes on PCA results, multivariate analysis of cell culture samples and speciation analysis in simulate samples respectively. Section 5.7 highlights direct speciation analysis of *Fe* by XANES while section 5.8 report on quantitative results of trace biometals for cancer diagnosis. Sections 5.9 and 5.10 presents the results on the multivariate calibration of trace biometals in urine and tissues. Section 5.11 report on SR-XRF quantitative analysis of trace biometals while section 5.12 gives results on correlation matrices of trace biometals for cancer diagnosis. Finally, sections 5.13, 5.14 and 5.15 report on the micro XRF imaging of trace biometals, prostate cancer stages of development in human and cancer diagnostic model respectively.

5.2 XRF spectra of biomedical samples

The Energy dispersive X-ray fluorescence (EDXRF), Total reflection X-ray fluorescence and Synchrotron radiation-based X-ray fluorescence spectral data for simulate, cell cultures, human tissue and urine are presented.

5.2.1 EDXRF spectra of simulate and human urine

Typical Rigaku EDXRF spectra of simulate and cancerous urine samples are as shown in Figure 5.1. The scatter peaks with a pronounced background continuum masked the weak fluorescence analyte signals. The enhanced background was attributed to absorption and enhancement effects of dark matrix by low Z elements (Pozebon *et al.*, 2017).

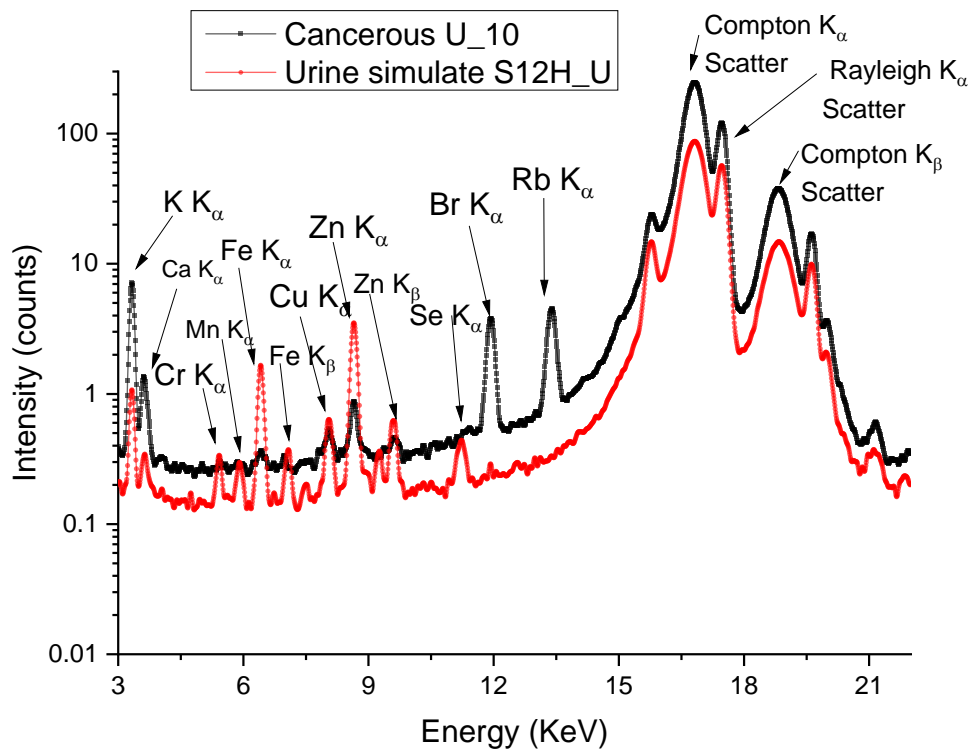


Figure 5.1: Typical EDXRF spectra of cancerous and simulate urine samples

The above spectra resulted to poor signal-to-noise ratio of the trace element of interest and relatively weak fluorescence profiles due to enhanced background. The relatively high background and the enhanced scatter posed a challenge in determination of low concentrations of trace biometals which are submerged in the background's noise (Grüner *et al.*, 2018). The energies of the detected X-rays were characteristic of the elements present while the corresponding intensities were representative of elemental concentrations (Van Grieken and Markowicz 2001).

Figure 5.2 shows overlapped EDXRF spectra of normal (NP_12) and cancerous (P_12) urine samples. The trace biometal (Cr, Mn, Fe, Cu, Zn and Se) characteristic profiles were weak with enhanced background and that the Compton scatter still dominated in both normal and cancerous urine spectra.

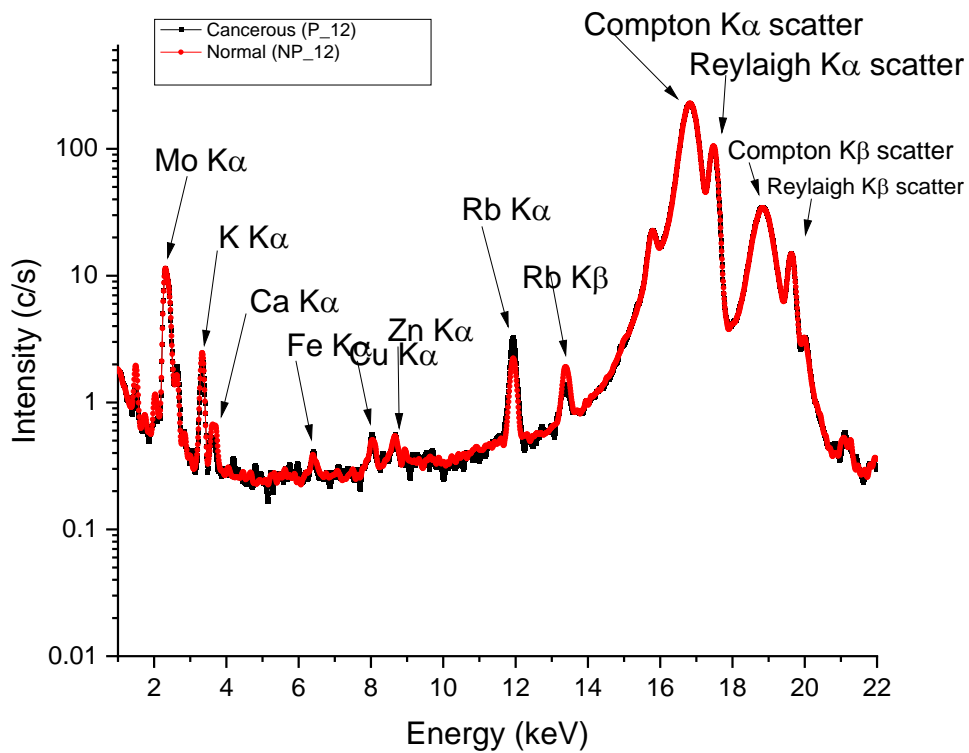


Figure 5.2 EDXRF spectra of cancerous and non-cancerous urine

In order to mine additional vital cancer diagnostic information, the spectral datasets were preprocessed by vector normalization and the high background and noise reduced from the characteristic spectral profiles. The determination of trace elements was therefore significantly affected hence the utility of multivariate machine learning models to mine vital cancer diagnostic information. The multivariate alterations of the weak biometals' profiles in human soft body model tissues and urine were correlated to the state of health thus used as parameters for cancer diagnostics.

5.2.2 TXRF spectra of urine samples

Figure 5.3 shows a typical TXRF spectra of cancerous urine samples with potential for mining qualitative and quantitative information for cancer diagnosis.

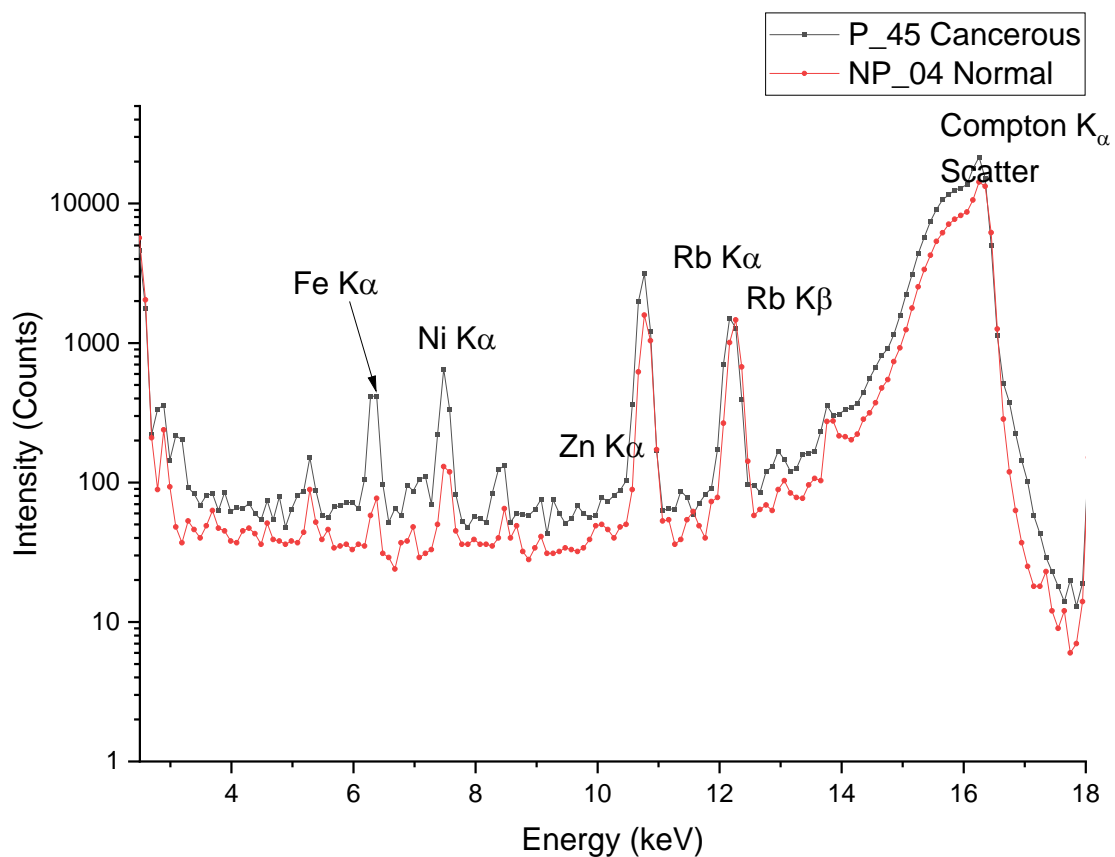


Figure 5.3 Typical TXRF spectra of cancerous urine samples

The TXRF reflective property on the sample substrate reduced the high spectral background (Carvalho *et al.*, 2007) thus significantly improved the signal to noise ratio of the biometal analytes. The biometals (Cr, Mn, Fe, Cu, Zn and Se) spectral fluorescence profiles together with the Compton Scatter as cancer biomarkers has the potential for early cancer diagnosis.

The TXRF sensitivity curve in Figure 5.4 shows how the detection limits were strongly dependent on the atomic number of the analytes.

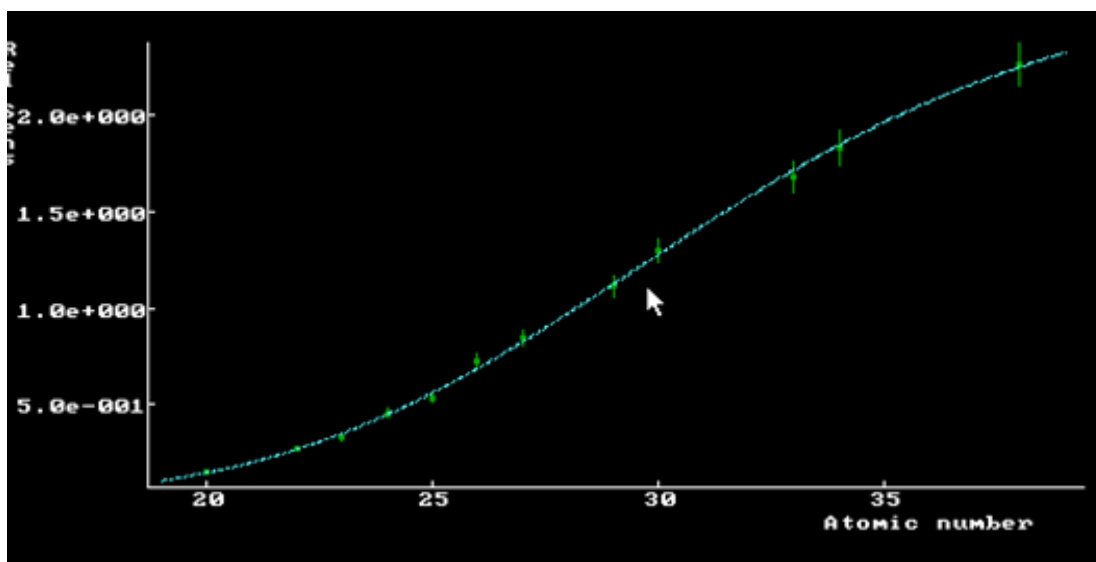


Figure 5.4 TXRF relative sensitivity graph of biometals.

This illustrated the greater sensitivity of TXRF analysis for multivariate modeling and exploratory analysis for acquisition of real time disease (cancer) diagnostic information for clinical purposes. The fluorescence peaks at higher detection limit for low Z elements were evident (Szoboszlai *et al.*, 2009).

5.2.3 SR-XRF spectra of cultured samples

Synchrotron radiation based XRF spectra in Figure 5.5 shows typical micro XRF fitted spectra at each pixel for V_D1 and DU_D1 cell cultures. It was apparent that the fluorescence signals were distinct with slightly enhanced scatter profiles with reduced background as a result of linear polarization of the X-ray beam. The prominent K_{α} line of Na was due to the sodium bicarbonate in the growth medium while the weak L lines of Mn, Fe, Cu and Se were also evident.

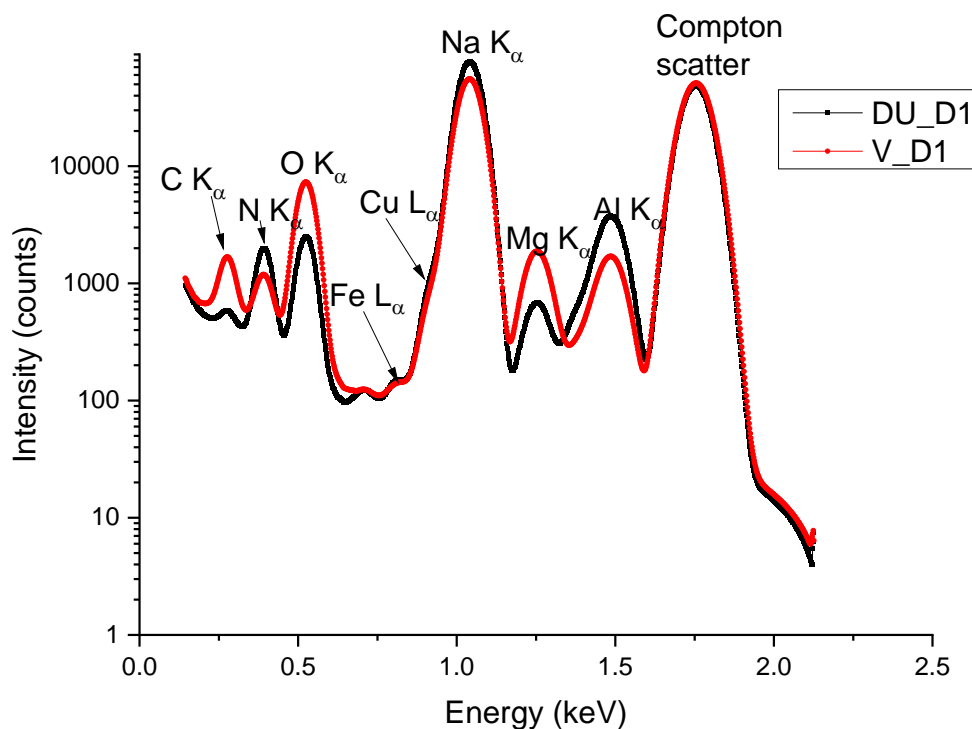


Figure 5.5 Micro X- ray fluorescence spectra of DU_D1 overlapped with V_D1 for day 1

The above complex spectra (due to the dark matrices) were characterized by low energy region dominated by low Z elements which overlapped the L-lines (for heavy-elements) fluorescence.

Figure 5.6 shows the SR-XRF based XRF spectrum of cultured of DU_D1 cells on a Zinc borosilicate microscopic cover slip substrate obtained at D08 beamline SESAME (Jordan).

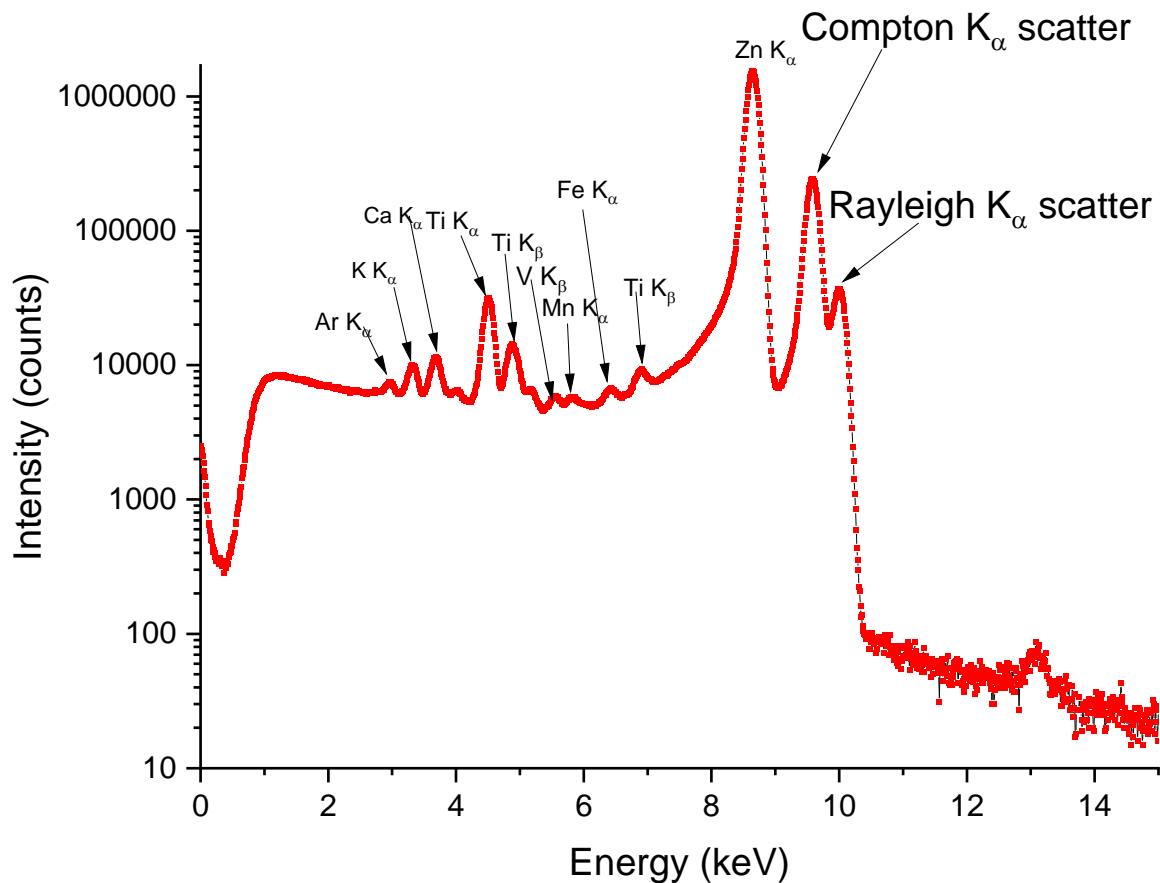


Figure 5.6 Synchrotron X-ray fluorescence spectra of DU_D1 cancerous cell culture

In the above SR based XRF spectra, the analytes signals are enhanced despite the complexity and extreme matrix effects with slightly enhanced scatter region. The enhanced Zn peak was due to its high concentration in Zinc borosilicate microscopic cover slip that was used as a substrate. The recovery of analyte signals of trace biometals from the relatively high spectral background calls for a multivariate approach.

5.3 Multivariate exploratory analysis of spectra data

Multivariate exploratory analysis of analyte fluorescence and scatter spectral profiles enabled determination of biometal diagnostic information. The biometals of interest in the fluorescence region in this context are linked to cancer development other than their vital roles in physiological and metabolic processes (Rose 2016). These were envisaged in possible classification of urine samples which was mainly as a result of fluorescence and Compton scatter profiles.

5.3.1 PCA of urine samples using EDXRF spectra

Figure 5.7 shows the utility of fluorescence region in an attempt to explore and classify prostate and urinary bladder urine samples based on PC1 and PC3 scores. For the entire spectral information, 68% variance was explained (65% and 3%) for PC1 and PC3 respectively.

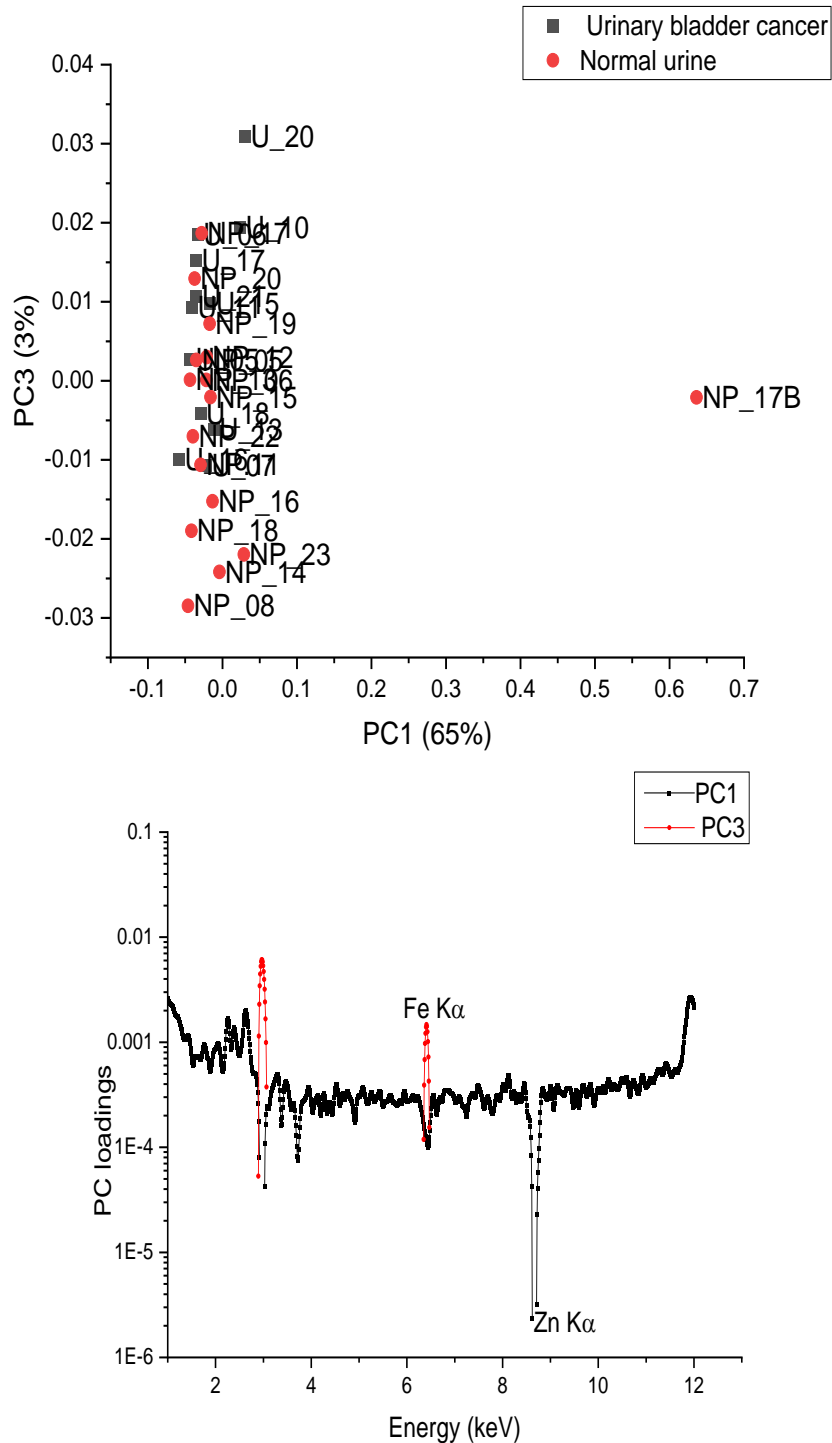


Figure 5.7 PC score and loadings plot utilizing fluorescence region of urinary samples

The utility of fluorescence peaks of the trace elements could not mine vital cancer diagnostic information from the cancerous and non-cancerous human urine for classification. From the corresponding loadings plot in Figure 5.7, it was evident that the vital diagnostic information for characterization was contained partially in the fluorescence (Fe and Zn) but could not differentiate the cancerous and normal healthy human urine. The utility of both fluorescence and Compton scatter in Figure 5.8, shows possible characterization of urine samples based on their matrix composition where clear clusters of cancerous and normal urine samples using PC1 and PC3 scores were evident with potential utility in cancer diagnosis.

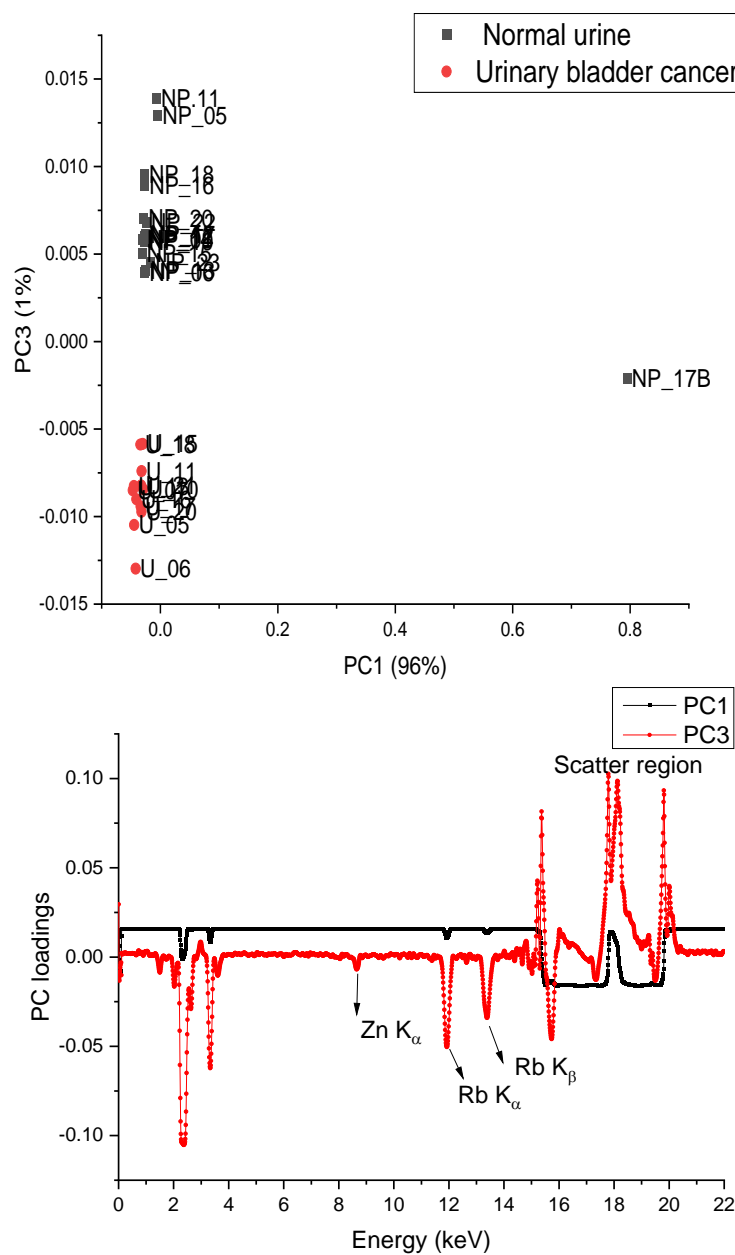


Figure 5.8 PC score and loadings plot of fluorescence and scatter of human urine

PCA enabled possible classification of cultured samples with explained variance of 97 % (96 % and 1 %) for PC1 and PC3 respectively which differentiated between normal and cancerous human urine. From the loadings plot, the fluorescence and scatter profiles in PC1 and PC3 separated cancerous and normal urine samples. Zn and the Compton scatter regions greatly influenced the above possible characterization of urine into cancerous and normal healthy clusters. The scatter region in this case had additional information vital for computing mean atomic number and the density of the sample (Goraieb *et al.*, 2007). The results clearly confirmed the fact that the scatter as a valuable additional biomarker due to morphological information for the identification of pathological disorders (cancer) in human.

5.3.2 PCA of normal and cancerous urine using TXRF data

Figure 5.9 shows PCA score plot in which urine samples were analyzed based on their matrix composition. The results were based on the analyte fluorescence of the biometals of interest and the corresponding Compton scatter for possible characterization of cancer and normal urine. PC1 and PC4 scores plot accounted for total explained variance of 73% with potential to characterize normal (NP) and cancerous (P) urine based on the fluorescence region.

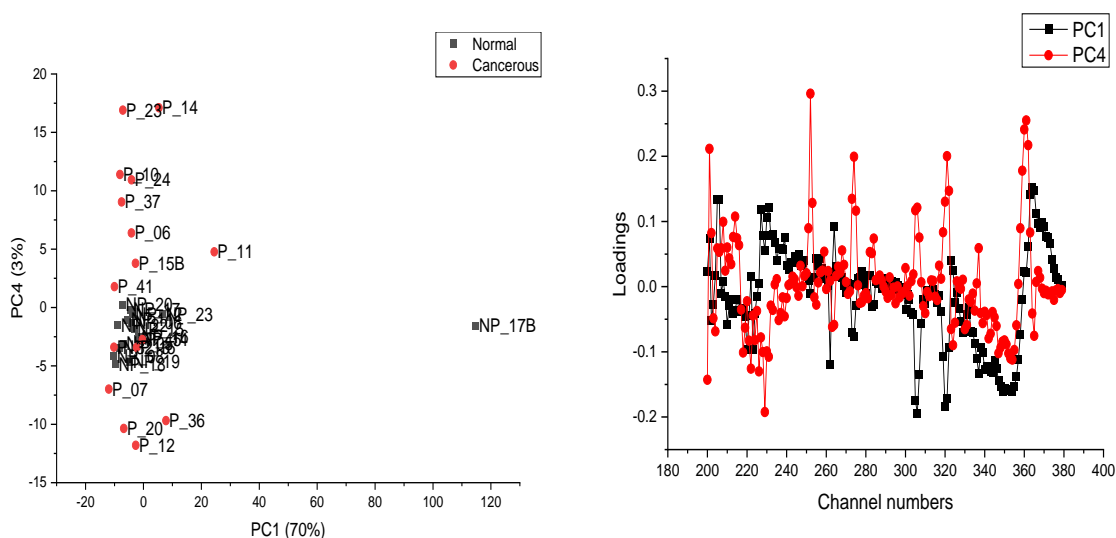


Figure 5.9 PC score and loadings plots of urine samples using the TXRF fluorescence region

The PC loadings plot indicated possible strong influence of trace biometals for urine characterization into cancerous and normal groups. Figure 5.10 shows PCA score plot of PC1 and PC2 at explained variance of 60% and 15% respectively using the fluorescence spectral profiles of prostate cancer urine samples with a potential for characterization of urine samples possibly due to different stages of prostate cancer development.

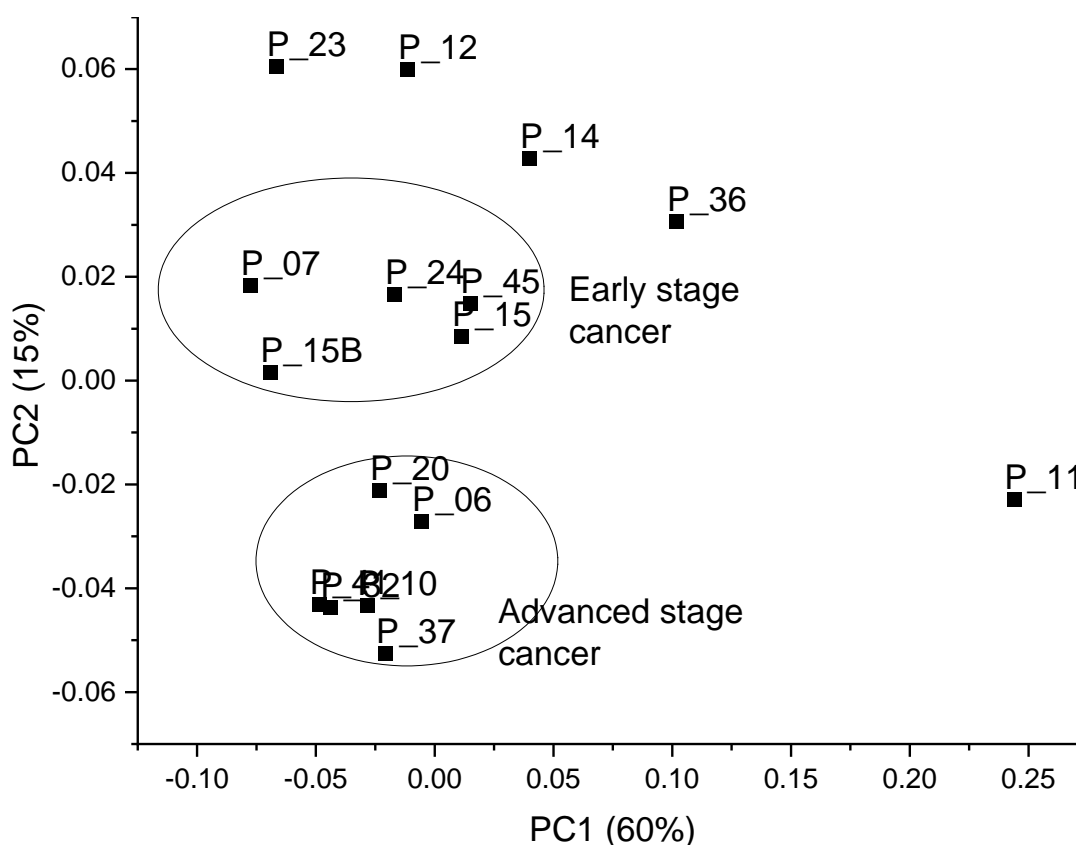


Figure 5.10 PC score plot of cancerous urine samples using the TXRF fluorescence region.

The exploratory analysis by PCA utilizing the fluorescence profiles indicated possible characterization of prostate cancer into two clusters probably due to different stage of development. The P_12, P_23, P_14 and P_36 samples were probably at intermediate stage of cancer development. In this context, PCA enabled extraction of information (variance) from large complex multivariate datasets for exploratory data analysis and predictive modeling for cancer diagnosis and stages of development.

5.4 Multivariate analysis of micro XRF images of the cell cultures

Machine learning exploratory techniques enabled multivariate exploration of the spectral images to unravel latent and complex patterns (due to the dark matrices) and their utility in determine biometal concentrations, speciation and their spatial distributions.

5.4.1 PCA of Micro XRF spatial image data of cell cultures

PCA reduced the dimensions of spectral spatial images for the extraction of subtle cancer diagnostic (distribution and semi-quantitative) information. The fitted spectral image profiles PCA score and loadings results in Figure 5.11 with an explained variance of 90 % (87 % and 3 % for PC1 and PC2 respectively) could not distinctively differentiate cancerous and normal cell cultures.

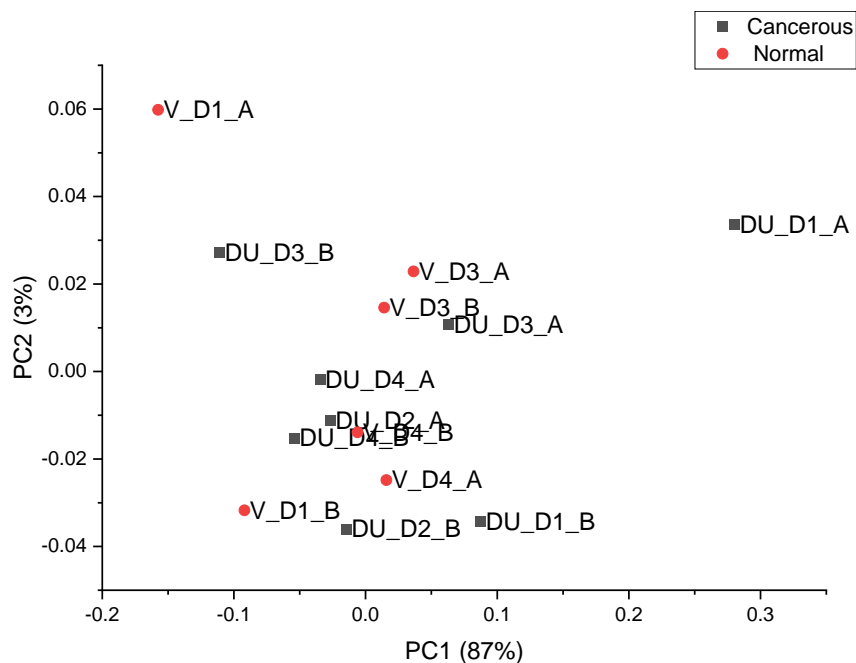


Figure 5.11: PCA score plot for cell cultures using the fitted fluorescence

The selected fluorescence of Cu and Fe spectral intensity data of cultured cancerous samples were further analyzed by PCA as shown in Figure 5.12. It's clear that the PCA at explained variance of 97% with PC1 (94%) and PC2 (3%) showed potential characterization of the various stages of cancer development.

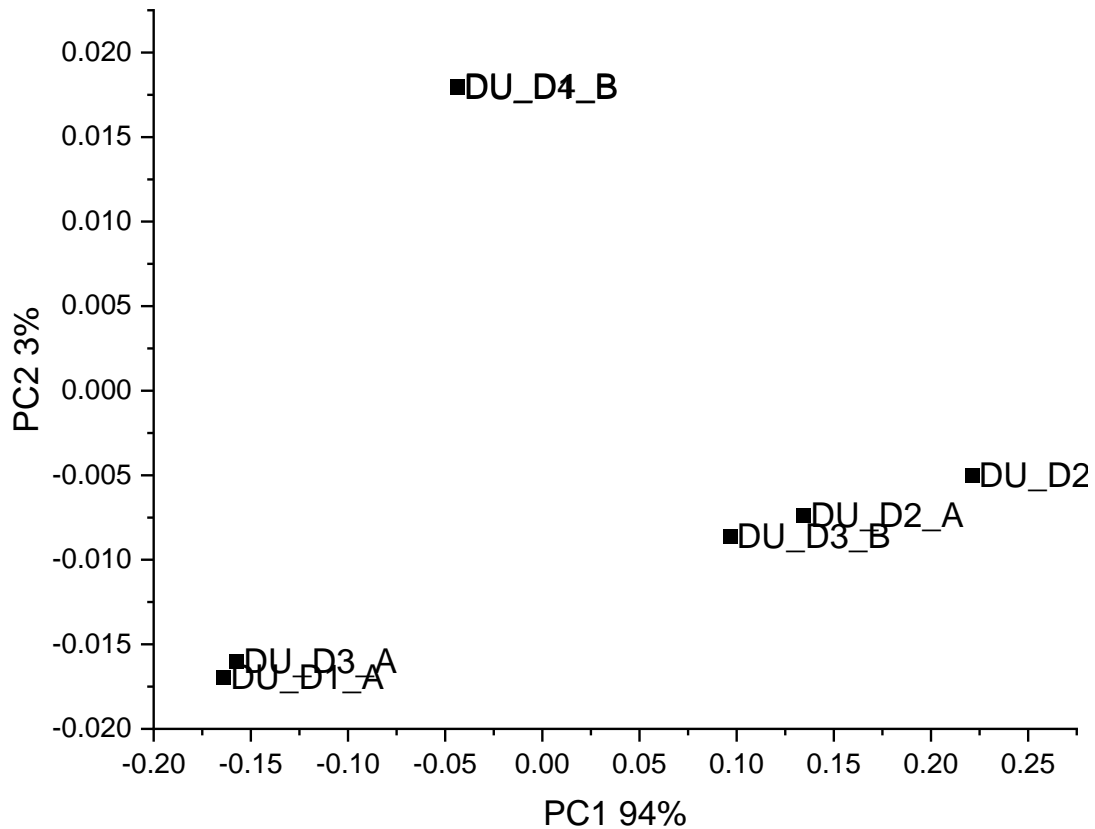


Figure 5.12 PCA Score plot for DU cell cultures of selected Fe and Cu spectral intensities

The utility of Cu and Fe enabled exploratory analysis of cancer into probable distinctive stages of development. The spatial distribution of Cu and Fe plays a vital function in oxygen metabolism despite Cu being a co-factor of Fe (Rocha *et al.*, 2019). Further, Fe is known to aid in blood supply to the malignant cells thus accelerates angiogenesis (Kwok and Richardson 2002). It therefore deduced that the spatial accumulation of Cu and Fe can be used to not only identify the presence of cancer but also have the potential to characterize cancer into various stages of development.

5.4.2 Multivariate exploratory analysis of cell cultures

The sample trace biometal multispectral image spatial distribution intensities (variables) were decomposed into scores and loadings. Principal component analysis (PCA) enabled link and correlation between trace biometals constituents. The cultured cells were not characterized as cancerous and non-cancerous as shown in Figure 5.13 by utility of spectral image data from entire spatial image profiles.

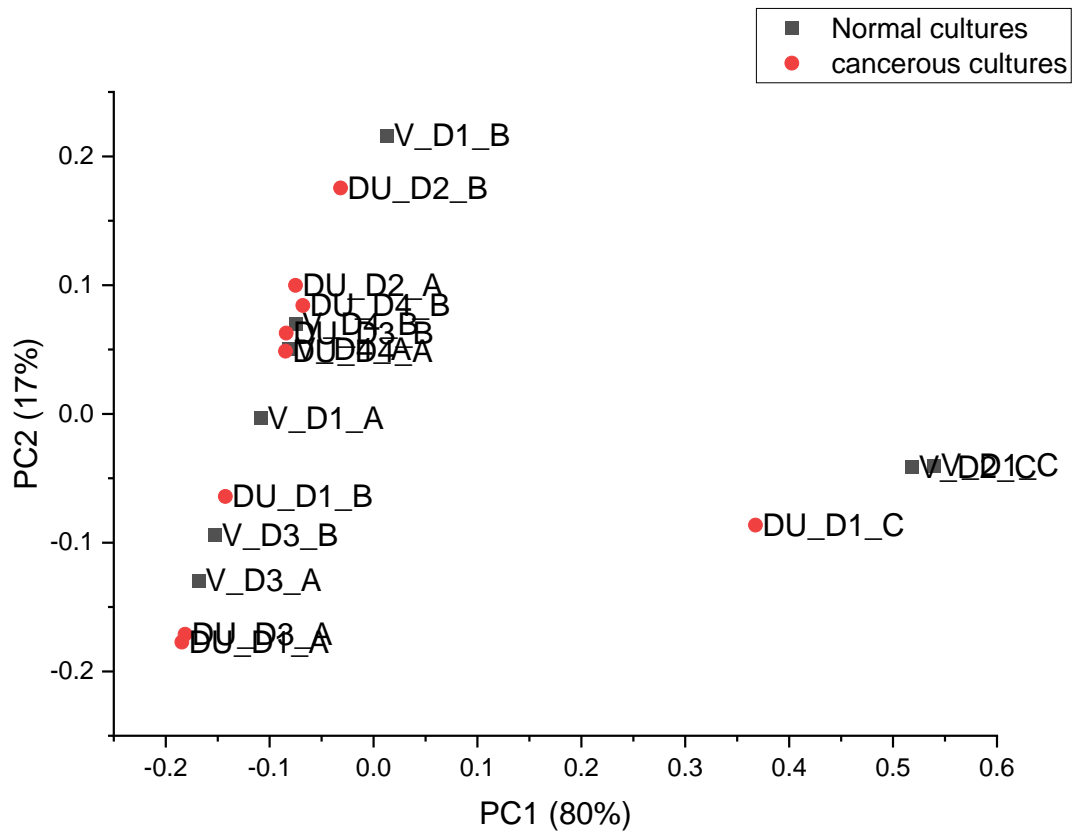


Figure 5.13: PCA Score plot for cell cultures using fitted spectral data.

The scatter region was utilized together with the fluorescence peaks for cancer stage characterization as shown in Figure 5.14 by utility of fitted fluorescence and scatter spectral profiles where the colors show possible stages of cancer development.

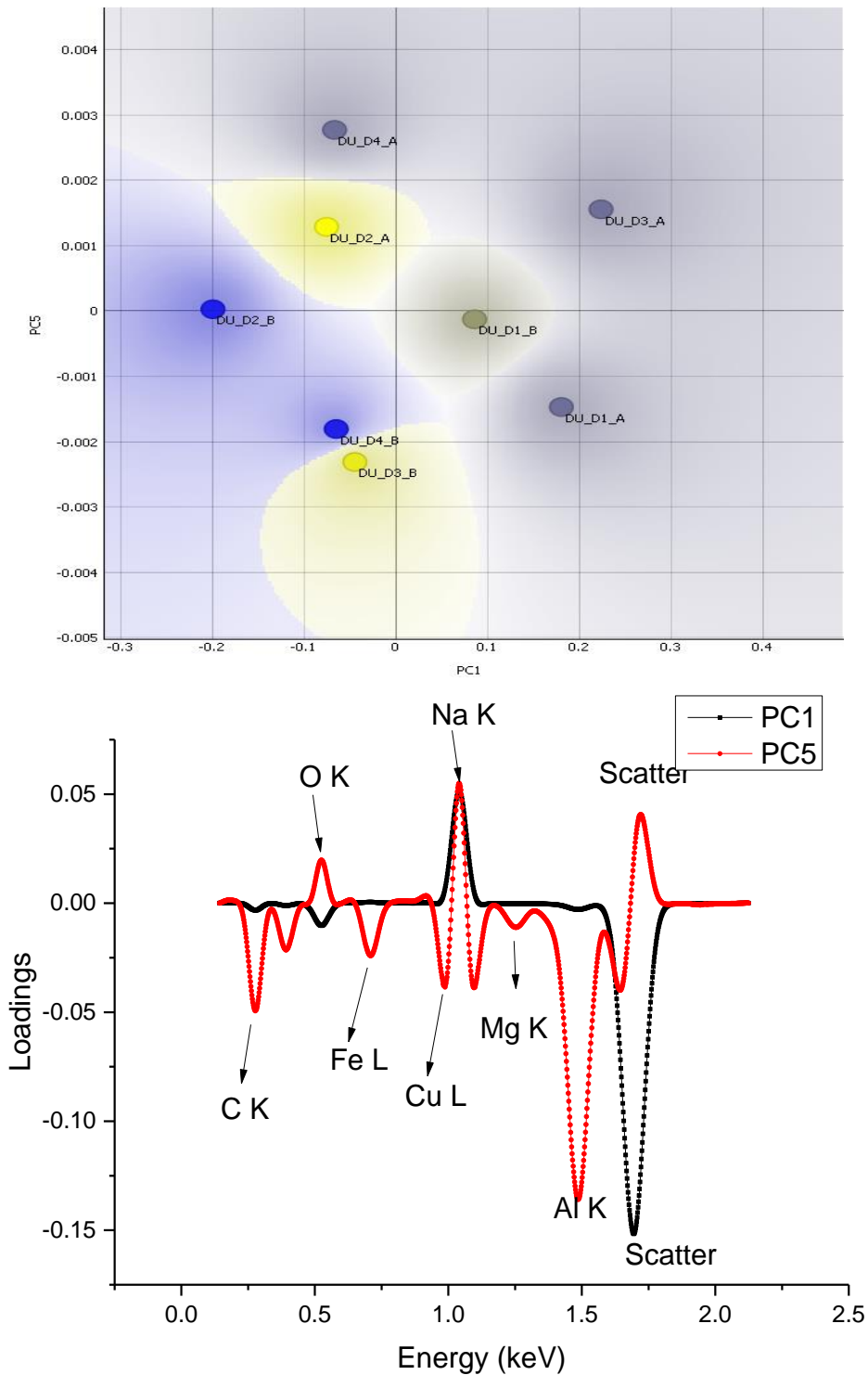


Figure 5.14: PCA score and loadings plot for DU cell cultures using fluorescence and scatter. PC scores were able to successfully identify 3 cancer stages at explained variance of 80% and 2% for PC1 and PC5 respectively. The simultaneous utility of both selected fluorescence and Scatter region enhanced possible detection of the 3 stages of cancer development.

The inclusion of scatter region improved the characterization of cultured samples but could not differentiate between the staging of DU_D4 and DU_D3 stages of cancer development.

This is due to the fact that the scatter peak contains further information pertaining to cell structure (morphology) that is an important property in cancer characterization.

From the loadings plot, Fe and Cu together with the Compton scatter biomarker spectral profile were partly responsible for the differentiation of the cultured samples into cancerous various stages of cancer development.

5.4.3 Hierarchical cluster analysis of urine samples

Hierarchical Cluster Analysis (HCA) using TXRF data enabled exploratory analysis urine samples using the Euclidean distance measure. HCA further demonstrated urine characterization through the agglomerative linkage method. The dendrogram (utilizing auto scaled values and Euclidean distance) on prostate datasets in Figure 5.15 clearly shows utility of fluorescence where two clusters corresponding to normal and cancerous urine were evident.

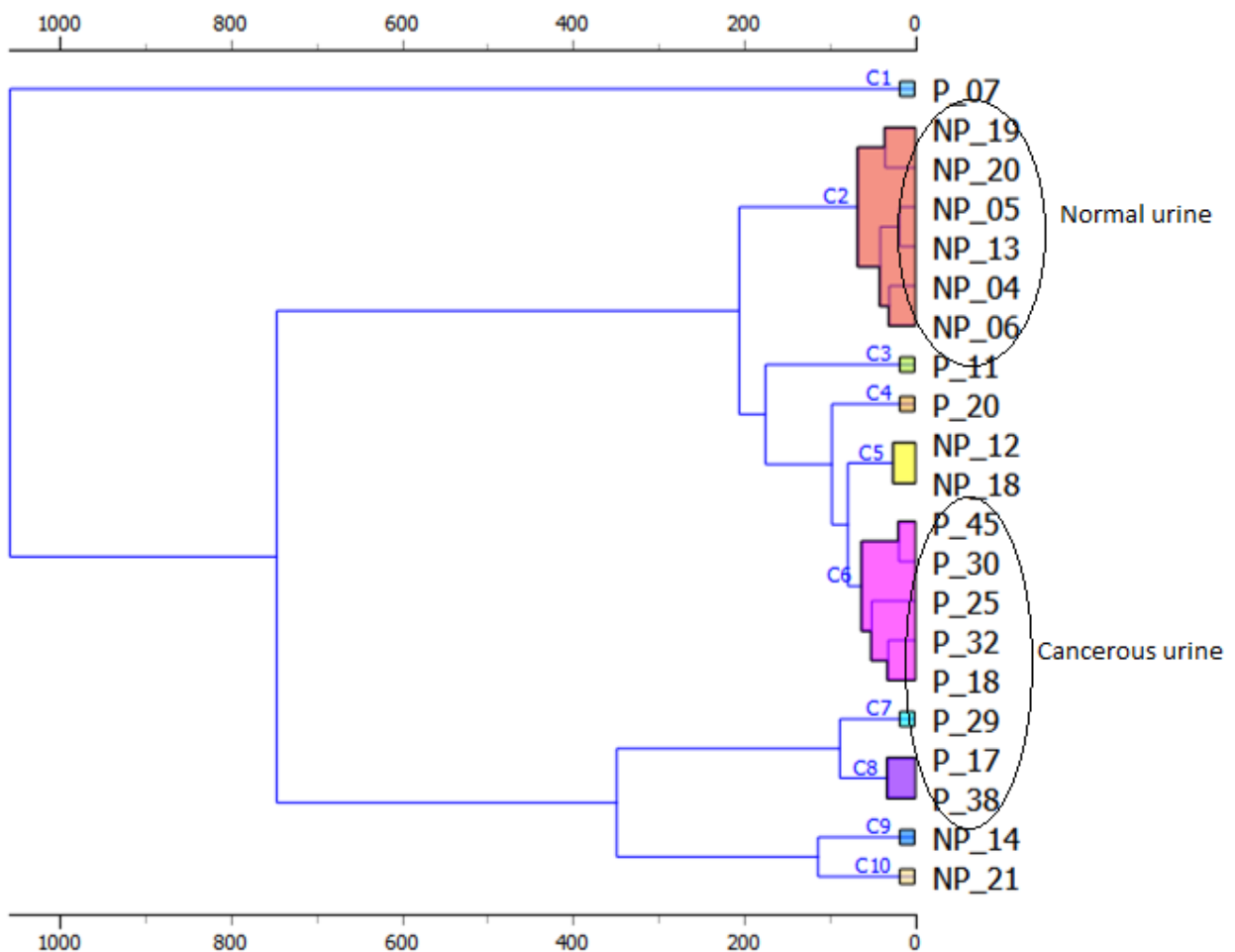


Figure 5.15 HCA classification of urine samples using fluorescence region

Exploratory analysis by HCA shows possible characterization urine samples into cancerous and healthy clusters. Further, it shows possible potential classification of cancerous urine samples into different groups (C6, C7 and C8) which depicts different stages of cancer. It's clear that trace elements (Fe, Cu, Mn, Zn and Se) in the fluorescence region were partially responsible for the above clustering for cancer diagnosis.

In Figure 5.16 where both the fluorescence and scatter regions were used, no clear cluster distinction between cancerous and normal urine was observed.

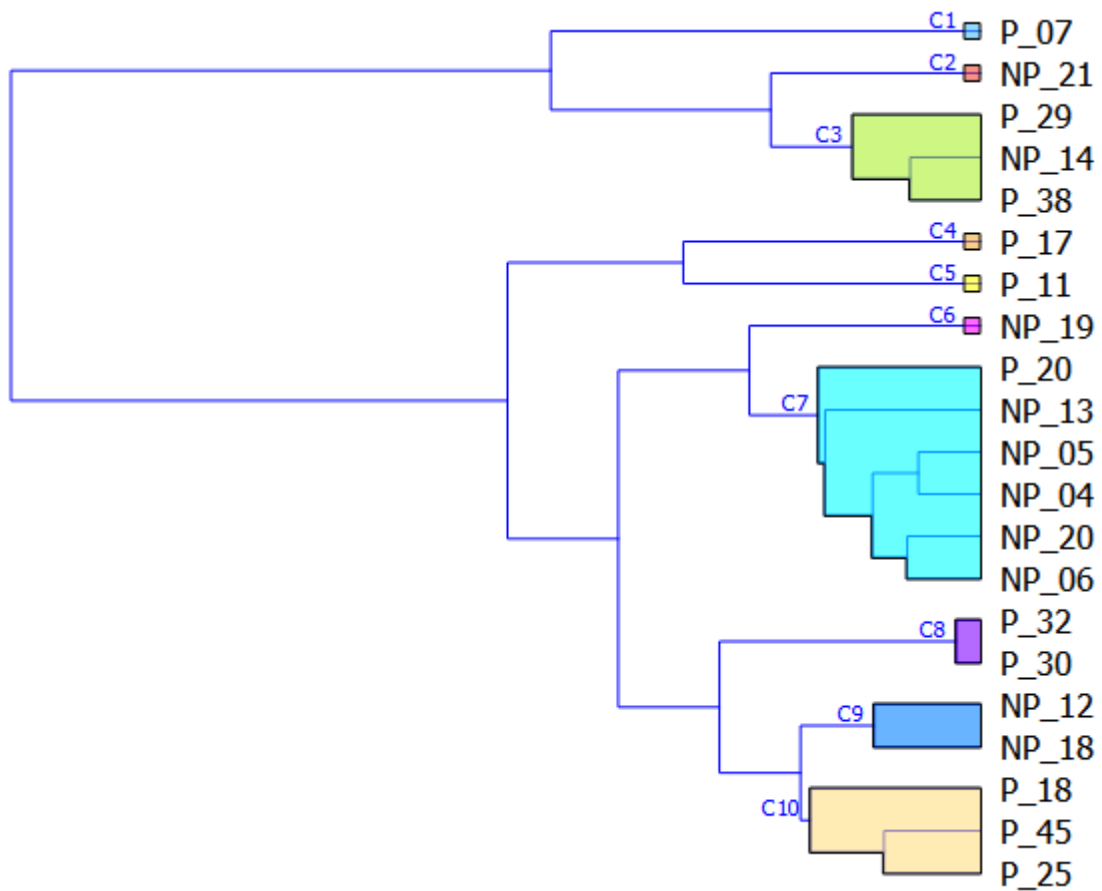


Figure 5.16 HCA analysis of urine samples using fluorescence and scatter region

It can be deduced that further utility of Compton scatter region hardly shows any characterization of urine samples as it is related to the cell morphology. The fluorescence region was further used to analyze prostate cancer urine samples as shown in Figure 5.17 where potential characterization of the stages of cancer development are evident.

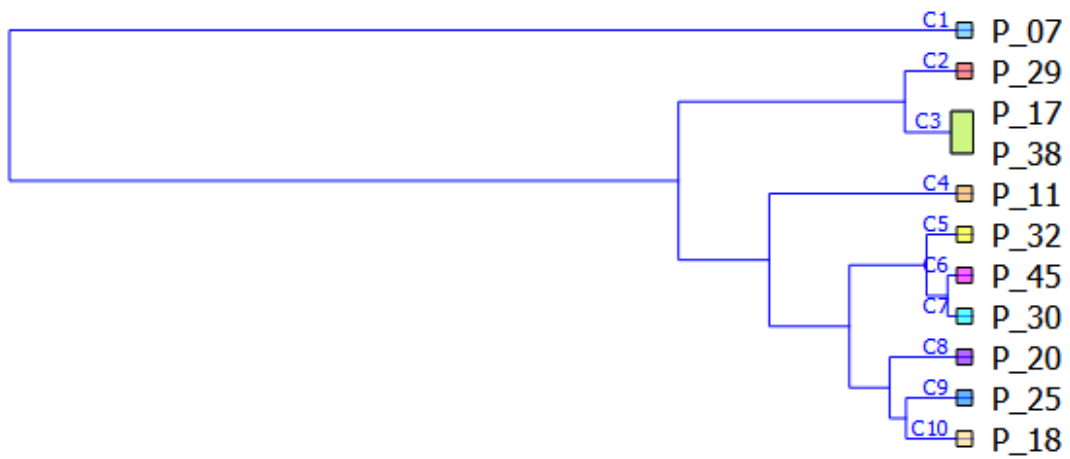


Figure 5.17 HCA characterization of cancerous prostate urine samples using fluorescence region

In this case, potential probable stages of cancer were as follows; early stage (C2 and C3), intermediate stage (C5-C7) and advanced stage (C8-C10) of cancer development.

5.5 Speciation analysis of cancerous and healthy urine samples

The trace element speciation analysis and modelling in complex biomedical samples was explored by multivariate analysis. The bulk additional chemical information of Fe and Cu enabled further investigation of the role of trace biometals in disease (cancer) diagnosis. KNN and SVM models enabled discrimination of cancerous and normal samples based on the training set of the simulate samples for indirect chemical speciation analysis.

Simulate sample models were calibrated and validated with satisfactory figures of merit using limited selected range of training set. The validated developed KNN and SVM chemical speciation of Fe and Cu are as shown in Table 5.1.

Table 5.1: Validation of speciation prediction model of Fe

Sampling type: stratified 20-fold cross validation scores				
Model	MSE	RMSE	MAE	R ²
KNN	0.1643	1.4311	0.3660	0.9450
SVM	0.2678	0.5175	0.5156	0.0819

K-Nearest Neighbours data were distributed into k different groups based on the similarity between points in terms of Euclidean distance metric. Figure 5.18 shows how K-Nearest Neighbours accurately enabled chemical speciation of Fe in simulate samples into higher and lower speciation samples.

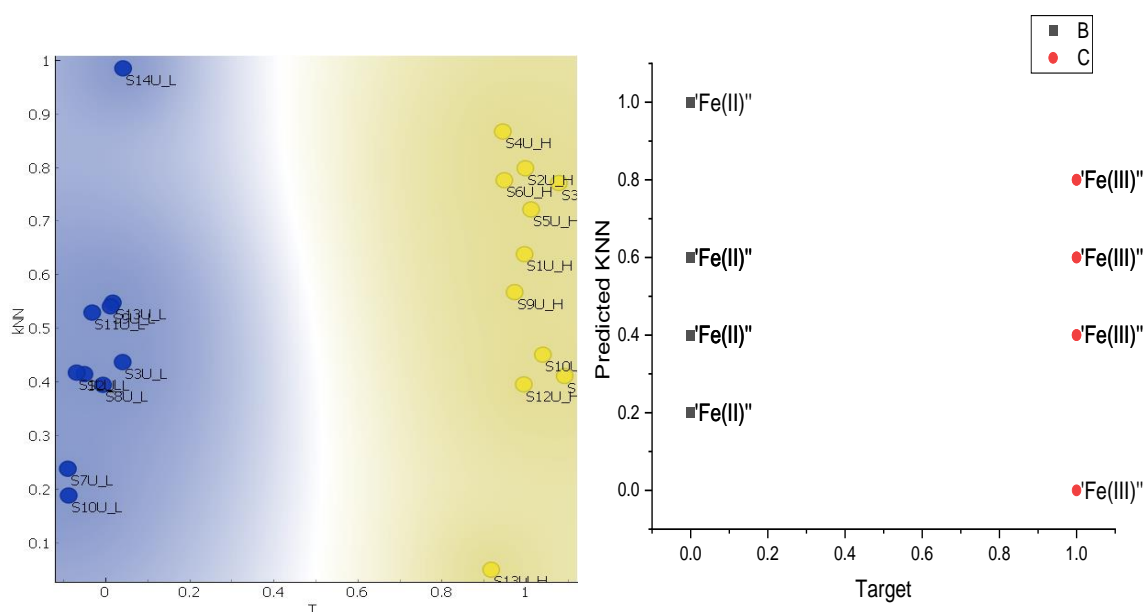


Figure 5.18 Predictive scatter plot of Fe speciation of simulate urine by KNN

The Fe speciation characterization of simulate samples were further modelled by support vector machines as shown in Figure 5.19 where distinct clusters of higher and lower speciation were evident.

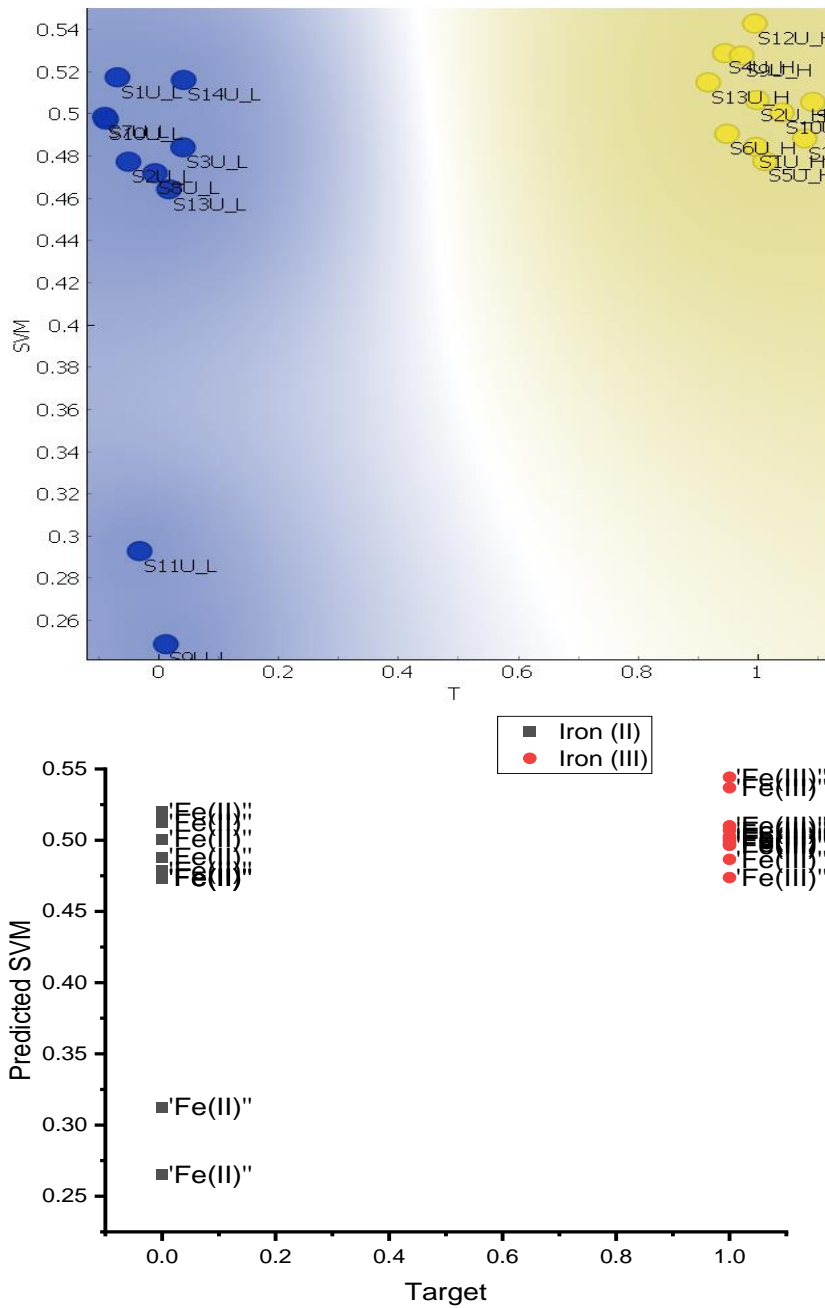


Figure 5.19 SVM classification of Fe speciation

SVM equally discriminated the two species of Fe as it is possible for proper discrimination of the oxidation states of Fe in urine simulates.

In addition, it was interesting to see that both higher and lower speciation Cu samples in Figure 5.20 were distinctively separated by KNN.

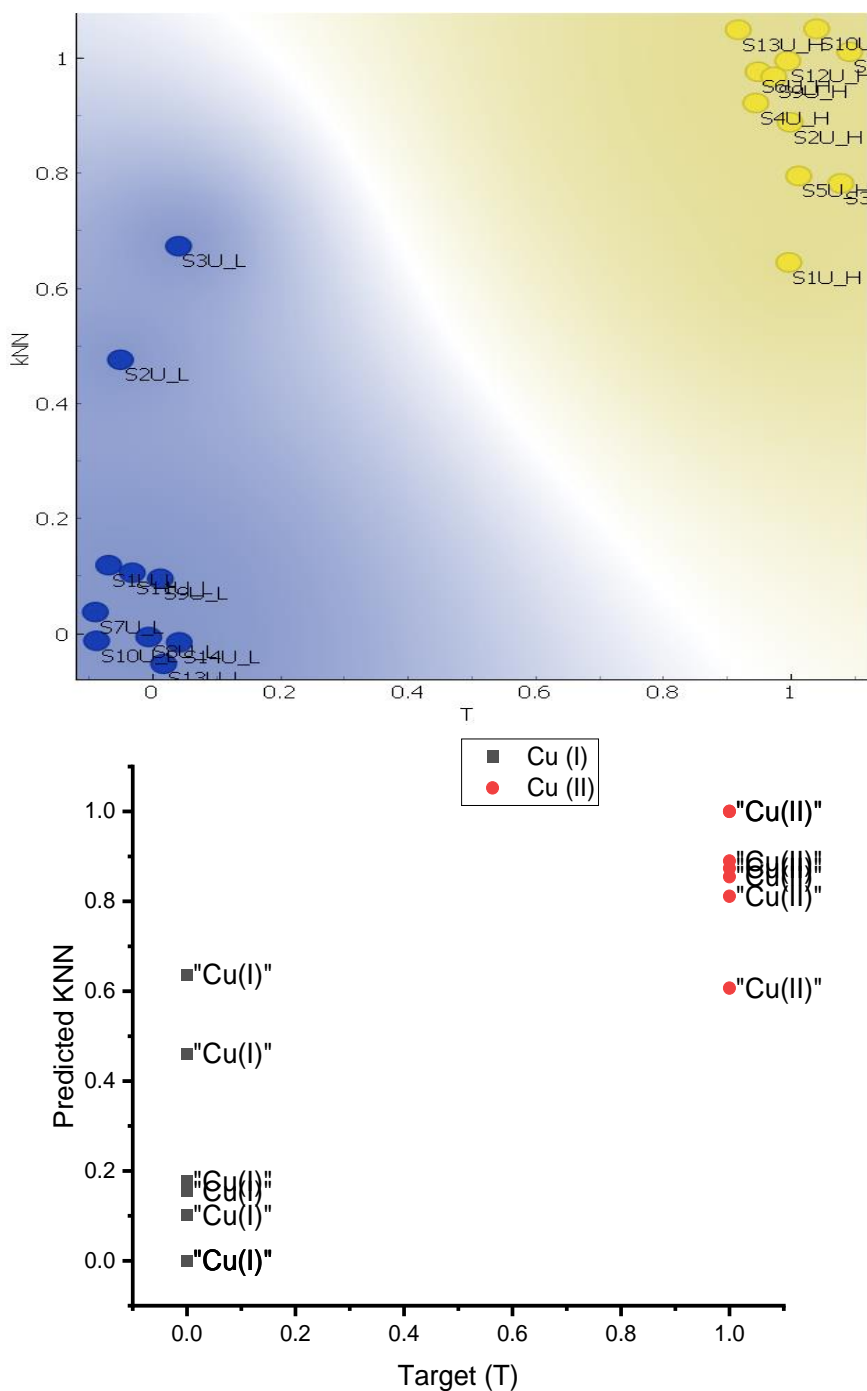


Figure 5.20 KNN speciation analysis of Cu in simulate samples

KNN distinctively characterized the two species of Fe as it is possible for analysis of the oxidation states of Fe in urine simulate.

The 2 oxidation states of Cu were also clearly characterized by SVM as shown in Figure 5.21 which indicated that Cu (I) was distinctively separated from Cu (II).

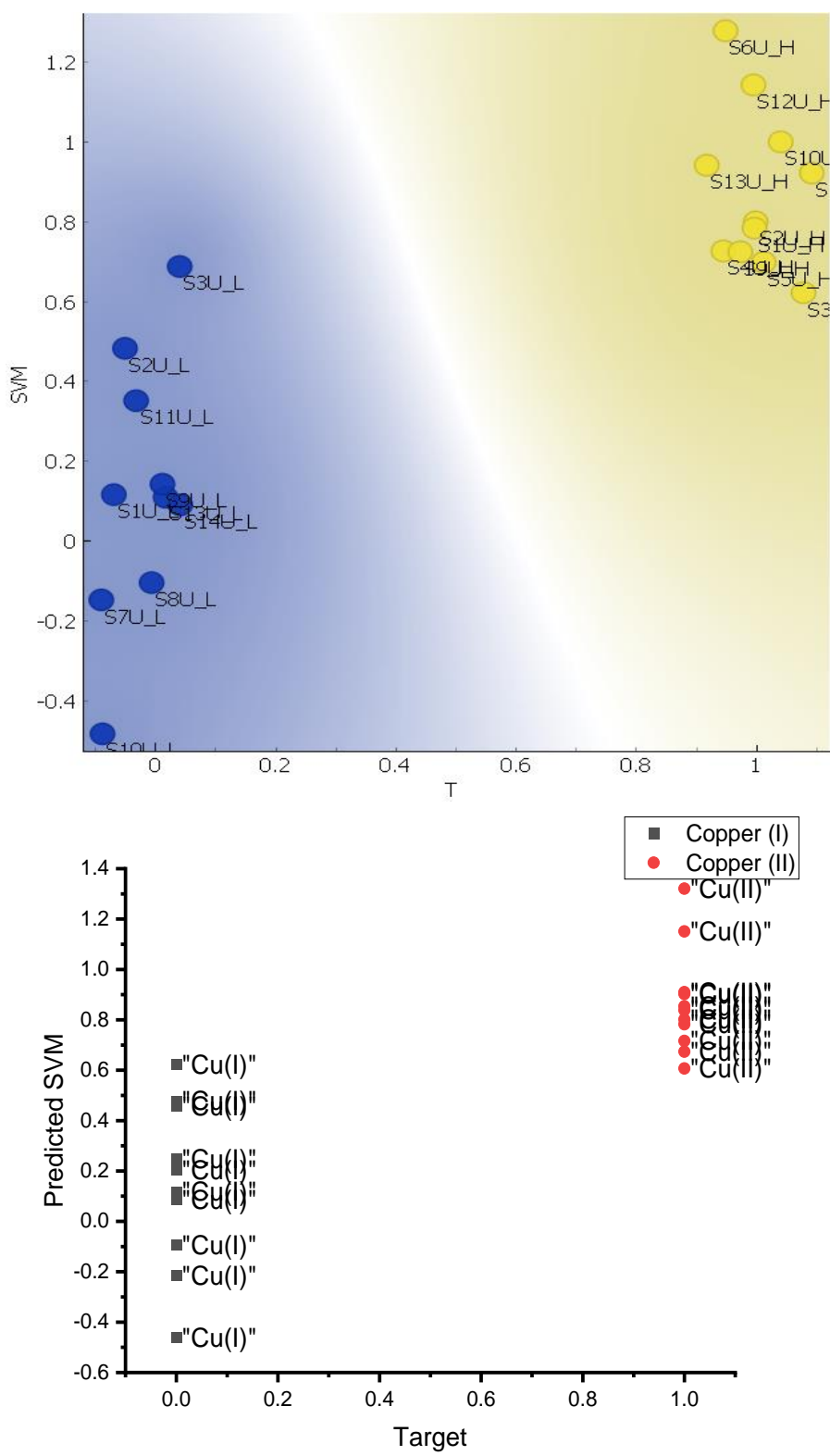


Figure 5.21 SVM speciation analysis of Cu in urine simulates

The additional information about chemical species of the elements enhanced clear investigation of the role of trace biometals in cancer diagnosis. From these results, the

changes in speciation can therefore be used as parameters for cancer diagnostic applicable at early stage of development.

5.5.1 Prediction of Fe speciation in urine test simulates by KNN model

KNN correctly distinguished urine samples containing Fe^{2+} and Fe^{3+} species as shown in Table 5.2 which predicted chemical speciation of Fe in urine simulate samples with a classification accuracy of 90%.

Table 5.2 Predicted Fe speciation in the test set of simulate samples

Sample	Species	Target Species	KNN
S13U_L	Fe^{2+}	0	0
S1U_L	Fe^{2+}	0	0
S9U_L	Fe^{2+}	0	1
S3U_L	Fe^{2+}	0	0
S9U_H	Fe^{3+}	1	1
S10U_H	Fe^{3+}	1	1
S14U_L	Fe^{2+}	0	0
S10U_L	Fe^{2+}	0	0
S13U_H	Fe^{3+}	1	1

The results obtained after show that the direct speciation is possible as the method was able to distinguish samples containing Fe and Cu chemical species. Further, KNN correctly classified the chemical speciation of Fe in the simulate samples thus potential for speciation analysis of Fe in real human urine for cancer and normal patients.

5.5.2 Prediction of chemical oxidation states of Fe and Cu in human urine

The validated models were used to determine the chemical oxidation states Fe and Cu in cancerous and normal urine. The method was able to distinguish one of these species using spectral region of interest of Fe and Cu thus can be used as parameters for cancer diagnosis. Table 5.3 shows how KNN predicted chemical speciation of Fe in cancerous and normal human urine samples.

Table 5.3 Predicted chemical speciation of Fe in cancerous and non-cancerous urine

Healthy Urine	KNN predicted	Cancerous Urine	KNN predicted
N_17	1	U_07	1
N_05	0	U_20	0
N_20	1	U_18	1
N_18	0	U_06	1
N_08	0	U_16	1
N_14	0	U_10	0
N_22	0	U_17	1
N_12	1	U_11	1
N_13	0	U_21	1
N_16	1	U_15	1
N_19	0	U_05	0
N_15	1	U_13	1

From the results, normal urine contained Fe (II) as compared to cancerous urine which was rich in Fe (III) which can be attributed to the role of Fe in oxidative reactions that damages the DNA strands for carcinogenesis. Table 5.4 shows the indirect determination of oxidation states of Cu in cancerous and normal human urine.

Table 5.4 Predicted chemical speciation of Cu in cancerous and healthy urine

Healthy urine	KNN Predicted	Cancerous urine	KNN Predicted
N_17	0	U_07	1
N_05	0	U_20	0
N_20	0	U_18	1
N_18	1	U_06	1
N_23	1	U_16	1
N_08	1	U_10	1
N_14	0	U_17	1
N_11	0	U_11	0
N_22	0	U_21	1
N_12	0	U_15	0
N_13	1	U_05	1
N_16	0	U_13	1
N_19	1		
N_15	0		

The urinary bladder cancer urine was rich in Cu^{2+} which demonstrated the possible production of free radicals for Fenton reactions (Rojas *et al.*, 1999) for DNA damage (Armendariz and Vulpe, 2003).

The KNN and SVM machine learning model-based speciation cancer diagnostic model for trace biometal speciation analysis and characterization of Fe and Cu speciation in urine samples was developed as shown in Figure 5.22.

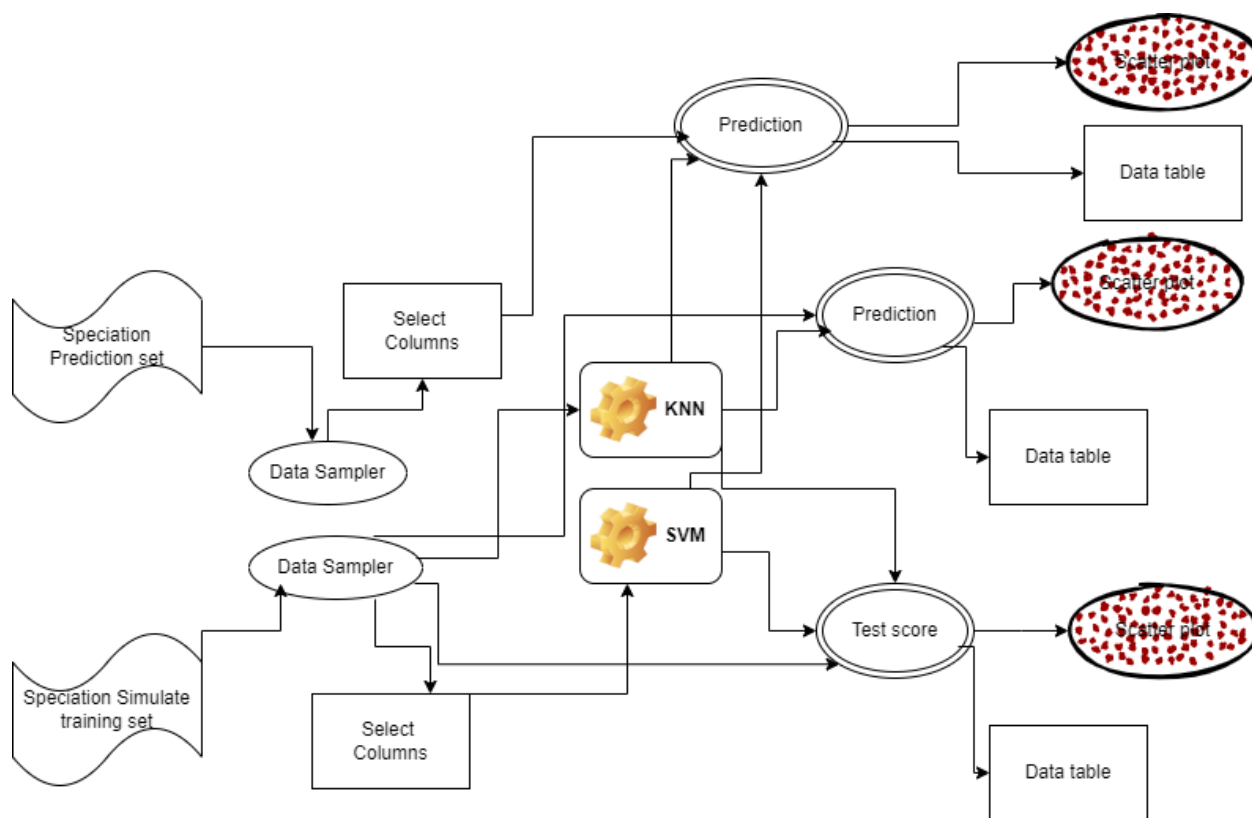


Figure 5.22 Quasar orange model for SVM and KNN speciation of Fe and Cu.

The above machine learning based method was robust and thus was able to distinguish samples containing just one chemical species using spectral region of interest. The method had the ability to indirectly determine chemical oxidation states of trace biometals via multivariate calibration modelling.

5.6 Direct chemical speciation of trace elements by XANES

The highly sensitive XANES enabled direct non-destructive determination of the chemical oxidation states (speciation) of Fe in simulate and urine samples for cancer diagnosis. The deconvoluted absorption spectral data by Athena software enabled direct determination of Fe chemical oxidation state. At the absorption edges, the specific chemical state of Fe was visible with the oxidation determined by comparison of its absorption K-edge position in the Fe calibration foil reference sample. Figure 5.23 shows the XANES typical spectrum for reference calibration sample.

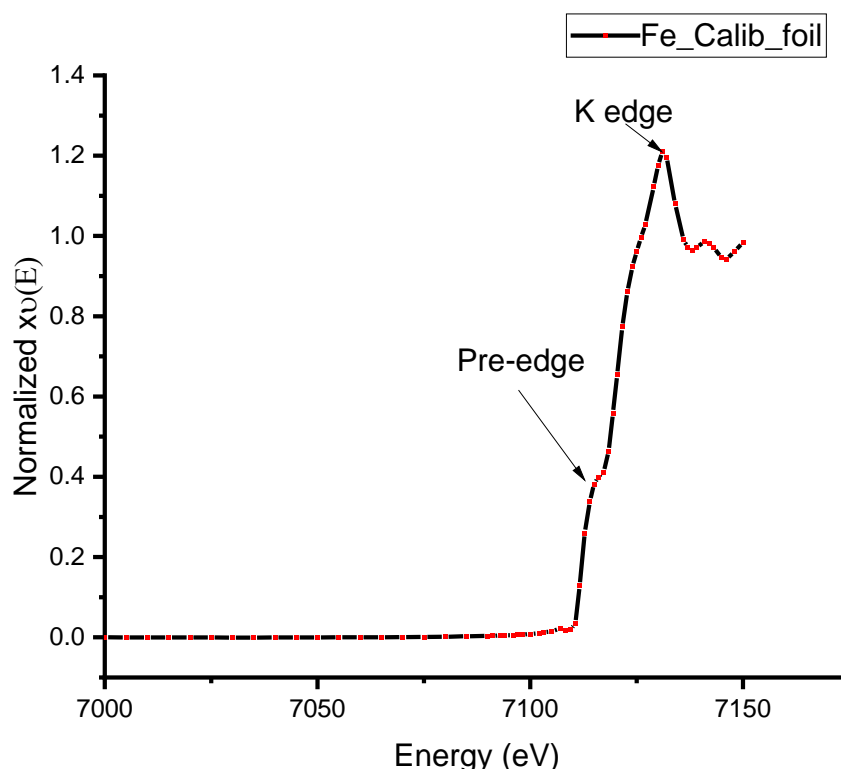


Figure 5.23 Normalized XANES of Fe calibration sample foil

The peak position was sensitive to the oxidation state and that the samples containing Fe^{2+} were distinctly different from Fe^{3+} in XANES spectrum. The pre-edge background was then removed and the edge jump normalized as shown in Figure 5.24 at Fe K edge and the background continuum removed for a pre-edge absorption at 7001-7200 eV in K-space, R-space and q-space which showed the potential for EXAFS analysis to determine the neighbouring atoms to Fe.

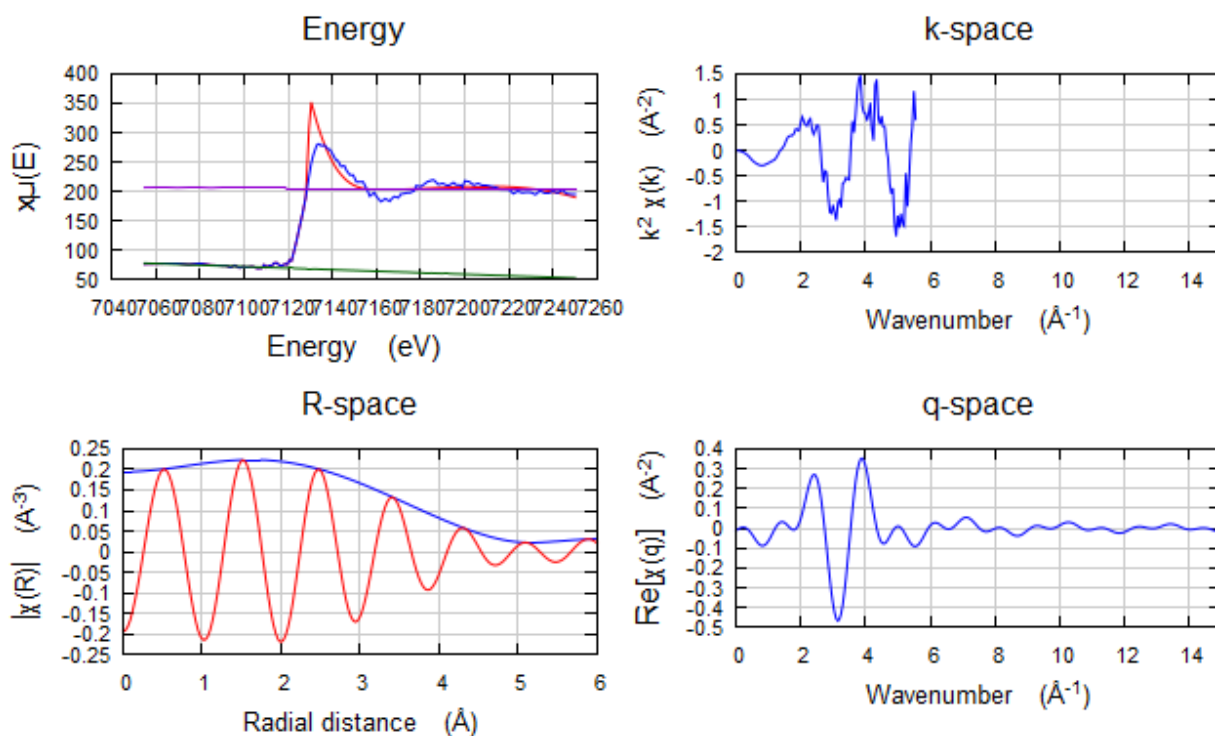


Figure 5.24 Typical EXAFS spectra in energy, k-space, R-space and q-space.

The EXAFS spectra indicated the energy spectrum smooth spline background from the pre-edge was removed to obtain the k-space, the k-space was Fourier transformed the amplitude to the R-space and q-space respectively. Figure 5.25 shows the XANES typical spectrum for Fe in the urinary bladder cancer sample with enhanced K-edge absorption.

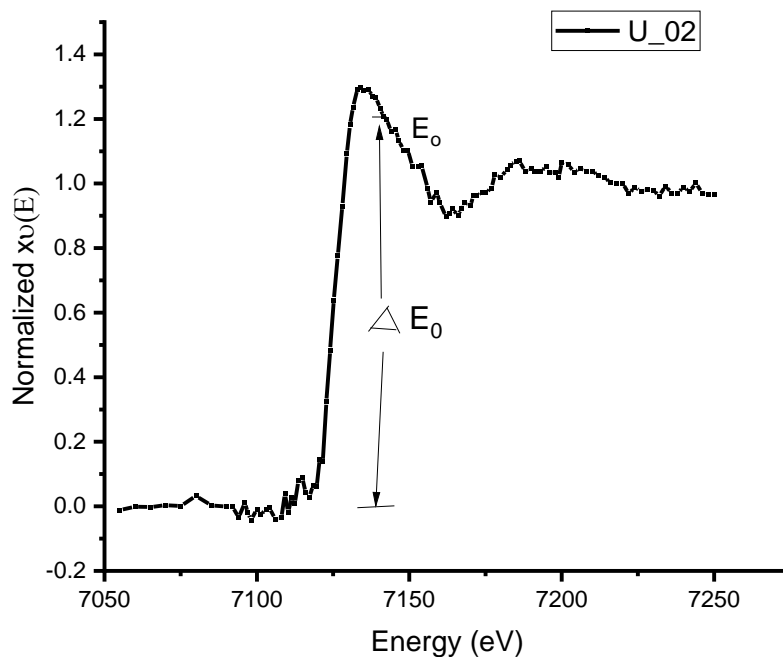


Figure 5.25 Normalized XANES of Fe in human urine

The Fe K-edge absorption edge slightly shifted towards higher energy when the chemical state of *Fe* changed as shown in the overlapped spectra in Figure 5.26 for the calibration standard and human urine (cancerous).

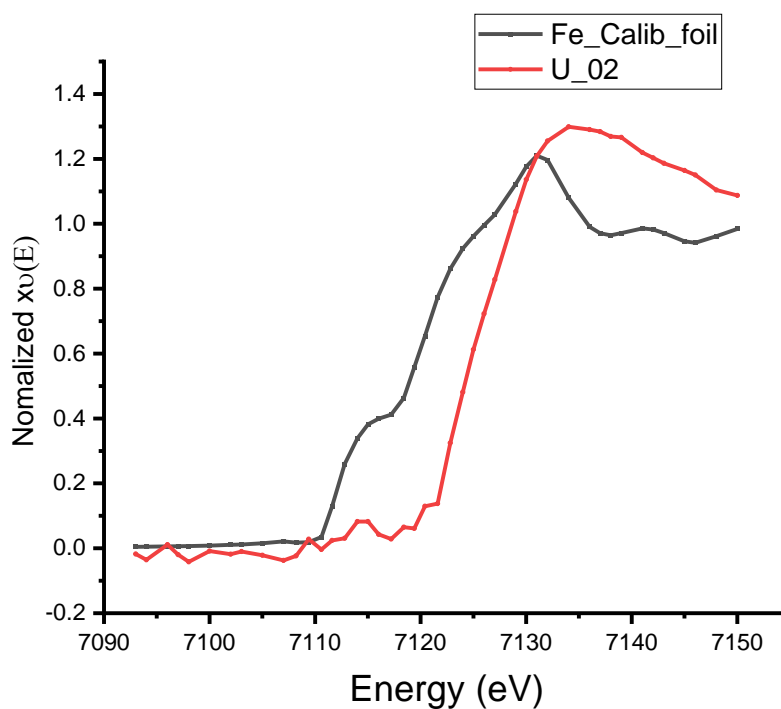


Figure 5.26 Overlapped spectra of reference Fe foil and urine sample

The significant difference in the shape of pre-edge of the Fe calibration foil and the urine sample was attributed to the shift in the pre-edge of Fe to the right in the urinary bladder cancer sample. The peak at the edge at 7123 eV was predominant compared to Fe in the reference standard foil where the peak was not visible. The XANES qualitative results indicate that all Fe was in trivalent state and that the pre-edge characteristic feature in Fe can be attributed to 1s transition to 4p.

The results further confirmed the results for the EDXRF indirect determination of the Fe oxidation state in simulate sample as shown in Fig. 5.36 in comparison with the Fe calibration foil in Figure 5.27.

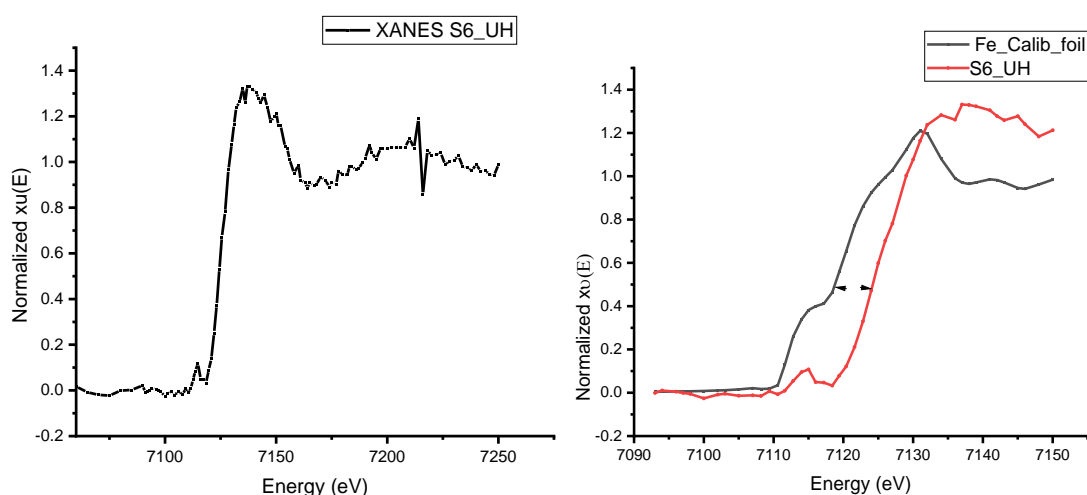


Figure 5.27 Normalized and overlapped XANES spectra of Fe in simulate and reference sample

For the simulate sample, the pre-edge absorption was intense and observed at less than 7113 eV where a small shift to the right was evident. The XANES analysis confirmed and validated the higher oxidation state of Fe^{3+} as predicted by the indirect speciation model using EDXRF spectral data.

Figure 5.28 shows the XANES typical spectra for Fe in cancerous human urine sample overlapped with Fe calibration foil standard.

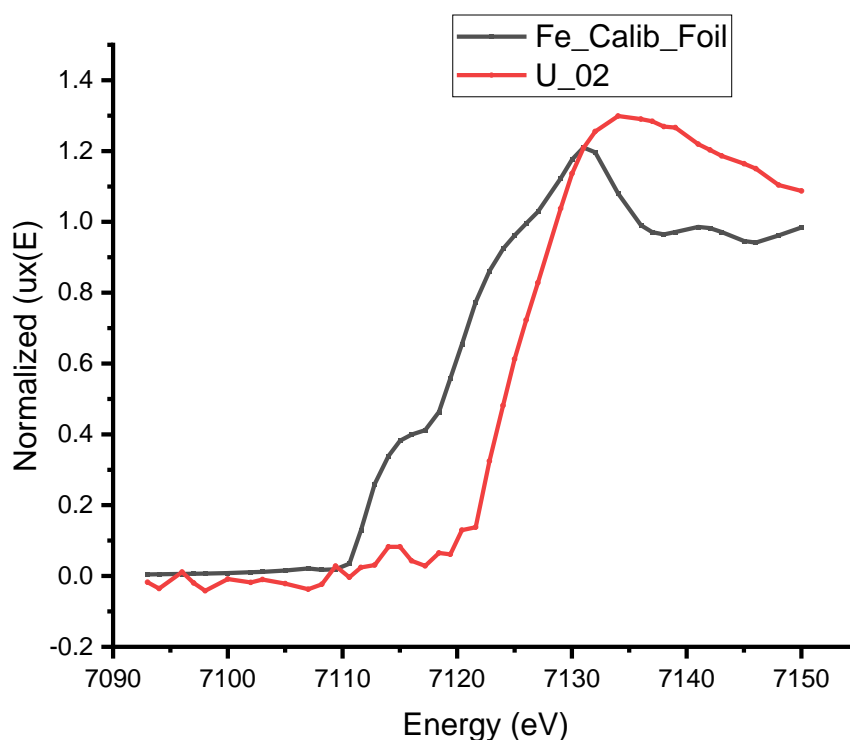


Figure 5.28 Overlapped spectra of reference Fe foil and cancerous human urine samples

The XANES spectra indicated that Fe was prevalent in trivalent state with an energy shift of 2.8 eV to the right. Cancerous sample (U-02) was confirmed to have Fe^{3+} which agreed with (Kwiatek *et al.*, 2005) probably due to its role in creation of new blood cells. Fe in this case promote carcinogenesis by redox reactions that results to production of free radicals which attack the DNA. Similarly, synchrotron-based XANES (Al- Ebraheem *et al.*, 2010) also found Fe^{3+} slightly higher than Fe^{2+} in ductal carcinoma breast tissues which was also linked to Fenton reactions for reactive oxygen species that induces DNA damage. The qualitative XANES results of Fe confirms ferritin as dominant chemical form for carcinogenesis (Gherase and Fleming 2019). In this regard, XANES was able to probe and validate the oxidation state of iron in the simulate samples thus potential utility of indirect speciation model in cancer diagnosis as it provides significant characteristic chemical oxidation state information for cancer diagnosis.

5.7 Quantitative analysis of trace elements in human urine

The quantitative analysis of trace biometals (Cu, Fe, Se, Zn and Mn) was determined to evaluate the statistical significance of trace elemental accumulation in normal and cancerous human tissues and urine. These indicated the potential role of these biometals in cancer diagnosis as they are important for biological and enzymatic processes.

5.7.1 EDXRF analysis of trace biometals in urine samples

The changes in trace biometal concentration in response to carcinogenesis were detected and characterized as a “fingerprint” of prostate cancer. The concentrations level descriptive statistics of trace biometals (Fe, Cu and Zn) in non-cancerous urine were obtained by fundamental parameter method as shown in Table 5.5.

Table 5.5 Trace biometals levels (ppm) in healthy urine

Cancerous	N	Mean	SD	Min	Max
Fe	16	12.66	±3.04	8.56	20.80
Cu	16	7.24	±1.24	4.86	9.56
Zn	16	5.59	±2.13	2.42	10.20

The concentration of Fe, Cu and Zn in cancerous are slightly higher than in the non-cancerous urine as shown in Table 5.6.

Table 5.6 Trace biometals levels (ppm) in cancerous prostate urine

Normal	N	Mean	SD	Min	Max
Fe	15	13.52	±2.83	9.34	18.90
Cu	15	8.02	±1.73	5.44	11.20
Zn	15	5.93	±2.26	3.41	12.40

In this study, alterations in concentration of Fe, Cu and Zn were evident in both prostate cancer and normal prostatic tissue. It's evident that the concentrations of Cu and Zn were alterations (8.02 ± 1.73 ppm and 5.93 ± 2.26 ppm) but with slightly higher level (13.52 ± 2.83 ppm) of Fe in prostate cancer. This result mimics a similar study (Zaichick *et al.*, 1997) where slightly high levels of Fe and Cu were evident in cancer tissues. In this case, trace biometals are actively involved in activation or inhibition of enzymatic reactions (Carvalho *et al.*, 2007) hence the damage of the DNA. The trace biometals (Cr, Mn and Se)

were not detected (Appendix VI) as they were below the detection limit of EDXRF which is less sensitive.

5.7.2 TXRF analysis of trace biometal levels in urine samples

Total reflection X-ray fluorescence (TXRF) quantitative analysis was achieved by utility of Ni as an internal standard to determine the concentrations of trace elements in cancerous as shown in Table 5.7. High levels of Fe and Cu were correlated to the onset and development of cancer as they are important in biological and metabolism processes that accelerate carcinogenesis (Kainth *et al.*, 2020).

Table 5.7 Concentration of cancerous urine by TXRF

Sample	Concentration of trace biometals (ppm)					
	Cr	Mn	Fe	Cu	Zn	Se
P_05	0.011 ±0.003	0.047 ±0.014	0.091 ±0.012	0.045 ±0.007	0.619 ±0.013	0.027 ±0.004
P_08	0.02 ±0.001	0.013 ±0.001	0.028 ±0.003	0.010 ±0.002	0.017 ±0.002	0.008 ±0.001
P_15	0.10 ±0.01	0.116 ±0.018	0.781 ±0.024	0.066 ±0.008	1.348 ±0.021	0.040 ±0.005
P_21	0.233 ±0.069	0.189 ±0.046	0.419 ±0.055	0.231 ±0.027	2.329 ±0.043	0.143 ±0.017
P_22	0.027 ±0.001	0.015 ±0.001	0.030 ±0.001	0.008 ±0.001	0.058 ±0.001	0.006 ±0.002
P_29	0.026 ±0.007	0.03 ±0.003	0.057 ±0.007	0.023 ±0.004	0.479 ±0.008	0.024 ±0.004
P_07	0.028 ±0.004	0.023 ±0.007	0.255 ±0.016	0.108 ±0.009	0.419 ±0.012	0.034 ±0.004
P_11	0.054 ±0.005	0.115 ±0.005	0.705 ±0.023	0.072 ±0.005	0.236 ±0.007	0.007 ±0.002
P_17	0.059 ±0.005	0.043 ±0.002	0.05 ±0.013	0.035 ±0.009	0.775 ±0.008	0.044 ±0.004
P_18	0.047 ±0.003	0.058 ±0.015	0.375 ±0.016	0.224 ±0.008	2.62 ±0.023	0.008 ±0.002
P_20	0.018 ±0.006	0.012 ±0.004	0.070 ±0.005	0.012 ±0.002	0.265 ±0.004	0.113 ±0.027
P_25	0.016 ±0.002	0.022 ±0.001	0.311 ±0.005	0.022 ±0.002	0.400 ±0.004	0.028 ±0.003

This study further indicates elevated levels of Zn in cancerous tissue samples for cell proliferation that supports tumor growth (Lee *et al.*, 2003) and cell division. Additionally, levels of Cu in cancerous samples were also evident due to increased generation of free

radicals in Fenton reaction and/or by interference with DNA repair processes (Yaman *et al.*, 2005). The results indicate that Fe, Cu and Zn concentrations were significantly ($P<0.05$) higher in cancerous than in healthy tissues. The significant ($P<0.05$) alteration of concentrations of Fe, Cu and Zn are possible trace biomarkers for early cancer diagnosis. The elevated levels of Fe, Cu, Zn and Se in cancerous compared can be attributed to increased cell division and proliferations. Table 5.8 shows the levels of trace biometals in non-cancerous (normal) prostate urinary samples where the low concentrations of trace biometals were evident.

Table 5.8 Concentration of trace biometals in 10 μ l normal urine by TXRF

Sample	Concentration of trace biometals (ppm)					
	Cr	Mn	Fe	Cu	Zn	Se
NC_05	0.015 ± 0.003	0.012 ± 0.001	0.109 ± 0.003	0.034 ± 0.003	0.067 ± 0.002	0.016 ± 0.001
NP_06	0.048 ± 0.002	0.041 ± 0.005	0.092 ± 0.008	0.039 ± 0.005	0.798 ± 0.008	0.029 ± 0.003
NP_07	0.205 ± 0.023	0.156 ± 0.012	0.219 ± 0.036	0.092 ± 0.042	0.845 ± 0.032	0.067 ± 0.008
NP_13	0.059 ± 0.007	0.039 ± 0.014	0.101 ± 0.011	0.030 ± 0.005	0.412 ± 0.009	0.026 ± 0.003
NP_04	0.176 ± 0.003	0.153 ± 0.013	0.801 ± 0.038	0.119 ± 0.015	1.079 ± 0.031	0.064 ± 0.010
NP_12	0.032 ± 0.011	0.023 ± 0.009	0.042 ± 0.005	0.016 ± 0.003	0.262 ± 0.005	0.013 ± 0.002
NP_14	0.058 ± 0.005	0.037 ± 0.003	0.071 ± 0.008	0.029 ± 0.004	0.309 ± 0.006	0.023 ± 0.003
NP_18	0.013 ± 0.008	0.01 ± 0.001	0.049 ± 0.004	0.013 ± 0.002	0.066 ± 0.003	0.006 ± 0.001
NP_19	0.014 ± 0.003	0.011 ± 0.002	0.029 ± 0.002	0.009 ± 0.001	0.153 ± 0.002	0.005 ± 0.001
NP_20	0.098 ± 0.027	0.085 ± 0.023	0.128 ± 0.021	0.068 ± 0.011	0.794 ± 0.020	0.028 ± 0.007
NP_23	0.021 ± 0.011	0.017 ± 0.009	0.018 ± 0.005	0.016 ± 0.003	0.81 ± 0.008	0.008 ± 0.002
NP_39	0.045 ± 0.005	0.028 ± 0.004	0.075 ± 0.006	0.022 ± 0.004	0.122 ± 0.005	0.016 ± 0.002

Elevated levels of Fe were also found in malignant tissues (Geraki *et al.*, 2004) probably due to its role as regulatory factor for angiogenesis and its catalytic role in the Fenton reaction that generates free radicals (Heath *et al.*, 2013) for tumor angiogenesis. Fe levels were elevated in cancerous tissue than normal urine which was in agreement with (Geraki *et al.*,

2004) who also found elevated Fe in malignant tissues probably due to its role in angiogenesis and catalytic role in Fenton reaction that results to increased requirement for DNA synthesis (Heath *et al.*, 2013) for tumor angiogenesis. Therefore, it is plausible that cancer development is associated with elevated levels Fe, Cu and Zn (Zaichick and Zaichick 2011).

5.8 Multivariate regression models for trace biometals in urine samples

The concentrations of the trace biometals by classical XRF was affected by the extreme matrix effects, low signal to noise ratio and weak analyte fluorescence signals. The EDXRFS spectral data of simulate samples were therefore used to develop multivariate calibration model for prediction of concentrations in unknown samples of similar matrices. The validated ANN and SVM calibration models were used to determine the concentration of the trace elements of interest in cancerous and normal human urine.

5.8.1 ANN and SVM regression models for prediction of biometals in urine samples

The trace elements of interest were directly quantified using the EDXRFS data sets in spite of the low resolution, enhanced matrix effects and noisy elemental spectra that is typical from complex light element matrices such as the biomedical samples (tissues and urine). ANN and SVM analysis were carried out using the principal components with highest variance (PC1 and PC2) in PCA. The ANN architecture (3:2:5) corresponding to number of neurons in the input, hidden and output layers, respectively was chosen based on low MSE attained for both training and validation sets. The training and testing of the networks were achieved using the calibration and test samples for validating the trained network with the model performance as shown in Table 5.9.

Table 5.9 ANN and SVM urine calibration model performance

		Model performance indices			
Trace Element		MSE	RMSE	MAE	R ²
Cu	ANN	6.48	2.55	2.07	0.77
	SVM	8.48	2.91	2.74	0.70
Zn	ANN	359.72	18.97	17.44	0.99
	SVM	25269.93	158.97	78.22	0.48
Fe	ANN	143.11	11.96	9.47	0.81
	SVM	400.84	20.02	10.58	0.46
Mn	ANN	0.51	0.72	0.60	0.99
	SVM	8.27	2.88	2.73	0.83
Cr	ANN	4.77	2.18	1.52	0.87
	SVM	7.04	2.65	2.42	0.80
Se	ANN	12.49	3.53	2.90	0.92
	SVM	11.67	3.42	3.06	0.93

The long training periods resulted to improved estimation ability with lower MSE attained at 2000 iterations for both ANN and SVM architecture. The predicted concentration verses the measured (known) concentration regression plots (Figure 5.29 to Figure 5.34) show a linear correlation on how the multivariate calibration models (SVM and ANN) were able to predict the concentrations of trace elements of interest (Cr, Fe, Cu, Mn, Zn and Se).

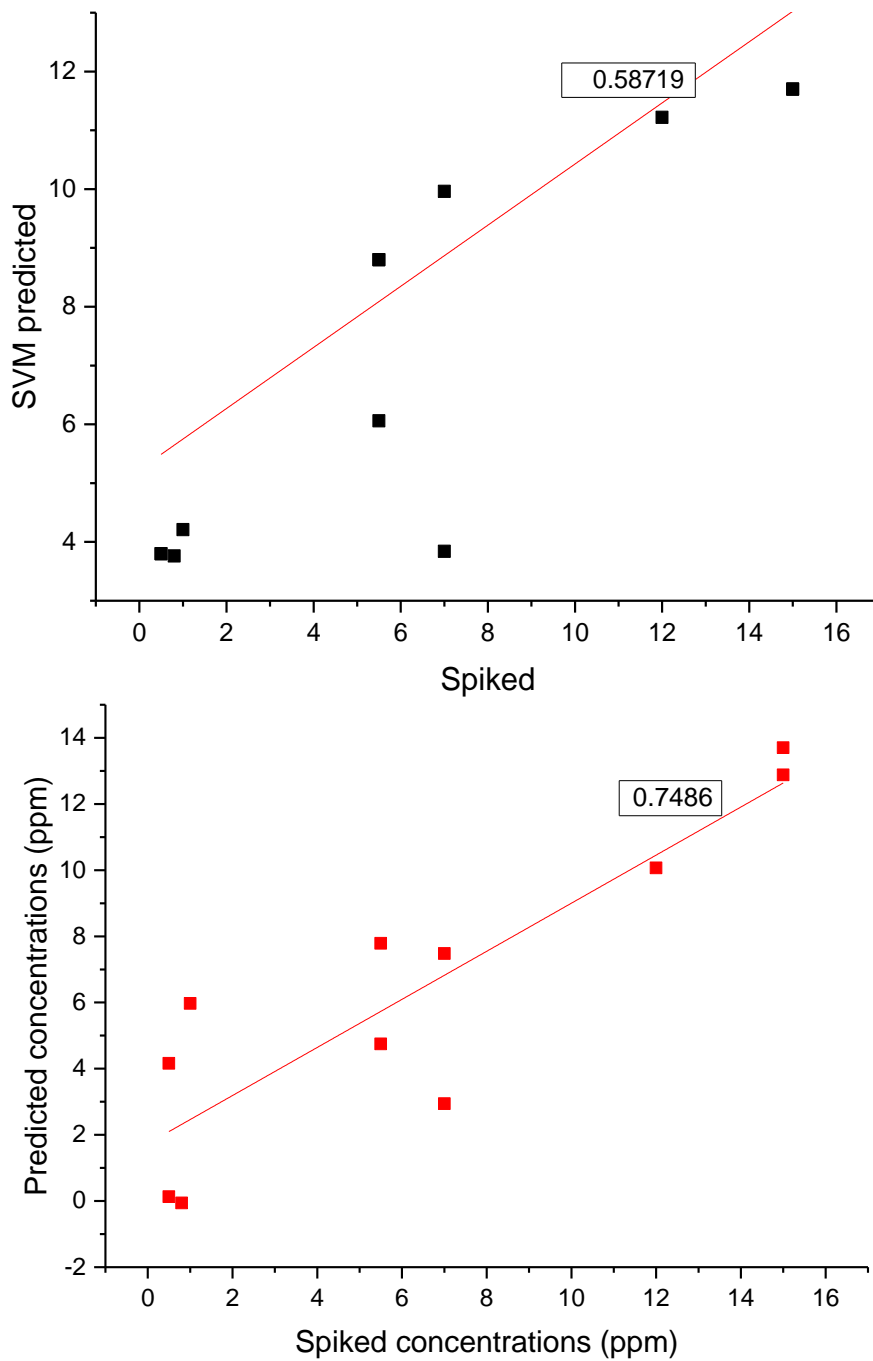


Figure 5.29 Simulate SVM and ANN model validation plots of Cu

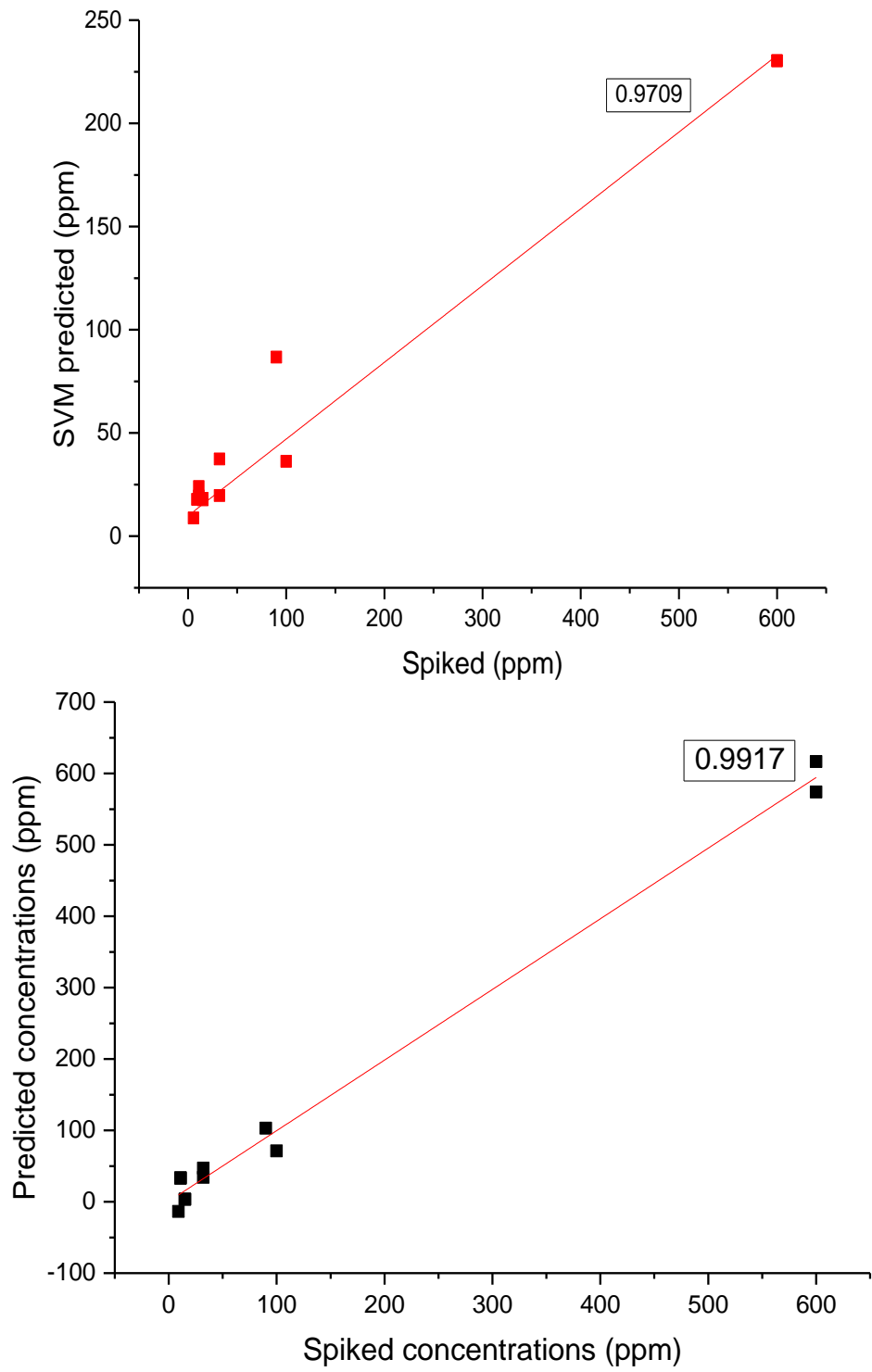


Figure 5.30 Simulate SVM and ANN model validation plots of Zn

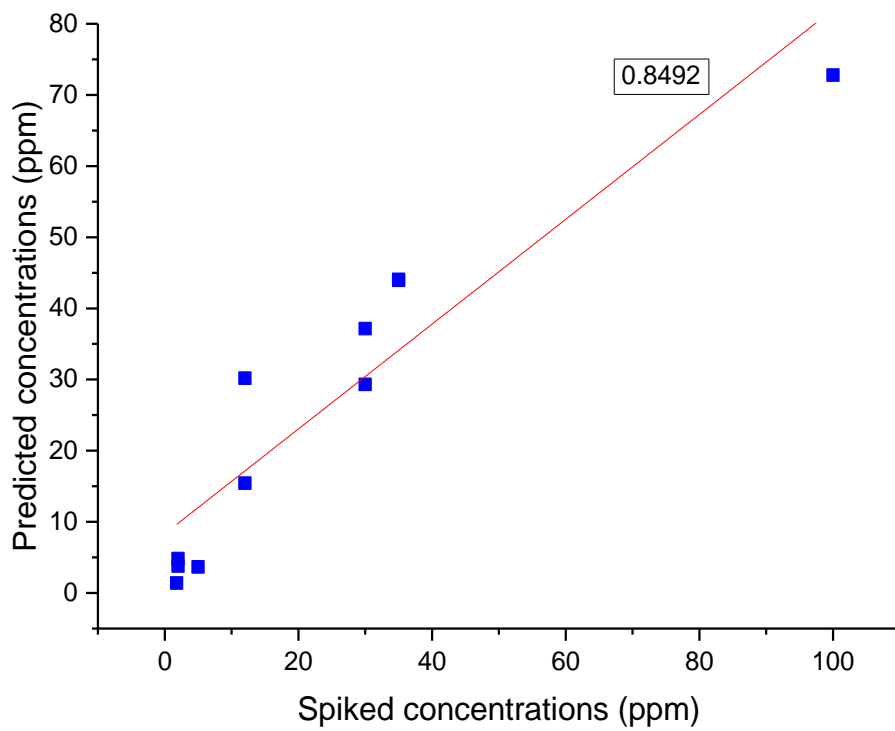
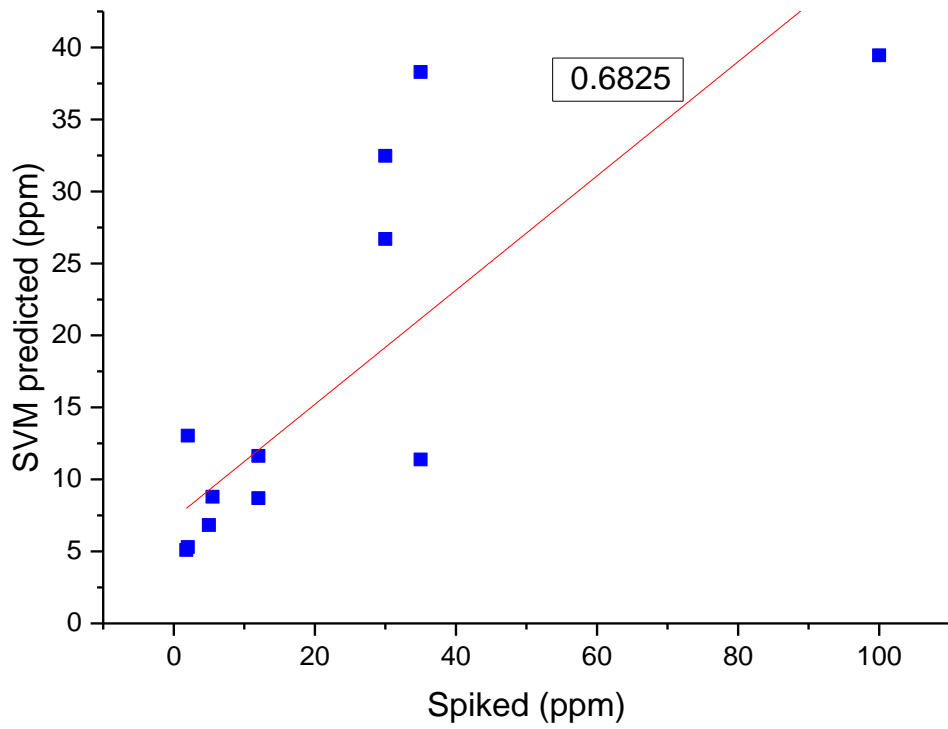


Figure 5.31 Simulate SVM and ANN model validation plots of Zn

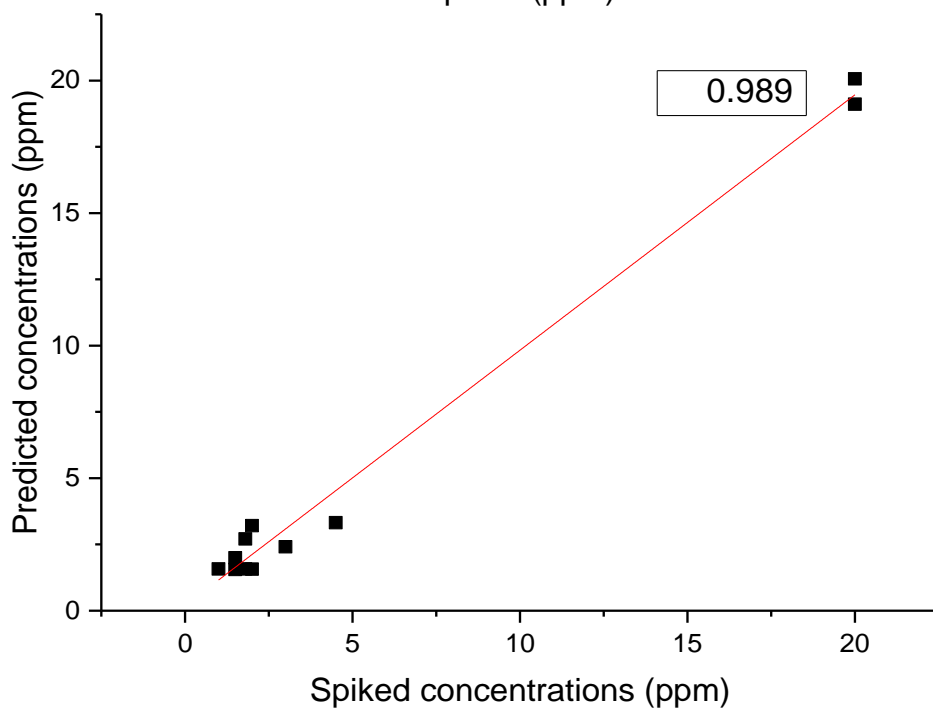
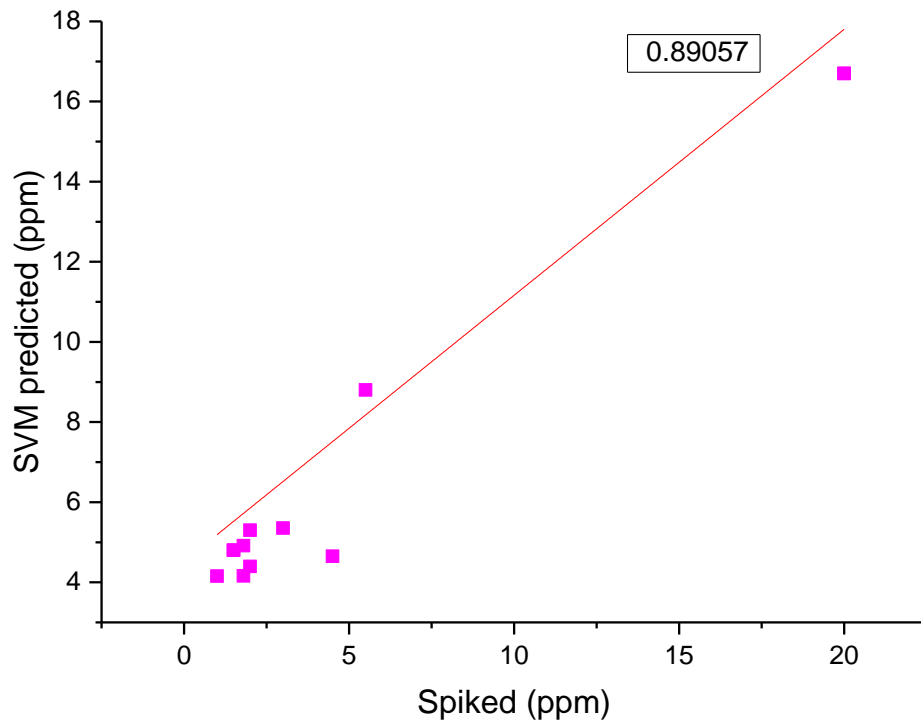


Figure 5.32 Simulate SVM and ANN model validation plots of Mn

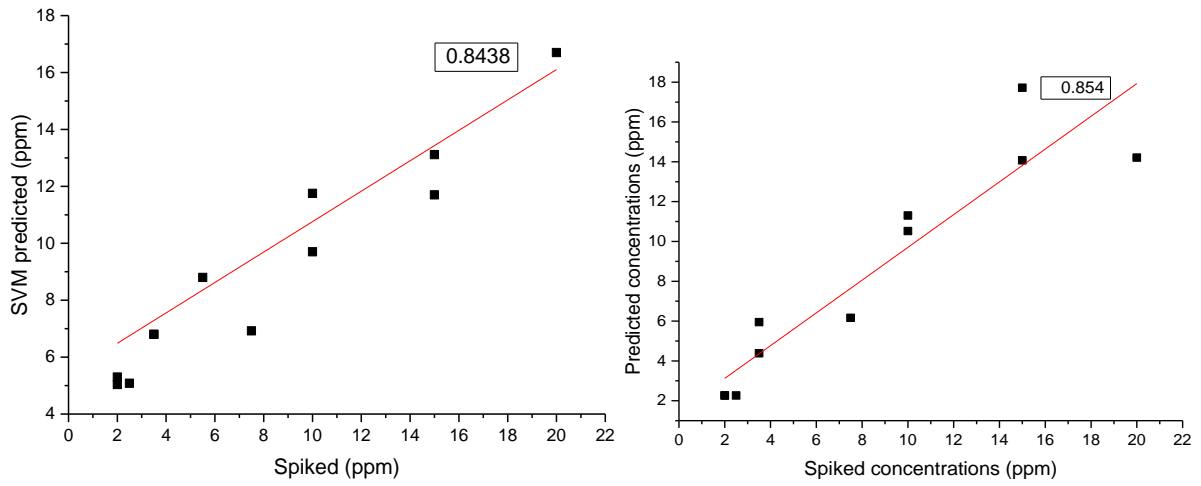


Figure 5.33 Simulate SVM and ANN model validation plots of Cr

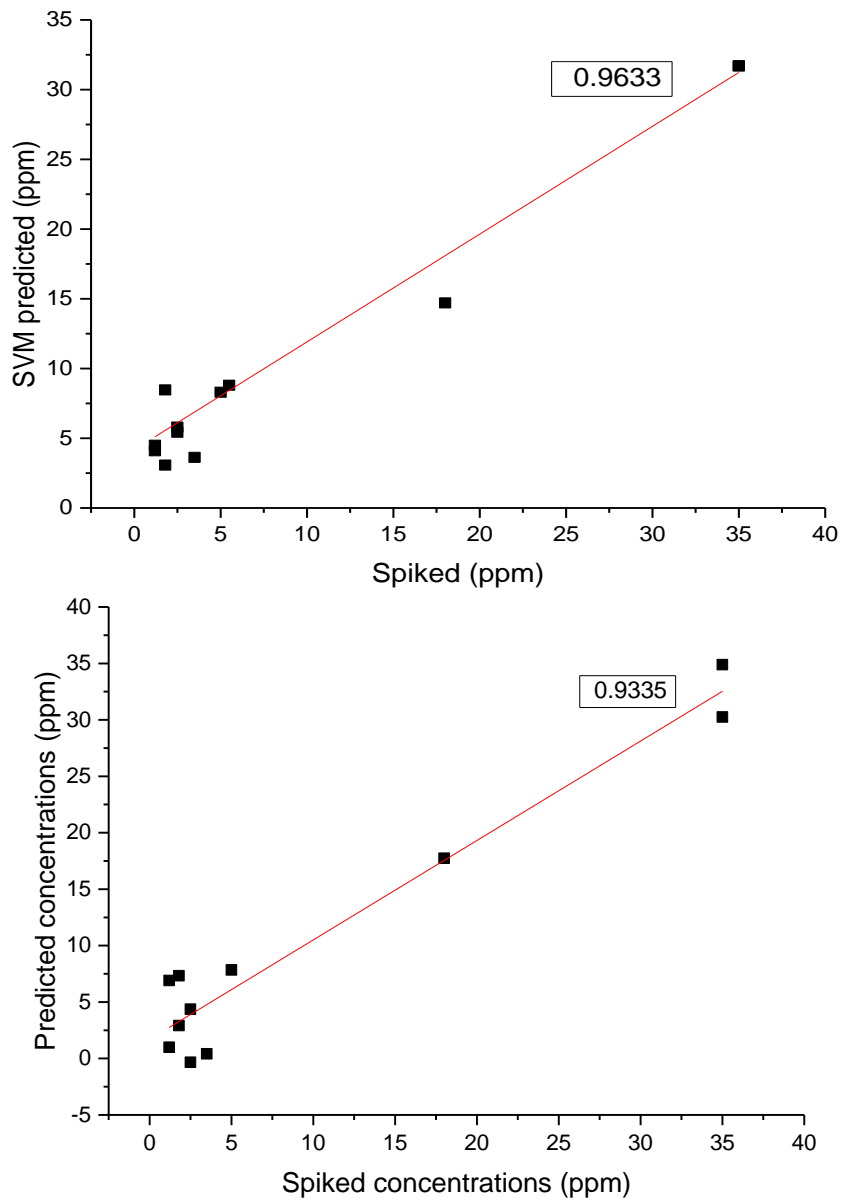


Figure 5.34 Simulate SVM and ANN model validation plots of Se

In general, the overall ANN model predictions of trace biometals levels was better than the SVM regression model where the close correlation between measured and predicted. 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).

5.8.2 Detection and quantification limits of trace biometals in simulate urine samples

In order to determine the sensitivity of the multivariate calibration strategy, the detection limits (*DL*) and quantitative limits of trace biometals in the simulate urine samples were calculated for the trace analytes (Cr, Mn, Fe, Cu, Zn and Se) as shown in Table 5.10.

Table 5.10 Detection limits (ppm) of elements in simulate urine

Element	Atomic number	K_{α} Energy (KeV)	Detection limit (ppm)	Quantitative limit (ppm)
Cr	24	5.41	1.56	2.56
Mn	25	5.89	2.14	4.67
Fe	26	6.40	2.23	5.11
Cu	29	8.05	2.05	3.76
Zn	30	8.64	2.98	4.81
Se	34	11.22	1.64	2.84

The concentrations of trace elements below the detection and quantitative limits in Table 5.10 were hardly achieved by the developed multivariate model towards cancer diagnostics in human urine.

5.8.3 Predicted concentrations in urine samples

The above validated model calibration curves were further utilized to determine the concentration of trace biometals in cancerous and normal human urine. Table 5.11 shows the ANN predicted elemental concentration in normal urine by ANN model based on the mean-centered and smoothened fluorescence and scatter spectral data.

Table 5.11 ANN Predicted concentrations of trace biometals in normal urine

Trace biometal concentrations (ppm)						
Nor	Fe	Mn	Cr	Cu	Se	Zn
NP_12	18.74±1.43	1.95±0.57	8.05±0.06	2.68±0.01	7.91±2.08	12.66±0.34
NP_20	9.18±4.33	1.95±0.67	8.06±0.07	3.14±0.04	7.34±2.22	12.43±0.06
NP_16	11.55±2.81	1.89±0.39	8.05±0.09	4.73±0.01	5.41±1.99	12.41±0.01
NP_11	10.80±1.49	1.58±0.19	8.07±0.07	2.68±0.03	4.71±2.25	12.49±0.12
NP_14	31.71±2.96	2.67±1.16	8.05±0.08	5.79±0.02	7.64±1.51	12.49±0.11
NP_22	16.37±1.93	2.14±0.72	8.06±0.07	5.40±0.03	6.13±2.10	12.65±0.36
NP_17B	22.22±0.65	1.39±0.04	8.16±0.09	3.62±0.10	6.70±0.24	11.89±0.66
NP_17	10.47±1.59	2.40±0.98	8.05±0.07	2.85±0.03	7.35±2.32	12.45±0.05
NP_19	24.57±4.53	1.51±0.18	8.06±0.07	2.86±0.03	5.58±2.11	12.43±0.04
NP_06	13.02±3.08	2.62±1.19	8.05±0.07	6.61±0.06	9.33±1.46	13.14±1.02
NP_08	21.13±1.45	2.29±0.92	8.05±0.08	5.94±0.06	10.16±1.08	12.53±0.18
NP_18	8.27±3.59	2.49±0.86	8.04±0.07	5.03±0.02	8.53±2.33	12.53±0.14
NP_15	8.21±3.75	2.34±0.99	8.07±0.06	6.58±0.07	6.35±1.11	12.72±0.44
	7.11±2.53	1.94±0.66	8.05±0.08	5.62±0.01	6.42±2.25	12.54±0.19
	8.78±1.89	1.79±0.49	8.06±0.07	4.35±0.02	5.98±2.09	12.61±0.26
	8.59±1.75	1.77±0.45	8.05±0.07	3.79±0.04	4.36±1.89	12.33±0.17

The levels of most trace elements can be linked to their functional role in metabolic processes. The concentrations of Fe, Cu, Zn and Se were all suppressed in normal urine specimens. Table 5.12 shows the ANN predicted elemental concentration in prostate cancer urine by ANN model. The elevated levels of Fe and Cu with suppressed level of Zn and Se could be a consequence of malignant transformation.

Table 5.12 ANN predicted concentrations of trace elements in cancerous urine

	Trace biometal concentrations \pm standard deviation (ppm)											
	Fe		Mn		Cr		Cu		Se		Zn	
P_24	49.83	± 2.52	1.58	± 0.28	8.05	± 0.07	10.51	± 0.11	5.88	± 0.49	12.13	± 0.97
P_23	54.97	± 1.84	4.11	± 0.14	7.04	± 0.08	18.09	± 0.23	3.14	± 1.06	13.23	± 1.14
P_15	24.13	± 2.94	1.61	± 0.36	8.05	± 0.07	10.49	± 0.04	4.81	± 1.59	11.37	± 0.07
P_36	12.02	± 2.98	1.49	± 0.19	7.05	± 0.09	7.84	± 0.25	3.29	± 1.85	10.11	± 0.47
P_11	6.26	± 1.62	1.63	± 0.29	8.06	± 0.06	9.05	± 0.09	6.75	± 2.08	12.19	± 0.64
P_37	11.50	± 1.07	1.37	± 0.08	9.04	± 0.08	9.06	± 0.05	6.12	± 2.02	11.71	± 0.40
P_10	13.18	± 1.50	1.87	± 0.58	8.04	± 0.09	10.46	± 0.17	7.93	± 2.47	12.54	± 0.18
P_07	32.25	± 2.55	3.72	± 0.22	8.05	± 0.07	11.66	± 0.12	3.35	± 2.59	12.40	± 0.02
P_41	33.71	± 3.64	2.94	± 0.88	6.05	± 0.07	16.91	± 0.01	6.62	± 2.20	11.88	± 0.67
P_12	36.93	± 2.09	2.69	± 1.15	8.05	± 0.07	11.91	± 0.01	8.93	± 2.06	12.57	± 0.23
P_45	25.19	± 3.64	2.26	± 0.46	7.05	± 0.07	9.63	± 0.01	4.59	± 1.51	13.65	± 0.31
P_14	56.53	± 5.28	1.51	± 0.26	8.05	± 0.08	9.45	± 0.07	7.89	± 2.49	10.33	± 0.14
P_20	32.70	± 1.92	1.54	± 0.24	6.04	± 0.07	10.01	± 0.01	9.63	± 2.27	12.74	± 0.45
P_06	18.76	± 2.91	2.20	± 0.43	8.05	± 0.09	8.44	± 0.05	6.84	± 1.98	10.38	± 0.04
P_32	6.16	± 1.51	3.67	± 0.34	8.04	± 0.07	6.69	± 0.19	4.43	± 2.12	12.56	± 0.18

Table 5.13 shows the ANN predicted elemental concentration in urinary bladder cancer urine by ANN model.

Table 5.13 ANN cancerous urine of trace elemental concentration in the urinary bladder

	Trace biometal concentrations \pm standard deviation (ppm)											
	Fe		Mn		Cr		Cu		Se		Zn	
U_17	20.14	± 4.39	3.20	± 0.60	8.05	± 0.05	7.82	± 0.03	9.81	± 2.66	11.34	± 0.10
U_18	18.86	± 3.61	2.46	± 1.02	8.04	± 0.07	8.11	± 0.04	8.95	± 2.61	11.49	± 0.09
U_21	25.31	± 2.44	1.81	± 0.54	8.03	± 0.06	7.62	± 0.04	9.19	± 2.42	12.83	± 0.58
U_10	16.94	± 1.14	2.21	± 0.78	8.05	± 0.07	7.20	± 0.01	9.54	± 2.43	10.44	± 0.06
U_06	15.97	± 3.69	2.12	± 0.79	8.04	± 0.07	8.75	± 0.11	10.61	± 2.63	12.39	± 0.04
U_05	16.69	± 2.02	2.89	± 1.07	8.04	± 0.08	6.28	± 0.03	8.97	± 2.12	11.09	± 0.93
U_20	42.85	± 2.49	2.13	± 0.61	8.06	± 0.08	7.74	± 0.04	9.59	± 2.05	11.37	± 0.06
U_13	14.29	± 1.66	2.07	± 0.71	8.05	± 0.07	7.22	± 0.01	7.06	± 2.09	12.58	± 0.22
U_16	22.83	± 3.89	1.82	± 0.62	8.07	± 0.07	6.33	± 0.06	7.39	± 2.14	11.67	± 0.37
U_15	16.44	± 1.72	2.19	± 0.76	8.05	± 0.06	7.61	± 0.01	6.06	± 2.19	10.55	± 0.18
U_07	19.54	± 2.60	2.13	± 0.81	8.06	± 0.05	8.38	± 0.02	7.07	± 2.22	10.73	± 0.49
U_11	21.12	± 3.09	1.89	± 0.56	8.04	± 0.07	7.74	± 0.01	7.76	± 2.36	12.66	± 0.35

It can be noted that there is elevated levels of trace biometals in cancerous urine compared to normal urine which may possibly be associated with metabolic processes (Baker *et al.*, 2002) as well as tumor development (Nguyen *et al.*, 2001). The identified mean imbalance of trace biometals in cancerous and healthy patients as shown in Table 5.14 serve as biomarkers for early diagnosis of cancer as the elements are involved in protein

synthesis, antioxidant defence immune function, and inhibition proliferation of cells (Silvera and Rohan 2007).

Table 5.14 Trace elements descriptive statistics in healthy urine

Trace biometal concentrations (ppm)					
	N	Mean	SD	Min	Max
Fe	16	14.42	±7.29	7.11	31.71
Mn	16	2.05	±0.39	1.39	2.67
Cr	16	8.06	±0.02	8.05	8.16
Cu	16	4.48	±1.42	2.68	6.61
Se	16	6.87	±1.60	4.36	10.16
Zn	16	12.52	±0.25	11.89	13.14

Table 5.15 Trace elements descriptive statistics in prostate cancer urine

Biometal	N	Mean	SD	Min	Max
Fe	15	27.61	±16.85	6.17	56.54
Mn	15	2.28	±0.93	1.37	4.11
Cr	15	7.65	±0.83	6.05	9.05
Cu	15	10.68	±3.09	6.69	18.09
Se	15	5.73	±2.53	0.19	9.64
Zn	15	11.29	±2.74	2.11	13.65

Table 5.16 Trace elements descriptive statistics in urinary bladder cancer urine

	N	Mean	SD	Min	Max
Fe	12	20.92	±7.58	14.29	42.85
Mn	12	2.24	±0.42	1.81	3.21
Cr	12	8.05	±0.01	8.05	8.06
Cu	12	7.57	±0.74	6.28	8.75
Se	12	8.50	±1.39	6.06	10.61
Zn	12	11.59	±0.84	10.44	12.83

The mean levels of trace elements in cancer patients were found to be significantly higher compared to their corresponding controls. Elevated levels of trace biomarkers (Fe and Cu) in prostate and urinary bladder cancer can be correlated to the onset and development of cancer. High levels of Fe has the ability to promote transformation of normal healthy cells to

neoplastic cells through oxidative stress (Shpyleva *et al.*, 2011). The high concentration of Fe in this study is in agreement with (Geraki *et al.*, 2004) who also found elevated Fe in malignant tissues probably due to its role in angiogenesis and due to its catalytic role in Fenton reaction that resulted to increased requirement for DNA synthesis (Heath *et al.*, 2013) for tumor angiogenesis. The elevated levels of Zn in cancerous tissue samples compared to the corresponding urine for cell proliferation that supports tumor growth (Lee *et al.*, 2003) and cell division. Additionally, levels of Cu in cancerous samples were also evident due to increased generation of free radicals in Fenton reaction and/or by interference with DNA repair processes (Yaman *et al.*, 2005). In contrast to Fe and Cu, Zn concentrations, the other trace biometals (Mn, Cr and Se) showed alteration levels in cancerous and normal urine.

A novel ANN and SVM multivariate models have been developed method as shown in Figure 5.35 for determination of the concentration of trace biometals in cancerous and normal urine at different stages of cancer development.

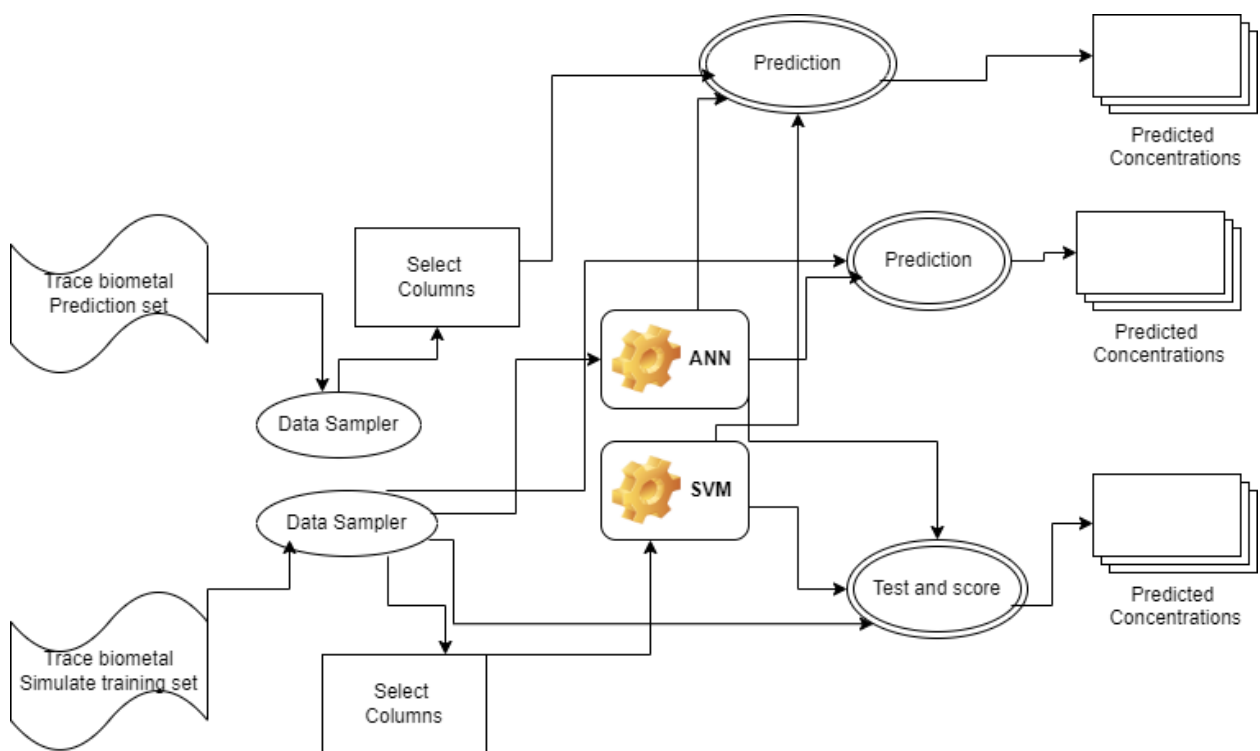


Figure 5.35 SVM and ANN multivariate calibration model

The developed model enabled determination of trace elements levels and their multivariate alterations in human body fluids (urine) that were correlated to the state of health. Concentrations of trace biometals further enabled determination of the various stages of

cancer development (Imam *et al.*, 2012); well differentiated (low grade), moderately differentiated (intermediate grade) and poorly differentiated (high grade).

5.8.4 Correlations of trace biometal levels

The correlations of trace biometal levels were evaluated by Pearson correlation as shown in Table 5.17 and Table 5.18 where latent vital relationships between the trace biometals and their role during normal and carcinogenic growth were determined.

Table 5.17 Correlation coefficients of trace biometals in normal urine

Element	Fe	Mn	Cr	Cu	Se	Zn
Fe	1					
Mn	0.0321	1				
Cr	0.2901	-0.5440	1			
Cu	-0.0024	0.6436	-0.2350	1		
Se	0.2532	0.6596	-0.1565	0.4107	1	
Zn	-0.2169	0.6148	-0.7153	0.5086	0.3832	1

The strong positive correlations between Mn and other trace elements in normal urine were evident, but a strong negative correlation was found between Cr and Zn (-0.7154). The observed strong positive Pearson correlation of Se and Mn can be attributed to the anti-carcinogenic role of Se in angiogenesis process in which new blood vessels are formed where Se tends to be protective against cancer.

Table 5.18 Correlation coefficients of trace biometals in prostate cancer urine

Element	Fe	Mn	Cr	Cu	Se	Zn
Fe	1					
Mn	0.1956	1				
Cr	-0.2459	-0.1658	1			
Cu	0.5915	0.5176	-0.4542	1		
Se	0.2219	-0.5505	-0.1029	0.0709	1	
Zn	0.3799	-0.2303	-0.2128	0.4880	0.5614	1

The positive correlations between Fe and Cu (0.5915), Cu and Mn (0.5176) were evident, but positive correlations between Fe and Zn (0.3799) and also Fe and Se (0.1029). This strong positive correlation of Fe with Zn and Se can be attributed to the exponential phase of cell division and proliferations resulting to metastasis. This is due to increased cellular activity (Raju *et al.*, 2006) and angiogenesis (Nasulewicz *et al.*, 2004).

5.9 Multivariate regression models for trace biometals in tissue simulates

The ANN architecture (3:3:5) corresponding to number of neurons in the input, hidden and output layers, respectively was chosen based on low MSE attained for both training and validation sets. The training and testing of the networks were achieved using the calibration and test samples for validating the trained network with the model performance as shown in Table 5.19.

Table 5.19 ANN and SVM tissue simulate model performance

Trace Element		Tissue calibration model performance			
		MSE	RMSE	MAE	R ²
Cu	ANN	3.77	1.77	1.38	0.98
	SVM	1963.35	44.31	34.06	0.82
Zn	ANN	150.24	12.26	8.69	0.96
	SVM	3471.70	58.92	42.96	0.49
Fe	ANN	2155.19	46.42	28.88	0.96
	SVM	26218.21	161.92	100.95	0.52
Mn	ANN	2.68	1.64	1.34	0.97
	SVM	1076.17	32.81	13.75	0.42
Se	ANN	37.65	6.14	5.25	0.72
	SVM	52.27	7.23	4.88	0.61
Cr	ANN	2.37	1.54	1.18	0.99
	SVM	396.96	19.92	12.19	0.22

The training resulted to improved estimation ability as shown in Table 5.20 above with lower MSE attained after 2200 and 2000 iterations for the architecture used in ANN and SVM respectively. The predicted against the measured regression plots in Figure 5.36 to Figure 5.41, show how the ANN model more accurately predicted the concentrations of Cu, Zn, Mn, Cr and Se in simulate tissues. The predicted concentrations show how the ANN model has a potential to be applied to real human tissue samples of similar matrix composition and its future utility in prediction of trace biometals in biomedical samples for cancer diagnosis.

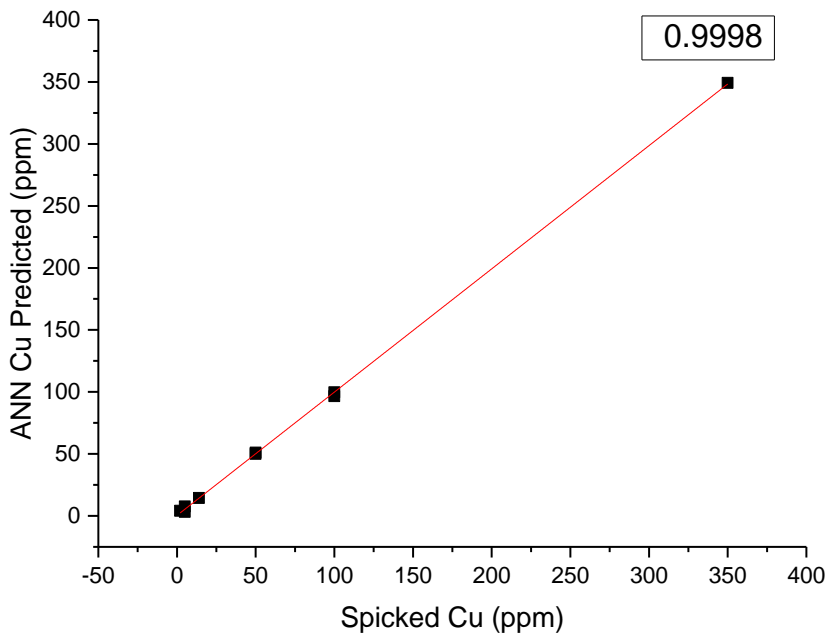
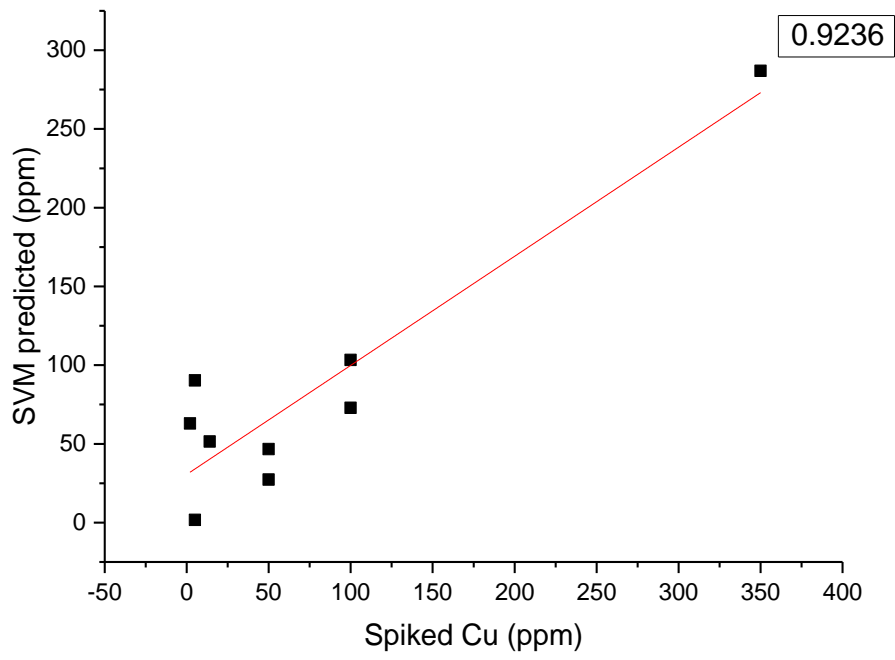


Figure 5.36 SVM and ANN regression plots of Cu

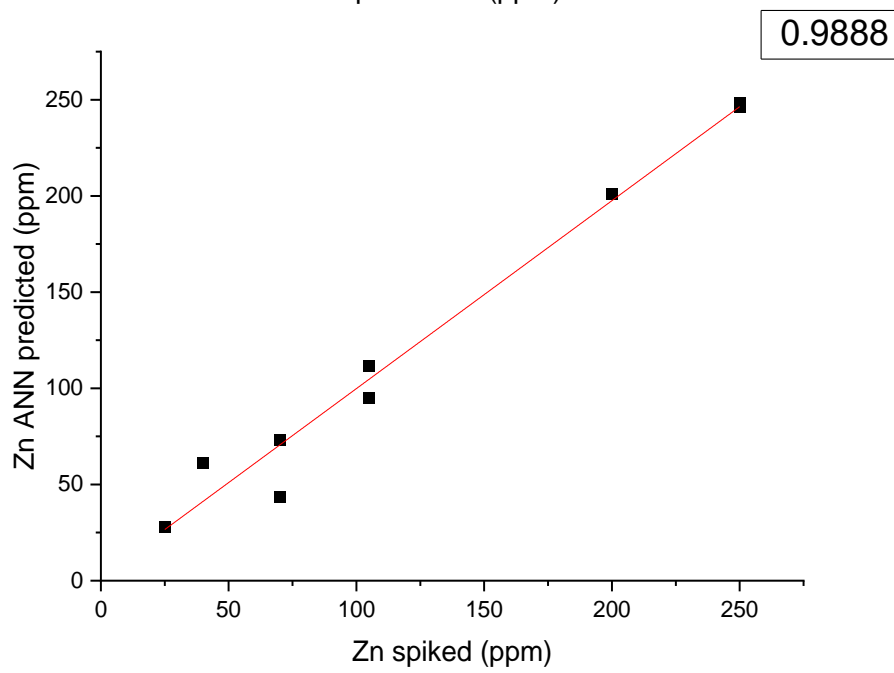
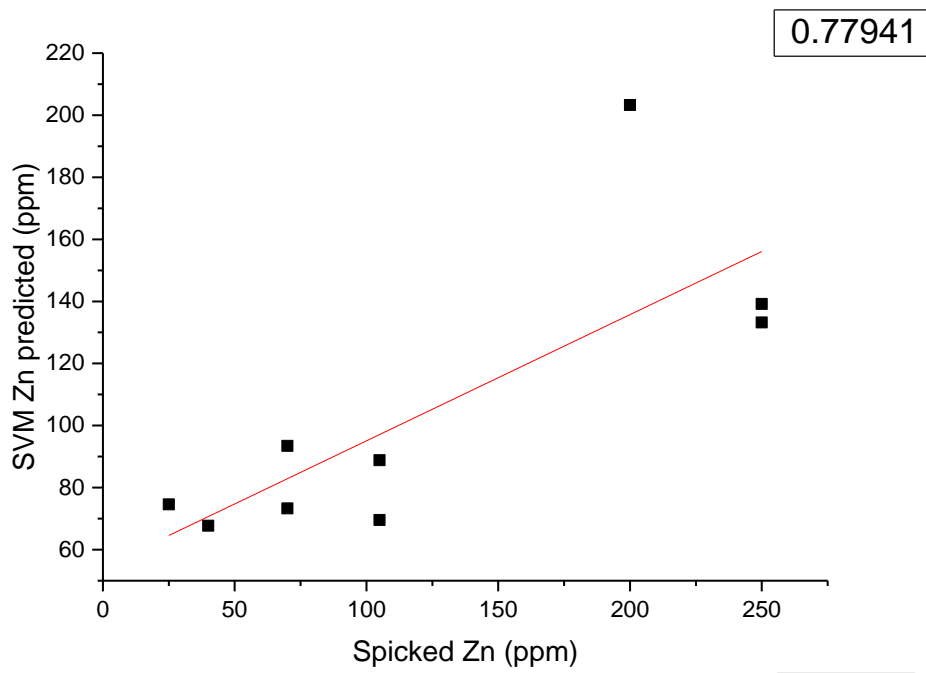


Figure 5.37 SVM and ANN regression plots of Zn

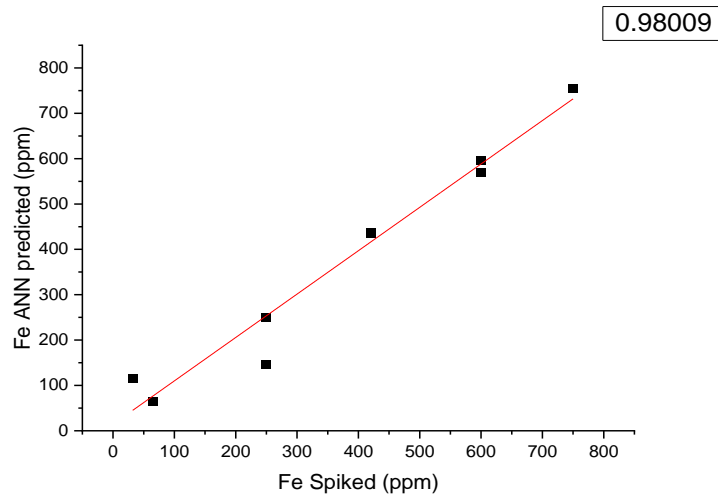
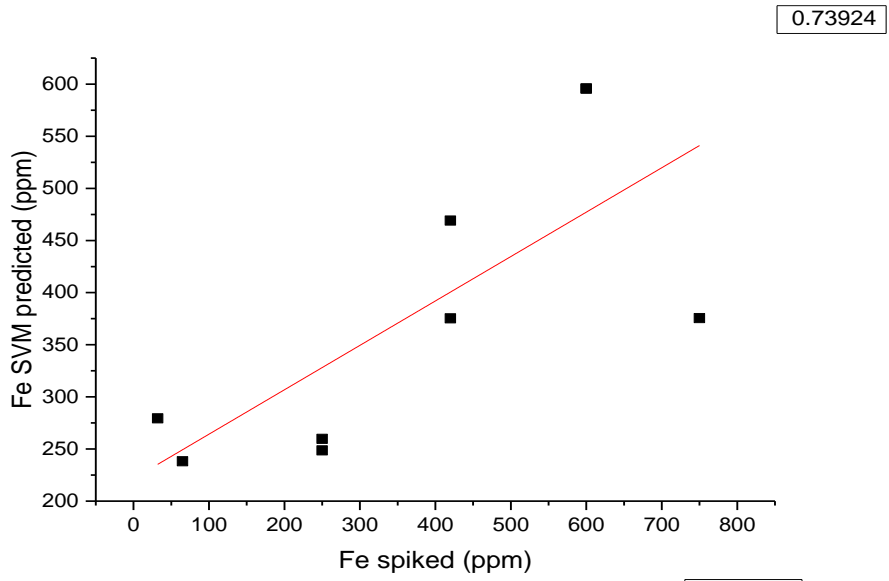


Figure 5.38 SVM and ANN regression plots of Fe

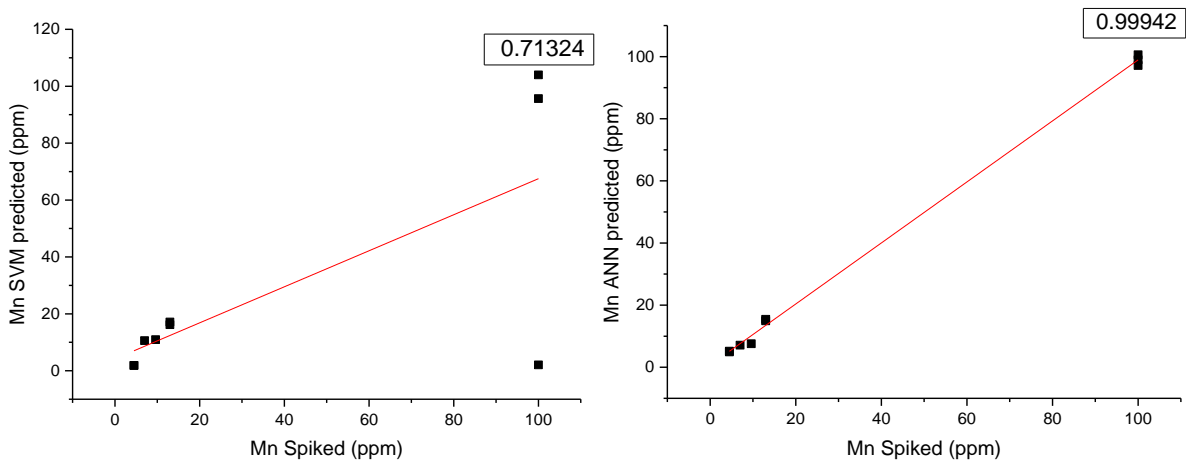


Figure 5.39 SVM and ANN regression plots of Mn

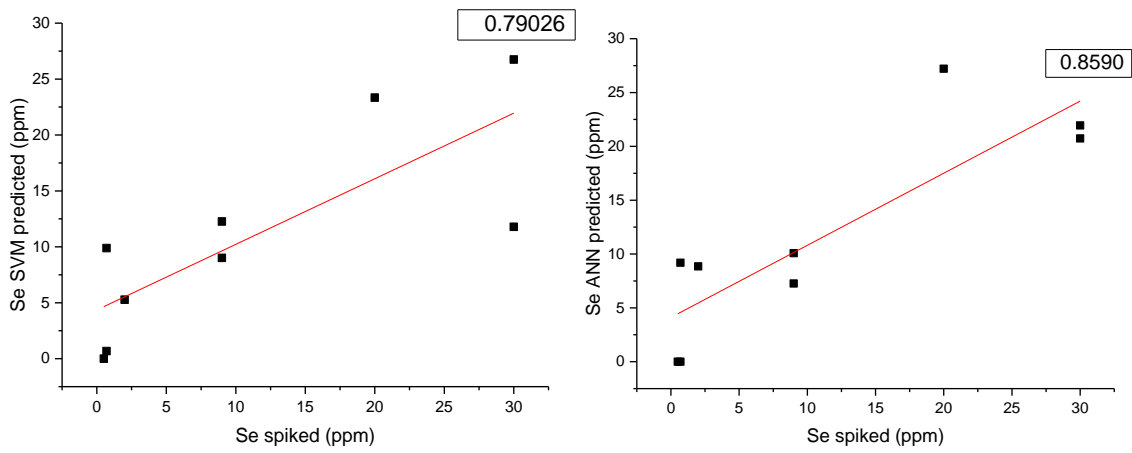


Figure 5.40 SVM and ANN regression plots of Se

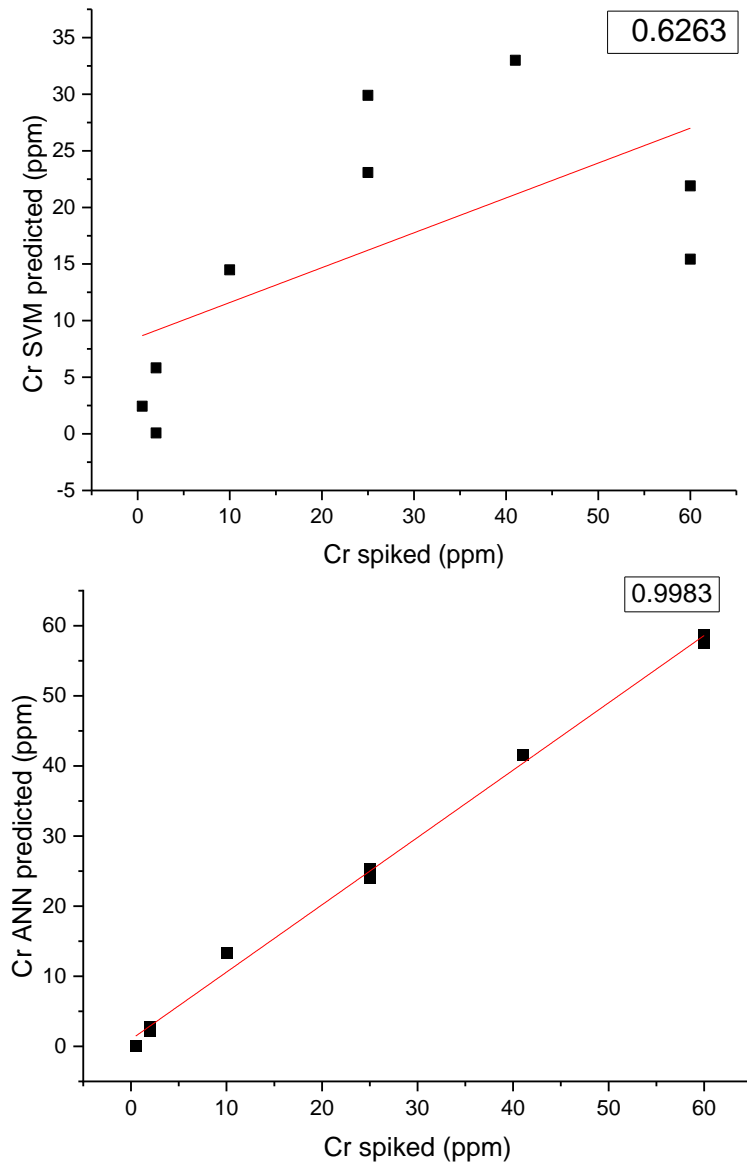


Figure 5.41 SVM and ANN regression plots of Cr.

The low R^2 values can be attributed to high background and the low spiked levels in simulate samples. These results indicate that the ANN models predicted accurate concentrations and can therefore be applied to analyze the concentrations of trace biometals in biomedical tissue samples.

5.9.1 Detection and quantification limits of trace biometals in simulate tissue samples

The ANN multivariate calibration model was used to determine the detection limits (DL) and quantification limits of trace biometals in simulate tissue samples as shown in Table 5.20.

Table 5.20 Detection limits (ppm) of trace elements in simulate tissue

Element	Atomic number	K_{α} Energy (KeV)	Detection limit (ppm)	Quantitative limit (ppm)
Cr	24	5.41	2.82	4.16
Mn	25	5.89	3.15	4.67
Fe	26	6.40	3.33	6.21
Cu	29	8.05	2.05	6.72
Zn	30	8.64	3.18	5.71
Se	34	11.22	1.94	2.99

The levels of trace biometals below the detection and quantitative limits in Table 5.20 were hardly achieved by the developed multivariate model towards cancer diagnostics in human tissue.

5.9.2 Predicted concentration in tissue biopsy samples

The SVM and ANNs validated models enabled determination of the trace biometals levels in human tissues. Table 5.21 and Table 5.22 show the predicted elemental concentration mapping between the inputs and target output in normal tissue biopsies based on the mean-centered and smoothed data for fluorescence and scatter regions.

Table 5.21 SVM prediction of trace biometal concentrations in normal tissues

	SVM Trace biometal concentrations (ppm)					
	Cr	Mn	Fe	Cu	Zn	Se
NP_07T	14.58±2.11	4.74±0.88	101.13±9.09	39.44±4.65	84.08±10.22	17.28±2.44
NP_20T	13.31±1.34	14.65±1.34	220.01±23.98	14.68±1.77	80.92±11.34	23.07±3.23
NP_06T	33.95±3.56	1.44±0.23	70.22±6.98	27.16±3.54	92.58±6.45	17.55±1.34
NP_23T	17.64±2.12	1.35±0.02	83.37±3.97	27.19±5.95	226.08±50.56	17.43±2.34
NP_12T	17.92±3.23	11.02±1.02	121.05±13.87	31.91±3.21	100.07±10.34	17.60±2.22
NP_19T	15.98±2.11	3.24±1.87	104.19±12.99	39.61±4.34	87.37±7.36	17.70±5.23
NP_14T	1.40±0.33	3.51±0.89	162.22±20.54	15.32±2.33	82.20±4.65	8.07±2.34
NP_22T	2.38±1.02	3.45±0.99	301.72±56.76	14.61±2.17	80.22±8.22	5.34±1.23

Table 5.22 ANN predicted Prediction of trace biometal concentrations in normal tissues

	ANN Trace biometal concentrations (ppm)					
	Cr	Mn	Fe	Cu	Zn	Se
NP_07T	3.20±0.33	4.54±0.45	428.09±57.02	18.34±3.21	60.46±5.55	3.53±1.34
NP_20T	16.11±2.54	31.99±5.43	169.69±8.96	41.46±8.22	60.59±3.56	5.40±1.11
NP_06T	13.75±1.34	4.64±2.11	327.59±89.09	18.3±4.98	60.88±8.44	22.69±6.43
NP_23T	10.31±2.67	4.54±1.43	184.67±10.45	18.34±3.56	243.38±17.32	21.82±2.87
NP_12T	10.92±3.54	4.35±2.88	494.34±79.23	22.67±1.33	83.88±8.34	4.22±1.45
NP_19T	15.29±1.22	4.54±1.87	781.78±97.93	22.15±2.87	65.41±9.43	22.06±3.23
NP_14T	12.28±3.18	4.58±1.32	516.81±67.27	34.33±7.89	60.66±5.23	3.59±0.27
NP_22T	10.16±1.37	4.84±1.76	772.56±102.65	38.99±6.75	60.35±7.21	3.72±0.99

Tables 5.23 shows and ANN predicted elemental concentration in cancerous tissue biopsies based on the mean-centered and smoothened data for fluorescence and scatter regions.

Table 5.23 ANN predicted Prediction of trace biometal concentrations in cancerous tissues

	ANN Trace biometal concentrations (ppm)					
	Cr	Mn	Fe	Cu	Zn	Se
P_30T	6.27±1.23	7.12±0.87	784.99±87.62	23.22±1.33	60.01±1.87	7.66±1.23
S1	12.28±1.66	7.07±0.91	1393.71±105.44	29.37±2.45	228.48±29.53	7.99±2.34
S5	12.27±3.56	7.26±1.25	1558.92±108.26	23.12±1.59	76.12±2.883	4.73±0.87
S14	12.27±2.87	23.97±3.65	834.74±67.21	19.72±2.33	110.51±7.71	22.66±3.59
S13	12.27±0.33	7.17±1.57	525.71±20.65	20.38±1.03	89.93±9.28	15.37±2.18
S11	10.89±1.22	7.71±1.02	491.86±36.63	31.73±2.19	72.47±2.52	4.92±0.24
S8	9.41±1.98	43.54±7.91	1183.00±96.57	18.34±0.93	199.81±20.69	4.78±1.73
S9	9.71±1.11	7.03±1.27	4661.32±229.54	23.056±2.65	209.89±43.82	22.94±2.61
S2	25.06±4.56	35.77±5.10	921.91±59.32	45.71±7.32	59.67±3.64	4.91±0.67
S15	18.53±3.59	6.92±1.93	667.36±33.28	18.34±2.11	128.03±18.32	12.35±4.29
S4	7.12±1.23	7.23±0.98	494.18±15.34	25.14±5.30	61.93±3.27	4.73±2.35
S7	11.66±1.45	7.24±1.05	1061.97±78.32	25.96±3.23	228.48±26.40	18.68±1.56
S10	12.28±2.69	7.23±0.84	497.88±56.12	32.75±1.65	102.41±14.28	5.26±0.27
S12	12.27±1.89	7.14±1.63	623.57±13.91	18.36±2.87	228.48±28.23	5.33±0.99
S3	10.79±2.14	8.87±1.11	351.95±15.22	18.54±4.26	59.67±2.45	4.86±1.49
S6	12.28±2.54	6.89±1.82	927.08±82.32	18.34±2.67	99.04±26.34	4.82±0.35

The ANN non-linear computational model was therefore reliable for the determination of trace biometals' levels in extremely complex matrices of biomedical samples. The mean concentrations of trace biometals in normal and cancerous tissues are as shown in Table 5.24 and Table 5.25;

Table 5.24 ANN descriptive values of trace biometal concentrations in normal tissues

	Trace biometal concentrations (ppm)			
	N	Mean	Min	Max
Cr	8	11.50±4.03	3.20	16.11
Mn	8	7.97±9.71	4.54	31.99
Fe	8	459.44±234.29	169.69	781.78
Cu	8	26.83±9.81	18.34	41.46
Zn	8	86.95±63.72	60.35	243.38
Se	8	10.8±9.39	3.53	22.69

Table 5.25 ANN descriptive values of trace biometal concentrations in cancerous tissues

Trace biometal concentrations (ppm)				
	N	Mean	Min	Max
Cr	16	12.21±4.36	6.27	25.06
Mn	16	12.38±11.52	6.89	43.54
Fe	16	1061.26±19.75	351.95	4661.32
Cu	16	24.50±7.44	18.34	45.71
Zn	16	125.93±68.04	59.67	228.48
Se	16	9.50±6.71	4.73	22.94

The results above show varying levels of trace biometals in the cancerous tissues in prostate tissue with cancer as compared to normal prostatic tissue. The significant alteration in concentration of trace biometals (Cr, Mn, Fe, Cu and Zn) indicate the presence of neoplasies in cancer cells thus a possible helpful diagnostic tool (Silva *et al.*, 2006). In this study, the levels of Fe support the hypothesis that there is a connection with both increased cellular activity and blood supply in the formation (Poletti *et al.*, 2004). The increase of trace elements levels in cancer samples compared to healthy samples can be justified by the increase of the cellular metabolism in cancer cells (Poletti *et al.*, 2004).

5.9.3 Correlation of trace biometals in tissue biopsy samples

The correlations between trace elements as shown in Table 5.26 and Table 5.27 indicates interactions between trace biometals in both normal and cancerous tissues.

Table 5.26 Trace biometal correlation coefficients in healthy tissues

Element	Cr	Mn	Fe	Cu	Zn	Se
Cr	1					
Mn	0.4619	1				
Fe	-0.0329	-0.4997	1			
Cu	0.3722	0.6026	0.1282	1		
Zn	-0.1176	-0.1672	-0.4590	-0.3827	1	
Se	0.3755	-0.2357	-0.1322	-0.5853	0.4548	1

Table 5.27 Trace biometal correlation coefficients in cancerous tissues

Element	Cr	Mn	Fe	Cu	Zn	Se
Cr	1					
Mn	0.3485	1				
Fe	-0.1112	-0.0299	1			
Cu	0.5280	0.2332	-0.0289	1		
Zn	-0.0960	0.0318	0.4379	-0.2191	1	
Se	-0.0424	-0.0805	0.5095	-0.2223	0.3617	1

The positive correlation between Fe and Zn was established in cancerous samples as compared to non-cancerous tissue biopsy samples. The trace biometals (Fe, Cu, Zn, Mn and Se) aid in maintenance of physiological homeostasis in organism proteins that are involved in cellular processes (Gruber *et al.*, 2020). The elevated correlation of Fe and Cu in cancerous tissues are consistent with previous studies (Carvalho *et al.*, 2007) as a result of increased cellular activity in cancerous tissues.

5.10 Synchrotron XRF quantitative analysis of trace biometals in cell cultures

Synchrotron radiation based quantitative analysis of the trace elements levels was determined using calibrated known simulate sample standards. The XRF fluorescence profiles were used to determine the concentration of Cr, Mn, Fe, Cu and Zn in normal cell cultures as shown in Table 5.28.

Table 5.28 Concentration of trace biometals in healthy normal cell cultures

Element	V_D1	V_D2	V_D3	V_D4	V_D5	V_D6
Cr	ND	12.2	8	ND	9.3	2.8
Mn	8.4	22.3	16.8	8.5	21.2	9.4
Fe	19.3	33.4	28.3	31.3	28.9	17.6
Cu	101.2	147.1	141.1	126.7	139.1	124.9
Zn	629.9	999.9	669.9	559.9	699.9	489.9

From these results, the concentration of Cu and Zn were elevated in normal cell cultures due to their normal physiological functions. The concentrations of trace biometals in cancerous cell cultures are as shown in Table 5.29 which serve as a vital biomarker for early cancer diagnosis.

Table 5.29 Concentration of trace biometals in cancerous cell cultures

Element	DU_D1	DU_D2	DU_D3	DU_D4	DU_D5	DU_D6
Cr	ND	10.2	66.5	9.1	25	22.7
Mn	2.9	20.5	78.3	16.3	29.6	28.1
Fe	26.1	43.4	77.2	43.3	41.2	38.6
Cu	121.4	143.8	346.9	139	170.1	166.9
Zn	729.9	869.9	659.9	929.9	1369.9	1139.9

Higher concentrations of trace biometals (Fe and Cu) were evident in cancerous cell cultures as compared to healthy cell cultures. The trace biometals play a significant role in protein synthesis, immune function, antioxidant defence and inhibition of cell proliferation (Silvera and Rohan 2007).

Concentration of trace biometals and their alterations at each stage of development of both normal and cancer cell cultures are as shown in Figure 5.42 to Figure 5.45.

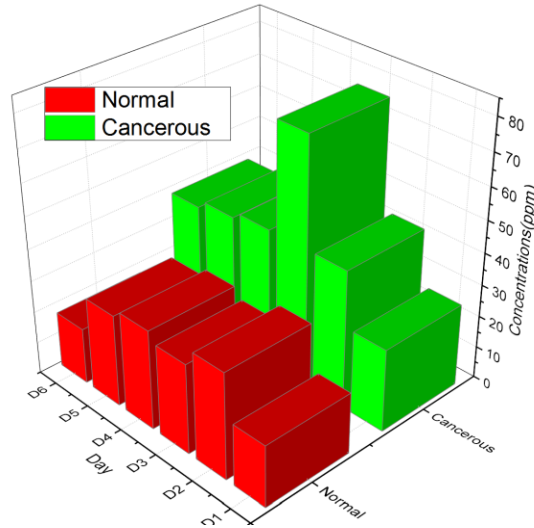


Figure 5.42 Bar graphs of Fe concentration in cancerous and normal cell cultures

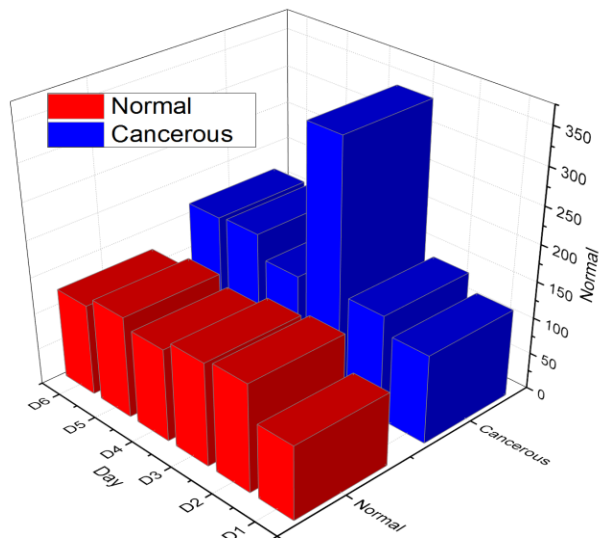


Figure 5.43 Bar graphs of Cu concentration in cancerous and normal cell cultures

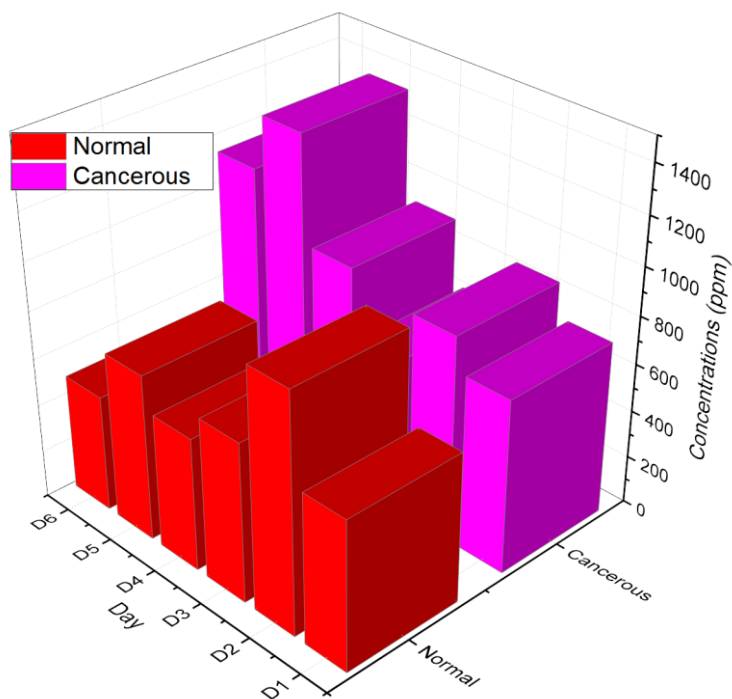


Figure 5.44 Bar graphs of Zn concentration in cancerous and normal cell cultures

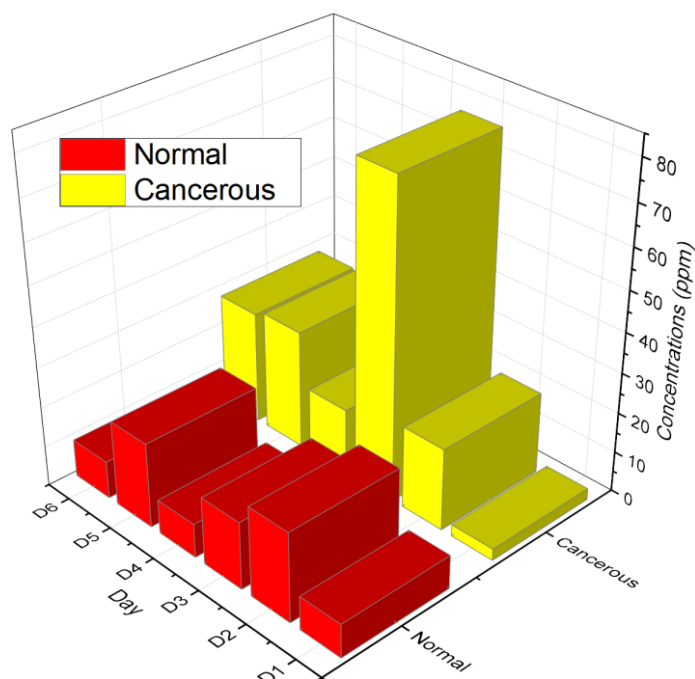


Figure 5.45 Bar graphs of Mn concentration in cancerous and normal cell cultures

The increasing levels of Fe with the stage of development was evident in cancerous cell cultures as compared to the normal cell cultures. In this study, high levels of Fe was implicated in its regulatory role for angiogenesis (Geraki *et al.*, 2004) and thus essential in

biochemical processes related proteins (ferroprotein) and hepcidin for carcinogenesis in cancer cells (Naidu *et al.*, 2019). The elevated levels of Zn is essential in biological samples enables amalgamation of RNA and DNA that increased apoptosis in malignant cells and inhibited cell growth (Naidu *et al.*, 2019). The implication of this apparent change between the concentration of trace biometals in cancerous and normal healthy cell cultures is due to increased cellular and enzymatic activities which alters the trace elemental composition (Geraki *et al.*, 2004).

5.11 Correlation of trace biometals

The impact cancer is not necessarily given by total element content alone, alteration in spatial distribution together with chemical speciation plays key role in carcinogenesis. Determination of spatial correlation between trace biometals revealed significant degree of correlation. Non-parametric Spearman's rank correlation test (Bárány *et al.*, 2002) was used to investigate the relationship between concentrations of trace biometals.

5.11.1 Correlations of trace biometals distributions

Correlations of morphologic and biochemical differences in conjunction with assessments based on chemically-specific spectroscopic techniques can aid in cancer diagnostics at early stages of development. Cancer alters biochemical levels in human cells where the correlations of trace biometals were linked to the state of health. Besides the concentrations of trace biometals, a correlation existed between trace biometals and carcinogenesis. Spearman correlation were determined for the spatial correlation between these pairs of trace elements ($p < 0.001$) to ascertain the importance of these elements in metabolic processes associated with tumor development. The distribution of trace biometals at cellular level in both cancerous and normal cell cultures were correlated where positive correlations was probably due to increase of a trace biometal as a consequence of increased concentration of another element due to exposure from the same metabolic interactions such as binding to the same proteins (Bárány *et al.*, 2002). Figure 5.46 compares the intensities of trace biometals in cell lines under study where strong correlation with Fe and Cu ($r=0.84$) indicated dominant presence in cancerous cell cultures, while Se had no correlation with other elements, although it is among the essential elements typical for the cells.

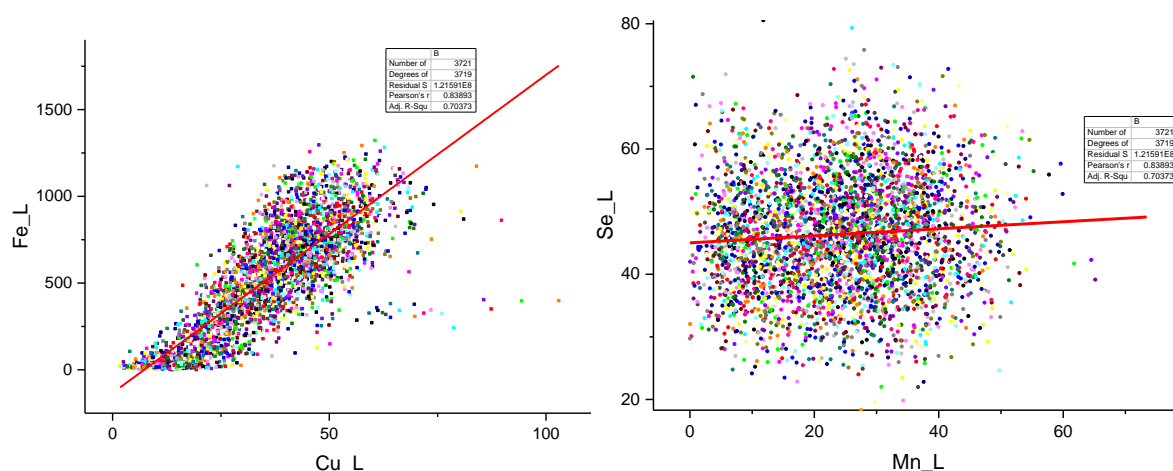


Figure 5.46 The correlation of Cu and Fe intensities in DU_D1 culture sample

A direct correlation between Fe-Cu and Mn-Se were determined as shown in Figure 5.47 where the colocalization of Fe and Cu was evident. The weak spectral signals from the XRF concentrations of trace element and their distribution in cell cultures were evident.

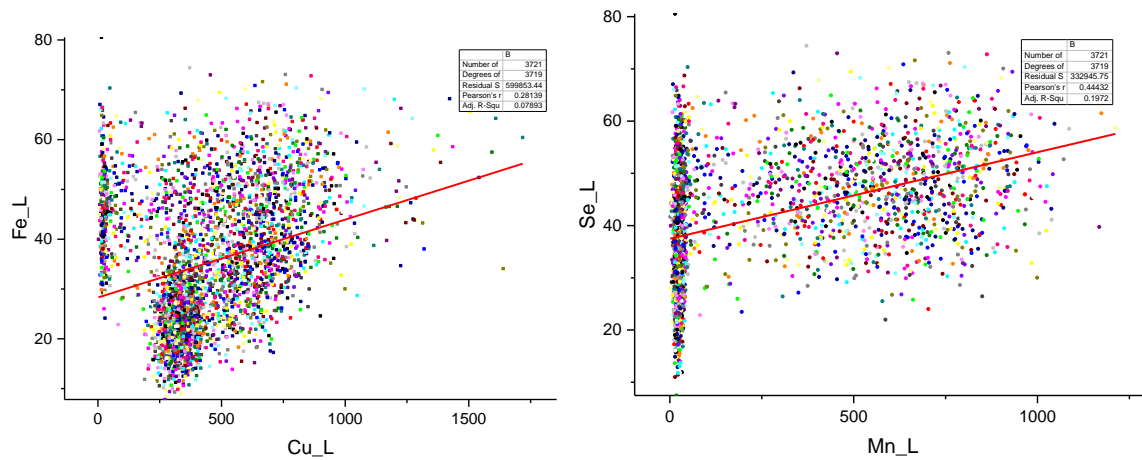


Figure 5.47 The correlation of Cu and Fe intensities in V_D1 culture sample

The correlation between trace biometal levels revealed significant diagnostic information for resolving weak spectral profiles in biomedical samples. The positive Pearson correlation coefficients could have been due to exposure from the same metabolic interactions (Bárány *et al.*, 2002). As shown in Table 5.30, Cu is strongly positively correlated to Fe (0.9225) in cancerous cell culture at advanced stage of development than at early stage (0.6045) and the trend is opposite to the corresponding normal cell culture.

Table 5.30 Correlations of trace biometals in cancerous cell cultures

Correlation coefficients					Correlation coefficients				
DU_D1	Mn	Fe	Cu	Se	DU_D2	Mn	Fe	Cu	Se
Mn	1.0000				Mn	1.0000			
Fe	0.9993	1.0000			Fe	0.6703	1.0000		
Cu	0.6068	0.6045	1.0000		Cu	0.8314	0.8389	1.0000	
Se	0.8445	0.8423	0.4723	1.0000	Se	0.0733	0.0997	0.0789	1.0000

Correlation coefficients					Correlation coefficients				
DU_D3	Mn	Fe	Cu	Se	DU_D4	Mn	Fe	Cu	Se
Mn	1				Mn	1			
Fe	0.7521	1			Fe	0.6295	1		
Cu	0.8236	0.9406	1		Cu	0.7509	0.9225	1	
Se	0.4388	0.5282	0.5858	1	Se	0.2931	0.4145	0.4527	1

The trace biometals summary statistics in Tables 5.31 and Table 5.32 show the correlations between trace elements expression in normal cell cultures.

Table 5.31 Correlations of trace biometals in healthy cell cultures

Correlation coefficients					Correlation coefficients				
V_D1	Mn	Fe	Cu	Se	V_D2	Mn	Fe	Cu	Se
Mn	1				Mn	1			
Fe	0.9997	1			Fe	0.9978	1		
Cu	0.7939	0.7931	1		Cu	0.5267	0.5215	1	
Se	0.9739	0.9734	0.7602	1	Se	0.8994	0.8948	0.4383	1

Table 5.32 Correlation coefficients of biomarkers in cancerous cultured cells

Correlation coefficients					Correlation coefficients				
V_D3	Mn	Fe	Cu	Se	V_D4	Mn	Fe	Cu	Se
Mn	1				Mn	1			
Fe	0.8317	1			Fe	0.9125	1		
Cu	0.8897	0.7606	1		Cu	0.2242	0.2030	1	
Se	0.6624	0.7165	0.7390	1	Se	0.7254	0.6793	0.1432	1

Based on these results, the correlation of trace biomarker concentration and distributions were established with stage of cancer development. Correlative localised 2D distribution of Fe and Cu increased with cancer staging Cu was relatively highly distributed in cancerous

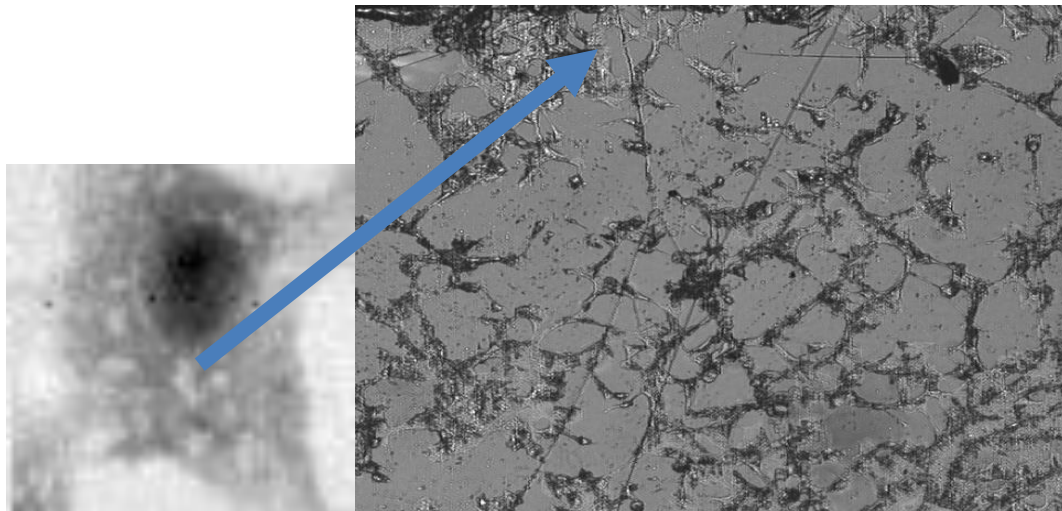
cultures than Fe. The correlations of trace elemental distributions indicate elevated concentrations in tumor cells of all samples, with the exception of Fe. Strong and weak positive correlations were evident in cancerous and normal cell cultures which indicate possible connection between these trace elements and carcinogenesis. Further, Mn as a trace biometal was strongly linearly correlated (Zabłocka-Słowińska *et al.*, 2018) to all other trace biometals in normal cell cultures thus contains important for morphological information of the cell structure.

The relatively strong correlation between Cu and Fe (0.9406) and (0.92252) in DU_D3 and DU_D4 respectively in cancerous compared to corresponding stages in normal cell cultures suggest their significance in carcinogenesis. In a similar study by micro XRF (Silva *et al.*, 2012), spearman correlation indicated a positive correlation of Cu with other elements over all area sample with statistical significance ($p < 0.001$) which suggest the association Cu to the tumor development. Cu aids in the onset and development of tumors while Fe plays a vital function in oxygen metabolism (Rocha *et al.*, 2019). A strong correlation between Fe and Cu indicated their vital role in cancer cells. Fe is known to aid in blood supply to the malignant cells thus accelerates angiogenesis (Kwok and Richardson 2002). Further, a lower correlation between Se and Mn was apparent probably due to their anti-carcinogenesis (Guo *et al.*, 2015). The strong correlation with Cu and Fe indicated their predominant presence in the cells, while Se had a weak correlation with Mn in cancerous cell than in normal cell culture probably due to the anti-carcinogenic nature of Se.

A weak correlation between Se and other trace biometals in cancerous cell probably indicated the anti-carcinogenic nature of Se due to its presence in the glutathione peroxidases that protect DNA from damage by oxygen radicals (Guo *et al.*, 2015, Xue *et al.*, 2021). Mn and Se had weak correlation with other trace biometals although they are essential elements typical for the normal cells. The strong correlation of Cu and Fe in cancerous cells indicated vital roles of these biometals in cancer development. The correlation between Mn and Se in cancerous and normal cell cultures are similar, whereas the correlation between Fe and Cu in normal cell cultures were different thus Fe and Cu had significant role in carcinogenesis.

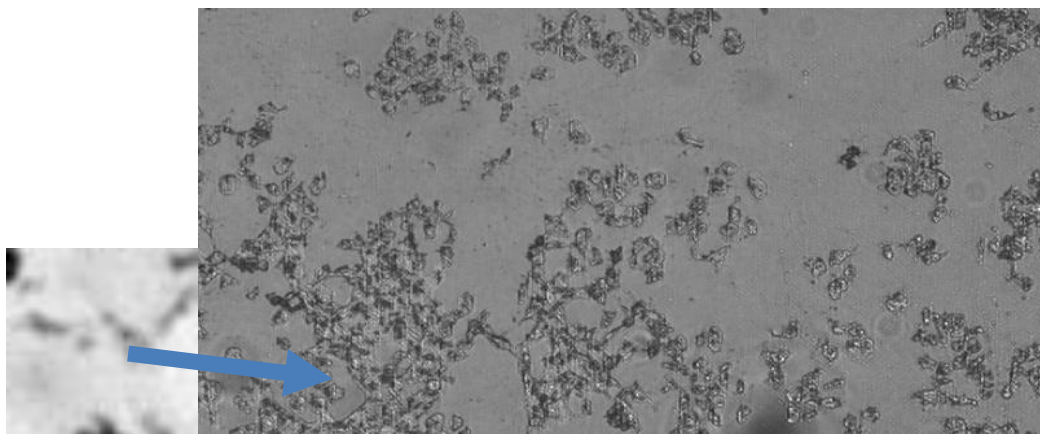
5.12 Imaging absorption of trace biometals in cell cultures

The typical illuminated pixel areas in Figure 5.48 and Figure 5.49 show where the strong self-absorption and differential phase contrasting imaging attributed to the dark matrix of light elements (H, C, O and N). Pixels of the multispectral images were used to generate spectral signatures utilizing multivariate machine learning techniques to identify cancerous and non-cancerous cell cultures.



36x36um

Figure 5.48 Scan region for V_D3 healthy cell culture



40x40um

Figure 5.49 Scan region for DU_D1 cancer cell culture

The obtained value at each pixel on image was proportional to the counts corresponding to the elemental x-ray lines at that point and normalized to the ring current. The cells were visible with strong self-absorption and differential phase contrast (Gianoncelli *et al.*, 2021).

5.12.1 Micro XRF 2D mapping of trace biometals in cell cultures

Direct rapid and non-destructive analysis enabled 2D mapping of biometals (Mn, Fe, Cu and Se) at high resolution for accurate diagnosis and characterization of cancer at an early stage of development. Micro-XRF (μ -XRF) determined microscopic localization of trace biometals in organic matrix within cell culture microenvironments. The X-ray photon background was subtracted and the trace elements fitted by PyMca for acquisition of RGB distribution maps.

Figure 5.50 shows high resolution trace biometal 2D distribution images of Mn, Fe, Cu and Se obtained in the control cell culture with 40 x 40 pixels at 10 μ m resolution.

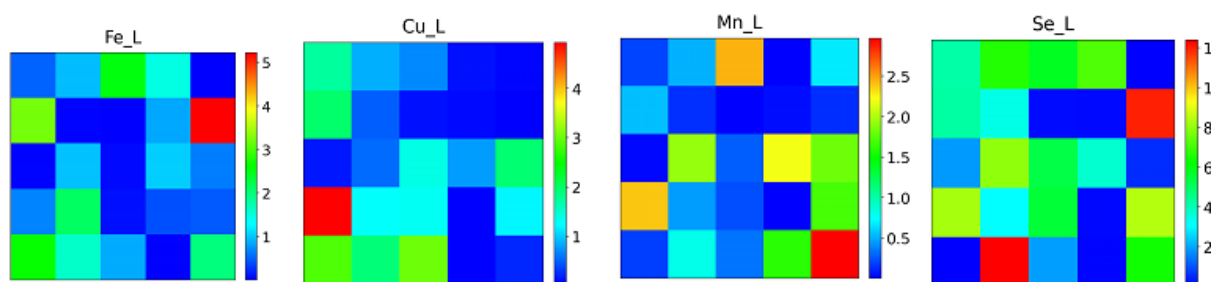


Figure 5.50 Micro XRF maps of Mn, Fe, Cu and Se in the growth media

The color bars indicated trace biometal contents for semi-quantitative determination in the irradiated area. The sparse distribution of the trace biometals were evident probably due to the fact that the media had not been seeded with cells.

The differences in 2D density distribution of trace biometals (Mn, Fe, Cu and Se) in cancerous cell cultures at different stages of development are as shown in Figure 5.51. The trace biometals (Mn, Fe, Cu and Se) distributions were detected in cancerous cell cultures where increased and co-localized distribution of Fe and Cu were observed at similar regions of the cell at various stages of development.

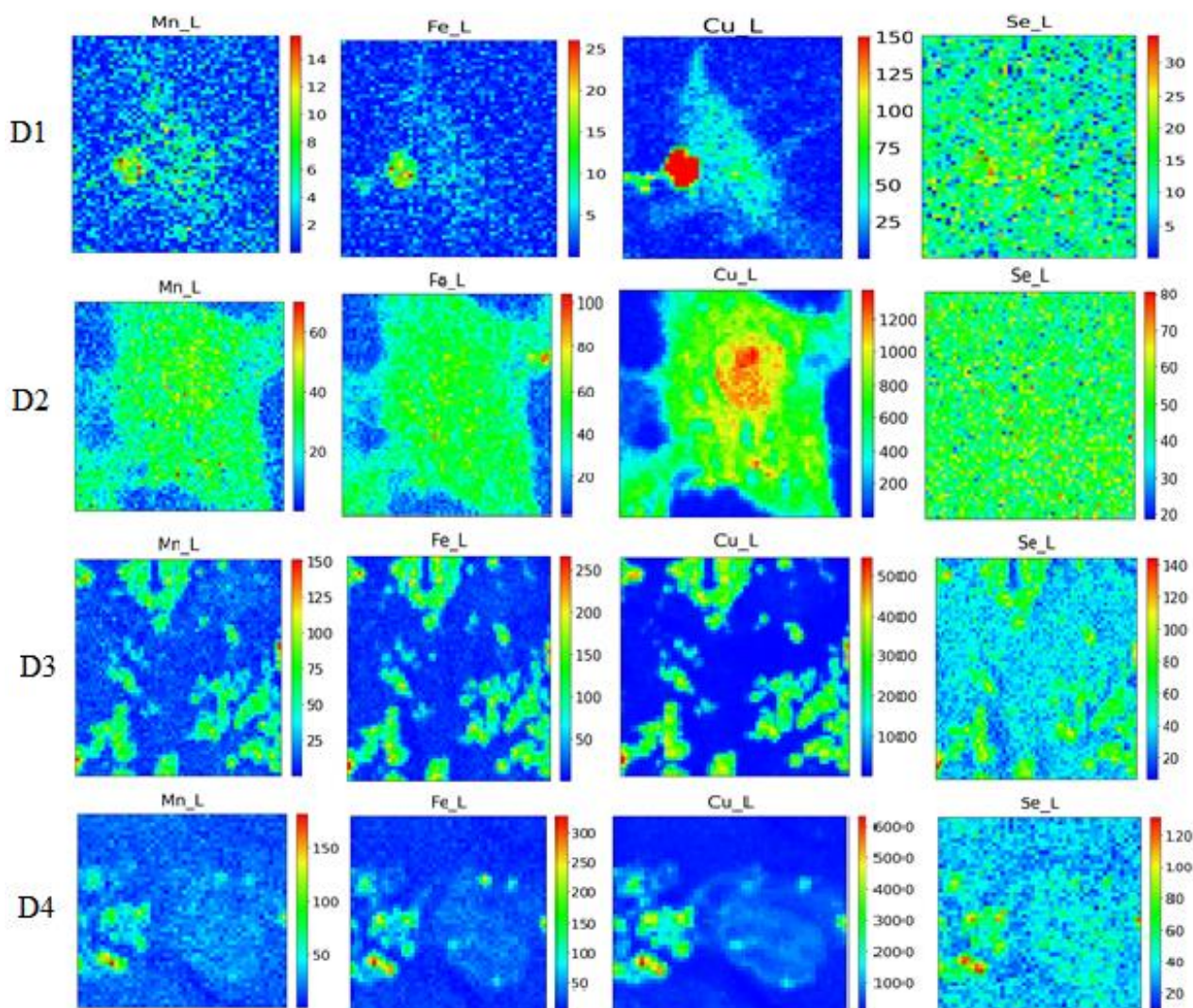


Figure 5.51 Micro XRF 2D maps of Mn, Fe, Cu and Se in DU145 cancer cell cultures

The 2D maps further indicated the spatial trace biometal (Mn, Fe, Cu and Se) distributions that showed elemental accumulation in cancerous cell cultures. The low inhomogeneous distribution of Se was attributed to the anti-carcinogenic nature of Se in reactive oxygen species for antioxidant enzymes (glutathione peroxidase) that inhibits the synthesis of tumour protein cells (Xue *et al.*, 2021). Further, high distribution of Cu was detected and localized in the cell as shown in DU_D2 which increased at different stages of cancer. The trace biometals (Fe and Cu) accumulated preferentially in cell nucleus of the cancer cell cultures probably due to their vital role in oxidative processes for DNA damage. Further, distinct distribution of Fe and Cu were observed in cancer cells with homogeneous distribution of Se compared to a non-homogeneous distribution of Mn.

The trace biometals (Fe, Cu, Mn and Se) distribution maps were also obtained in Vero monolayer non-cancerous cell cultures as shown in Figure 5.50.

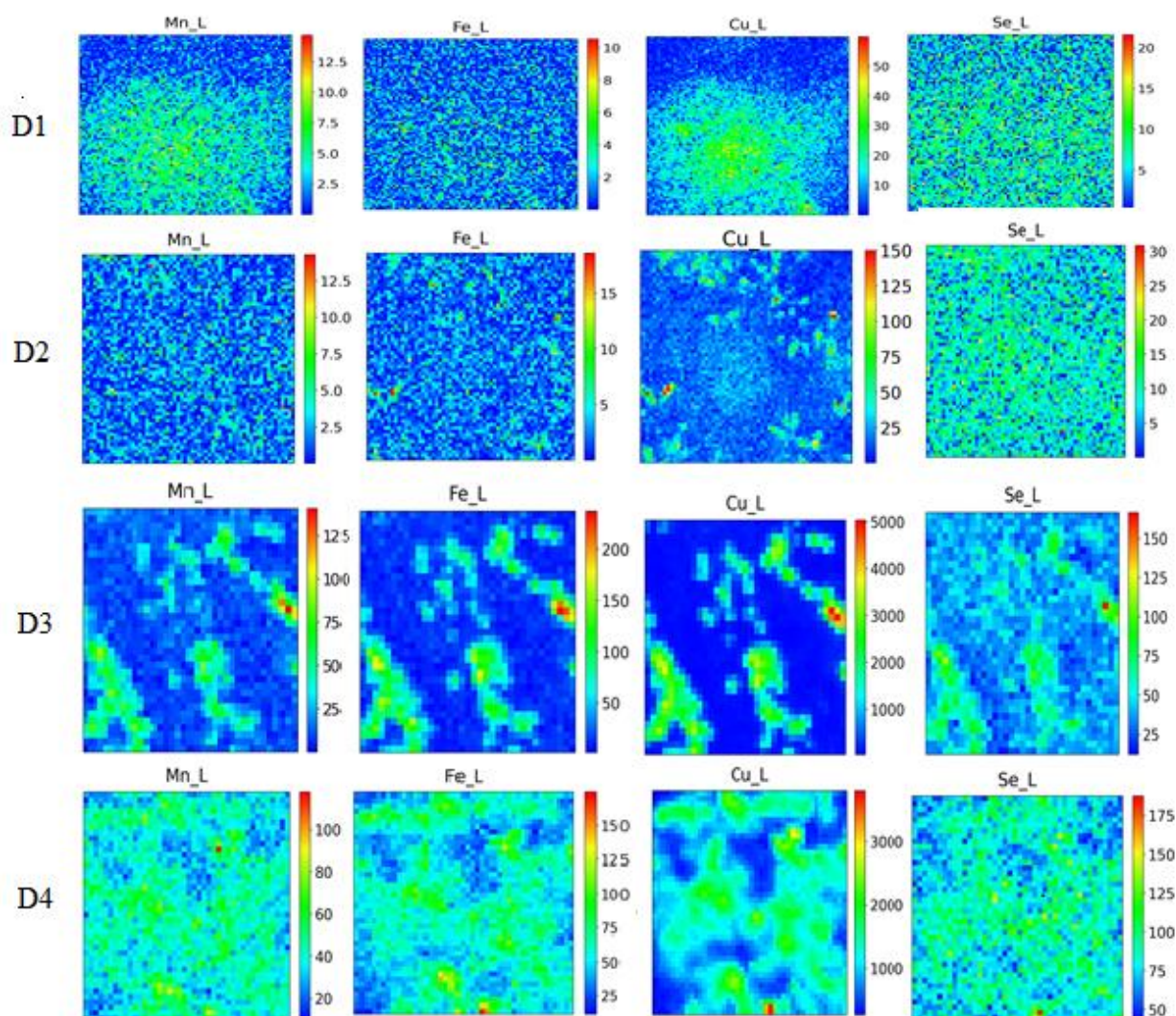
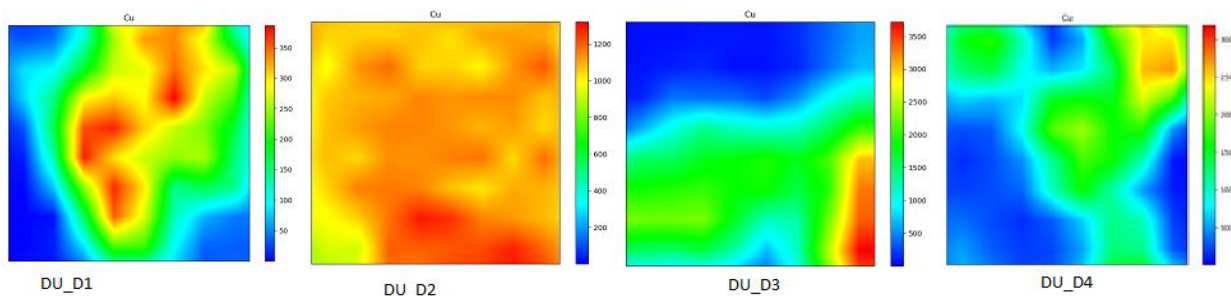


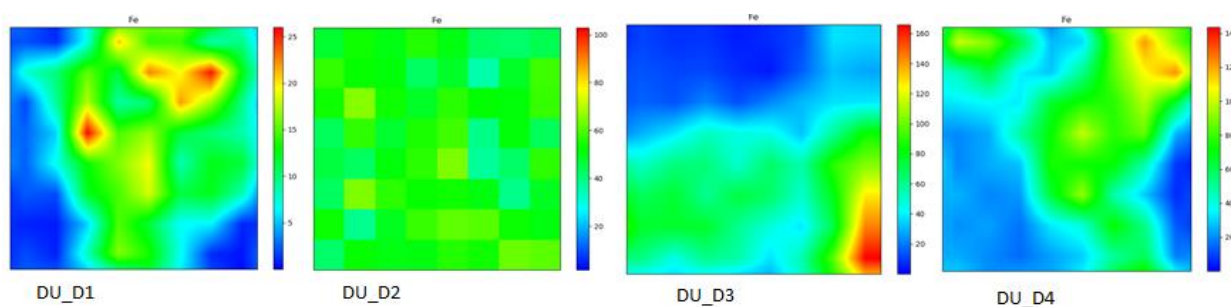
Figure 5.52 Micro XRF 2D maps of Mn, Fe, Cu and Se in Vero healthy cell cultures

The sparse spatial distributions of analyte biometals in the above 2D maps was correlated to normal cell growth. A decline in distribution of Fe in the non-cancerous cells with an increase in Cu distribution were evident. Further, reduced distributions of Fe and Cu as compared to the cancerous cell cultures was also evident in normal culture cells with inhomogeneous distributions of Fe and Cu were found throughout the various stages of development.

Figure 5.53 shows the 2D map of both Fe and Cu show increased accumulation in cancerous cell cultures due to cell proliferation which may be related to angiogenesis and oxidation processes for free radical generation that accelerates carcinogenesis (Tapia *et al.*, 2003).



(a) Cu



(b) Fe

Figure 5.53 Spatial distribution of Cu and Fe at different stages of development in cancer cells.

The spatial distribution Fe and Cu in Figure 5.51 were co-localized in same regions of the tumor cells with a significant correlation map between them ($p < 0.001$). The homogenous and elevated distribution of Fe and Cu were as a result of their probable co-localization in cancerous cell cultures especially in the perinuclear/cytoplasmic region of the cells. This was attributed to their role in angiogenic and oxidative processes in tumor tissues (Tapia *et al.*, 2003).

The significant difference in distribution of trace biometals were observed in cancerous and normal cell cultures. The localization of trace biometals in both healthy and cancerous cells together with (their correlations suggest potential carcinogenic mechanism for cancer diagnostics. Fe and Cu intensities in DU145 cells were higher in cancerous than in normal Vero cell cultures at similar stage of development suggesting their vital role in metabolic processes associated with carcinogenesis. The distributions of Mn did not show much significant distribution alterations in both cancerous and normal cell cultures. The alterations in levels of trace biometals due to biochemical changes in tumors were linked to changes in cellular structures (Silva *et al.*, 2012), thus spatial distribution of trace elements enabled their correlation with various stages of cancer.

In a similar study (Matsuyama *et al.*, 2009), Cu and Zn were co-localized with high concentration of Fe in the nucleus. Synchrotron based μ -XRF show similar results for 2D distribution of trace element (Ca, Fe, Cu and Zn) in normal and malignant breast tissues (Silva *et al.*, 2012). The qualitative results indicated that Ca-Zn and Fe-Cu correlations can be linked to malignant transformation in tissues. The changes in distribution of Fe and Cu at different stages of development presumably indicate their vital role in carcinogenesis and therefore can be used as potential early cancer biomarkers.

5.12.2 The quantitative distribution of trace biometals in cell cultures

The semi-quantitative analysis by mass fraction distribution of trace biometals from 2D micro XRF data were determined based on elemental density distributions. The semi-quantitative analytical results of trace biometals in normal cell cultures are as shown in Table 5.33.

Table 5.33 Semi-quantitative distributions of trace elements in healthy Vero cell cultures

Biometal	Mass Fraction distribution			
	V_D1	V_D2	V_D3	V_D4
Mn	0.0071	0.0005	0.0599	0.1402
Fe	0.0015	0.0086	0.0764	0.1265
Cu	0.0601	0.0683	0.6431	1.0450
Se	0.0109	0.0099	0.0276	0.0385

The corresponding semi-quantitative results of trace biometals in cancerous cell cultures are as shown in Table 5.34.

Table 5.34 Semi-quantitative distributions of trace elements in cancerous cell cultures

Biometal	Mass Fraction distribution			
	DU_D1	DU_D2	DU_D3	D_D4
Mn	0.0011	0.2647	0.0068	0.1061
Fe	0.0043	0.2445	0.0169	0.0959
Cu	0.0499	1.5420	0.3788	0.9444
Se	0.0111	0.0597	0.0229	0.0204

The semi-quantitative accumulation of Cu in concentration as in 2D images of cancerous cell cultures were higher than Fe as compared to corresponding normal cell cultures. Further, the Mn quantitative distributions were almost undetected (very low) in both cancerous and non-

cancerous cell cultures. The distribution of trace biometals in cell microenvironments is an important tool for early diagnosis of disease (Ortega *et al.*, 2009).

5.12.3 PC-ANN classification of cell cultures

The principal components with the greatest information (variance) of selected fluorescence of trace biometals (Mn, Fe, Cu and Se) and Compton scatter used as input data in ANN classification model enabled predictive classification modelling of cancer stage of development. Figure 5.54 shows the PC-ANN results using the principal components of selected fluorescence of Fe and Cu (PC1 and PC2) at a total explained variance of 92% as input data in ANN classification model.

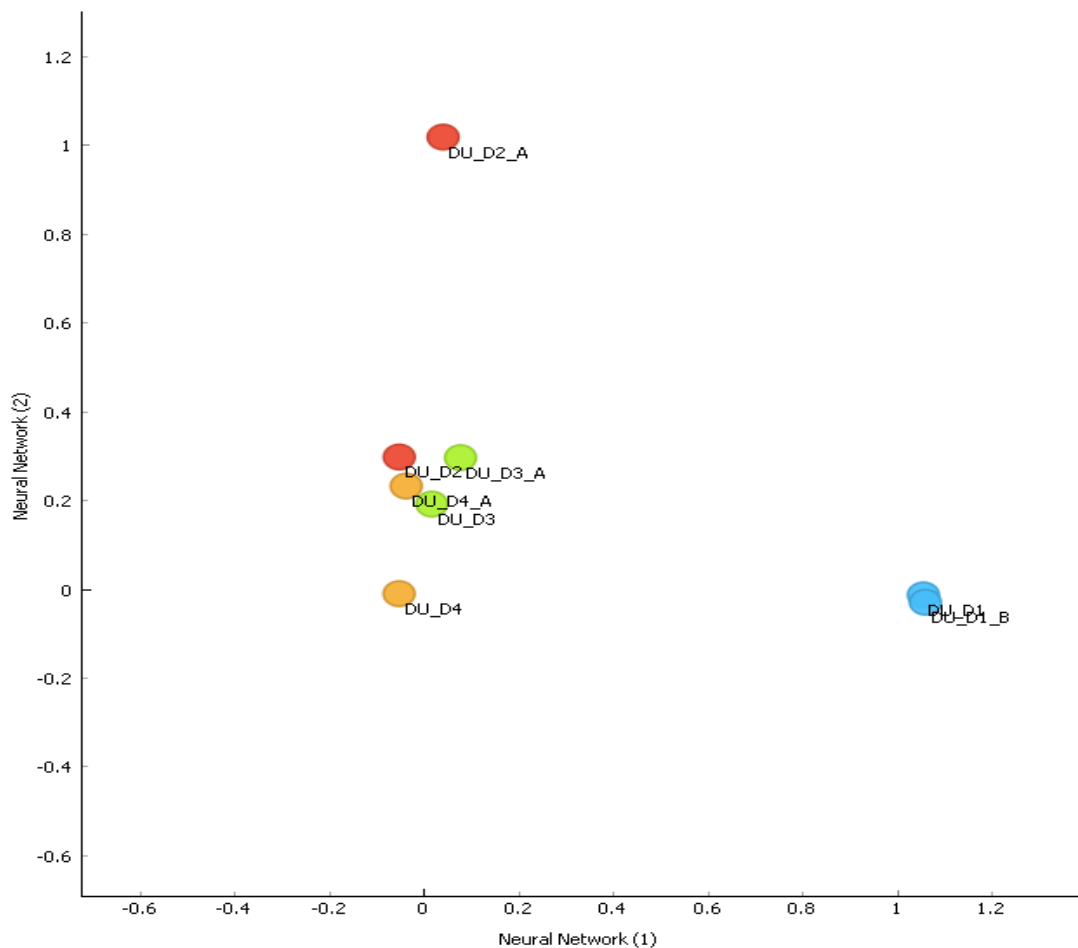


Figure 5.54 PC-ANN analysis of cancerous cells using the entire fluorescence region

The entire fluorescence region could not differentiate the various stages of cancer development and thus the need to utilize selected fluorescence regions of interest. Furthermore, the 2 dominant principal components score of selected fluorescence of Fe and

Cu with a total explained variance of 97% as input data matrix in ANN enabled accurate classification of cultured cells into cancerous and normal groups. The selected biomarker fluorescence spectral profiles were partially responsible for the differentiation of the cultured samples into cancerous and normal groups. These enabled the visualization of cancerous and normal cultures as shown in Figure 5.55 into two distinct clusters (cancerous and normal groups).

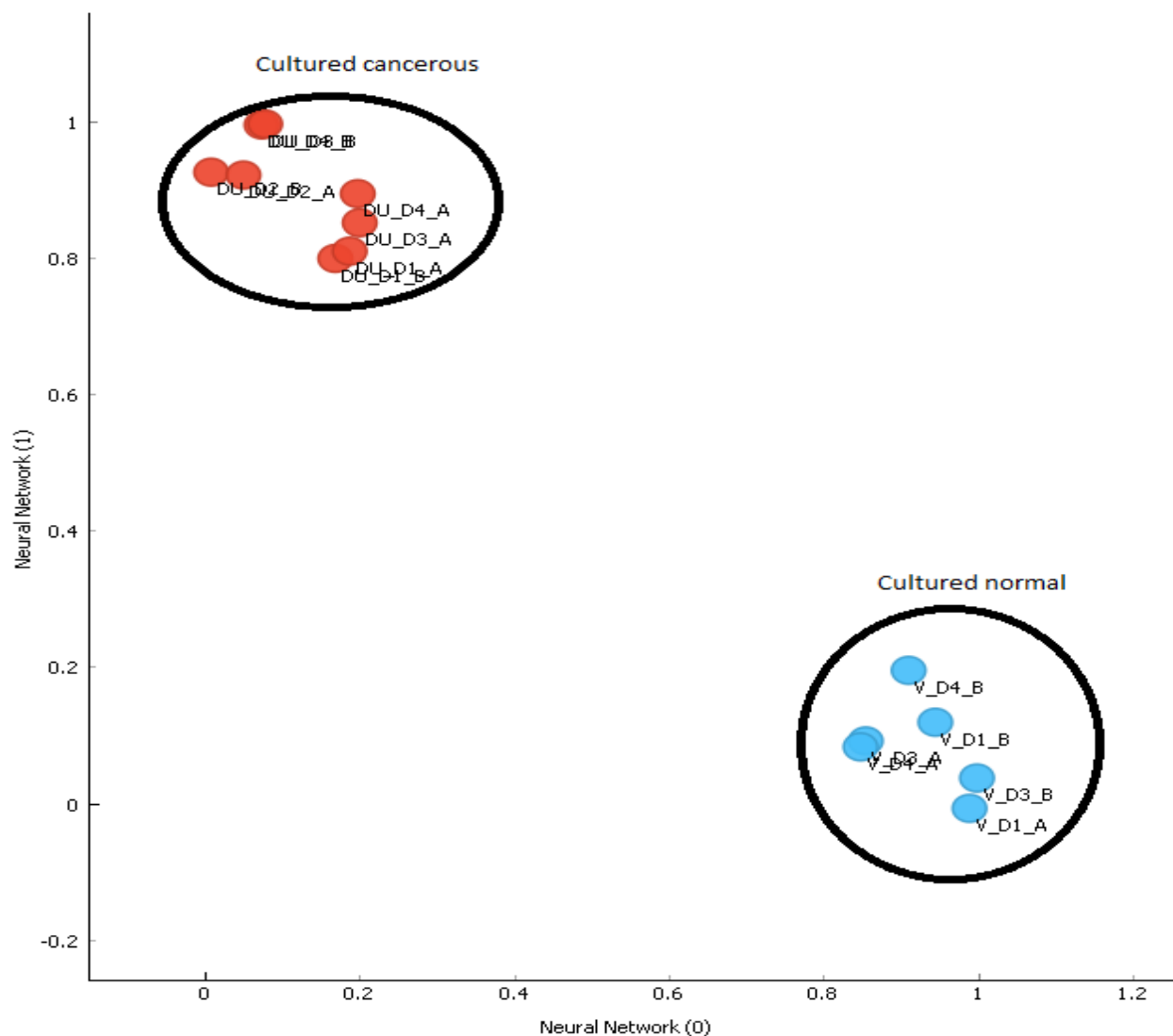


Figure 5.55 ANN fluorescence (Fe and Cu) scatter plot of cultured cells

The utility of Fe and Cu enabled characterization of cultured cancer cells where Cu it induced cell growth proliferation while Fe was greatly involved in oxygen metabolism (Rocha *et al.*, 2019) for cancer cells. Further, Fe aided in blood supply to the malignant cells thus

accelerated angiogenesis (Kwok and Richardson 2002). It was therefore deduced that Cu and Fe are potential biomarkers to not only identify the presence of cancer but also characterize cancer stages of development. The fluorescence and Compton scatter principal components characterized the 3 stages (early, intermediate and advanced) of cancer development with no clear distinction between DU_D3 and DU_D4 stages of cancer development as shown in Fig. 5.56. This was achieved at with area under the curve (AUC) of 0.958, classification accuracy (CA) of 0.875 and precision of 0.917 as the scatter peak contained additional information of cell structure (morphology) for cancer characterization.

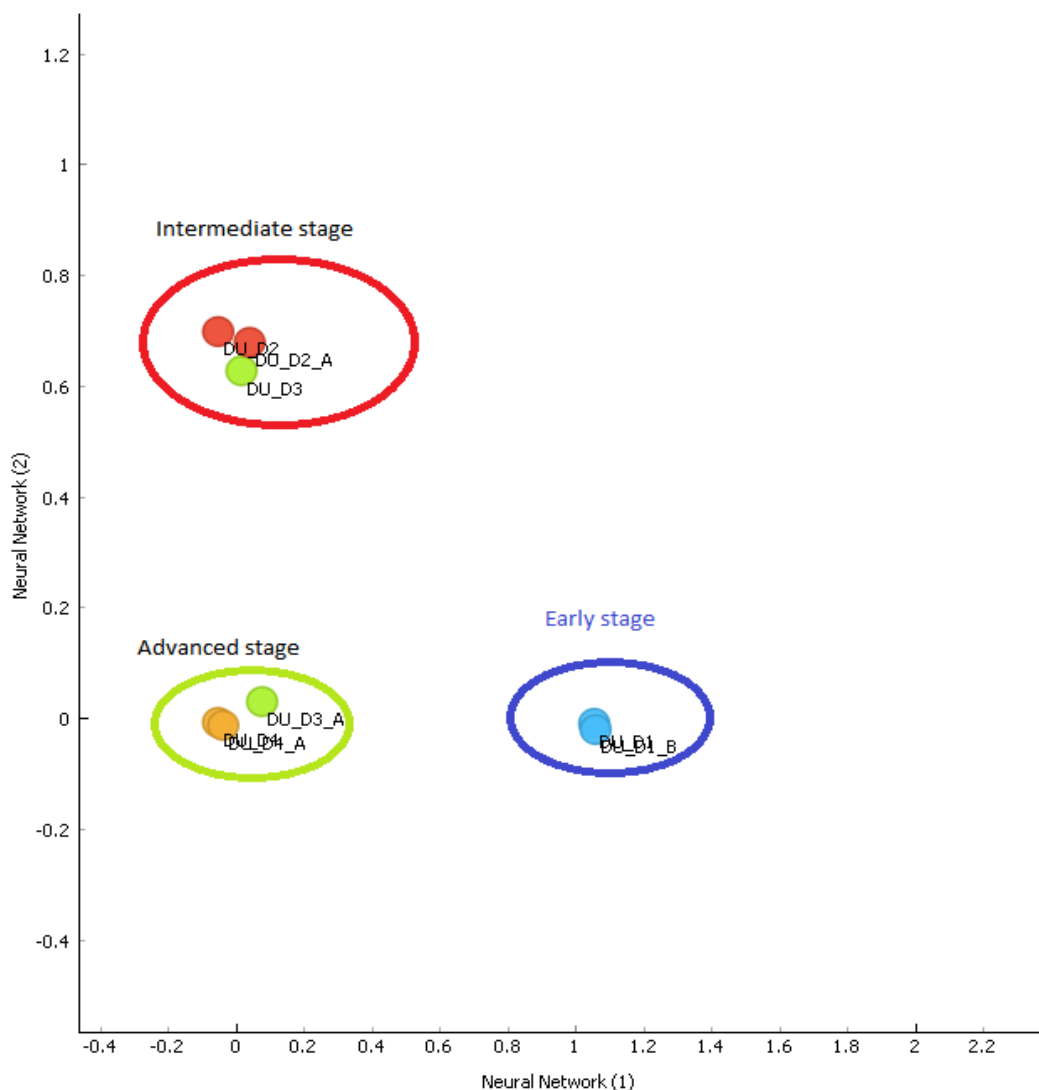


Figure 5.56 ANN scatter plot of cancerous cultured cells by fluorescence and Compton scatter

The inclusion of Compton scatter to the selected fluorescence further improved the exploratory analysis. In this regard, the results clearly indicated that the 2D distribution

profiles of trace biometals (Mn, Fe, Cu and Se) as potentially valuable cancer biomarkers for characterization of cell cultures and cancer severity for a reliable alternative cancer diagnostic method. The distribution of Fe and Cu change in response to cancer development staging thus they were responsible for the differentiation of the cultured cancerous samples into 3 stages of development. The probable stages of cancer development were early (well differentiated), intermediate (moderately differentiated) and advanced (poorly differentiated) stages of cancer development. In this case, Fe and Cu are unique cancer biomarkers for culture cells characterization as they were able to clearly distinguish the various stages of cancer development based on cell morphology.

5.12.4 PC-ANN classification of tissue biopsies

The tissue biopsy samples were characterized into clusters based on the fluorescence and scatter spectral profiles with area under the curve (AUC) of 0.904, classification accuracy (CA) of 0.781 and precision of 0.848 but with no clear distinction between cancerous and normal tissue biopsies as shown in Figure 5.57.

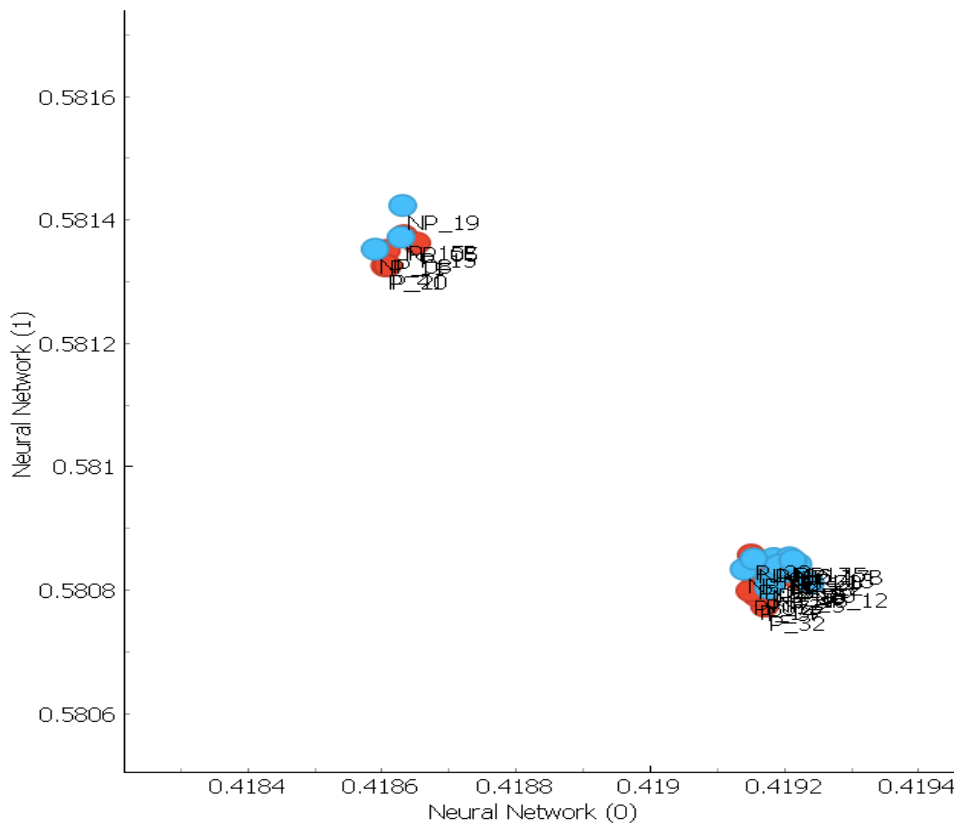


Figure 5.57 ANN scatter plot of tissue biopsies by fluorescence and scatter

The misclassification was probably due to false-negative or false-positive histopathological results where the patients were diagnosed at different stages of development of prostate

cancer that could not be resolved by ANN. Nonetheless, PCA reduced the ANN data dimensions to characterize the stages of cancer development of prostate cancer based on the selected fluorescence profiles of Mn, Fe, Cu, Zn and Se and Compton scatter region as shown in Fig. 5.58.

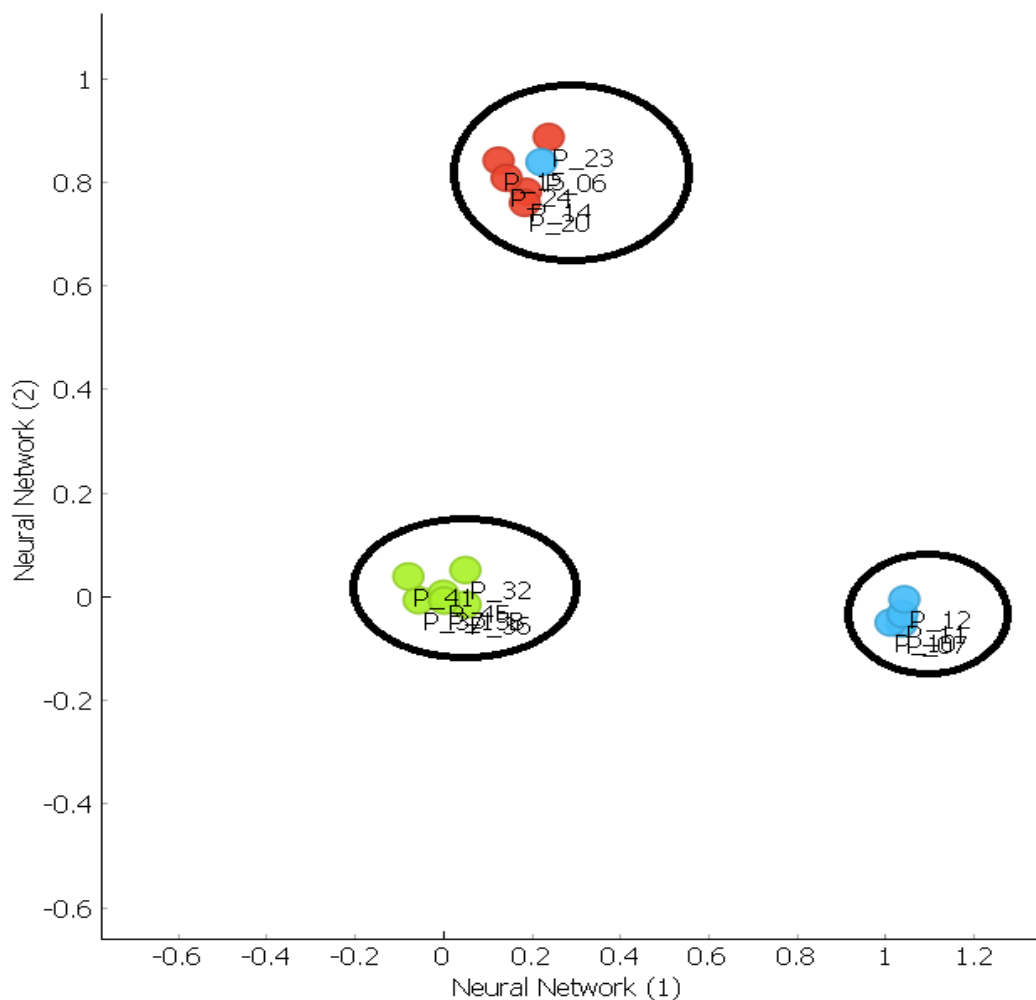


Figure 5.58 ANN classification of cancerous prostate urine by selected fluorescence and scatter

Prostate cancer urine samples were further classified into 3 stages of cancer development with AUC of 0.971, classification accuracy (CA) of 0.988 and precision of 0.948 which agreed with histopathologic biopsy results. Sample P_23 was misclassified into intermediate stage probably due to contamination during sampling. Machine learning based analytical approach thus distinguished between early and advanced stages of cancer development with additional intermediate stage utilizing the selected fluorescence and Compton scatter.

5.13 Trace biometal concentrations in prostate cancer stages

The trace biometals' levels of prostate cancer urine based on the above clusters in Figure 5.56 obtained using the validated ANN model are as shown in Table 5.35.

Table 5.35 Concentrations of trace elements in cancerous prostate urine

		Mean elemental concentrations (ppm)											
Stage	Sample	Fe		Mn		Cr		Cu		Se		Zn	
Intermediate prostate cancer	P_24	49.83	±2.52	1.58	±0.28	8.05	±0.08	10.51	±0.11	5.88	±0.50	12.13	±0.97
	P_23	54.97	±1.84	4.11	±0.14	7.05	±0.07	18.10	±0.22	3.14	±1.06	13.23	±1.14
	P_15	24.13	±2.94	1.61	±0.37	8.05	±0.05	10.50	±0.04	4.81	±1.57	11.37	±0.07
	P_14	56.54	±5.29	1.51	±0.26	8.05	±0.07	9.45	±0.07	7.89	±2.50	10.33	±0.14
	P_20	32.70	±1.92	1.54	±0.24	6.05	±0.02	10.01	±0.02	9.64	±2.27	12.74	±0.45
	P_06	18.76	±2.91	2.20	±0.43	8.05	±0.03	8.44	±0.05	6.84	±1.98	10.38	±0.04
Advanced Prostate cancer	P_41	33.71	±3.64	2.94	±0.88	6.05	±0.05	16.91	±0.02	6.62	±2.20	11.88	±0.67
	P_37	11.51	±1.08	1.37	±0.09	9.05	±0.06	9.06	±0.05	6.13	±2.02	11.71	±0.40
	P_45	25.19	±3.64	2.26	±0.46	7.05	±0.08	9.63	±0.01	4.59	±1.51	13.65	±0.31
	P_15	24.13	±2.94	1.61	±0.37	8.05	±0.06	10.50	±0.02	4.81	±1.57	11.37	±0.07
	P_32	6.17	±1.51	3.67	±0.34	8.04	±0.07	6.70	±0.19	4.44	±2.11	12.56	±0.19
	P_36	12.02	±2.98	1.50	±0.19	7.06	±0.11	7.84	±0.25	3.29	±1.85	10.11	±0.47
Early prostate cancer	P_12	36.93	±2.10	2.70	±1.15	8.05	±0.27	11.91	±0.01	8.93	±2.06	12.57	±0.23
	P_11	6.26	±1.63	1.63	±0.29	8.06	±0.06	9.05	±0.09	6.75	±2.08	12.19	±0.64
	P_10	13.18	±1.50	1.87	±0.58	8.04	±0.02	10.46	±0.17	7.93	±2.47	12.54	±0.18
	P_07	32.25	±2.55	3.72	±0.22	8.05	±0.22	11.66	±0.12	3.35	±2.60	12.40	±0.02

The above results show varying levels of trace biometals in the cancerous prostate urine with the concentration of Fe and Cu decreasing with staging of cancer in cancerous urine which is consistent with previous studies (Carvalho *et al.*, 2007). This study further indicated high levels of Fe and Cu at early stage of development due to rapid cell division and proliferation that supports tumor growth (Lee *et al.*, 2003). Additionally, significant decrease in levels of Cu were also evident due to lag phase with probably decreased generation of free radicals in Fenton reaction (Yaman *et al.*, 2005).

5.13.1 Relative values of trace biometal concentrations in cancerous human urine

The relative values in the levels of trace elements in malignant urine indicate the potential of using these ratios as discriminant factor for cancer characterization. Table 5.36 shows the relative values (ratios) of trace biometals in characterized prostate cancer samples.

Table 5.36 Ratios of trace biometal levels in cancerous samples

Relative values of trace biometal concentrations							
Stage	Sample	Fe/Cu	Fe/Se	Fe/Zn	Cu/Se	Cu/Zn	Se/Zn
Intermediate Prostate Cancer	P_24	4.74	8.47	4.11	1.79	0.87	0.48
	P_23	3.04	17.51	4.15	5.76	1.37	0.24
	P_15	2.30	5.02	2.12	2.18	0.92	0.42
	P_14	5.98	7.17	5.47	1.20	0.91	0.76
	P_20	3.27	3.39	2.57	1.04	0.79	0.76
	P_06	2.22	2.74	1.81	1.23	0.81	0.66
	Mean		3.59	7.38	3.37	2.20	0.95
Advanced Prostate Cancer	P_41	1.99	5.09	2.84	2.55	1.42	0.56
	P_37	1.27	1.88	0.98	1.48	0.77	0.52
	P_45	2.62	5.49	1.85	2.10	0.71	0.34
	P_15	2.30	5.02	2.12	2.18	0.92	0.42
	P_32	0.92	1.39	0.49	1.51	0.53	0.35
	P_36	1.53	3.65	1.19	2.38	0.78	0.33
	Mean		1.77	3.75	1.58	2.03	0.86
Early Prostate Cancer	P_12	3.10	4.14	2.94	1.33	0.95	0.71
	P_11	0.69	0.93	0.51	1.34	0.74	0.55
	P_10	1.26	1.66	1.05	1.32	0.83	0.63
	P_07	2.77	9.63	2.60	3.48	0.94	0.27
	Mean		1.92	4.02	1.74	1.90	0.86

The increased mean ratio of Fe/Se in prostate cancer tissues was due to increased level of Fe and a corresponding decreased level of Se in cancer tissues which can be used to differentiate prostate cancer stages of development cancer. It can further be deduced that increased ratios of Fe/Cu (> 7.0) and Cu/Se (> 6.0) and alteration in ratios of Fe/Zn, Fe/Se and Cu/Zn are

potential biomarkers for advancement of cancer from early to advanced stage of development. The ratios could be considered as useful parameters to characterize cancer stage of development which unravels the significance of the trace elemental ratios in metabolic processes associated with the development of the tumor. The relative values of trace biometals are valuable diagnostic markers and thus should be considered as tracers for the identification of pathological disorders in human fluid (urine).

5.14 Cancer diagnostic model

The above qualitative and quantitative results were used to develop diagnostic model (Figure 5.59) for prediction of the presence of cancer, its type and its stage that can subsequently be used in early diagnosis of cancer. The application of EDXRF, TXRF, μ -XRF and XANES enabled determination of a potential cancer diagnostic model based on alterations, levels, speciation and distribution of trace elements (Mn, Fe, Cu, Zn and Se) in tissues and urine for cancer diagnostics.

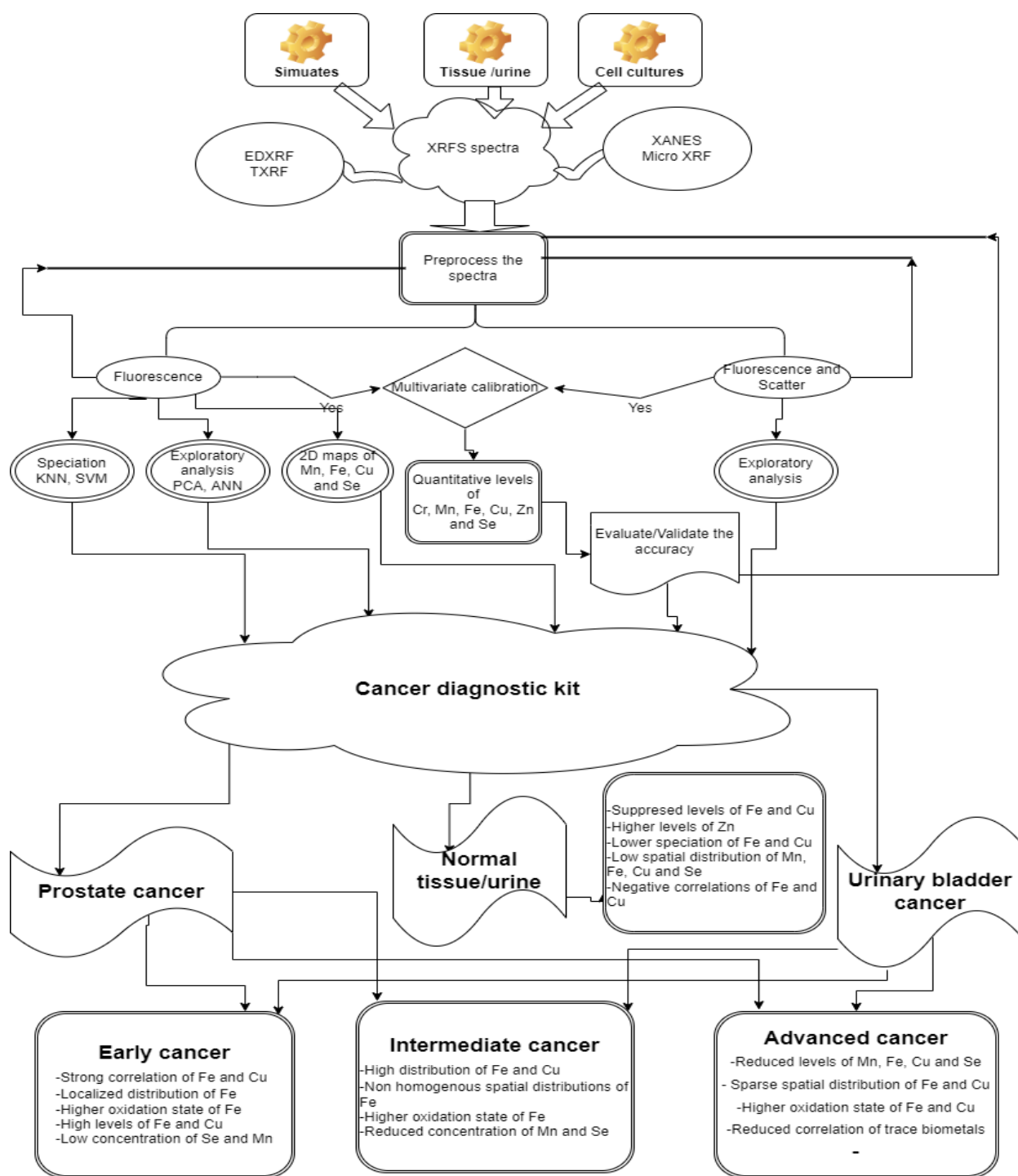


Figure 5.59: Machine learning aided XRFS cancer diagnostic model.

The above developed model was not only able to characterize cancer but also determine staging of cancer. These was associated with subtle alterations in the levels, distribution and speciation of trace biometals due to their active role in various enzymes and metabolic processes catalyzing chemical interactions in the cells. PC-ANN enabled differentiation between normal and cancerous cells as well as differentiation of the various stages of cancer

development. The results indicate that 2D mapping of trace biometals can be used for early diagnosis of cancer in individual cells with potential for clinical application. The above nested machine learning strategies not only indicated a potential method for the utilization of both fluorescence but also the Compton scatter.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

The study was aimed at developing the utility of machine learning aided XRF analysis by simultaneously qualitative and quantitative analysis of trace biometals in human tissue, urine and cell cultures for early cancer diagnosis. EDXRF, TXRF, μ -XRF and XANES were used to mine quantitative, speciation and distribution information of trace elements in tissues and urine for cancer diagnostics at early stages of development. The utility of hardly discernible fluorescence peaks and Compton Scatter peaks in tissues cell cultures and urine samples realized a rapid non-invasive method for characterization of cancer at an early stage of development.

Machine learning enabled data reduction for recovery of weak signals from high background and overlapped peaks. In this regard, PCA enabled reduction of data dimensions for possible classification. KNN and SVM determined the oxidation state of Fe and Cu in both human tissues and fluids while ANN and SVM were utilized for multivariate calibration to determine the concentration of trace biometals in human tissue biopsy and urine for cancer diagnosis. The cancerous tissue biopsy samples were characterized as cancerous and non-cancerous by PCA and HCA based on the fluorescence and Compton scatter peaks. In this study, a decline in distribution of Fe in the cancer cell cultures was noted with an increase in Cu localization in the cultured cancer cells. Further, multivariate modeling and exploratory capability were achieved by PCA enabled ANN that characterized cancer into well differentiated, moderately differentiated and poorly differentiated stages of cancer development.

The strong distribution correlation of Cu and Fe in cancerous cell cultures indicated vital roles Cu in cancer development. Further, weak correlation of Se with other trace biometals was evident despite Se being essential element typical for the normal cells. The results indicated predictable associations of trace biometals fluorescence together with the Compton scatter as cancer biomarkers with potential to be used for cancer detection, differentiation and characterization especially at the local stage of development. The levels, speciation and distribution patterns in human tissues and urine can be used as biomarkers for cancer severity characterization besides distinguishing normal and cancerous human body tissues and fluids. The levels and distribution of trace biometals and their speciation in human body tissue and

urine are therefore potential cancer biomarkers based on the correlations and patterns of trace elements in body fluids and tissues. Further, the concentrations, alterations, correlations and ratios of trace biometals in human tissues and urine can be used as biomarkers for cancer diagnostics at the local stage of development. An alternative and complementary approach for early-stage cancer screening has been developed and its subsequent application in a clinical set-up has a potential to reduce cancer morbidity and mortality rates.

6.2 Recommendations

This work highlights that trace biometals can be potential early biomarkers for early cancer diagnosis but in order to better understand the role of the elements and its changes, further analysis with large sample size is hereby highly recommended to reiterate and consolidate the current findings. Additional information on diet, drugs and biographical information of the patient that may affect the biometal profiles (levels, speciation and 2D mapping) need to be incorporated in the model before clinical trials for potential application of the method for cancer diagnosis.

REFERENCES

- Castillo-Michel, H., A. G Diaz-Sanchez, A. Martinez-Martinez and B. Hesse (2016). "Investigations of sulfur chemical status with synchrotron micro focused X-ray fluorescence and X-ray absorption spectroscopy." *Protein and Peptide Letters* **23**(3): 291-299.
- Al-Ebraheem, A., J. Goettlicher, K. Geraki, S. Ralph and M. Farquharson (2010). "The determination of zinc, copper and iron oxidation state in invasive ductal carcinoma of breast tissue and normal surrounding tissue using XANES." *X-Ray Spectrometry* **39**(5): 332-337.
- Alexandre, T. L. and M. I. Bueno (2006). "Classification of some species, genera and families of plants by x-ray spectrometry." *X-Ray Spectrometry* **35**(4): 257-260.
- Angeyo, K., S. Gari, A. Mustapha and J. Mangala (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**(11): 2596-2601.
- Aquilanti, G., M. Giorgetti, R. Dominko, L. Stievano, I. Arčon, N. Novello and L. Olivi (2017). "Operando characterization of batteries using X-ray absorption spectroscopy: advances at the beamline XAFS at synchrotron Elettra." *Journal of Physics D: Applied Physics* **50**(7): 074001.
- Aquilanti, G., M. Giorgetti, R. Dominko, L. Stievano, I. Arčon, N. Novello and L. J. J. o. P. D. A. P. Olivi (2017). "Operando characterization of batteries using x-ray absorption spectroscopy: advances at the beamline XAFS at synchrotron Elettra." **50**(7): 074001.
- Aruoma, O., B. Halliwell, E. Gajewski and M. J. J. o. B. C. Dizdaroglu (1989). "Damage to the bases in DNA induced by hydrogen peroxide and ferric ion chelates." **264**(34): 20509-20512.
- Aslam, R. and S. Neubauer (2013). "Dairy foods, milk, calcium, and risk of prostate cancer." *Oncology Nutrition Connection* **21**(1): 3-10.
- Ates, A., O. Simsek, B. Ertugral and M. J. S. Ertugrul (2008). "Trace elemental analysis of mitral valves by EDXRF." **22**(1): 57-62.
- Baffes, P. T. (1989). "NETS Users's Guide, Version 2.0 of NETS." *Technical Re-port JSC-23366, NASA, Software Technology Branch, Lyndon B. Johnson Space Center.*
- Baker, E., T. Stephenson, M. Reed and N. Brown (2002). "Expression of proteinases and inhibitors in human breast cancer progression and survival." *Molecular Pathology* **55**(5): 300.
- Banaś, A., W. Kwiatek and W. Zajac (2001). "Trace element analysis of tissue section by means of synchrotron radiation: the use of GNU PLOT for SRIXE spectra analysis." *Journal of Alloys and Compounds* **328**(1): 135-138.
- Bárány, E., I. A. Bergdahl, L.-E. Bratteby, T. Lundh, G. Samuelson, A. Schütz, S. Skerfving and A. Oskarsson (2002). "Relationships between trace element concentrations in human blood and serum." *Toxicology letters* **134**(1-3): 177-184.

Beckhoff, B., B. Kanngießer, N. Langhoff, R. Wedell and H. Wolff (2007). *Handbook of practical X-ray fluorescence analysis*, Springer Science & Business Media.

Beltrán, B. G., V. Ramos-Sanchez, D. Chávez-Flores, R. Rodríguez-Maese and E. Palacio (2020). "Total Reflection X-Ray Fluorescence Spectroscopy (TXRF) Method Validation: Determination of Heavy Metals in Dietary Supplements." *Journal of Chemistry* **2020**.

Benninghoff, L., D. Von Czarnowski, E. Denkhaus and K. Lemke (1997). "Analysis of human tissues by total reflection X-ray fluorescence. Application of chemometrics for diagnostic cancer recognition." *Spectrochimica Acta Part B: Atomic Spectroscopy* **52(7)**: 1039-1046.

Bethesda, M. (1989). "Tissue substitutes in radiation dosimetry and measurement." *International Commission on Radiation Units and Measurements (ICRU)*.

Bilderback, D. H., P. Elleaume and E. Weckert (2005). "Review of third and next generation synchrotron light sources." *Journal of Physics B: Atomic, molecular and optical physics* **38(9)**: S773.

Bortoleto, G. G., L. C. M. Pataca and M. I. M. Bueno (2005). "A new application of X-ray scattering using principal component analysis—classification of vegetable oils." *Analytica chimica acta* **539(1-2)**: 283-287.

Brereton, R. G. (2003). *Chemometrics: data analysis for the laboratory and chemical plant*, John Wiley & Sons.

Brown, C. W. and S.-C. Lo (1998). "Chemical information based on neural network processing of near-IR spectra." *Analytical Chemistry* **70(14)**: 2983-2990.

Bueno, M. I. M. S., M. T. Castro, A. M. de Souza, E. B. S. de Oliveira and A. P. Teixeira (2005). "X-ray scattering processes and chemometrics for differentiating complex samples using conventional EDXRF equipment." *Chemometrics and intelligent laboratory systems* **78(1)**: 96-102.

Bueno, M. I. M. S., M. T. Castro, A. M. de Souza, E. B. S. de Oliveira and A. P. Teixeira (2005). "X-ray scattering processes and chemometrics for differentiating complex samples using conventional EDXRF equipment." *Chemometrics and Intelligent Laboratory Systems* **78(1-2)**: 96-102.

Carter, E. A., B. S. Rayner, A. I. McLeod, L. E. Wu, C. P. Marshall, A. Levina, J. B. Aitken, P. K. Witting, B. Lai and Z. Cai (2010). "Silicon nitride as a versatile growth substrate for microspectroscopic imaging and mapping of individual cells." *Molecular Biosystems* **6(7)**: 1316-1322.

Carvalho, M., J. Brito and M. Barreiros (1998). "Study of trace element concentrations in human tissues by EDXRF spectrometry." *X-Ray Spectrometry: An International Journal* **27(3)**: 198-204.

Carvalho, M., J. Brito and M. Barreiros (1998). "Study of trace element concentrations in human tissues by EDXRF spectrometry." *X-Ray Spectrometry* **27**(3): 198-204.

Carvalho, M., T. Magalhães, M. Becker and A. Von Bohlen (2007). "Trace elements in human cancerous and healthy tissues: A comparative study by EDXRF, TXRF, synchrotron radiation and PIXE." *Spectrochimica Acta Part B: Atomic Spectroscopy* **62**(9): 1004-1011.

Castillo-Michel, H. A., C. Larue, A. E. P. Del Real, M. Cotte and G. Sarret (2017). "Practical review on the use of synchrotron based micro-and nano-X-ray fluorescence mapping and X-ray absorption spectroscopy to investigate the interactions between plants and engineered nanomaterials." *Plant Physiology and Biochemistry* **110**: 13-32.

Chakraborty, B., R. Kaustubha, A. Hegde, A. Pereira, W. Done, R. Kirilin, A. Moghaddamjoo, A. Georgakis, C. Kotropoulos and P. Xafopoulos (2000). "Bishop, CM, Neural Networks for Pattern Recognition, Oxford University Press, New York, 1995. Carreira-Perpiñán M., Mode-Finding for Mixtures of Gaussian Distributions, IEEE transaction on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, November 2000, 1318-1323." *IEEE transaction on Pattern Analysis and Machine Intelligence* **22**(11): 1318-1323.

Chang, D.-H. and S. Islam (2000). "Estimation of soil physical properties using remote sensing and artificial neural network." *Remote Sensing of Environment* **74**(3): 534-544.

Chen, Z., W. M. Gibson and H. Huang (2008). "High definition x-ray fluorescence: principles and techniques." *X-ray Optics and Instrumentation* **2008**.

Christensen, C., J. Cutler and D. Christensen (2004). "Using X-ray absorption near edge structure (XANES) spectroscopy to determine selenium oxidation states in animal mineral supplements and feeds." *Canadian journal of animal science* **84**(2): 171-175.

Cochran, W. G. (2007). *Sampling techniques*, John Wiley & Sons.

Culjat, M. O., D. Goldenberg, P. Tewari and R. S. Singh (2010). "A review of tissue substitutes for ultrasound imaging." *Ultrasound in medicine & biology* **36**(6): 861-873.

Cunha, D., O. Oliveira, C. Pérez and M. Poletti (2006). "X-ray scattering profiles of some normal and malignant human breast tissues." *X-Ray Spectrometry* **35**(6): 370-374.

da Silva, M. P., O. L. A. D. Zucchi, A. Ribeiro-Silva and M. E. Poletti (2009). "Discriminant analysis of trace elements in normal, benign and malignant breast tissues measured by total reflection X-ray fluorescence." *Spectrochimica Acta Part B: Atomic Spectroscopy* **64**(6): 587-592.

Davis, J. M. (2014). "Support Vector Machines for Classification and Quantitative Analysis." *Microscopy and Microanalysis* **20**(S3): 656-657.

De Leeuw, J. (2011). "Nonlinear principal component analysis and related techniques."

De Oliveira, L., A. Antunes and M. Bueno (2010). "Direct chromium speciation using X-ray spectrometry and chemometrics." *X-Ray Spectrometry* **39**(4): 279-284.

Deming, S., Y. Michotte, D. L. Massart, L. Kaufman and B. Vandeginste (1988). *Chemometrics: a textbook*, Elsevier.

Di Cicco, A., G. Aquilanti, M. Minicucci, E. Principi, N. Novello, A. Cognigni and L. Olivi (2009). *Novel XAFS capabilities at ELETTRA synchrotron light source*. Journal of Physics: Conference Series, IOP Publishing.

Eriksson, L., J. Trygg and S. Wold (2014). "A chemometrics toolbox based on projections and latent variables." *Journal of Chemometrics* **28**(5): 332-346.

Fahlman, S. E. (1988). *An empirical study of learning speed in back-propagation networks*, Carnegie Mellon University, Computer Science Department Pittsburgh, PA, USA.

Fahrni, C. J. (2007). "Biological applications of X-ray fluorescence microscopy: exploring the subcellular topography and speciation of transition metals." *Current opinion in chemical biology* **11**(2): 121-127.

Ferreira, C. C., R. Ximenes, C. A. B. Garcia, J. W. Vieira and A. F. Maia (2010). *Total mass attenuation coefficient evaluation of ten materials commonly used to simulate human tissue*. Journal of Physics: Conference Series, IOP Publishing.

Finney, L. A. and Q. Jin (2015). "Preparing adherent cells for X-ray fluorescence imaging by chemical fixation." *Journal of visualized experiments: JoVE*(97).

Fraga, C. G. and P. I. Oteiza (2002). "Iron toxicity and antioxidant nutrients." *Toxicology* **180**(1): 23-32.

Freitas, R. P., C. Calza, T. A. Lima, A. Rabello and R. T. Lopes (2010). "EDXRF and multivariate statistical analysis of fragments from Marajoara ceramics." *X-Ray Spectrometry* **39**(5): 307-310.

Frenkel, K. (1992). "Carcinogen-mediated oxidant formation and oxidative DNA damage." *Pharmacology & therapeutics* **53**(1): 127-166.

Garg, V. K., A. L. Srivastav, M. K. Tiwari, A. Sharma and V. S. Kanwar (2021). "Synchrotron based X-ray fluorescence for trace elemental analysis of industrial sludge." *Journal of Environmental Treatment Techniques* **9**(1): 192-195.

Gecit, I., S. Kavak, H. Demir, M. Gunes, N. Pirinççi, C. Cetin, K. Ceylan, E. Benli and I. Yildiz (2011). "Serum trace element levels in patients with bladder cancer." *Asian Pac J Cancer Prev* **12**: 3409-3413.

Geraki, K., M. Farquharson and D. Bradley (2002). "Concentrations of Fe, Cu and Zn in breast tissue: a synchrotron XRF study." *Physics in Medicine & Biology* **47**(13): 2327.

Geraki, K., M. Farquharson and D. Bradley (2003). "X-ray fluorescence and energy dispersive x-ray diffraction for the quantification of elemental concentrations in breast tissue." *Physics in Medicine & Biology* **49**(1): 99.

Geraki, K., M. Farquharson, D. Bradley and R. Hugtenburg (2004). "A synchrotron XRF study on trace elements and potassium in breast tissue." *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **213**: 564-568.

Gershenson, C. (2003). "Artificial neural networks for beginners." *arXiv preprint cs/0308031*.

Gherase, M. R. and D. E. Fleming (2019). "Probing Trace Elements in Human Tissues with Synchrotron Radiation." *Crystals* **10**(1): 12.

Gherase, M. R. and D. E. Fleming (2020). "Probing Trace Elements in Human Tissues with Synchrotron Radiation." *Crystals* **10**(1): 12.

Gianoncelli, A., V. Bonanni, G. Gariani, F. Guzzi, L. Pascolo, R. Borghes, F. Billè and G. Kourousias (2021). "Soft X-ray Microscopy Techniques for Medical and Biological Imaging at TwinMic—Elettra." *Applied Sciences* **11**(16): 7216.

Gianoncelli, A., B. Kaulich, R. Alberti, T. Klatka, A. Longoni, A. De Marco, A. Marcello and M. Kiskinova (2009). "Simultaneous soft X-ray transmission and emission microscopy." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **608**(1): 195-198.

Gianoncelli, A., G. Kourousias, L. Merolle, M. Altissimo and A. Bianco (2016). "Current status of the TwinMic beamline at Elettra: a soft X-ray transmission and emission microscopy station." *Journal of Synchrotron Radiation* **23**(6): 1526-1537.

Girotti, A. and J. Thomas (1984). "Damaging effects of oxygen radicals on resealed erythrocyte ghosts." *Journal of Biological Chemistry* **259**(3): 1744-1752.

Gonçalves, E. M., C. Â. Ventura, T. Yano, M. L. R. Macedo and S. C. Genari (2006). "Morphological and growth alterations in Vero cells transformed by cisplatin." *Cell biology international* **30**(6): 485-494.

Goraieb, K., T. L. Alexandre and M. I. M. Bueno (2007). "Sucrose inversion monitored by x-ray scattering and chemometrics." *X-Ray Spectrometry: An International Journal* **36**(4): 241-246.

Gray, M. A., J. A. Centeno, D. P. Slaney, J. W. Ejniak, T. Todorov and J. N. Nacey (2005). "Environmental exposure to trace elements and prostate cancer in three New Zealand ethnic groups." *International journal of environmental research and public health* **2**(3): 374-384.

Greaves, E., L. M. Parra, A. Rojas and L. Sajo-Bohus (2000). "Determination of platinum levels in serum and urine samples from pediatric cancer patients by TXRF." *X-Ray Spectrometry* **29**(5): 349-353.

Griffiths, K., D. Prezioso, A. Turkes and L. J. Denis (2007). The prevention of prostate cancer. *Prostate Cancer*, Springer: 33-63.

Gruber, A., R. Müller, A. Wagner, S. Colucci, M. V. Spasić and K. Leopold (2020). "Total reflection X-ray fluorescence spectrometry for trace determination of iron and some

additional elements in biological samples." *Analytical and bioanalytical chemistry* **412**(24): 6419-6429.

Grüner, F., F. Blumendorf, O. Schmutzler, T. Stauer, M. Bradbury, U. Wiesner, T. Rosentreter, G. Loers, D. Lutz and B. Richter (2018). "Localising functionalised gold-nanoparticles in murine spinal cords by X-ray fluorescence imaging and background-reduction through spatial filtering for human-sized objects." *Scientific reports* **8**(1): 1-12.

Guidotti, T., J. McNamara and M. Moses (2008). "The interpretation of trace element analysis in body fluids." *Indian Journal of Medical Research* **128**(4): 524.

Guo, C.-H., S. Hsia, M.-Y. Shih, F.-C. Hsieh and P.-C. Chen (2015). "Effects of selenium yeast on oxidative stress, growth inhibition, and apoptosis in human breast cancer cells." *International journal of medical sciences* **12**(9): 748.

Halliwell, B. and O. I. Aruoma (1991). "DNA damage by oxygen-derived species Its mechanism and measurement in mammalian systems." *FEBS letters* **281**(1-2): 9-19.

Halliwell, B. and J. Gutteridge (1984). "Oxygen toxicity, oxygen radicals, transition metals and disease." *Biochemical journal* **219**(1): 1.

Halliwell, B. and J. M. J. B. j. Gutteridge (1984). "Oxygen toxicity, oxygen radicals, transition metals and disease." **219**(1): 1-14.

Harfouche, M., M. Abdellatif, Y. Momani, A. Abbadi, M. Al Najdawi, M. Al Zoubi, B. Aljamal, S. Matalgah, L. U. Khan and A. Lausi (2022). "Emergence of the first XAFS/XRF beamline in the Middle East: providing studies of elements and their atomic/electronic structure in pluridisciplinary research fields." *Journal of Synchrotron Radiation* **29**(4): 1107-1113.

Heath, J. L., J. M. Weiss, C. P. Lavau and D. S. Wechsler (2013). "Iron deprivation in cancer—potential therapeutic implications." *Nutrients* **5**(8): 2836-2859.

Heermann, P. D. and N. Khazenie (1992). "Classification of multispectral remote sensing data using a back-propagation neural network." *IEEE Transactions on geoscience and remote sensing* **30**(1): 81-88.

Henrich, A., P. Hoffmann, H. Ortner, T. Greve and H. Itzel (2000). "Non-invasive identification of chemical compounds by energy dispersive X-ray fluorescence spectrometry, combined with chemometric methods of data evaluation." *Fresenius' journal of analytical chemistry* **368**(2): 130-138.

Hernández-Caraballo, E. A. and L. M. Marco-Parra (2003). "Direct analysis of blood serum by total reflection X-ray fluorescence spectrometry and application of an artificial neural network approach for cancer diagnosis." *Spectrochimica Acta Part B: Atomic Spectroscopy* **58**(12): 2205-2213.

Imam, H., R. Mohamed and A. A. Eldakrouri (2012). "Primary study of the use of laser-induced plasma spectroscopy for the diagnosis of breast cancer."

Imtiaz, M., A. Rashid, P. Khan, M. Memon and M. Aslam (2010). "The role of micronutrients in crop production and human health." *Pak. J. Bot* **42**(4): 2565-2578.

INAGAKI, K., T. UMEMURA, H. MATSUURA and H. HARAGUCHI (2000). "Speciation of trace elements, binding and non-binding with proteins in human blood serum, by surfactant-mediated HPLC with element-selective detection by ICP-MS." *Analytical sciences* **16**(8): 787-788.

Jalilehvand, F. (2006). "Sulfur: not a "silent" element any more." *Chemical Society Reviews* **35**(12): 1256-1268.

Jiménez, E., A. Rojas and E. Greaves (2001). "Determination of Zn/Cu ratio and oligoelements in serum samples by total reflection X-ray fluorescence spectrometry for cancer diagnosis." *Spectrochimica Acta Part B: Atomic Spectroscopy* **56**(11): 2195-2201.

Jones, M. W., G. Mallmann, J. L. Wykes, J. Knafelc, S. E. Bryan and D. L. Howard (2020). "Iterative energy self-calibration of Fe XANES spectra." *Journal of synchrotron radiation* **27**(1): 207-211.

Kainth, H. S., D. Khandelwal, R. Singh, G. Singh and S. Puri (2020). Role of Trace Elements in Breast Cancer and Their Characterization Using X-Ray Fluorescence Techniques. Trace Elements and Their Effects on Human Health and Diseases, IntechOpen.

Kaniu, M., K. Angeyo, A. Mwala and M. Mangala (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.

Kaulich, B., D. Bacescu, J. Susini, C. David, E. Di Fabrizio, G. Morrison, P. Charalambous, J. Thieme, T. Wilhein and J. Kovac (2006). *TwinMic: a European twin X-ray microscopy station commissioned at ELETTRA*. 8th Int. Conf. X-ray Microscopy.

Kawakami, T., A. Ide-Ektessabi, K. Sugimura, Y. Kitamura, A. Gotoh and T. Shirakawa (2003). *Investigations on Distribution and Quantity of Zinc in Prostate Cancer Using Synchrotron Radiation Microbeams*. AIP Conference Proceedings, American Institute of Physics.

Khan, L. U., G. H. da Silva, A. M. de Medeiros, Z. U. Khan, M. Gidlund, H. F. Brito, O. Moscoso-Londono, D. Muraca, M. Knobel and C. A. Perez (2019). "Fe₃O₄@ SiO₂ nanoparticles concurrently coated with chitosan and GdOF: Ce³⁺, Tb³⁺ luminophore for bioimaging: toxicity evaluation in the zebrafish model." *ACS Applied Nano Materials* **2**(6): 3414-3425.

Khuder, A., M. Bakir, J. Karjou and M. K. Sawan (2007). "XRF and TXRF techniques for multi-element determination of trace elements in whole blood and human hair samples." *Journal of Radioanalytical and Nuclear Chemistry* **273**(2): 435-442.

Klockenkämper, R. and A. von Bohlen (2014). "Worldwide distribution of Total Reflection X-ray Fluorescence instrumentation and its different fields of application: A survey." *Spectrochimica Acta Part B: Atomic Spectroscopy* **99**: 133-137.

Koch, H. J., E. R. Smith, N. F. Shimp and J. Connor (1956). "Analysis of trace elements in human tissues. I. Normal tissues." *Cancer* **9**(3): 499-511.

Kong, Y., S. Akatsuka, Y. Motooka, H. Zheng, Z. Cheng, Y. Shiraki, T. Mashimo, T. Imaoka and S. Toyokuni (2022). "BRCA1 haploinsufficiency promotes chromosomal amplification under Fenton reaction-based carcinogenesis through ferroptosis-resistance." *Redox Biology*: 102356.

Kourousias, G., L. Pascolo, P. Marmorato, J. Ponti, G. Ceccone, M. Kiskinova and A. Gianoncelli (2015). "High-resolution scanning transmission soft X-ray microscopy for rapid probing of nanoparticle distribution and sufferance features in exposed cells." *X-Ray Spectrometry* **44**(3): 163-168.

Kowalik, C. and J. W. Einax (2006). "Modern chemometric data analysis—methods for the objective evaluation of load in river systems." *Acta hydrochimica et hydrobiologica* **34**(5): 425-436.

Kowalik, C. and J. W. Einax (2006). "Modern chemometric data analysis—methods for the objective evaluation of load in river systems." *CLEAN—Soil, Air, Water* **34**(5): 425-436.

Kowalski, B. and C. Bender (1972). "Pattern recognition. Powerful approach to interpreting chemical data." *Journal of the American Chemical Society* **94**(16): 5632-5639.

Kubala-Kukuś, A., D. Banaś, J. Braziewicz, U. Majewska, M. Pajek, J. Wudarczyk-Moćko, G. Antczak, B. Borkowska, S. Gózdź and J. Smok-Kalwat (2014). "Analysis of copper concentration in human serum by application of total reflection X-ray fluorescence method." *Biological trace element research* **158**(1): 22-28.

Kubala-Kukuś, A., M. Arabski, I. Stabrawa, D. Banaś, W. Różański, M. Lipiński, U. Majewska, J. Wudarczyk-Moćko, J. Braziewicz and M. J. X. R. S. Pajek (2017). "Application of TXRF and XRPD techniques for analysis of elemental and chemical composition of human kidney stones." **46**(5): 412-420.

Kwiatk, W., A. Banaś, M. Gajda, M. Gałka, B. Pawlicki, G. Falkenberg, T. J. J. o. a. Cichocki and compounds (2005). "Cancerous tissues analyzed by SRIXE." **401**(1-2): 173-177.

Kwiatk, W., A. Hanson, C. Paluszkiwicz, M. Gałka, M. Gajda and T. Cichocki (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**(1): 83-87.

Kwiatk, W. M., T. Drewniak, M. Gajda, M. Gałka, A. L. Hanson and T. Cichocki (2002). "Preliminary study on the distribution of selected elements in cancerous and non-cancerous kidney tissues." *Journal of trace elements in medicine and biology* **16**(3): 155-160.

Kwok, J. C. and D. R. Richardson (2002). "The iron metabolism of neoplastic cells: alterations that facilitate proliferation?" *Critical reviews in oncology/hematology* **42**(1): 65-78.

- Lee, R., W. Woo, B. Wu, A. Kummer, H. Duminy and Z. Xu (2003). "Zinc accumulation in N-methyl-N-nitrosourea-induced rat mammary tumors is accompanied by an altered expression of ZnT-1 and metallothionein." *Experimental Biology and Medicine* **228**(6): 689-696.
- Leitão, R., A. Palumbo Jr, P. Souza, G. Pereira, C. Canellas, M. Anjos, L. Nasciutti and R. Lopes (2014). "Elemental concentration analysis in prostate tissues using total reflection X-ray fluorescence." *Radiation Physics and Chemistry* **95**: 62-64.
- Lin, C.-N., L.-H. Wang and K.-H. Shen (2009). "Determining urinary trace elements (Cu, Zn, Pb, As, and Se) in patients with bladder cancer." *J Clin Lab Anal* **23**(3): 192-195.
- Liotti, E., A. Lui, T. Connolley, I. Dolbnya, K. Sawhney, A. Malandain, M. Wilson, M. Veale, P. Seller and P. J. S. r. Grant (2015). "Mapping of multi-elements during melting and solidification using synchrotron X-rays and pixel-based spectroscopy." **5**: 15988.
- Looney, C. G. (1996). "Advances in feedforward neural networks: demystifying knowledge acquiring black boxes." *IEEE Transactions on Knowledge and Data Engineering* **8**(2): 211-226.
- Lu, J., J. Guo, Q. Wei, X. Tang, T. Lan, Y. Hou and X. Zhao (2022). "A Matrix Effect Correction Method for Portable X-ray Fluorescence Data." *Applied Sciences* **12**(2): 568.
- Luo, L. (2006). "Chemometrics and its applications to x-ray spectrometry." *X-Ray Spectrometry* **35**(4): 215-225.
- Luo, L., C. Guo, G. Ma and A. Ji (1997). "Choice of optimum model parameters in artificial neural networks and application to X-ray fluorescence analysis." *X-Ray Spectrometry: An International Journal* **26**(1): 15-22.
- Mahugija, J. A., Z. S. Kasenya and K. F. Kilulya (2018). "Levels of heavy metals in urine samples of school children from selected industrial and non-industrial areas in Dar es Salaam, Tanzania." *African health sciences* **18**(4): 1226-1235.
- Majewska, U., D. Banaś, J. Braziewicz, A. Kubala-Kukuś, M. Pajek, I. Sychowska, J. Wudarczyk-Moćko, G. Antczak, B. Borkowska and S. Gózdź (2014). "X-ray fluorescence techniques in medical applications: reference values of elements in human serum, urine and hair." *Acta Physica Polonica A* **125**(4): 864-868.
- Marini, F. (2010). "Classification methods in chemometrics." *Current Analytical Chemistry* **6**(1): 72-79.
- Marini, F., R. Bucci, A. Magrì and A. Magrì (2008). "Artificial neural networks in chemometrics: History, examples and perspectives." *Microchemical journal* **88**(2): 178-185.
- Marini, F. J. C. A. C. (2010). "Classification methods in chemometrics." **6**(1): 72-79.
- Mark, H. (2001). "Data analysis: multilinear regression and principal component analysis." *PRACTICAL SPECTROSCOPY SERIES* **27**: 129-184.

Markowicz, A. (2011). "An overview of quantification methods in energy-dispersive X-ray fluorescence analysis." *Pramana* **76**(2): 321-329.

Markowicz, A. A., R. Van Grieken and A. Markowicz (1993). "Handbook of X-ray Spectrometry." *Marcal Dekker, New York*.

Markowicz, A. J. P. (2011). "An overview of quantification methods in energy-dispersive X-ray fluorescence analysis." **76**(2): 321-329.

Marmorato, P., G. Ceccone, A. Gianoncelli, L. Pascolo, J. Ponti, F. Rossi, M. Salomé, B. Kaulich and M. Kiskinova (2011). "Cellular distribution and degradation of cobalt ferrite nanoparticles in Balb/3T3 mouse fibroblasts." *Toxicology letters* **207**(2): 128-136.

Massart, D. L., B. Vandeginste, S. Deming, Y. Michotte and L. Kaufman (1988). "Chemometrics: a textbook."

Mastelaro, V. R. and E. D. Zanotto (2018). "X-ray absorption fine structure (XAFS) studies of oxide glasses—a 45-year overview." *Materials* **11**(2): 204.

Matsuyama, S., M. Shimura, H. Mimura, M. Fujii, H. Yumoto, Y. Sano, M. Yabashi, Y. Nishino, K. Tamasaku and T. Ishikawa (2009). "Trace element mapping of a single cell using a hard x-ray nanobeam focused by a Kirkpatrick-Baez mirror system." *X-Ray Spectrometry: An International Journal* **38**(2): 89-94.

Mitchell, T. M. (1997). *Machine learning*, McGraw-hill New York.

Mitchell, T. M. (1999). "Machine learning and data mining." *Communications of the ACM* **42**(11): 30-36.

Morona, F., F. R. Dos Santos, A. M. Brinatti and F. L. Melquiades (2017). "Quick analysis of organic matter in soil by energy-dispersive X-ray fluorescence and multivariate analysis." *Applied Radiation and Isotopes* **130**: 13-20.

Nagata, N., P. G. Peralta-Zamora, R. J. Poppi, C. A. Perez and M. I. M. J. X. R. S. A. I. J. Bueno (2006). "Multivariate calibrations for the SR-TXRF determination of trace concentrations of lead and arsenic in the presence of bromine." **35**(1): 79-84.

Naidu, B. G., P. Sarita, G. N. Raju and M. Tiwari (2019). "Multivariate analysis of trace elemental data obtained from blood serum of breast cancer patients using SRXRF." *Results in Physics* **12**: 673-680.

Nasulewicz, A., A. Mazur and A. Opolski (2004). "Role of copper in tumour angiogenesis—clinical implications." *Journal of Trace Elements in Medicine and Biology* **18**(1): 1-8.

Navdaev, A. and J. A. Eble (2011). "Components of cell-matrix linkage as potential new markers for prostate cancer." *Cancers* **3**(1): 883-896.

Nguyen, M., J. Arkell and C. J. Jackson (2001). "Human endothelial gelatinases and angiogenesis." *The international journal of biochemistry & cell biology* **33**(10): 960-970.

Northrup, P., A. Leri and R. Tappero (2016). "Applications of "tender" energy (1-5 keV) X-ray absorption spectroscopy in life sciences." *Protein and peptide letters* **23**(3): 300-308.

Northrup, P., A. Leri, R. J. P. Tappero and P. Letters (2016). "Applications of "tender" energy (1-5 keV) X-ray absorption spectroscopy in life sciences." **23**(3): 300-308.

Ogunlewe, J. and D. Osegbe (1989). "Zinc and cadmium concentrations in indigenous blacks with normal, hypertrophic, and malignant prostate." *Cancer* **63**(7): 1388-1392.

Okonda, J., H. Angeyo, A. Dehayem-Kamadjeu and A. Rogena (2022). "Chemometrics aided energy dispersive X-ray fluorescence and scattering for diagnostic modeling of trace biometals as cancer biomarkers." *Applied Radiation and Isotopes*: 110489.

Olivieri, A. C., G. M. Escandar and A. M. De La Peña (2011). "Second-order and higher-order multivariate calibration methods applied to non-multilinear data using different algorithms." *TrAC Trends in Analytical Chemistry* **30**(4): 607-617.

Ortega, R., G. Deves and A. Carmona (2009). "Bio-metals imaging and speciation in cells using proton and synchrotron radiation X-ray microspectroscopy." *Journal of the Royal Society Interface* **6**(suppl_5): S649-S658.

Pao, Y. (1989). "Adaptive pattern recognition and neural networks."

Pascolo, L., B. Bortot, N. Benseny-Cases, A. Gianoncelli, G. Tosi, B. Ruozi, C. Rizzardi, E. De Martino, M. A. Vandelli and G. M. Severini (2014). "Detection of PLGA-based nanoparticles at a single-cell level by synchrotron radiation FTIR spectromicroscopy and correlation with X-ray fluorescence microscopy." *International journal of nanomedicine* **9**: 2791.

Platz, E. A. and K. J. Helzlsouer (2001). "Selenium, zinc, and prostate cancer." *Epidemiologic reviews* **23**(1): 93-101.

Podgoczyk, M., W. Kwiatek, W. Zajac, J. Dulińska-Litewka, E. Welter and D. Grolimund (2009). "Zinc in native tissues and cultured cell lines of human prostate studied by SR-XRF and XANES." *X-Ray Spectrometry: An International Journal* **38**(6): 557-562.

Poletti, M., O. Gonçalves, C. Pérez and S. Magalhaes (2004). "A preliminary study of the distribution of trace elements in healthy and neoplastic breast tissues with synchrotron radiation X-ray fluorescence." *Radiation Physics and Chemistry* **71**(3-4): 975-976.

Porcaro, F., S. Roudeau, A. Carmona and R. J. T. T. i. A. C. Ortega (2018). "Advances in element speciation analysis of biomedical samples using synchrotron-based techniques." **104**: 22-41.

Pozebon, D., G. Scheffler and V. Dressler (2017). "Recent applications of laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) for biological sample analysis: a follow-up review." *Journal of Analytical Atomic Spectrometry* **32**(5): 890-919.

Prasad, A. S. (2013). *Essential and toxic element: trace elements in human health and disease*, Elsevier.

Pushie, M. J., I. J. Pickering, M. Korbas, M. J. Hackett and G. N. George (2014). "Elemental and chemically specific X-ray fluorescence imaging of biological systems." *Chemical Reviews* **114**(17): 8499-8541.

Putnam, D. F. (1971). "Composition and concentrative properties of human urine."

Qi, H., J. Wang, X. Zhang and Y. Wang (2015). "Matrix effect of Fe and Ca on EDXRF analysis of Ce concentration in bayan obo ores." *Guang pu xue yu Guang pu fen xi= Guang pu* **35**(12): 3510-3513.

Raju, G. N., P. Sarita, M. R. Kumar, G. R. Murty, B. S. Reddy, S. Lakshminarayana, V. Vijayan, P. R. Lakshmi, S. Gavarasana and S. B. Reddy (2006). "Trace elemental correlation study in malignant and normal breast tissue by PIXE technique." *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **247**(2): 361-367.

Rao, A. J. O. J. o. H. and A. Sciences (2005). "Trace element estimation—Methods & clinical context." **4**(1).

Ren, Y., Z. Zhang, Y. Ren, W. Li, M. Wang and G. Xu (1997). "Diagnosis of lung cancer based on metal contents in serum and hair using multivariate statistical methods." *Talanta* **44**(10): 1823-1831.

Rocha, K. J., R. G. Leitão, E. O. Barros, M. A. Oliveira, C. G. Canellas, M. J. Anjos, L. E. Nasciutti and R. T. Lopes (2019). "Microbeam X-ray fluorescence mapping of Cu and Fe in human prostatic carcinoma cell lines using synchrotron radiation." *Brazilian Journal of Radiation Sciences* **7**(2A).

Romanenko, S. V. and A. Stromberg (2007). "Modelling of analytical peaks: peaks modifications." *Analytica chimica acta* **581**(2): 343-354.

Rose, J. (2016). *Trace elements in health: a review of current issues*, Butterworth-Heinemann.

Rosenberg, M. G. (2002). *Supramolecular carbene chemistry*, State University of New York at Binghamton.

Rumelhart, D. E., G. E. Hinton and J. L. McClelland (1986). "A general framework for parallel distributed processing." *Parallel distributed processing: Explorations in the microstructure of cognition* **1**(45-76): 26.

Sartoros, C. and E. D. Salin (1997). "Pattern recognition for sample classification using elemental composition—application for inductively coupled plasma atomic emission spectrometry." *Journal of Analytical Atomic Spectrometry* **12**(8): 827-831.

Schmeling, M. J. P. S. R. (2019). "Total reflection X-ray fluorescence." **4**(7).

Schubert, H. L., K. S. Wilson, E. Raux, S. C. Woodcock and M. J. Warren (1998). "The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase." *Nature Structural & Molecular Biology* **5**(7): 585-592.

Shmaefsky, B. R. (1990). "Artificial urine for laboratory testing." *The American Biology Teacher* **52**(3): 170-172.

Shpyleva, S. I., V. P. Tryndyak, O. Kovalchuk, A. Starlard-Davenport, C. Vasyl'F, F. A. Beland and I. P. Pogribny (2011). "Role of ferritin alterations in human breast cancer cells." *Breast cancer research and treatment* **126**(1): 63-71.

Sichangi, E. K., H. K. Angeyo and A. Dehayem-Massop (2019). "Trace metal biomarker based Cancer diagnostics in body tissue by energy dispersive X-ray fluorescence and scattering (EDXRFS) spectrometry." *Spectrochimica Acta Part B: Atomic Spectroscopy*: 105635.

Silva, M., A. Tomal, A. Conceição and M. Poletti (2006). "Determination of trace elements in normal and neoplastic breast human tissues using SR-XRF I: experimental procedure."

Silva, M., A. Tomal, C. Perez, A. Ribeiro-Silva and M. E. Poletti (2009). "Determination of Ca, Fe, Cu and Zn and their correlations in breast cancer and normal adjacent tissues." *X-Ray Spectrometry: An International Journal* **38**(2): 103-111.

Silva, M., A. Tomal and M. Poletti (2006). "Determination of trace elements in normal and neoplastic breast human tissues using SR-XRF II: results and analysis Brazilian Synchrotron Light Laboratory Activity Report LNLS."

Silva, M. P., D. F. Soave, A. Ribeiro-Silva and M. E. Poletti (2012). "Trace elements as tumor biomarkers and prognostic factors in breast cancer: a study through energy dispersive x-ray fluorescence." *BMC Research Notes* **5**(1): 194.

Silva, M. P. d., D. Silva, A. Conceição, A. Ribeiro-Silva and M. E. Poletti (2013). "Role of Ca, Fe, Cu and Zn in breast cancer: study by X-ray fluorescence techniques and immunohistochemical analysis." *X-Ray Spectrometry* **42**(4): 303-311.

Silva, M. P. d., D. M. d. Silva, A. Ribeiro-Silva and M. E. Poletti (2012). *Correlations of trace elements in breast human tissues: Evaluation of spatial distribution using μ -XRF*. AIP Conference Proceedings, American Institute of Physics.

Silvera, S. A. N. and T. E. Rohan (2007). "Trace elements and cancer risk: a review of the epidemiologic evidence." *Cancer Causes & Control* **18**(1): 7-27.

Silversmit, G., B. Vekemans, F. E. Brenker, S. Schmitz, M. Burghammer, C. Riekel and L. Vincze (2009). "X-ray fluorescence nanotomography on cometary matter from Comet 81P/Wild2 returned by Stardust." *Analytical Chemistry* **81**(15): 6107-6112.

Skalnaya, M. G. and A. V. J. P. H. o. T. S. U. Skalny, Tomsk (2018). "Essential trace elements in human health: a physician's view." **224**.

Su, J. M. F., L. Perlaky, X. N. Li, H. C. E. Leung, B. Antalffy, D. Armstrong and C. C. Lau (2004). "Comparison of ethanol versus formalin fixation on preservation of histology and RNA in laser capture microdissected brain tissues." *Brain pathology* **14**(2): 175-182.

Sung, H., J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray (2021). "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." *CA: a cancer journal for clinicians* **71**(3): 209-249.

Suzuki, S., Y. Okada and S. Hirai (2004). "Determination of trace elements in certified reference material of human urine by instrumental neutron activation analysis." *Bunseki Kagaku (Japan Analyst)* **53**(2): 109-112.

Szalóki, I., A. Gerényi, F. Fodor, G. Radócz, V. Czech and L. Vincze (2021). "Improved Micro-X-ray Fluorescence Confocal Imaging of Two-Dimensional Distribution of Arsenic Concentration in Cucumber Hypocotyls Using Synchrotron Radiation." *Analytical Chemistry* **93**(34): 11660-11668.

Szoboszlai, N., Z. Polgári, V. G. Mihucz and G. Záray (2009). "Recent trends in total reflection X-ray fluorescence spectrometry for biological applications." *Analytica chimica acta* **633**(1): 1-18.

Tan, C. and H. Chen (2011). "Screening of prostate cancer by analyzing trace elements in hair and chemometrics." *Biological trace element research* **144**(1-3): 97-108.

Tapia, L., M. Suazo, C. Hödar, V. Cambiazo and M. González (2003). "Copper exposure modifies the content and distribution of trace metals in mammalian cultured cells." *Biometals* **16**(1): 169-174.

Tariq, S. R., A. Ejaz, T. Mahmud and A. R. Tariq (2016). "Distributive Variability of Selected Trace Elements in the Blood Samples of Leukemia Patients." *Journal of Heavy Metal Toxicity and Diseases*.

Teebor, G. W., R. J. Boorstein and J. Cadet (1988). "The reparability of oxidative free radical mediated damage to DNA: a review." *International journal of radiation biology* **54**(2): 131-150.

Terzano, R., M. A. Denecke, G. Falkenberg, B. Miller, D. Paterson, K. J. P. Janssens and A. Chemistry (2019). "Recent advances in analysis of trace elements in environmental samples by X-ray based techniques (IUPAC Technical Report)." **91**(6): 1029-1063.

Thissen, U., B. Üstün, W. J. Melssen and L. M. Buydens (2004). "Multivariate calibration with least-squares support vector machines." *Analytical chemistry* **76**(11): 3099-3105.

Toplak, M., G. Birarda, S. Read, C. Sandt, S. Rosendahl, L. Vaccari, J. Demšar and F. Borondics (2017). "Infrared orange: connecting hyperspectral data with machine learning." *Synchrotron Radiation News* **30**(4): 40-45.

Van Espen, P. (2002). "Handbook of X-ray Spectrometry 2nd Edition, Revised and Expanded." *Ch4. Spectrum Evaluation, Marcell Dekker*: 249-252.

- Van Grieken, R. and A. Markowicz (2001). *Handbook of X-ray Spectrometry*, CRC press.
- Wagner, K.-H. and H. Brath (2012). "A global view on the development of non communicable diseases." *Preventive medicine* **54**: S38-S41.
- Wei, J., N. Ni, L. Zhang and Y. Gao (2018). "Early candidate biomarkers in urine of Walker-256 lung metastasis rat model." *bioRxiv*: 306050.
- Wilson, D. R. and T. R. Martinez (2000). "Reduction techniques for instance-based learning algorithms." *Machine learning* **38**(3): 257-286.
- Wróbel, P. M., Ł. Chmura, M. M. Grzelak, Z. Stęgowski, M. Lankosz, D. Adamek, R. Jach, A. Migliori and A. G. Karydas (2019). "Towards histopathological analysis based on X-ray fluorescence elemental imaging supported by multivariate analysis-Case study of ovarian cancers." *Spectrochimica Acta Part B: Atomic Spectroscopy* **155**: 4-11.
- Xu, L., L. Wencong, J. Shengli, L. Yawei and C. Nianyi (2006). "Support vector regression applied to materials optimization of sialon ceramics." *Chemometrics and intelligent laboratory systems* **82**(1-2): 8-14.
- Xu, Y., S. Zomer and R. G. Brereton (2006). "Support vector machines: a recent method for classification in chemometrics." *Critical Reviews in Analytical Chemistry* **36**(3-4): 177-188.
- Xue, H., R. Qiao, L. Yan, S. Yang, Y. Liang, Y. Liu, Q. Xie, L. Cui and B. Cao (2021). "The Correlation Between Potential "Anti-Cancer" Trace Elements and the Risk of Breast Cancer: A Case-Control Study in a Chinese Population." *Frontiers in oncology* **11**.
- Xue, H., R. Qiao, L. Yan, S. Yang, Y. Liang, Y. Liu, Q. Xie, L. Cui and B. Cao (2021). "The Correlation Between Potential "Anti-Cancer" Trace Elements and the Risk of Breast Cancer: A Case-Control Study in a Chinese Population." *Frontiers in oncology*: 3086.
- Yaman, M., D. Atici, S. Bakirdere and I. Akdeniz (2005). "Comparison of trace metal concentrations in malign and benign human prostate." *Journal of medicinal chemistry* **48**(2): 630-634.
- Yap, C., R. Ayala and P. Wobrauschek (1988). "Quantitative trace element determination in thin samples by total reflection x-ray fluorescence using the scattered radiation method." *X-Ray Spectrometry* **17**(5): 171-174.
- Yin, L. I., J. I. Trombka and S. M. Seltzer (1989). "A pattern recognition approach in X-ray fluorescence analysis." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **277**(2-3): 619-626.
- Yoshinaga, J., A. Chatterjee, Y. Shibata, M. Morita and J. S. Edmonds (2000). "Human urine certified reference material for arsenic speciation." *Clinical chemistry* **46**(11): 1781-1786.
- Zabłocka-Słowińska, K., S. Płaczkowska, A. Prescha, K. Pawełczyk, I. Porębska, M. Kosacka, L. Pawlik-Sobecka and H. Grajeta (2018). "Serum and whole blood Zn, Cu and Mn

profiles and their relation to redox status in lung cancer patients." *Journal of Trace Elements in Medicine and Biology* **45**: 78-84.

Zaichick, S. and V. Zaichick (2011). "The Br, Fe, Rb, Sr, and Zn contents and interrelation in intact and morphologic normal prostate tissue of adult men investigated by energy-dispersive X-ray fluorescent analysis." *X-Ray Spectrometry* **40**(6): 464-469.

Zaichick, V. and S. Zaichick (2016). "Prostatic Tissue Levels of 43 Trace Elements in Patients with Prostate Adenocarcinoma." *Cancer and Clinical Oncology* **5**(1): 79.

Zaichick, V. Y., T. Sviridova and S. Zaichick (1996). "Zinc concentration in human prostatic fluid: normal, chronic prostatitis, adenoma and cancer." *International urology and nephrology* **28**(5): 687-694.

Zaichick, V. Y., T. Sviridova and S. Zaichick (1997). "Zinc in the human prostate gland: normal, hyperplastic and cancerous." *International urology and nephrology* **29**(5): 565-574.

Zhang, P., C. A. Georgiou and V. Brusic (2017). "Elemental metabolomics." *Briefings in Bioinformatics*: bbw131.

Zhang, R., L. Li, Y. Sultanbawa and Z. P. Xu (2018). "X-ray fluorescence imaging of metals and metalloids in biological systems." *American journal of nuclear medicine and molecular imaging* **8**(3): 169.

APPENDICES

Appendix I: Informed consent explanation

Introduction (Researcher statement)

My name is Okonda Justus, a PhD (Physics) post graduate student in Department of Physics at the University of Nairobi. I would like to introduce to you a research study that I am conducting; with the aim of giving you relevant information that may help you make an informed decision on whether or not you are willing to voluntarily participate in the study.

Research title: Diagnostic Analysis and Modeling of Trace Biometals and Speciation by Machine Learning Based X-Ray Fluorescence and Scatter in selected Adult Genital-Urinary Tract Cancers

Brief information on cancer

Cancer as a major non-communicable disease is among the leading cause of death in Kenya after infectious and cardiovascular diseases. Genital urinary cancer cases have been on the rise in Kenya with resultant loss of lives and resources during treatment.

Purpose of the study

The study is aimed to help in the detection of genital urinary tract cancers at the local stage stage of development to enhance effective mode of treatment.

Shipment of the samples outside Kenya

Analytical techniques outside Kenya will need shipment of the samples to external laboratory for analysis

Benefits

The study will contribute in giving a complementary method of cancer diagnosis, hence provide guidance in developing interventions for the prevention and control of the known non-communicable infections associated with cancer. The results obtained from his study will help policy makers in making informed decisions best suited in the management of cancer.

Risks

The sample collection will not require any invasive procedure, it is painless. Therefore there is no risk of any nature whatsoever.

Voluntarism

Participants in this study shall volunteer without any coercion. You may wish to decline participate or to respond to any question in the questionnaire that you are not comfortable

with or even terminate the interview at will without any condition whatsoever. You may also withdraw from the study at any time you wish should you change your mind about participating without any loss of health care benefits to which you are entitled in the hospital.

Do you have any question concerning the above explained information?

Yes

No

Are you willing to participate in this study?

Yes

No

If yes sign in below

Consent by the patient

I (Mr/Mrs/Miss/Dr/Prof).....consent number
..... has read or been explained the information provided above. All my questions have been addressed. I do hereby give my informed consent to be part of the study fully aware of the benefits. I am aware that I can withdraw from this study without loss of any benefits of quality of clinical services and care to which I am entitled in this hospital. I understand that the results of these tests shall be used for research work only and be kept confidential.

Participants **Signature** **/Thumb**
print.....**Date**.....

Doctor /Nurse
(Witness)..... **Signature****Date**

Principals' investigator
(Witness)..... **Signature****Date**

Contact information

For any question please contact the Principal Investigator on;

Researcher contact

Okonda J. Justus

Cell Phone 0724 569 590

OR

UON contacts

Department of Human Pathology

Tel. +254-2-7263000 ext 43769,

+254-2-2725102

Supervisor:

Prof. Emily Rogena - Cell Phone 0721674647

Chairman of

(KNH/UON/ERC) – Telephone No. 2726306-9. Ext. 44102

Appendix II: Urine collection and infection control procedure

1. Once the participant has accepted to participate and is eligible for the study, he/she will be given two urine specimen containers (one to be used in the evening and the other in the early morning)
2. He or she will wash and dry his hands before getting into the toilet.
4. He or she will then collect the first void urine specimen in clearly labeled polypropylene bottles and cock and clearly labeled without any contamination.
5. After 24 hours (next day), he will again collect the second void of urine in polypropylene bottles by using the instructions provided and cock the bottles to avoid contamination.

Study questionnaire

All consenting participants will be required to fill the questionnaire before sample collection
Study number Date.....

Social demographic Characteristics;

Age.....years

Kindly tick one of the appropriate choices given

Marital status.....

Single Married Divorced Widowed

Level of education

None Primary Secondary College

Employment

Employed Unemployed Self employed

- You are healthy.
- You have prostate cancer.
- You have cervical cancer.
- You have urinary bladder cancer.

Medical history

Are you on medication? Yes No

Kwa Kiswahili;

Kiambatisho 1

Utangulizi: Maelezo ya Mtafiti

Jina langu ni Okonda J. Justus. Mimi ni mwanafunzi wa Shahada ya Uzamifu katika Chuo Kikuu cha Nairobi. Ningependa kukupa maelezo ya kimsingi kuhusu kiini cha utafiti wangu ninaoufanya. Maelezo haya ndio yatakayoweza kukufanya uamue iwapo utaweza kukubali kujitolea kwa hiari kushiriki katika utafiti huu au utakataa kushiriki.

Research Title: Diagnostic Analysis and Modeling of Trace Biometal Speciation by Machine Based X-Ray Fluorescence and Scattering in Selected Adult Genital Urinary Tract Cancer
(Mada ya utafiti)

Maelezo Kwa Ufupi Kuhusu Kansa

Kansa kama ugonjwa usioambukizwa na ni nambari tatu katika kusababisha vifo miongoni mwa mataifa

ambayo uchumi wake ni wa chini au wastani. Maradhi yanaongoza ni yale yanayoambukizwa na ya moyo. Nchini Kenya kuna ongezeko la visa vya wale wanaougua Kansa ya viungo vya uzazi. Swala hili linasababisha kutumika kwa rasilimali nyingi kugharamia matibabu ya waathiriwa na hata kusababisha vifo.

Lengo La Utafiti (*Purpose of Research*)

Utafiti huu unalenga kuwaidia katika kugundua maambukizi ya kansa katika hatua zake za mwanzo ili kushauri namna ya kupata matibabu mwafaka.

Manufaa (*Benefits*)

Utafiti huu utachangia katika kupata mbinu kamilishani za kubainisha ugonjwa wa kansa. Haya yatatuongoza katika kuzindua mbinu za kutumia kukabili na kuthibiti magonjwa ambayo hayaambukizani na yanayohusishwa na ugonjwa wa kansa. Matokeo ya utafiti huu yataweza kusaidia waunda sera kufanya maamuzi yafaayo ambayo yatathibiti ugonjwa wa kansa.

Mashaka (*Risks*)

Utaratibu wa kukusanya sampuli ya kufanyia utafiti hautahitaji mtafiti kuhisi uchungu wowote unaoweza kusababisha kuugua maradhi yoyote. Kwa hivyo utafiti huu hautaweza kusababisha mashaka/hatari ya aina yoyote ile.

Kujitolea kwa hiari (*Voluntarism*)

Washiriki katika utafiti huu watakuwa wa kujitolea bila ya kushurutishwa. Unaweza kukosa kushiriki katika utafiti huu, kukosa kujibu swali lolote lile iwapo huridhishwi nalo au hata ukasitisha mahojiano haya wakati wowote ule bila ya masharti yoyote. Unaweza kujitoa

katika utafiti huu wakati wowote ule ukibadili msimamo wako kuhusu ushiriki wako bila ya kupoteza manufaa ya kiafya ambayo umekuwa ukipata katika hospitali hii

Una swali lolote kuhusu maelezo haya ambayo umepewa?

Ndio

La

Je, umejitolea kushiriki katika utafiti huu?

Kama ndio tia sahihi ukifwata maelezo yafuatayo:

Kiambatisho II:

Hojaji ya kujitolea (*Consent Form*)

Kujitolea kwa mgonjwa (*Consent by the patient*)

Mimi

Bw/Bi/Dr./Prof.....Nambari.....

.....

Nimesoma/nimesomewa/nimeelezwa na mwezangu ujumbe ulio katika kiambitisho, nimepata fursa ya kuuliza maswali kwa namna ambayo imeweza kuniridhisha. Ninajitolea kushiriki katika utafiti huu nikijua kuwa ninaweza kujiondoa bila ya kupoteza manufaa ambayo nimekuwa nikipokea kwa kuwa mgonjwa katika hospitali hii. Ninafahamu na kuelewa kuwa matokeo ya utafiti huu yatakuwa siri na yatatumiwa tu kwa minajili ya lengo la utafiti.

Sahihi ya Mshiriki/Kidole gumba

Tarehe.....

Sahihi ya Daktari/Muuguzi.....Tarehe

(*anayeshuhudia*)

Sahihi ya Mtafiti..... Tarehe.....

Mawasiliano kwa Mtafiti

Iwapo una swali lolote kuhusiana na utafiti huu tafadhali wasiliana na mtafiti kupitia;

Okonda J. Justus

Nambari ya simu **0724569590**

Au

Chuo Kikuu cha Nairobi

Kitengo cha Magonjwa ya Binadamu

Nambari +254-2-7263-000-43769

+254 -2- 272-5102

Ama

Msimamizi

Prof. Emily Rogena

0721 674 647

Mwenyekiti wa Idara ya **KNH/UON/ERC**

2726306-44102

Maswali ya Utafiti (Study Questionnaires)

Washiriki wote ambao wamejitolea kwa hiari kushiriki katika utafiti huu wanahitajika kukusanya data kutoka kwako

Nambari ya utafiti.....

Tarehe

.....

Sifa za kidemokrafiki za kijamii

Umri (miaka)

Tafadhali weka alama mwafaka panapohitajika/panapofaa

Hadhi ya ndoa

Mseja (*single*)

Nimeoa (oleka)

Nimetaliki

Nimefiwa

Kiwango cha masomo

Sina Kisomo

Shule ya Msingi

Shule ya upili

Vyuo

Ajira

Nimeajiriwa

Sina Kazi

Nimejiajiri

Historia ya Matibabu

Uko na kansa

Naugua kansa ya tezi kibofu

Sina Kansa

Naugua kansa ya kizazi

Naugua kansa ya kibofu cha mkojo

Kiambatisho III

Utaratibu wa kukusanya mkojo na kuzuia maambukizi ya maradhi

1. Mshiriki atakayekubali kushiriki katika utafiti huu na ametimiza masharti atapewa chupa mbili za kukusanyia kielelezo cha mkojo. Chupa ya kwanza aitumie jioni na ya pili aitumie asubuhi na mapema.
2. Mshiriki aweze kunawa mikono kabla ya kwenda choo kukusanya kielelezo cha mkojo.
3. Mkojo uweze kukusanywa katika chupa safi angavu.
4. Baada ya saa ishirini na nne, mshiriki aweze kutoa kielelezo cha mkojo wa pili ambao utatiwa katika chupa ya pili huku maagizo yaliyowekwa yakizingatiwa. Chupa hii ifunikwe ili kuzuia sampuli kuchafuka.

Standard operating procedure (SOP) for retrieved tissue blocks handling and processing

1. The retrieved blocks shall be sectioned into 5, 10 and 15 microns using microtome and the sections floated in warm water.
2. The sections will then be picked on a slide and placed in a warm oven for 15 minutes so to adhere to the slide.
3. The sections on the slide will then be stained using standard Haematoxylin.
4. Upon drying of the stained tissue sections, X-ray fluorescence of the tissues will be done.

Appendix III: Calibration set design for simulate tissue samples

Analyte	Simulate sample concentration (ppm)								
	S1T ¹	S2T ¹	S3T ¹	S4T ¹	S5T ¹	S6T ¹	S7T ¹	S8T ¹	S9T ¹
<i>Cu</i> ²⁺	2.0	10.0	14.0	20.0	35.0	30.0	27.0	40	50
<i>Zn</i> ²⁺	25	300	200	450	10	80	30	95	70
<i>Fe</i> ³⁺	32	24	65	80	180	200	300	350	420
<i>Mn</i> ⁴⁺	9.6	1	100	5	1.5	50	3	1.7	13
<i>Cr</i> ³⁺	0.5	15	10	0.6	1.5	0.8	0.3	1	2
<i>Se</i> ⁴⁺	0.5	1	2	1.6	5	10	3.5	7.5	9

Analyte	Simulate sample concentration (ppm)					
	S10T ¹	S11T ¹	S12T ¹	S13T ¹	S14T ¹	S15T ¹
<i>Cu</i> ⁺	100	120	150	5	350	200
<i>Zn</i> ²⁺	105	150	55	250	40	130
<i>Fe</i> ²⁺	250	5	400	600	750	800
<i>Mn</i> ²⁺	100	90	2.5	4.5	7	3.5
<i>Cr</i> ³⁺	25	20	38	60	41	50
<i>Se</i> ⁴⁺	0.7	1.2	15	30	20	12

Sample size; 15

Appendix IV: Calibration set design for simulate urine samples

Analyte	Simulate sample concentration (ppm)								
	S1U ¹	S2U ¹	S3U ¹	S4U ¹	S5U ¹	S6U ¹	S7U ¹	S8U ¹	S9U ¹
<i>Cu</i> ⁺	0.5	2	1	2.5	0.8	5	0.7	10	7
<i>Zn</i> ²⁺	11	70	100	10	90	200	300	500	32
<i>Fe</i> ²⁺	2	1.5	5	20	1.8	2.3	3.3	2.5	30
<i>Mn</i> ²⁺	2	5	1	0.5	3	2.4	0.8	15	20
<i>Cr</i> ³⁺	2	3	7.5	4.7	2.5	6.0	5	8	10
<i>Se</i> ⁴⁺	1.8	2.7	3.5	3	5	1	2	6	1.2

Analyte	Simulate sample concentration (ppm)					
	S10U ¹	S11U ¹	S12U ¹	S13U ¹	S14U ¹	S15U ¹
<i>Cu</i> ⁺	15	20	3	5.5	12	25
<i>Zn</i> ²⁺	600	40	30	15	9	5
<i>Fe</i> ²⁺	12	45	80	35	100	60
<i>Mn</i> ²⁺	1.5	3.5	2.5	1.8	4.5	25
<i>Cr</i> ³⁺	15	12	1.5	3.5	20	30
<i>Se</i> ⁴⁺	2.5	10	15	35	18	20

Sample size; 15

Appendix V: TXRF analysis of concentration data of 20µl urine of prostate sample

	Cr	Mn	Fe	Cu	Zn	Se
P_07	0.050	0.040	0.144	0.054	0.442	0.020
	0.009	0.004	0.013	0.007	0.010	0.004
P_11	0.035	0.031	0.089	0.057	0.228	0.027
	0.003	0.004	0.008	0.005	0.006	0.003
P_17	0.032	0.030	0.201	0.045	0.322	0.021
	0.009	0.003	0.008	0.004	0.006	0.002
P_18	0.074	0.098	0.299	0.136	2.470	0.054
	0.014	0.012	0.012	0.007	0.018	0.004
P_20	0.065	0.184	0.179	0.072	0.486	0.046
	0.003	0.008	0.014	0.007	0.013	0.003
P_25	0.022	0.023	0.091	0.027	0.085	0.009
	0.002	0.002	0.005	0.002	0.002	0.001
P_30	0.000	<0.024	0.137	0.060	0.551	0.021
	0.000	0.003	0.006	0.005	0.006	0.003

Appendix VI: EDXRF concentrations of trace biometals in normal and cancerous urine samples

Healthy	Mn	Fe	Cu	Zn	Se	Cancerous	Mn	Fe	Cu	Zn	Se
NP.05	ND	12.6	8.31	6.06	ND	P_O6	ND	17.2	11.2	6.74	ND
NP.06	ND	11.4	6.35	4.05	ND	P_O8	ND	9.68	6.83	5.84	ND
NP.08	ND	12	6.02	5.28	ND	P_10	ND	9.87	6.26	3.42	ND
NP.11	ND	8.68	6.59	10.2	ND	P_11	ND	13.2	9.46	6.53	ND
NP.12	ND	14.7	8.32	5.64	ND	P_12	ND	14.4	8.16	6.26	ND
NP.13	ND	14.7	7.64	7.71	ND	P_14	ND	9.34	5.63	3.74	ND
NP.14	ND	11.8	7.31	6.13	ND	P_15	ND	16.6	7.44	4.92	ND
NP.15	ND	12.3	7.09	4.95	ND	P_20	ND	14.4	8.26	6.47	ND
NP.16	ND	13.43	4.86	2.42	ND	P_23	ND	14.2	9.42	8.34	ND
NP.17	ND	9.12	7.03	4.78	ND	P_24	ND	11.6	6.36	5.09	ND
NP.18	ND	8.56	5.75	3.44	ND	P_32	ND	13.3	7.81	4.3	ND
NP.19	ND	15	9.56	9.77	ND	P_36	ND	11.1	5.44	3.41	ND
NP.20	ND	10	6.76	4.54	ND	P_37	ND	13.7	8.44	6.23	ND
NP.22	ND	13.6	8.56	4.46	ND	P_41	ND	15.3	8.86	5.05	ND
NP.23	ND	13.9	8.82	6.02	ND	P_45	ND	18.9	10.7	12.4	ND
						P_15	ND	20.8	6.82	3.91	ND

Appendix VII: TXRF concentration of trace biometals in 10 μ l prostate cancer

	Cr	Mn	Fe	Cu	Zn
P_07	0.255	0.023	0.255	0.108	0.419
	0.047	0.007	0.016	0.009	0.012
P_11	0.054	0.115	0.705	0.072	0.236
	0.005	0.011	0.023	0.005	0.007
P_17	<0.053	<0.043	0.048	0.029	0.672
	0	0	0.009	0.006	0.013
P_18	<0.047	0.058	0.382	0.233	2.67
	0	0.015	0.017	0.008	0.021
P_20	<0.014	<0.012	0.071	0.012	0.261
	0	0	0.005	0.002	0.005
P_25	<0.025	0.022	0.177	0.011	0.204
	0	0.006	0.007	0.003	0.005
P_30	<0.054	0.045	0.111	0.022	0.613
	0	0.01	0.011	0.002	0.012

RHP52UJOB1-

CASH SALE



Business Unit/IGU: BINDERT-
P. O. Box 68241-00200 Nairobi, Tel: 020 491 3910
E-mail: unes@unonbi.ac.ke
Website: www.unes.co.ke
Fb: uneslimited Twitter: uneslimited

M/S JUSTUS JULIUS OKONDA Date 25.08.23

QTY	PARTICULARS	SHS.	CTS.
4	PRINTING & BINDING @ 3,000 each	12,000	00
E&O.E	28988 TOTAL	12,000	00

Stamp: RECEIVED SECTION 25 AUG 2023

GOODS ONCE SOLD WILL NOT BE RE-ACCEPTED
Portfolio: Consultancy & Training | Hospitality | Bookstores
Dental Care | Radiology | Bindery | Funeral Services
ISO 9001:2015 CERTIFIED

0724569590-

UNIVERSITY OF NAIROBI
Towards World Class Excellence

P.O. BOX 30197-00100 NAIROBI TEL-318262
Website: www.uonbi.ac.ke



POSTGRADUATE STUDENT ID CARD

OKONDA JUSTUS JULIUS

Reg. No:180/51695/2017

ID/PP NO:24571014

DOCTOR OF PHILOSOPHY IN PHYSICS

FACULTY OF SCIENCE AND TECHNOLOGY

Holder's Sign

Director Graduate School Sign

Issued : 21-OCT-2022 Expires : 21-OCT-2023

I80/51695/2017



9 7 9 9 9 8 3 7



This card is the property of the University of Nairobi and should be
surrendered on request or expiry.



Contents lists available at ScienceDirect

Spectrochimica Acta Part B: Atomic Spectroscopy

journal homepage: www.elsevier.com/locate/sab

Feasibility for early cancer diagnostics by machine learning enabled synchrotron radiation based micro X-ray fluorescence imaging of trace biometals as cancer biomarkers

J.J. Okonda^{a,*}, H.K. Angeyo^a, A. Dehayem-Kamadjeu^b, A.E. Rogena^c

^a Department of Physics, University of Nairobi, Kenya

^b Alan Alda Center for Communicative Science, Stony Brook University, New York, United States of America

^c Department Human Pathology, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Kenya

ARTICLE INFO

Keywords:

Micro X-ray fluorescence imaging
Machine learning
Trace biometals
Cell culture
Cancer diagnostics
Biomarkers

ABSTRACT

Trace quantitative spectroscopic imaging has the potential to provide the location and distribution of trace biometals in a biological sample for disease (cancer) diagnostics. However, spatial qualitative analysis of the trace biometals for cancer diagnostics remains a challenge due to the complex biological matrices that result in weak analyte signals and intricate multivariate relationships between the analyte spatial distribution and disease (cancer) state. In this study, principal component analysis (PCA)-enabled artificial neural networks (ANN) for micro X-ray fluorescence for simultaneous determination of biometal (Mn, Fe, Cu and Se) spatial profiles as biomarkers for cancer diagnosis in model human cell cultures (DU145 and Vero). The cell lines were cultured on silicon nitride membranes and micro XRF analysis carried out at TwinMic beamline, Elettra synchrotron source at beam excitation energy of 1.7 keV. These enabled 2D mapping of Mn, Fe, Cu and Se using cylindrical beam dimensions of ≈ 690 nm in stepper motor-controlled step sizes of $0.6\ \mu\text{m}$ with dwell times of 10 s per pixel. Python multichannel analyzer (PyMca) software was used for spectral deconvolution and determination of 2D maps of the trace biometals. Principal component analysis (PCA) reduced the data dimensions for optimal artificial neural networks (ANN) exploratory modelling of cancer pathogenesis stages utilizing the pixel spectral profile of the trace biometals (Mn, Fe, Cu, Se). The 2D spatial distribution maps of the trace biometals revealed high spatial correlations between Cu and Fe ((0.941) and (0.923)) in DU_D3 and DU_D4 in cancerous compared to normal cell culture stages (V_D3 and V_D4) at 0.661 and 0.203 respectively. Utilizing PC1 and PC2 scores from selected fluorescence L (L_{α} and L_{β}) lines of Fe and Cu, ANN distinctively classified the cell cultures into cancerous and healthy groups. Further, the selected fluorescence L-lines of Fe, Cu and Compton scatter spectral profiles enabled ANN classification of cancerous cultured cells into early, intermediate and advanced stages. The study has provided proof of concept for early diagnosis of cancer based on the multivariate alterations and spatial distribution of the trace biometals as cancer biomarkers.

1. Introduction

Soft X-rays can penetrate through varying thickness of biomedical samples [1] for qualitative and quantitative analysis of elements. Synchrotron radiation source provides a highly polarizing and stable monochromatic emission of high flux and low divergence [2] beam of soft X-rays valuable for cellular and sub-cellular elemental imaging in complex biological systems [3]. Synchrotron radiation X-ray fluorescence (SR-XRF) further combines the flexibility of variable incident X-ray energy and advanced sample manipulation for the realization of

sample analysis with low detection limits [4] and micro-analysis [5]. The tunability of synchrotron sources further improves the spatial resolution of quantitative imaging [6]. In this regard, synchrotron radiation in microscopy domain has the potential for determination of spatial distributions of trace biometals in individual cells for disease (cancer) diagnostics *via* biochemical mechanisms at reduced spectral background and minimum scatter [7].

The low Z elements in biomedical samples are difficult to detect in fluorescence mode thus leads to low fluorescence yield and absorption contrast. The trace biometals are therefore masked in the high

* Corresponding author at: Department of Physics, The University of Nairobi, P.O. Box 30197, Nairobi, Kenya.

E-mail address: okondajustus@uonbi.ac.ke (J.J. Okonda).

<https://doi.org/10.1016/j.sab.2023.106671>

Received 8 May 2022; Received in revised form 30 March 2023; Accepted 4 April 2023

Available online 7 April 2023

0584-8547/© 2023 Elsevier B.V. All rights reserved.

Table 1

Spectral feature selection of L (L_{α} and L_{β}) lines for multivariate exploratory analysis.

Biometal	Line series	Energy range (keV)
Mn	L	0.625–0.639
Fe	L	0.692–0.712
Cu	L	0.915–0.935
Se	L	1.360–1.431

background and are not easily studied by soft X-rays to provide biochemical levels and distribution [1]. Each cell has characteristic biochemical composition that changes in response to pathological stimuli which impairs the immune system and results in changes in cellular chemical and structural profiles. The essential trace biometals such as Mn, Fe, Cu and Se maintain homeostatic balance and regulate enzymatic reactions [1]; therefore changes in their concentrations as well as their spatial distributions alter the cellular functions [8]. Distribution and alterations of trace biometals such as Mn, Fe, Cu and Se together with their correlative localization in cellular microenvironments can be used as ‘biomarkers’ for early cancer diagnosis [9]. According to 2019 World Health Organization (WHO) statistics, cancer is ranked as among the major cause of death for the population below 70 years in 112 out of 183 countries [10]. Prostate cancer is currently the most commonly diagnosed cancer in men [11] and is globally the sixth most common cancer [12]. Early diagnosis of cancer remains a challenge due to the limitations of conventional diagnostic methods.

SR-XRF enabled analysis of trace element levels in healthy and cancerous serum human breast [13] where significantly high concentrations of Cu, Fe, and Pb but low concentrations of Mn, Ti, Cr and Zn were evident. Synchrotron radiation total reflection X-ray fluorescence (SR-TXRF) determined the concentration of Fe, Zn, Cu and Rb in human prostate cancer tissue [14] where high levels of Fe and Zn were evident. Synchrotron radiation induced X-ray emission (SRIXE) quantified Fe, K, Zn and Cu in normal and cancerous breast tissue where high concentrations of Fe in cancerous tissue were recorded [15]. Synchrotron

radiation micro-XRF enabled spatial distributions of Cu and Fe in human prostate spheroid cells using DU145 (cancerous) and RWPE-1 (normal) cell lines supplemented with zinc chloride [16]. In this study, inhomogeneous distributions of both Cu and Fe were observed in all the spheroids due to their significant regulatory role in carcinogenesis. Turyanskaya et al. (2021) used laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) based micro-XRF spectrometry to analyze both healthy and diseased human bones and cartilage [7] where Ca, P, Zn, Pb and Sr were highly spatially distributed. Further quantitative analysis through matrix-matched reference standards indicated that Ca and P levels were elevated (120 ppm and 80 ppm respectively) in bone and cartilage. The *in vivo* toxicity of zebra fish embryos has been assessed through analysis of bio-distribution of nanoparticles by synchrotron radiation X-ray fluorescence (SR-XRF) imaging [17]. Micro XRF imaging enabled the segmentation of elemental distribution images in stromal and neoplastic ovarian tumours and healthy tissue [18]. In this study, the non-negative matrix factorization and K-means clustering and were used for image segmentation for characterization of stromal, tumour or unclassified areas utilizing elemental (P, Mg and Zn) distribution maps as potential biomarkers for cancer diagnosis.

The above SR-based XRF methods are not only limited by weak spatial and spectral signals but also difficulties in analysis and interpretation of highly-dimensional data for cancer diagnostics. Effective cancer management calls for early detection and diagnosis preferably at the cellular level with complex cellular microenvironments with subtle spatial distribution of the trace biometals. Cancer diagnostics at the local stage of development based on pixel spatial distribution of trace biometals as biomarkers [19] can be achieved by SR-XRF quantitative imaging at sub-micrometer resolution with low detection limits at cellular levels. However, individual pixels may contain biometal contributions from other different sources [20] as well as tissue morphological information hence the need to utilize multivariate techniques for the spectral data analysis.

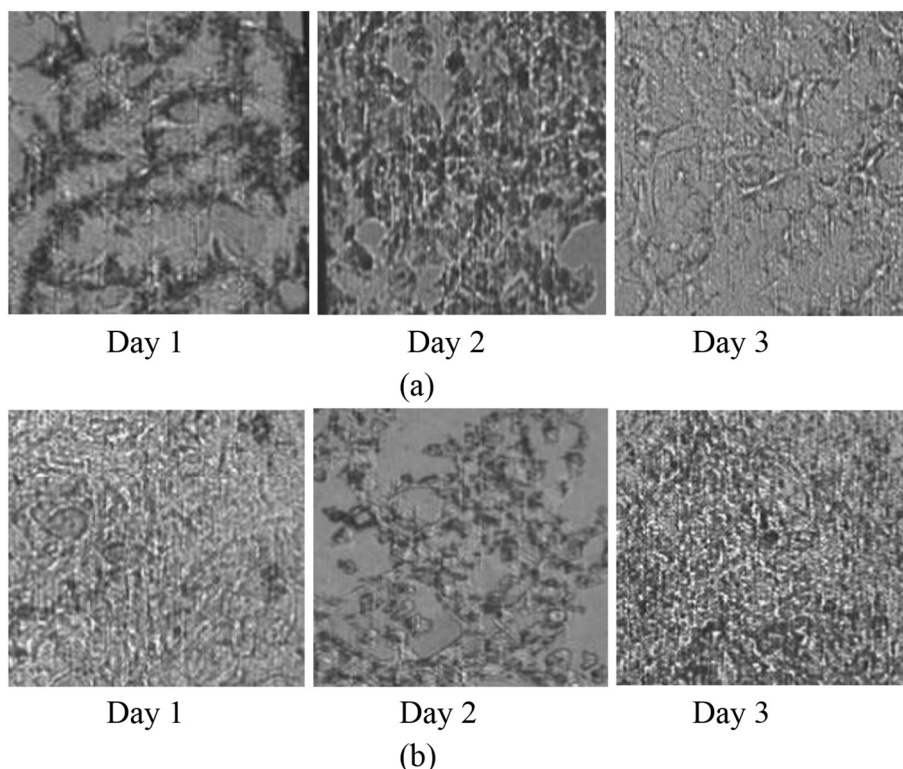


Fig. 1. Visible light microscopy images (10 \times) for (a) cancerous and (b) normal cell cultures.

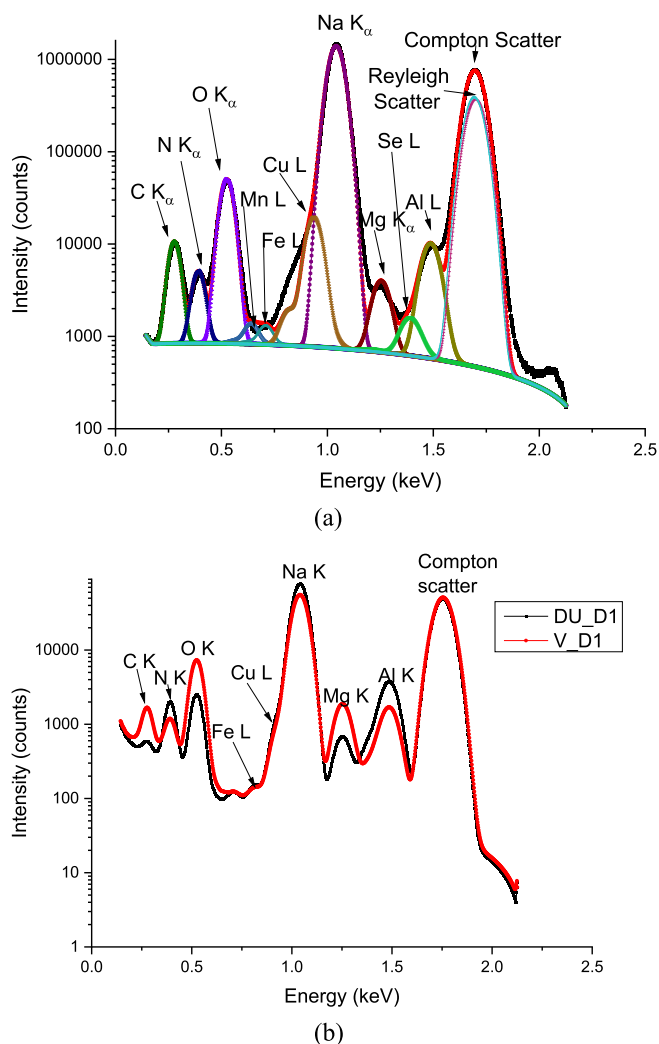


Fig. 2. X-ray fluorescence spectra of (a) DU_D1 and (b) overlapped DU_D1 and V_D1.

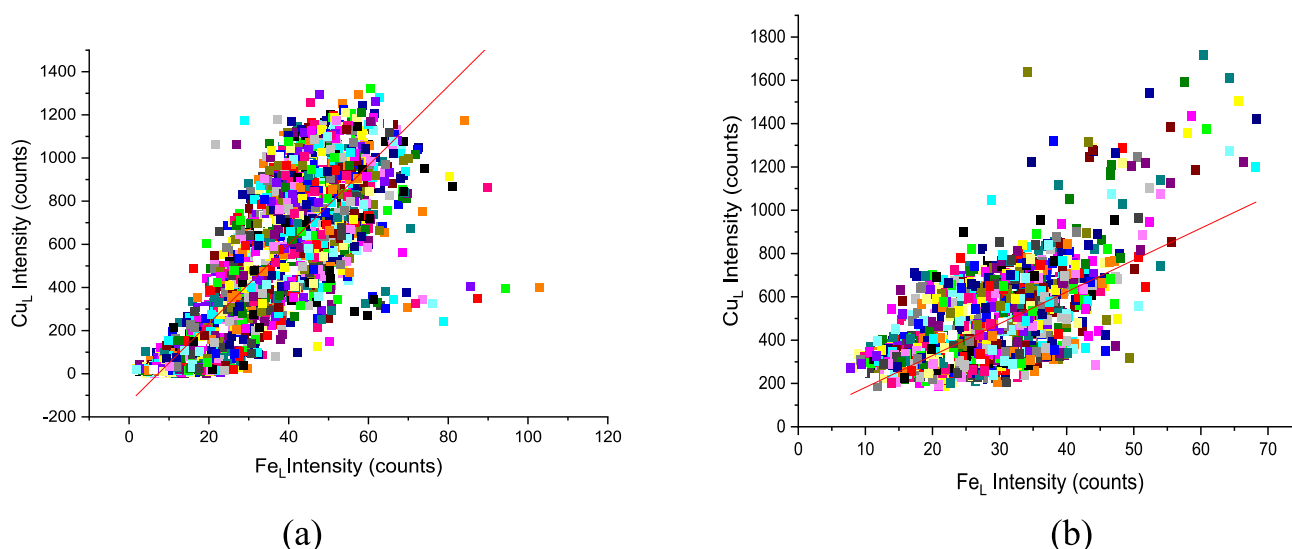


Fig. 3. Correlation of Fe and Cu L line intensities in (a) DU_D2 and (b) V_D2 cell cultures.

1.1. Machine learning

The identity and distribution of elements in the spectral images are limited by the complexity of cell structure and bio environment which complicate trace quantification and diagnostic interpretation of the spectral and spatial profiles. Further, deconvolution of L-lines and multivariate data visualization can be a challenge in differentiating several cell constituents as their spectral characteristics tend to overlap which constitute a multivariate analytical problem requiring the application of novel image data processing techniques. Machine learning (ML) enabled μ -XRF has the potential for cancer diagnosis [21] at the local stage of development. Multivariate Principal Component Analysis (PCA) can reduce the complexity (removes redundant variables) of micro-XRF data sets and retains the most important and original information [22] potential to mine information about the biochemical composition of the samples. PCA can be utilized in reconstructing spatial images back to their original stack to extract latent features [23]. Neural networks (NNs) have the potential to model and train human brain neurons to explore, recognize and characterize complex patterns [24] observed in the PC scores.

In this study, machine learning (ML) enabled μ -XRF was aimed to elucidate the spatial distribution of trace biometals (Mn, Fe, Cu and Se) as biomarkers for cancer diagnostics, which was realized from pixel-based spectral datasets. SR-XRF micro-imaging spectrometry combined with machine learning (ML) has shown the potential for mining trace Mn, Fe, Cu and Se markers towards early cancer diagnosis, based on their quantitative distribution and correlations in cell cultures (DU145 and Vero cell lines).

2. Materials and methods

2.1. Cell culture and sample preparation

Optically transparent silicon nitride (100 nm thick) windows were used as substrate for monolayer cell culturing [25] as they promote robust cell growth [26]. About 1 ml aliquot of prostate cancer cell line (DU_145) and healthy cell line (Vero) suspensions in culture growth media at a density of 1×10^6 cells per ml were seeded on the silicon nitride windows in cell culture tubes. The cells were then cultured in 10% fetal bovine serum and 1% of an antibiotic-antimycotic mixture (1%; Sigma Aldrich) [27] in Dulbecco's modified Eagle's medium (DMEM). Two days (48 h) after seeding, the monolayer formed. About 2 ml of fresh DMEM growth media with 10% fetal bovine serum, about 10

Table 2
Trace biometal correlation coefficients in cancer (DU) and healthy (V) cultured cells.

	Mn	Fe	Cu	Se		Mn	Fe	Cu	Se
DU_D1					V_D1				
Mn	1				Mn	1			
Fe	0.999	1			Fe	1	1		
Cu	0.607	0.605	1		Cu	0.794	0.793	1	
Se	0.844	0.842	0.472	1	Se	0.974	0.973	0.76	1
DU_D2					V_D2				
Mn	1				Mn	1			
Fe	0.67	1			Fe	0.998	1		
Cu	0.831	0.839	1		Cu	0.527	0.521	1	
Se	0.073	0.1	0.079	1	Se	0.899	0.895	0.438	1
DU_D3					V_D3				
Mn	1				Mn	1			
Fe	0.752	1			Fe	0.832	1		
Cu	0.824	0.941	1		Cu	0.890	0.661	1	
Se	0.439	0.528	0.586	1	Se	0.662	0.716	0.739	1
DU_D4					V_D4				
Mn	1				Mn	1			
Fe	0.629	1			Fe	0.912	1		
Cu	0.751	0.923	1		Cu	0.224	0.203	1	
Se	0.293	0.414	0.453	1	Se	0.725	0.679	0.143	1

$p > 0.05$.

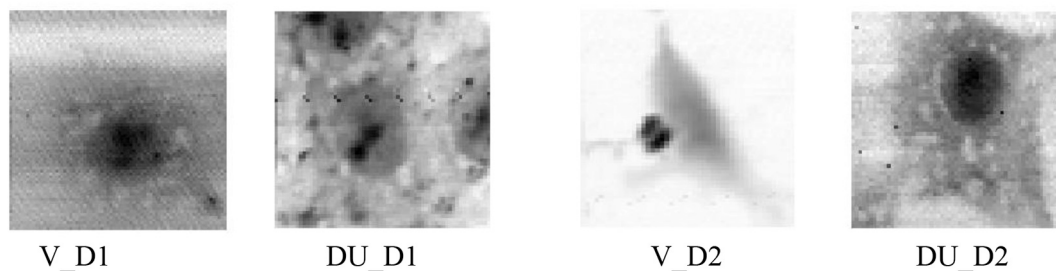


Fig. 4. XRF absorption images for normal (V_D1 and V_D2) and cancerous (DU_D1 and DU_D2) cell cultures.

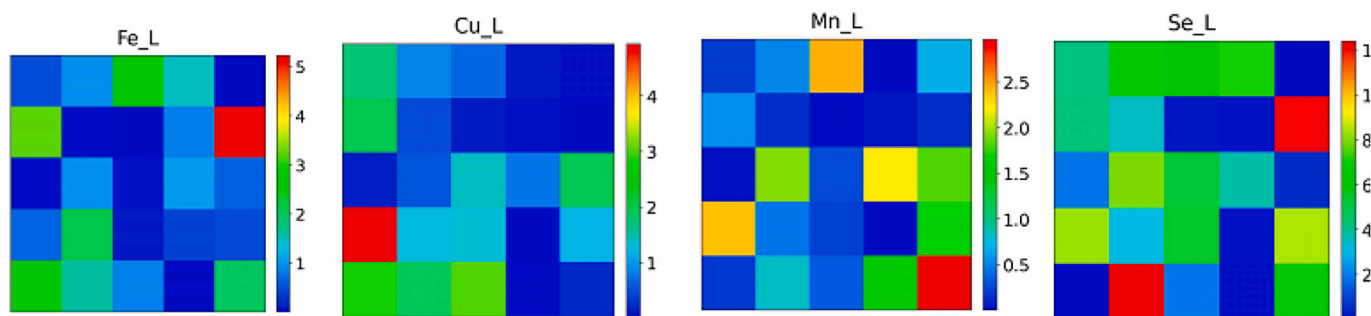


Fig. 5. μ -XRF maps of Mn, Fe, Cu and Se in the culture growth media for day 1.

ml penicillin-streptomycin, 10 mM HEPES and 1 mM sodium hydrogen carbonate [28] were added. The cells were then cultured at 37 °C at the Kenya Medical Research Institute (KEMRI) cell culture laboratory and harvested at intervals of 1 day (24 h) for 4 consecutive days so as to mimic the various stages of normal (healthy) and cancerous cell development. The monolayer cells were harvested in duplicate, fixed in 4% paraformaldehyde (Sigma Aldrich) and cells cleaned with phosphate-buffered saline (PBS) prior to μ -XRF analysis. The cells on silicon nitride membranes were photographed (10 \times) under optical light microscope and mounted on the spectrometer stage.

2.2. Micro-XRF analysis

Low energy μ -XRF analysis of the cell cultures was performed at the TwinMic beamline of the Elettra Synchrotron in Trieste (Italy). The microscope was set in a scan transmission mode with the X-ray beam of spot size 690 nm focused on the sample plane through a zone plate (\varnothing 600 μ m, zone width: 50 nm). Despite the fact that radiation damage due to soft X-ray was less significant [1], it was reduced through minimization of the radiation dose in μ -XRF analysis of the cell cultures. The incident photon energy was set to 1.7 keV with a photon flux of 2.5 \times

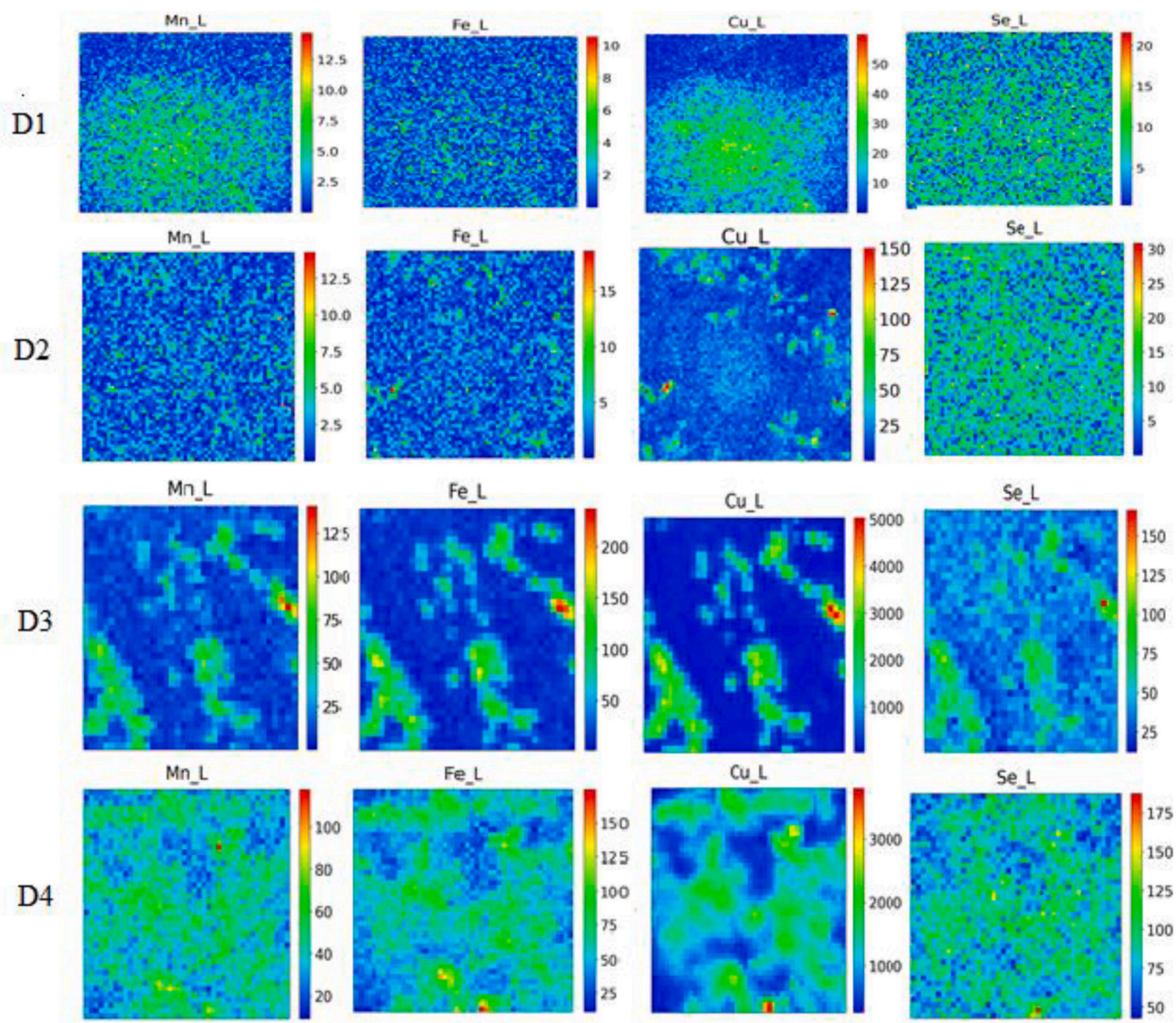


Fig. 6. μ -XRF maps of Mn, Fe, Cu and Se in Vero normal cell cultures.

10^{11} photons/s. The samples were raster-scanned perpendicularly to the incident X-ray beam with a step size of 600 nm where a spatially resolving detector (CCD camera) recorded the transmitted X-ray photons absorption images. The incident beam was set perpendicular to the sample surface facing the symmetrically mounted 8 silicon drift detectors (SDD) with their axes placed at 20° from the sample plane to record the emitted XRF fluorescence from the specimen in an energy range of 180–2200 eV [29]. The irradiation dwell time for the μ -XRF measurements was 10 s per pixel; and each image was acquired for 8 h. Scanning areas of the samples were selected and observed using a microscope and the BL15U saved the XRF spectrum data at each pixel. Elemental distributions were achieved by windowing on the biometal lines of interest in the XRF spectra. The windows can be applied during data collection or during data analysis since the full XRF spectrum is saved for each pixel [30].

2.3. Micro XRF data analysis

Micro-XRF pixel spectra of the trace biometals (Mn, Fe, Cu and Se) were utilized to retrieve 2D maps of their distribution in the cell

cultures. The multispectral image datasets were fitted and deconvoluted into 2D spatial distribution images by PyMca [31] of the trace biometals [32] which enabled the determination of net peak areas of the L (L_α and L_β) lines for the trace biometals of interest. The spatial distribution of L line spectral intensity data was then utilized to explore correlations between the analyzed trace biometals to unravel existing (if any) synergistic or antagonistic biochemical interactions between the biometals. The trace biometals correlation test coefficients were used to investigate the relationship between the spectral distribution of trace biometals in cancerous and normal cell cultures where $P < 0.05$ values were considered statistically significant [33].

The trace biometal distribution and alterations profiles together with their correlative alterations in cellular microenvironments were mined in selected pixel spectral regions as potential 'biomarkers' for diagnosis of cancer. The L-lines pixel spectral regions of interest for trace biometals were selected as shown in Table 1.

2.4. Multivariate analysis of spectral data

Multivariate analysis has the capability to reduce data dimensions of

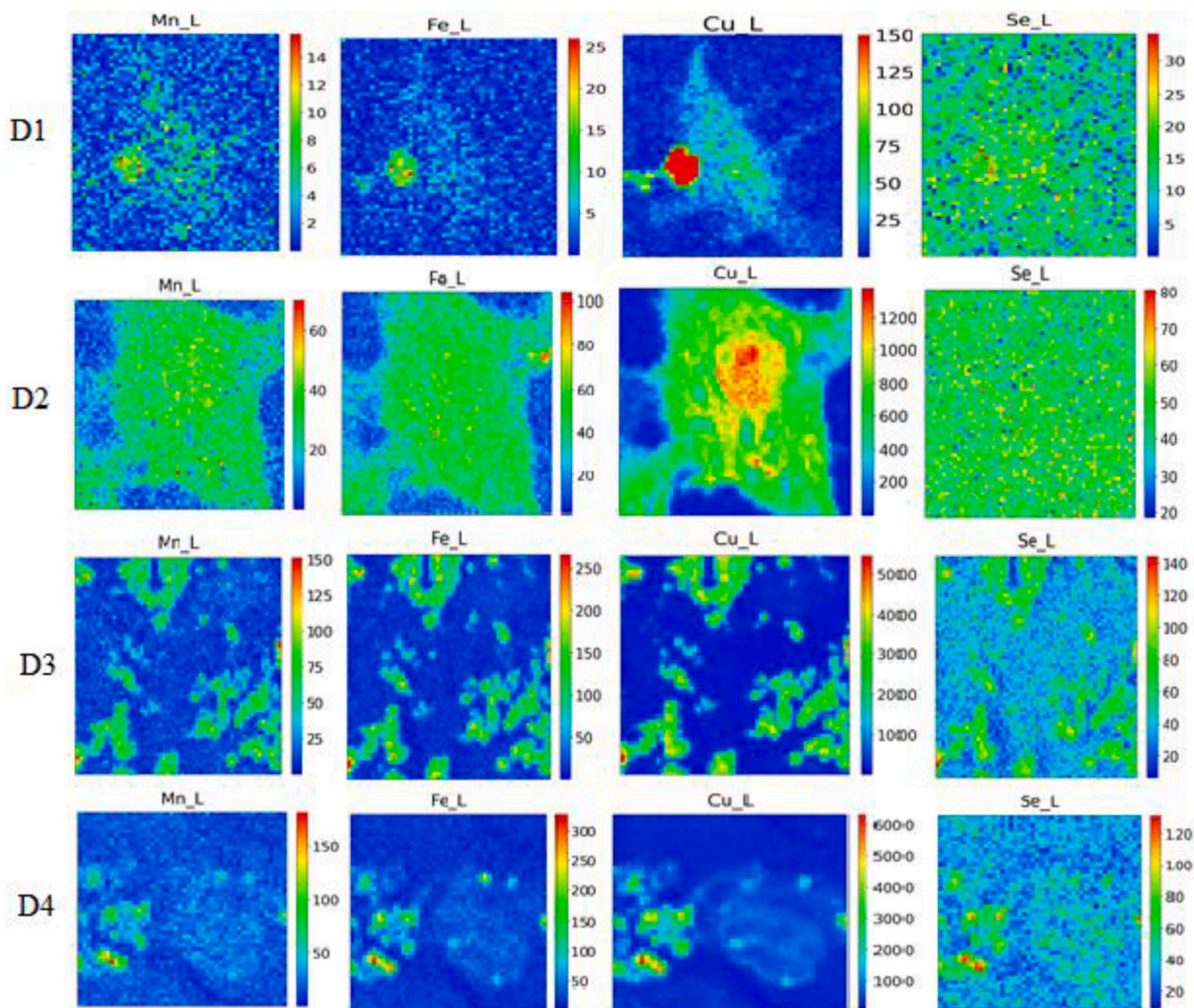


Fig. 7. μ -XRF maps of Mn, Fe, Cu and Se in DU145 cancer cell cultures for day 1 to 4.

major variations in samples under investigation [19]. Prior to multivariate analysis, the pixel spectral intensities datasets were auto-scaled, vector normalized and the high background subtracted. The PCA and ANN multivariate exploratory analysis were performed using quasar orange and origin software. PCA was used to reduce the dimension of the pixel spatial spectral data [34] which enabled visualization of the relationship between samples characterized by the multiple measured trace biometals spatial pixel intensity variables. PCA thus extracted underlying cancer diagnostic information data due to its ability to reduce the correlated pixel spatial intensity variables [35] by retaining the most relevant information. In the PCA domain, the original data matrix X_c is normally decomposed to a score matrix Y together with the loading matrix T and residuals E by subsequent orthogonal subtraction from the largest variation in spectral data [36] as shown in Eq. (1);

$$X_c = YT^T + E \quad (1)$$

The ANN model consists of processing nodes called neurons for exploratory and multivariate calibration. ANN algorithms receive inputs and process them to obtain an output modelled on the structure and behaviour of neurons in the human brain [24] and can be trained for complex pattern recognition and classification [37]. In ANN supervised

approach, known class labels were used for supervising the learning process of the architecture and iterative flow of data through training using the training sets part of the data to determine the weight values of the network [38]. The input data (matrix) were the scores from principal component analysis (PCA) while the targets were given as the various stages of cancer development/ pathogenesis in the ANN model for exploratory classification. Supervised forward neural networks (FNN) approach [39] with 3 layers; input (3 neurons), hidden (5 neurons) and output (7 neurons) were used to reduce the root mean square (RMS) error and the 3-fold cross-validation used to evaluate the classification model for 200 iterations.

The overall performance of ANN depended on the transfer function and the network architecture [40] as shown in Eq. (2);

$$y = f\left(\sum_i w_i x_i + w_0\right) \quad (2)$$

where y is the output of the neuron, x_i is the input and w_0 is an offset term.

To obtain outputs, an activation function was applied to the inputs connected to the weights by neurons. The Rectified Linear Unit (ReLU)

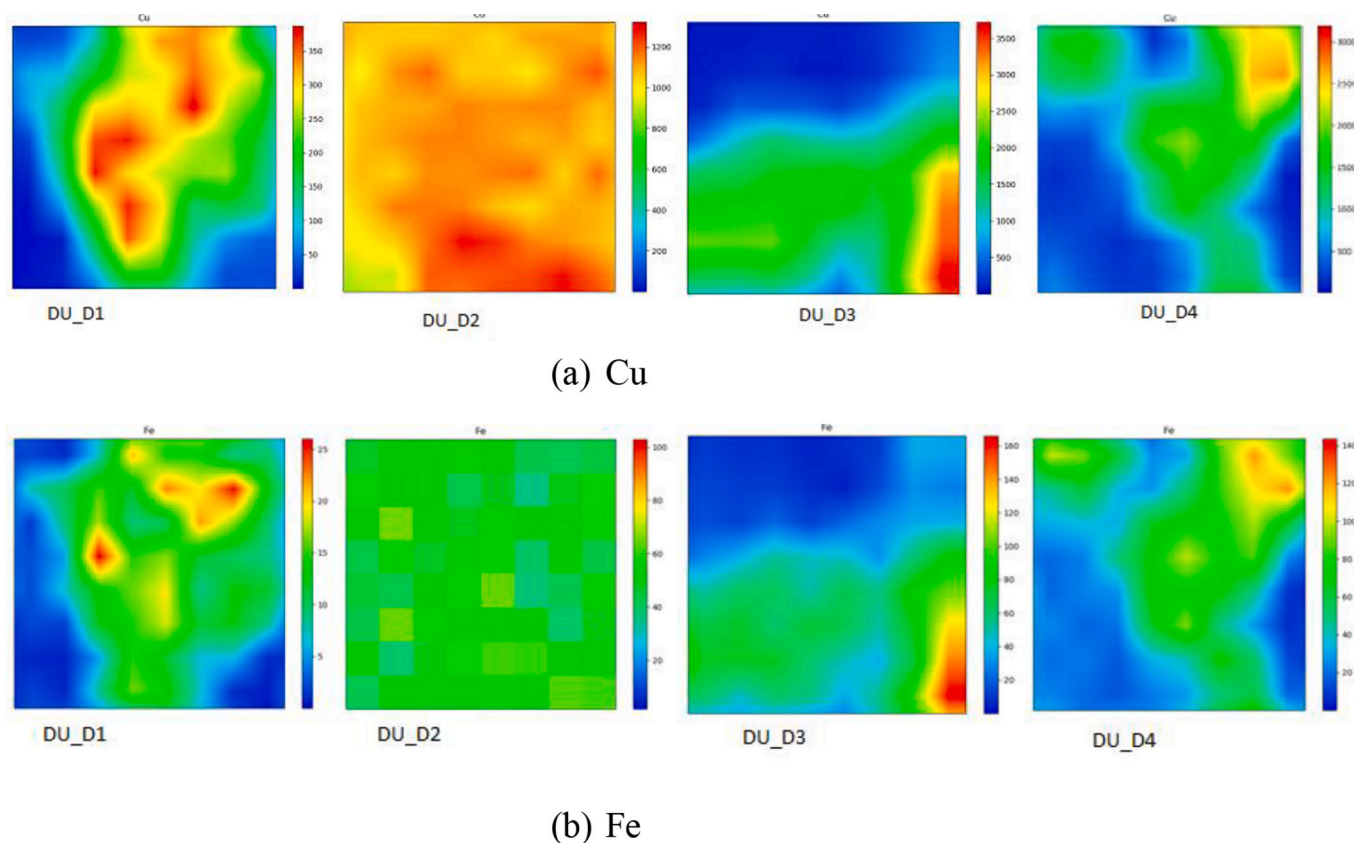


Fig. 8. μ -XRF maps of (a) Cu and (b) Fe in DU145 cancer cell cultures.

activation function was employed for classification where the optimization architecture reduced the complexity of the neural networks thus improving their classification ability. ANNs simulated the structure and behaviour of neurons in the human brain [24] where pattern recognition was achieved through adjusting ANN parameters to minimize the root mean square (RMS) error.

3. Results and discussion

Pathological advancement resulted in chemical and structural changes in the cultured cells that altered the micro XRF spectra generated which can be used as sensitive markers for cancer. The microscopy images of the various stages of cancerous (DU 145) and healthy (Vero) cell cultures are as shown in Fig. 1. The presence of cell residuals (growth media and dead cells residues) is evident and that fixation indeed preserved the biochemical and morphological properties of the cells in both cancerous and healthy samples.

From the above images, the morphological transformations and alterations [41] of DU145 and Vero monolayers could not differentiate cancerous and healthy cells as the cells in colonies that were not distinct.

Fig. 2 shows the typical micro XRF spectrum of DU_D1 acquired on the first day where an enhanced K_{α} line of Na was noted which is due to the presence of sodium bicarbonate in the growth medium or probably the high levels of Na in irradiated areas corresponding to the cell nucleus [42]. The elevated background radiation masked the weak spectral profiles of the L lines of trace biometals of interest (Mn, Fe, Cu, Se).

The low energy region 0–1.25 keV was dominated by the K-shell of low Z elements like Na which overlapped the L-shell lines of analyte trace biometal fluorescence peaks as shown in Fig. 2 (a). Further, the overlapped fitted spectral intensities against the energy of both cancerous (DU_D1) and healthy (V_D1) cell cultures in Fig. 2 (b) were visually difficult to differentiate based on analytes of interest for cancer diagnostics thus the potential utility of multivariate machine learning

tools (PCA and ANN) to mine cancer diagnostic information.

3.1. Correlation between trace biometals

Micro SR-XRF analysis of 2D distribution maps of trace Fe, Cu, Mn and Se in both cancerous and healthy cell cultures were explored for correlation by Spearman's method. The Spearman's correlation scatter plots in Fig. 3 show a comparison of the spatial distribution of Fe and Cu in both cancerous and healthy cell cultures.

The correlations were indicative of trace biometals distributions in cells correlated with the pathological status of the cell lines. The enhanced positive correlation of Fe-Cu in cancerous cell cultures was due to the increase of Fe as a consequence of the increased Cu levels probably bound to the same proteins [33] implying that the concentration of Fe and Cu were probably higher in cancerous compared to the corresponding healthy cultured cells. Table 2 shows the resultant correlations matrices of analyte trace biometal interrelations in both cancerous and healthy cell cultures.

The cancerous and healthy cell cultures show significant ($P < 0.05$) and positive correlations as indicated in the triangular matrices of Table 2. The correlation of Fe-Cu increases from day 1 (D1) to day 4 (D4) in cancerous as compared to healthy cell cultures. The relatively strong positive correlation of Cu-Fe; 0.941 and 0.923 in DU_D3 and DU_D4 respectively compared to corresponding correlations in normal cell cultures (0.661 and 0.203) is probably due to their role in the pathogenesis [43] of cancer. The results further indicate as concluded elsewhere [34] that Mn is strongly positively correlated to other trace biometals in both cancerous and normal cell cultures due to its vital role in the morphological structure of the cell. On the other hand, Se shows strong positive correlations in healthy cell cultures but a weak correlation in cancerous cells for the 10 days probably due to its anti-carcinogenic role [44,45]. For instance, increased concentration of Se in glutathione peroxidases tends to protect the DNA damage by oxygen

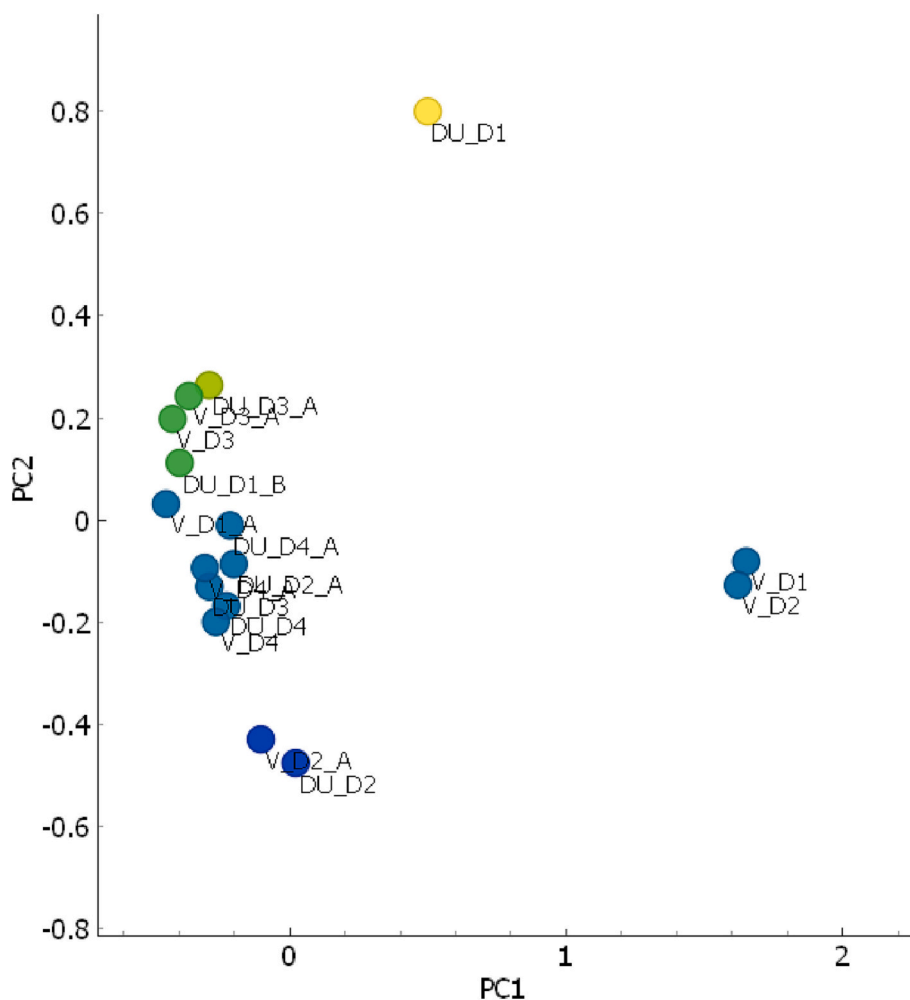


Fig. 9. PCA score plot of cell cultures using whole fluorescence and scatter.

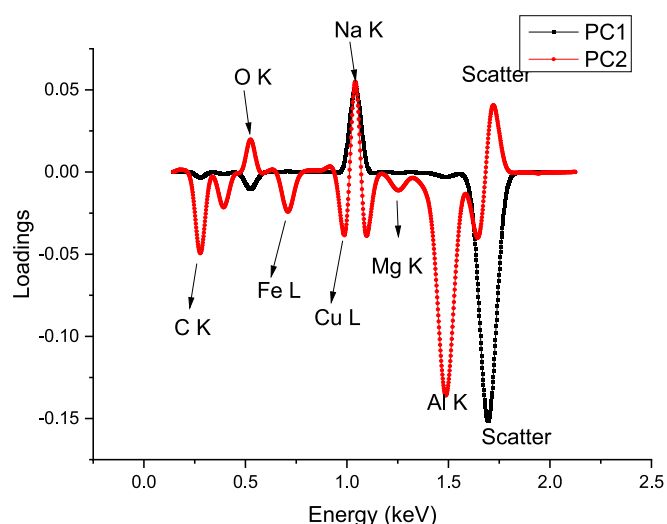


Fig. 10. PC loadings plot of cultured cells by fluorescence and scatter spectral.

radicals [46].

3.2. Distribution maps of the trace biometals

The spatial fluorescence intensity maps were obtained from the

typical illuminated pixel areas as shown in Fig. 4. The cells were visible with strong self-absorption and differential phase contrasting images [42] partially attributed to the dark matrix of light elements (H, C, O and N).

The 2D micro XRF images were acquired in the above absorption microscopy regions with high absorption signals to retain essential information in the low-energy μ -XRF maps [42].

The fluorescence intensity maps provide information on the trace biometal distribution in the cultured cells irradiated by the beam at each scan. Fig. 5 shows 2D distribution maps of Mn, Fe, Cu and Se in the culture growth media.

The sparse spatial distribution of trace biometals in the culture media that was used for culturing both the cancerous and healthy cells was evident. Fig. 6 shows the spatial distributions in the Vero healthy cell cultures in corresponding to their emitted characteristic fluorescence lines from day 1 to 4.

The results show low distribution of trace biometals (Mn, Fe, Cu and Se) in healthy Vero cell culture compared to the cancerous cell cultures in Fig. 7. This can be linked to their functional role in metabolic processes of normal cell growth as a result of their probable co-localization in cancerous cell cultures especially in the perinuclear/cytoplasmic region of the cells. Altered distribution of Mn (12.5–150) and Se (20–175) with a slight increase in the distribution of Fe (150 and 200) and Cu (3000 and 5000) on day 3 (D3) and day 4 (D4) respectively were evident in the healthy cell cultures. The non-homogenous distributions of the trace biometals suggest their active role in angiogenesis [47] for normal cell growth. The sparse distribution of Se was probably due to its

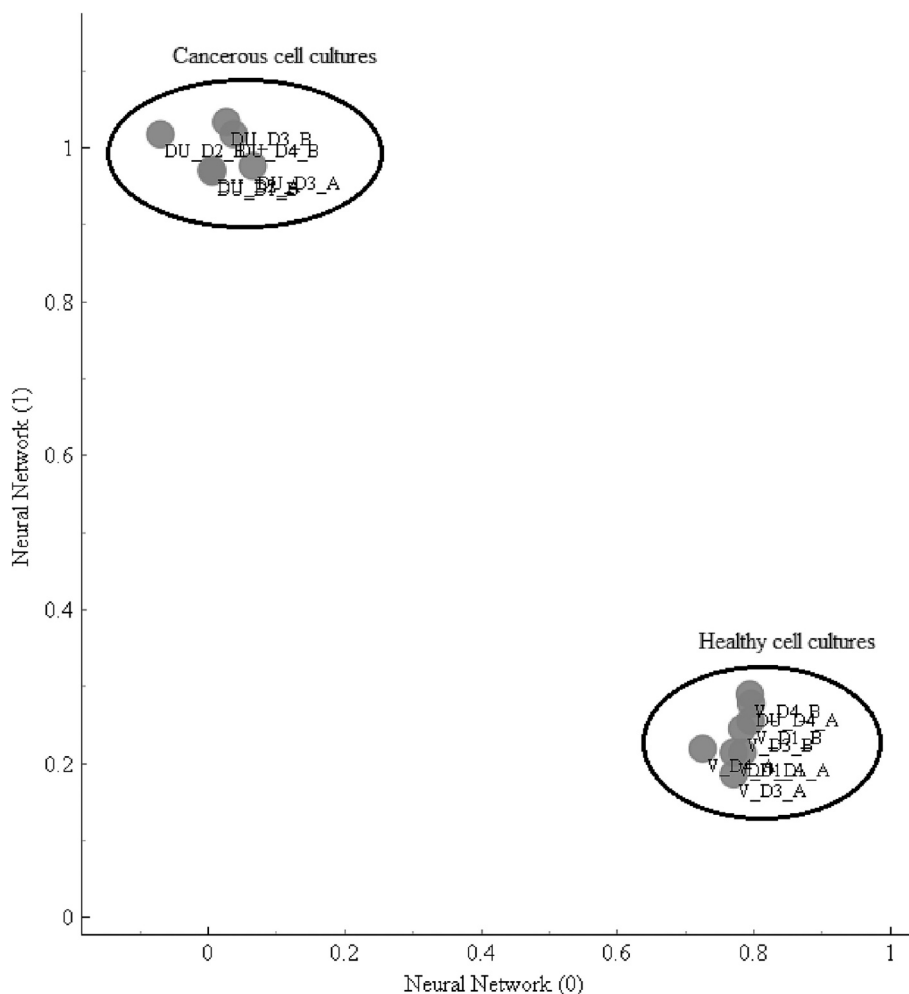


Fig. 11. ANN characterization of cultured cells by PC1 and PC2 of selected μ -XRF fluorescence image data.

essential role in selenoproteins in humans and subsequent promotion of apoptosis [40].

Fig. 7 shows the corresponding 2D trace biometal spatial distributions in cancerous cell cultures. The dense spatial distributions of Fe (25–300) and Cu (150–6000) at different stages of development were evident due to their active role in carcinogenesis.

The low homogeneous distribution of Se and Mn in cancer cells as a result of their suppressed distributions were evident. These can be attributed to the anti-carcinogenic nature of Se involved in reactive oxygen species for antioxidant enzymes [48] (glutathione peroxidase) that inhibits the synthesis of tumour protein cells and DNA [49]. This prevents tumour formation and metastasis through the neutralization of free radicals.

The cellular spatial accumulation of Cu and Fe in cancerous cells harvested at different stages of development are as shown in Fig. 8 (a) and (b) respectively.

The homogenous and elevated distribution of Fe and Cu were as a result of their probable co-localization in cancerous cell cultures especially in the perinuclear [42] region of the cells as compared to the nucleus. Similar major alterations in the distribution of Cu and Fe were observed where Cu was relatively highly distributed in cancerous DU-145 cell as compared to Fe. This is due to high co-localization in cancerous cell cultures especially in the cytoplasmic region of the cells. Cu is also a co-factor of Fe for the generation of reactive oxygen species that are linked to carcinogenesis [19]. The catalytic role of Fe and Cu in the Fenton reaction results in the increased requirement for DNA synthesis [50].

3.3. Multivariate analysis of micro XRF images of the cell cultures

The complex multivariate correlations between trace biometals and their corresponding spectral patterns were discernible by machine learning. Multivariate techniques namely PCA and ANN enabled multivariate exploration of the spectral images to unravel latent patterns and relationships and the biometal concentrations and their spatial distributions. PCA enabled the mining of reduced data dimensions as input data in ANN models to probe the correlational patterns to inform cancer diagnosis.

3.3.1. Principal component analysis (PCA) of the cell culture micro XRF images

PCA decomposed the “pre-processed” datasets into 5 principal components (PCs) that contained spectral variations of each collected pixel spectral data. In this case, PCA reduced the dimensions of spectral images as it exploited the variance to mine maximum cancer diagnostic information. The 2D spectral intensities were filtered by PCA into a new data matrix with the principal components [51].

Fig. 9 shows exploratory analysis of cell cultures using the pixel spectra with an explained variance of 94% (90% and 4% for PC1 and PC2 respectively). The utility of the whole spectra data could not distinctively differentiate cancerous and normal cell cultures and characterize the various stages of cancer development depicted by the different colour codes.

The corresponding loadings plot in Fig. 10 indicated that the fluorescence of major and trace biometals together with Compton scatter

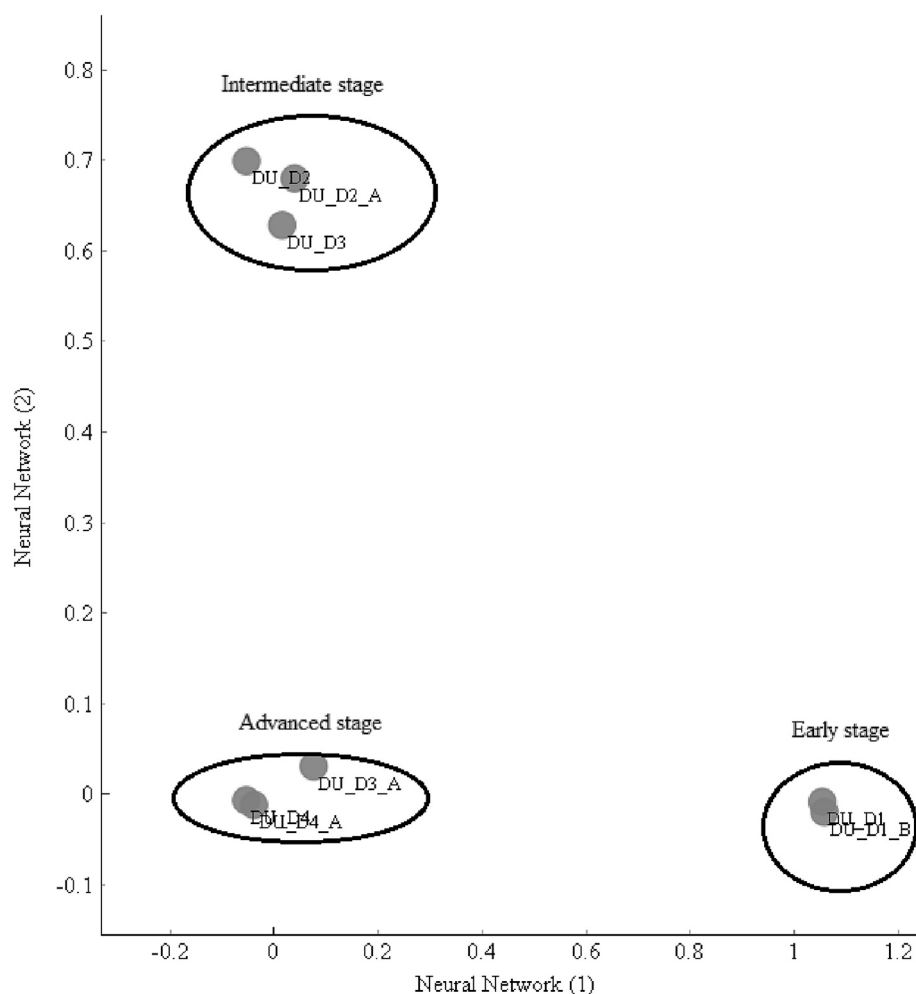


Fig. 12. ANN classification of cultured cancer cells by selected μ -XRF fluorescence and scatter image data.

regions seemingly contain vital information potential for cancer diagnosis thus need for further investigation with advanced multivariate techniques.

The PC1 and PC2 loadings showed how the variance was attributed to Fe and Cu for both positive and negative values. From the above loadings plot, the fluorescence part of the spectra and scatter peaks in this study show potential biomarkers that can be utilized for characterization of cancer into various stages of development.

3.3.2. ANN classification of cell cultures

Supervised ANN model enabled multivariate modelling of the trace biometal distribution profiles for early diagnosis of cancer. The principal components of selected L fluorescence lines of Fe and Cu (PC1 and PC2) at a total explained variance of 92% were used as input data in the ANN model as PCA reduced the background intensity to mine diagnostic information in the selected analytes as indicated in Table 1.

The selected Fe and Cu biomarker fluorescence spectral profiles were partly responsible for the differentiation of the cultured samples into cancerous and normal groups. These enabled the visualization of cancerous and normal cultures as shown in Fig. 11 into two distinct clusters (cancerous and healthy).

The spatial distribution of Cu and Fe in cells are therefore potential biomarkers for cancer diagnostic based on cellular trace metal biomarkers rather than cellular morphology. The biometals act as a catalyst for hydroxyl radical production linked carcinogenesis and angiogenesis processes [52]. Cu also aids in the onset and development of tumours while Fe plays a vital function in oxygen metabolism [16]. Further, Fe is

known to aid in blood supply to the malignant cells thus accelerating angiogenesis [53].

The inclusion of the Compton scatter region in the ANN model improved the characterization performance in the cultured cancerous cells. Three clusters for different stages (early, intermediate and advanced) of cancer development were now clearly visible with ANN input data at an explained variance of 88% (82% and 6% for PC1 and PC2 respectively) as shown in Fig. 12.

The inclusion of Compton scatter to the selected fluorescence further improved the exploratory analysis in XRF 2D mapping. In this regard, ANN classified cultured cancerous cells into 3 stages of cancer severity with no clear distinction between DU_D3 and DU_D4 stages of cancer development. This can be based on the fact that the Compton scatter peak contains further information about to cell structure (morphology) which is significant in the characterization of various stages of cancer development. ANN has therefore indicated potential for prediction and diagnosis of cancer at early stage of development that can supplement the conventional cancer diagnostic techniques in a clinical setup.

4. Conclusions

We have demonstrated a feasible spectroscopic diagnostic method for cancer utilizing low-energy μ -XRF. The SR micro-XRF beam enabled 2D imaging of trace biometals at high spatial resolution. The biometal profiles were detected, mapped and correlated to cancer to illustrate the associations of trace biometal distribution potential for early cancer diagnostics. The distribution, as well as correlations of trace elements

(Mn, Fe, Cu and Se), were altered in response to cancer development and thus were responsible for the characterization of the stages of cancer development. The distribution of Fe was low in the cancer cell cultures but with a corresponding increase in distribution of Cu. The strong correlation of Cu and Fe in cancerous cell culture suggests their vital diagnostic role in cancer development. Weak correlations of Se with other trace bimetallics were evident in advanced stages of cancer cell cultures. Utility of Cu and Fe L-lines selected spatial fluorescence profiles in PCA enabled accurate ANN classification of cultured cells into cancerous and healthy groups. The inclusion of the Compton scatter region to the selected fluorescence further characterized cancer into early, intermediate and advanced stages. The results indicated predictable associations (2D fluorescence distribution profiles) of trace biometallics (Mn, Fe, Cu and Se) in conjunction with Compton Scatter as potentially valuable biomarkers for cancer diagnosis applicable at an early stage of development.

Ethical approval

This research study has been ethically approved (P604/10/2017) by the Kenyatta National Hospital/University of Nairobi Ethics Research Committee.

CRediT authorship contribution statement

J.J. Okonda: Writing – original draft, Methodology, Formal analysis, Writing – review & editing. **H.K. Angeyo:** Writing – original draft, Methodology, Formal analysis, Writing – review & editing. **A. Dehayem-Kamadjeu:** Visualization, Conceptualization, Writing – review & editing. **A.E. Rogena:** Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that there is no conflict of interest in this work.

Data availability

Data will be made available on request.

Acknowledgments

This study was performed at TwinMiC beamline in Elettra - Sincrotrone Trieste (Italy), proposal reference number 20205073. The authors acknowledge Elettra for the beam time access and beam scientific team; Alessandra Gianoncelli, Valentina Bonann and Gianluca Gariani for their support in sample analysis and data analysis during the beam time. The authors further acknowledge the financial support from ICTP-ELETTRA Users Programme and National Research Fund (NRF) Kenya. We also acknowledge the Swedish International Development Cooperation Agency (SIDA) through the International Science Programme (ISP), Uppsala University for additional financial support (IPPS KEN: 04) towards this study. We finally acknowledge the technical staff of Kenya Medical Research Institute (KEMRI) for their support during the cell culture and sample preparation stages for this research work.

References

- J.F. Collingwood, F. Adams, Chemical imaging analysis of the brain with X-ray methods, *Spectrochim. Acta B At. Spectrosc.* 130 (2017) 101–118.
- H.S. Kainth, D. Khandelwal, R. Singh, G. Singh, S. Puri, Role of trace elements in breast cancer and their characterization using X-ray fluorescence techniques, in: *Trace Elements and their Effects on Human Health and Diseases*, IntechOpen, 2020.
- A. Gianoncelli, L. Vaccari, G. Kourousias, D. Cassese, D. Bedolla, S. Kenig, P. Storici, M. Lazzarino, M. Kiskinova, Soft X-ray microscopy radiation damage on fixed cells investigated with synchrotron radiation FTIR microscopy, *Sci. Rep.* 5 (1) (2015) 1–11.
- G. Silversmit, B. Vekemans, F.E. Brenker, S. Schmitz, M. Burghammer, C. Riekel, L. Vincze, X-ray fluorescence nanotomography on cometary matter from comet 81P/Wild2 returned by stardust, *Anal. Chem.* 81 (15) (2009) 6107–6112.
- B. Lai, W. Yun, D. Legnini, Y. Xiao, J. Chrzas, P. Viccaro, V. White, S. Bajjkar, D. Denton, F. Cerrina, Hard x-ray phase zone plate fabricated by lithographic techniques, *Appl. Phys. Lett.* 61 (16) (1992) 1877–1879.
- P.M. Kopittke, T. Punshon, D.J. Paterson, R.V. Tapper, P. Wang, F.P.C. Blamey, A. van der Ent, E.J.P.P. Lombi, Synchrotron-based X-ray fluorescence microscopy as a technique for imaging of elements in plants 178 (2) (2018) 507–523.
- A. Turyanskaya, S. Smetaczek, V. Pichler, M. Rauwolf, L. Pernecky, A. Roschger, P. Roschger, P. Wobruschek, A. Limbeck, C. Strel, Correlation of μ XRF and LA-ICP-MS in the analysis of a human bone-cartilage sample, *J. Anal. At. Spectrom.* 36 (2021) 1512–1523.
- A. Ektesabi, S. Shikine, N. Kitamura, M. Rokkum, C. Johansson, Distribution and chemical states of iron and chromium released from orthopedic implants into human tissues, *X-Ray Spectrom.* Int. J. 30 (1) (2001) 44–48.
- R. Ortega, G. Deves, A. Carmona, Bio-metals imaging and speciation in cells using proton and synchrotron radiation X-ray microspectroscopy, *J. R. Soc. Interface* 6 (suppl 5) (2009) S649–S658.
- H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J. Clin.* 71 (3) (2021) 209–249.
- K. Rocha, R. Leitão, E. Oliveira-Barros, M. Oliveira, C. Canellas, M. Anjos, L. Nasciutti, R. Lopes, Elemental distribution by synchrotron X-ray microfluorescence of prostate 3D cell culture, *X-Ray Spectrom.* 48 (5) (2019) 476–481.
- T. Pischon, H. Boeing, S. Weikert, N. Allen, T. Key, N.F. Johnsen, A. Tjønneland, M. T. Severinsen, K. Overvad, S. Rohrmann, Body size and risk of prostate cancer in the European prospective investigation into cancer and nutrition, *Cancer Epidemiol. Prevent. Biomark.* 17 (11) (2008) 3252–3261.
- B.G. Naidu, P. Sarita, G.N. Raju, M. Tiwari, Multivariate analysis of trace elemental data obtained from blood serum of breast cancer patients using SRXRF, *Res. Phys.* 12 (2019) 673–680.
- R. Leitão, A. Palumbo Jr., P. Souza, G. Pereira, C. Canellas, M. Anjos, L. Nasciutti, R. Lopes, Elemental concentration analysis in prostate tissues using total reflection X-ray fluorescence, *Radiat. Phys. Chem.* 95 (2014) 62–64.
- H. Sky-Peck, Trace metals and neoplasia, *Clin. Physiol. Biochem.* 4 (1) (1986) 99–111.
- K.J. Rocha, R.G. Leitão, E.O. Barros, M.A. Oliveira, C.G. Canellas, M.J. Anjos, L. E. Nasciutti, R.T. Lopes, Microbeam X-ray fluorescence mapping of Cu and Fe in human prostatic carcinoma cell lines using synchrotron radiation, *Brazil. J. Radiat. Sci.* 7 (2A) (2019).
- L.U. Khan, G.H. da Silva, A.M. de Medeiros, Z.U. Khan, M. Gidlund, H.F. Brito, O. Moscoso-Londono, D. Muraca, M. Knobel, C.A. Perez, Fe₃O₄@ SiO₂ nanoparticles concurrently coated with chitosan and GdO₃: Ce³⁺, Tb³⁺ luminophore for bioimaging: toxicity evaluation in the zebrafish model, *ACS Appl. Nano Mat.* 2 (6) (2019) 3414–3425.
- P.M. Wróbel, L. Chmura, M.M. Grzelak, Z. Stegowski, M. Lankosz, D. Adamek, R. Jach, A. Migliori, A.G. Karydas, Towards histopathological analysis based on X-ray fluorescence elemental imaging supported by multivariate analysis-case study of ovarian cancers, *Spectrochim. Acta B At. Spectrosc.* 155 (2019) 4–11.
- J. Okonda, H. Angeyo, A. Dehayem-Kamadjeu, A. Rogena, Chemometrics aided energy dispersive X-ray fluorescence and scattering for diagnostic modeling of trace biometallics as cancer biomarkers, *Appl. Radiat. Isot.* 190 (2022), 110489.
- C.K. Egan, S.D. Jacques, R.J. Cernik, Multivariate analysis of hyperspectral hard X-ray images, *X-Ray Spectrom.* 42 (3) (2013) 151–157.
- J.F. McCarthy, K.A. Marx, P.E. Hoffman, A.G. Gee, P. O'neil, M.L. Ujwal, J. Hotchkiss, Applications of machine learning and high-dimensional visualization in cancer detection, diagnosis, and management, *Ann. N. Y. Acad. Sci.* 1020 (1) (2004) 239–262.
- S. Romanenko, A. Stromberg, Modelling of analytical peaks: peaks modifications, *Anal. Chim. Acta* 581 (2) (2007) 343–354.
- R.G. Brereton, *Applied Chemometrics for Scientists*, John Wiley & Sons, 2007.
- J. Khan, J.S. Wei, M. Ringner, L.H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C.R. Antonescu, C. Peterson, Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks, *Nat. Med.* 7 (6) (2001) 673–679.
- S. Corezzi, L. Urbanelli, P. Cloetens, C. Emiliani, L. Helfen, S. Bohic, F. Elisei, D. Fiochetto, Synchrotron-based X-ray fluorescence imaging of human cells labeled with CdSe quantum dots, *Anal. Biochem.* 388 (1) (2009) 33–39.
- L.A. Finney, Q. Jin, Preparing adherent cells for X-ray fluorescence imaging by chemical fixation, *J. Visualiz. Exp. JoVE* 97 (2015).
- M. Podgorczyk, W. Kwiatek, W. Zajac, J. Dulińska-Litewka, E. Welter, D. Grolimund, Zinc in native tissues and cultured cell lines of human prostate studied by SR-XRF and XANES, *X-Ray Spectrom.* Int. J. 38 (6) (2009) 557–562.
- L. Pascolo, B. Bortot, N. Benseny-Cases, A. Gianoncelli, G. Tosi, B. Ruozzi, C. Rizzardi, E. De Martino, M.A. Vandelli, G.M. Severini, Detection of PLGA-based nanoparticles at a single-cell level by synchrotron radiation FTIR spectromicroscopy and correlation with X-ray fluorescence microscopy, *Int. J. Nanomedicine* 9 (2014) 2791.
- A. Gianoncelli, G. Kourousias, A. Stofa, B. Kaulich, Recent developments at the TwinMic beamline at ELETTRA: An 8 SDD detector setup for low energy X-ray fluorescence, in: *Journal of Physics: Conference Series*, IOP Publishing, 2013.
- L. Lu, S. Tian, H. Liao, J. Zhang, X. Yang, J.M. Labavitch, W. Chen, Analysis of metal element distributions in rice (*Oryza sativa* L.) seeds and relocation during

- germination based on X-ray fluorescence imaging of Zn, Fe, K, Ca, and Mn, *PLoS One* 8 (2) (2013), e57360.
- [31] R.G. Leitão, C. Santos, A. Palumbo, P. Souza, G. Pereira, M. Anjos, L. Nasciutti, R. Lopes, Distribution of Fe, Cu and Zn in cellular spheroid derived human prostate tumor cells by synchrotron X-ray fluorescence, *IEEE Trans. Nucl. Sci.* 60 (2) (2013) 758–762.
- [32] V. Solé, E. Papillon, M. Cotte, P. Walter, J. Susini, A multiplatform code for the analysis of energy-dispersive X-ray fluorescence spectra, *Spectrochim. Acta B At. Spectrosc.* 62 (1) (2007) 63–68.
- [33] E. Bárány, I.A. Bergdahl, L.-E. Bratteby, T. Lundh, G. Samuelson, A. Schütz, S. Skerfving, A. Oskarsson, Relationships between trace element concentrations in human blood and serum, *Toxicol. Lett.* 134 (1–3) (2002) 177–184.
- [34] T. Mitchell, *Machine Learning*, 1997.
- [35] I.T. Jolliffe, A note on the use of principal components in regression, *Appl. Stat.* (1982) 300–303.
- [36] R.G. Brereton, *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, John Wiley & Sons, 2003.
- [37] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.
- [38] M.A. Rahman, R. Chandren Muniyandi, D. Albashish, M.M. Rahman, O.L. Usman, Artificial Neural Network with Taguchi Method for Robust Classification Model to Improve Classification Accuracy of Breast Cancer vol. 7, *PeerJ Computer Science*, 2021, p. e344.
- [39] R.C. Tomas, A.J. Sayat, A.N. Atienza, J.L. Danganan, M.R. Ramos, A. Fellizar, K. I. Notarte, L.M. Angeles, R. Bangaol, A. Santillan, Detection of breast cancer by ATR-FTIR spectroscopy using artificial neural networks, *PLoS One* 17 (1) (2022), e0262489.
- [40] M. Khayatzaheh Mahani, M. Chaloosi, M. Ghanadi Maragheh, A.R. Khanchi, D. Afzali, Comparison of artificial neural networks with partial least squares regression for simultaneous determinations by ICP-AES, *Chin. J. Chem.* 25 (11) (2007) 1658–1662.
- [41] E.M. Gonçalves, C.Á. Ventura, T. Yano, M.L.R. Macedo, S.C. Genari, Morphological and growth alterations in Vero cells transformed by cisplatin, *Cell Biol. Int.* 30 (6) (2006) 485–494.
- [42] A. Gianoncelli, V. Bonanni, G. Gariani, F. Guzzi, L. Pascolo, R. Borghes, F. Billè, G. Kourousias, Soft X-ray microscopy techniques for medical and biological imaging at TwinMic—Elettra, *Appl. Sci.* 11 (16) (2021) 7216.
- [43] J.M. Matés, F.M. Sánchez-Jiménez, Role of reactive oxygen species in apoptosis: implications for cancer therapy, *Int. J. Biochem. Cell Biol.* 32 (2) (2000) 157–170.
- [44] H. Xue, R. Qiao, L. Yan, S. Yang, Y. Liang, Y. Liu, Q. Xie, L. Cui, B. Cao, The correlation between potential “anti-cancer” trace elements and the risk of breast cancer: a case-control study in a Chinese population, *Front. Oncol.* (2021) 11.
- [45] C.-H. Guo, S. Hsia, M.-Y. Shih, F.-C. Hsieh, P.-C. Chen, Effects of selenium yeast on oxidative stress, growth inhibition, and apoptosis in human breast cancer cells, *Int. J. Med. Sci.* 12 (9) (2015) 748.
- [46] G. Schrauzer, Anticarcinogenic effects of selenium, *Cell. Mol. Life Sci. CMLS* 57 (13) (2000) 1864–1873.
- [47] L. Tapia, M. Suazo, C. Hödar, V. Cambiazo, M. González, Copper exposure modifies the content and distribution of trace metals in mammalian cultured cells, *Biomaterials* 16 (1) (2003) 169–174.
- [48] M. Goodman, R. Bostick, Ö. Küçük, D.P. Ve Jones, Clinical trials of antioxidants as cancer prevention agents: past, present, and future, *Free Radic. Biol. Med.* 51 (2011) 1068–1084.
- [49] H. Xue, R. Qiao, L. Yan, S. Yang, Y. Liang, Y. Liu, Q. Xie, L. Cui, B. Cao, The correlation between potential “anti-cancer” trace elements and the risk of breast cancer: a case-control study in a Chinese population, *Front. Oncol.* (2021) 3086.
- [50] J.L. Heath, J.M. Weiss, C.P. Lavau, D.S. Wechsler, Iron deprivation in cancer—potential therapeutic implications, *Nutrients* 5 (8) (2013) 2836–2859.
- [51] S. Aida, T. Matsuno, T. Hasegawa, K. Tsuji, Application of principal component analysis for improvement of X-ray fluorescence images obtained by polycapillary-based micro-XRF technique, *Nucl. Instrum. Methods Phys. Res., Sect. B* 402 (2017) 267–273.
- [52] M. Farquharson, K. Geraki, G. Falkenberg, R. Leek, A. Harris, The localisation and micro-mapping of copper and other trace elements in breast tumours using a synchrotron micro-XRF system, *Appl. Radiat. Isot.* 65 (2) (2007) 183–188.
- [53] J.C. Kwok, D.R. Richardson, The iron metabolism of neoplastic cells: alterations that facilitate proliferation? *Crit. Rev. Oncol. Hematol.* 42 (1) (2002) 65–78.



Chemometrics aided energy dispersive X-ray fluorescence and scattering for diagnostic modeling of trace biometals as cancer biomarkers

J.J. Okonda^{a,*}, H.K. Angeyo^a, A. Dehayem-Kamadjeu^b, A.E. Rogena^c

^a Department of Physics, University of Nairobi, Kenya

^b Alan Alda Center for Communicative Science, Stony Brook University, New York, United States

^c Department Human Pathology, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Kenya.P.O. Box, 30197-00100, Nairobi, Kenya

ARTICLE INFO

Keywords:

Energy dispersive X-ray fluorescence
Multivariate chemometrics
Trace biometals
Biomarkers
Compton scatter
Cancer diagnosis

ABSTRACT

Direct analysis of biometals in biomedical samples by energy dispersive X-ray fluorescence (EDXRF) for disease diagnostics has hardly been fully explored due to dark matrix analytical challenges. In this study, we exploited multivariate chemometrics modeling of cancer diagnostics in model human tissue simulates and cultures using selected biometals' (Mn, Fe, Cu, Zn and Se) fluorescence and Compton scatter profiles. PCA successfully reduced the correlated data dimension to uncorrelated datasets for the characterization of the cell cultures. Artificial neural network (ANN) enhanced the classification of cancer staging and the development of a multivariate calibration strategy for the quantification of trace elements. ANN characterized cancer into early, intermediate, and advanced stages of development. Low concentrations of Fe (101 ± 28 ppm), Zn (59 ± 4 ppm) and Cu (21 ± 1 ppm) were evident in SV10 due to the lag phase stage of cancer development. Further, strong correlation (0.976) was evident in early-stage cancer between Zn and Se but with strong negative correlations between Mn and Se (-0.973) and between Mn and Zn (-0.900) probably due to their antioxidant effects. The results show predictable and systematic associations between the concentrations of Fe, Cu, Zn, Se and Mn as cancer biomarkers with the potential to be used for cancer diagnosis at the early stage of development.

1. Introduction

Energy dispersive X-ray fluorescence (EDXRF) spectroscopy analytical method has the potential to identify and quantify elements by measuring their spectral energy and intensity. In practice, EDXRF spectra of dark matrix-dominated samples (biological specimens) consist of both fluorescence lines and scatter peaks as well as pronounced Bremsstrahlung. The Bremsstrahlung masks the weak analyte fluorescence signals corresponding to the often low (ppb – a few ppm) trace biometals present in biomedical samples. This limits direct and rapid EDXRF analysis of trace elements in cell cultures due to organic matrices (Angeyo et al., 1998) that stimulated the development of chemometrics aided energy dispersive X-ray fluorescence and scatter (EDXRFS) spectrometry (Okonda et al., 2017). The conventional EDXRF spectrometry utilizes the most intense and interference-free peaks to perform quantification. In this study, EDXRFS exploits the X-ray Compton scatter peaks in addition to fluorescence spectral regions for multivariate quantitative analysis and implicit predictive modeling for model human tissues (Sichangi et al., 2019).

Measurable biometals may be regarded as biomarkers that can demonstrate the presence of malignancy as well as predict the development of cancer (Hinestrosa et al., 2007). Atomic absorption spectroscopy (AAS) indicated high levels of Cu but low levels of Zn, Fe and Se in bladder cancer as compared to the non-cancerous patients (Goyal et al., 2006). Particle-induced X-ray emission (PIXE) enabled the determination of the concentrations of Fe and Zn in benign, malignant and normal breast tissues (Vatankhah et al., 2003). Their levels were significantly higher in malignant compared to benign tissues probably due to their role as cofactors in enzymes. Inductively coupled plasma mass spectrometry (ICP-MS) and neutron activation analysis (NAA) (Zaichick and Zaichick 2016) found significantly high levels of Cr, Mn and Fe, but low concentrations of Se and Zn in prostate adenocarcinoma and normal prostate tissues respectively. EDXRF spectrometry enabled determination of Zn and Cu in human tissues where a positive correlation with age was established (Carvalho et al., 1998). Total reflection X-ray fluorescence (TXRF) has been utilized in analysis of the elemental levels of Ca, Ti, Fe, Cu and Zn in digested human breast tissue (normal and malignant) where multivariate discriminant analysis (MDA)

* Corresponding author.

E-mail address: okondajustus@uonbi.ac.ke (J.J. Okonda).

<https://doi.org/10.1016/j.apradiso.2022.110489>

Received 14 May 2022; Received in revised form 21 September 2022; Accepted 24 September 2022

Available online 28 September 2022

0969-8043/© 2022 Elsevier Ltd. All rights reserved.

enabled classifications of human tissues into cancerous and non-cancerous (da Silva et al., 2009) groups. TXRF further enabled the determination of trace biometals; Fe, Br, Cu and Zn in normal and cancerous breast tissues in different age groups (Mirji et al., 2018). The concentrations Cu and Zn were significantly higher in healthy breast tissue compared to the malignant breast tissue in the samples of in samples different age groups. Multivariate chemometric modeling by soft independent class analogy (Farquharson et al., 2013) was used to classify human tissue samples where a significant ($p < 0.01$) increase in concentrations of K, Rb and Zn were found in the tumor samples that correlated with tumor reported in the sample. PCA has the potential to condense original data to a new dataset with few variables of most relevant information (*principal components*) and thereby increasing information mining (Einax et al., 1997; De Groot et al., 2001; Nagata et al., 2006). PCA enhanced exploratory analysis (De Oliveira et al., 2010) that enabled classification of tissue samples into cancerous and non-cancerous groups. Energy-dispersive x-ray diffraction (EDXRD) (Farquharson and Geraki 2004) analysis of Fe, Cu and Zn by multivariate PLS and PCA differentiated between normal and cancerous breast tissues at classification accuracy of 71% and 60%. The X-ray fluorescence and coherent scatter determined concentrations of Fe, Cu and Zn for breast cancer diagnostics (Farquharson et al., 2013). Further utility of Multivariate soft independent modeling of class analogy in this study enabled model classification of breast tissue samples.

Some of the above methods are invasive and their interpretations in cancer diagnosis are mostly limited to simple comparisons of concentration levels, which is unreliable for early cancer diagnosis. The sample composition is paramount for the successful development of the calibration curve and the FP method is limited by uncertainties due to mass absorption coefficients and also fluorescence yields (Angeyo et al., 2012) of trace elements which result in limited analytical precision of unknown sample composition. Direct rapid but accurate trace metal determination and multivariate exploratory modeling for the detection and characterization of cancer at the early stages of development remains a challenge. Localized cancers are curable but once they metastasize they tend to be lethal (Speransky 2019). Despite the complex correlations and multivariate relationships among trace biometals in human body tissue, chemometrics approaches can mine vital diagnostic information from such complex spectra (Jose and Neil 2009).

PCA enables visualization of the latent relationship between samples and variables via scores and loadings plots (Wentzell and Hou 2012). Principal components analysis (PCA) has the potential to reduce the complexity of EDXRFS spectral data to mine vital pathological information for disease (cancer) diagnostics. ANN with its multivariate capability has the potential for classification and prediction application for spectral analysis (Ferreira et al., 2008). The datasets are first pre-processed by mean-centering and normalized prior to PCA (Angeyo et al., 2012). The analytical approach not only rapidly quantifies the trace biometals (Cu, Mn, Se, Zn, and Fe) directly in tissue but also characterizes cancer including at early stages (which is not possible via histopathological microscopy). Further, the method reliably explores latent multivariate alterations of these trace metal biomarkers as metrics in cancer diagnostics.

In this study, we focused on the analysis of those trace metals inherent in cellular architecture (biometals) as opposed to general trace elements that may include those from occupational exposure which have also been linked to cancer (Mulware 2013). The ultimate goal was to exploit the power of multivariate chemometric enabled EDXRFS spectrometry to diagnose and characterize cancer using both fluorescence and the Compton scatter for exploratory analysis of trace biometals in soft body tissues.

2. Materials and method

2.1. Sample preparation procedure

Two sets of samples were prepared for analysis.

2.1.1. Simulated tissues

Synthetic standard human tissue simulated samples were prepared (Okonda et al., 2017) from highly purified (analytical grade) paraffin wax as the base matrix. The blank paraffin wax base matrix was spiked with multi-elemental stock solutions analytes (Mn, Fe, Cu, Zn and Se) within the range (Banaš et al., 2001; Kwiatek et al., 2004; Silva et al., 2012) in which they occur in human body tissue; Mn (1.0–9.0 ppm), Fe (12–120 ppm), Cu (5–33 ppm), Zn (5.0–195 ppm) and Se (1–6 ppm). The simulate samples were chemically similar to tissue cell cultures for an accurate and precise qualitative and quantitative analysis of trace biometals for cancer diagnostics.

2.1.2. Tissue cell cultures

A total of 20 samples were cultured (10 cancerous (SV1-SV10) and 10 non-cancerous (SL1-SL10)) at Kenya Medical Research Institute (KEMRI). The Vero P₃₃ cells and healthy human lymphocytes (normal/non-cancerous cells) enriched in 3 mL growth media were separately cultured at 37°C and 97% humidified atmosphere of 5% CO₂ (Podgórzyk et al., 2009). The lymphocyte cultures were microscopically examined on daily basis to check for contamination and periodically supplied with 1 mL fresh sterilized growth medium at intervals of 3 days. The Vero P₃₃ cancer cells were maintained by pouring off the used growth media and trypsinized before adding 3 mL of fresh growth media. During culture harvesting, about 5 mL of peripheral blood mononuclear cells (PBMCs) suspension were centrifuged to obtain 100 µL pellets while Vero P₃₃ cells were trypsinized and incubated for 5–10 min to detach the cells from the surface of the culture flask before being centrifuged to obtain 100 µL pellet. Subsequent harvesting was done at intervals of every 24 h for 10 days to mimic the various stages of cancer development in humans. The Vero P₃₃ and the PBMC pellets were then labeled as SV and SL respectively and stored at –80°C to minimize morphological damage. About 10 µL of each of the above samples was pipetted onto a 2 µm Mylar foil and dried at a temperature of 25 °C in a biosafety cabinet to form a thin sample substrate for EDXRFS spectrometric analysis.

2.2. Energy dispersive X-ray fluorescence and scatter (EDXRFS) spectrometric analysis

The EDXRFS spectral profiles from the above samples were acquired using Shimadzu EDX 800 HS spectrometer with a Rhodium X-ray tube. The samples were rapidly directly irradiated for 50 s in air under an applied voltage and current of 50 kV and 0.001–1 mA respectively. The spectral data were obtained through the multi-channel analyzer in the range 0–40 keV at a resolution of 160 eV at Mn K_α (5.9 keV) peak at a dead time of <25%. At these conditions, the fluorescence and Compton scatter spectral profiles were acquired for biometal quantification (Okonda et al., 2017) via multivariate calibration and subsequent multivariate chemometrics modeling. In order to determine the accuracy of EDXRFS analytical method, certified reference material (NIST Oyster tissue 1566b) was prepared and analyzed in duplicate under same irradiation conditions.

2.3. Multivariate chemometrics modeling for cancer diagnostics

A typical EDXRFS spectrum of a diseased (cancer) tissue sample is mathematically a complex multivariate dataset. The multivariate chemometrics techniques enable the separation of characteristic and non-characteristic information from the spectra (Kessler et al., 2002). This

calls for dimension reduction and appropriate multivariate interpretation to yield accurate diagnostic information. In this work, chemometrics techniques (PCA and ANN) were used for multivariate exploratory analysis and calibration towards quantitative analysis for tissue characterization (Meglen 1992). PCA and ANN have the potential to extract spectral information for exploratory analysis (Waddell et al., 2004). Selected multiple fluorescence signatures of analyte biometals in Table 1 were used for the development of a multivariate calibration strategy for trace analysis and modeling for cancer diagnostics.

The simulated tissue and Oyster tissue (certified reference standard) spectral fluorescence and K_{α} Compton scatter region (channel numbers 950–1000) were selected. Prior to analysis, the datasets were pre-processed through mean-centering and normalization. The Quasar orange software package (Toplak et al., 2021) was utilized for advanced data evaluation and chemometrics (PCA and ANN) analysis.

PCA extracted underlying cancer diagnostic information from the complex multivariate EDXRFS data. The correlated variables of the original data (Jolliffe 1982) were reduced to important uncorrelated variables with the most relevant information (principal components). The data matrix X was decomposed to a score matrix t and the loading matrix p as shown in equation (1) (Brereton 2003; Geladi 2003).

$$X = t_1p_1 + t_2p_2 + \dots + t_ap_a + E \tag{1}$$

Matrix E contained the residuals and t described the sample position in the new coordinate system. The new axis is visualized in the loading matrix p where PC loadings vectors give the hidden variable profiles while the scores reveal information on the elemental concentrations.

Principal component scores were further used as inputs in the ANN model for the classification of cancer staging while the ANN multivariate calibration technique was based on standard calibration curves (Ferreira et al., 2008). ANN backpropagation techniques algorithm (McClelland et al., 1986) was constructed using the simulated tissue biometal concentrations. The input data (matrix) for each spectral dataset for the measured independent variable was utilized to build the ANN calibration strategy. The datasets were further split into 3 subsets; training, testing and validation (Okonda et al., 2017). The neurons in the hidden layer were chosen based on the root mean square error (RMS) and the training performance of each network where training was repeated several times until the best results of low MSE for acceptable performance. Forward neural networks (FNN) supervised approach (Tomas et al., 2022) with 3 layers; input (3 neurons), hidden (5 neurons) and output (7 neurons) to minimize the root mean square (RMS) error. The 3-fold cross-validation with ReLu activation function was used to evaluate the classification model at 200 iterations.

3. Results and discussion

Typical overlaid EDXRFS spectra of simulated and cultured tissue samples are as shown in Fig. 1. Besides the fluorescence lines and scatter peaks, a pronounced background continuum (Bremsstrahlung) masked the weak fluorescence analyte signals of trace biometals hence poor

Table 1
Analyte spectral feature selection.

Element	Actual (or Expected) Fluorescence Peak	Energy (keV)	Channel Number (Spectrum region of interest)
Mn	K_{α}	5.90	287–296
	K_{β}	6.49	320–329
Fe	K_{α}	6.40	315–324
	K_{β}	7.06	347–356
Cu	K_{α}	8.05	397–406
	K_{β}	8.91	440–449
Zn	K_{α}	8.64	428–437
	K_{β}	9.57	474–483
Se	K_{α}	11.22	561–568
	K_{β}	12.50	618–627

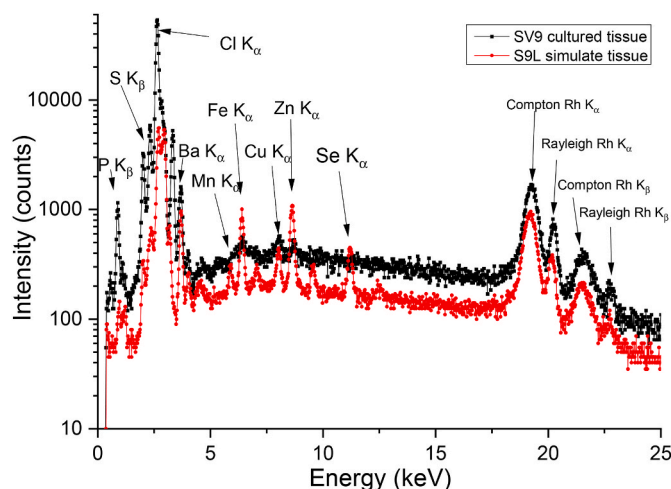


Fig. 1. Typical overlaid EDXRFS spectra of model tissue samples acquired at 200s.

detection limits and signal-to-noise ratio (SNR) of analytes (Kaniu et al., 2012).

The diagnostic trace biometal peaks are not clearly discernible for selected fluorescence (4.5–12.0 keV) against the pronounced background, thus their recovery and trace quantitative deconvolution remains a challenge. These limits the direct and rapid EDXRF analysis of trace biometals in biomedical sample matrices for disease (cancer) diagnostic.

3.1. Multivariate characterization of cell culture samples

Multivariate exploratory analysis of the fluorescence and scatter spectral profiles enabled exploratory analysis and quantitative determination of biometal utilizing multivariate chemometric techniques (PCA and ANN).

3.2. Principal component analysis of cell culture samples

Fig. 2 shows the utility of entire EDXRFS spectral features in the PCA model in which had no potential for exploring the various stages of cancer development. The modeling was also not specific to the studied trace metal biomarkers.

The score plot had an explained variance of 90% (80% and 10% for PC1 and PC2 respectively) with hardly any clear classifications. There is need for further exploration of the ability of both fluorescence and Compton scatter in chemometrics EDXRFS for diagnostic modeling of cancer. From the corresponding loadings plot, the fluorescent peaks of the trace biometals together with Compton scatter are possible potential biomarkers to be exploited for cancer diagnosis.

Fig. 3 shows a PCA score and corresponding loadings plots for modeling of cancer development (early and advanced stages of cancer) based on the selected fluorescence (4.5–12.0 keV) signatures of trace biometals.

The above PCA score plot with 97% (82% and 15% for PC1 and PC2 respectively) of the total explained variance classified the cultured samples into 2 major distinct clusters; early cancer (SV1, SV2, SV3 SV4 and SV5) and advanced cancer (SV6, SV7, SV8 and SV10). The positive PC1 loadings plot shows the score plot clusters are mostly influenced by Mn, Fe, Cu, and Zn but not Se.

3.2.1. ANN classification of cultured cancerous samples

The classification of the various stages of cancer development was further achieved by the utility of PC scores as input data in ANN characterization. The 2 dominant PCs of selected fluorescence and the

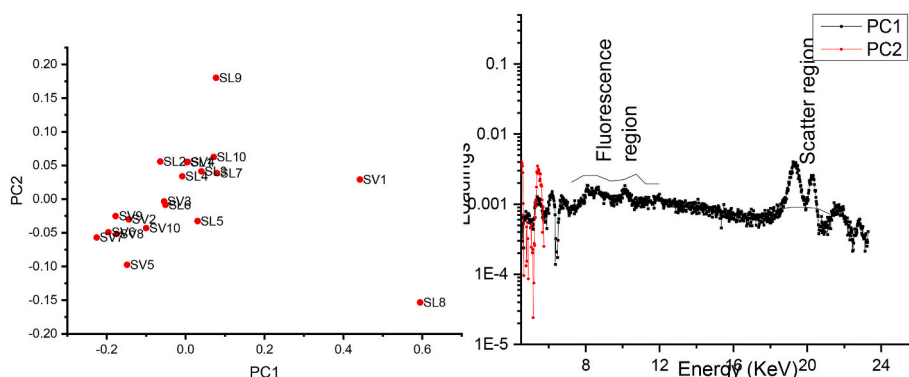


Fig. 2. PCA score and loadings plots of cultured tissue samples utilizing the entire spectra.

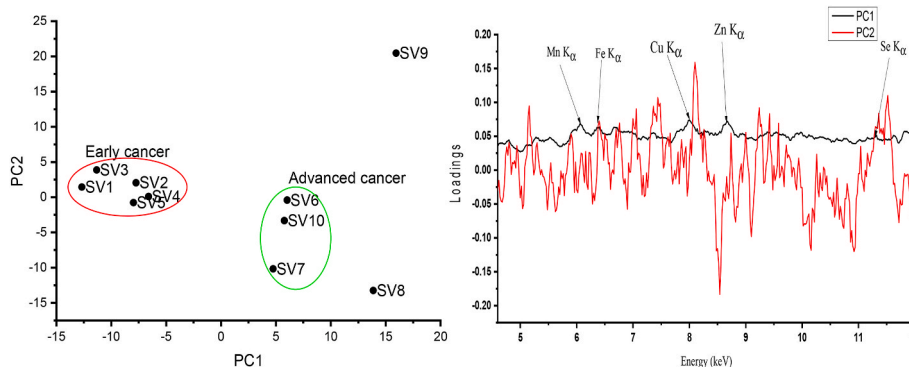


Fig. 3. Selected fluorescence PCA score and loading plots for cultured cancerous samples.

Compton scatter with a total explained variance of 91% were used as input data matrix in ANN. The cultured cancerous samples were characterized into various stages of development as shown in Fig. 4. In this case, the scatter region contained the main distinguishing feature in cancer staging, which indicates high concentration of low-Z elements (O, C and H) in the cultured samples.

The potential applicability of our method in Fig. 4 which shows a clear characterization of cancerous cultured samples into early, intermediate and advanced stages of development. The vital cancer

characterization information was partially contained in the scatter region (Kessler et al., 2002) due to low-Z elements, which were exploited by multivariate ANN. This was attributed to the difference in the cell structure at various stages of cancer development. The scatter (Compton) region due to low-Z elements reveals further information on the mean atomic number and the density of the sample (Compton and Allison 1935; Kunzendorf 1972) which influence tissue cellular morphology (Malenica et al., 2021).

3.2.2. Concentration of trace biometals

The calibration strategy developed for direct determination of the concentrations of trace biometals (Okonda et al., 2017) was used for quantification of trace biometals in certified reference material (NIST Oyster tissue 1566b) in order to test the accuracy of the calibration models as shown in Table 2.

The results in Table 2 above demonstrated the potential of chemometric aided EDXRFS spectrometry method in direct rapid trace biometals analysis for cancer diagnostics modeling. The application of our method for the quantitative determination of trace biometals in Oyster tissue was within the acceptable limits of accuracy. This was attributed to the potential of chemometric machine learning through robust multivariate modeling for quantitative analysis of trace biometals in

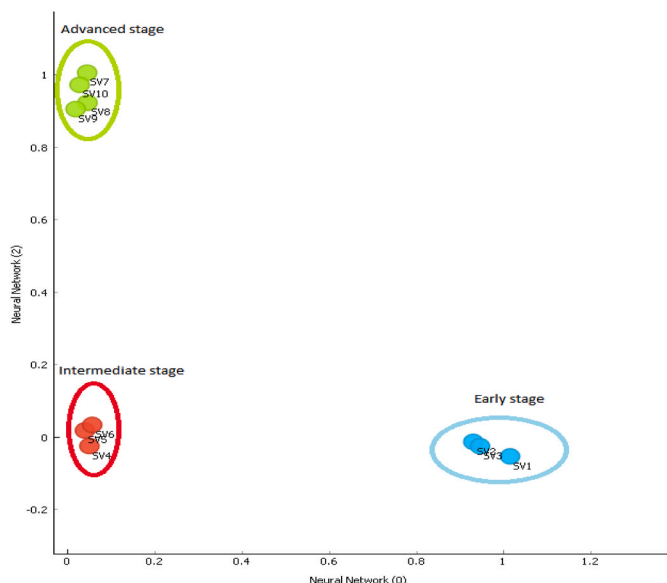


Fig. 4. Selected fluorescence and Compton scatter.

Table 2 EDXRFS concentrations of trace biometals in Oyster tissue (NIST 1566b).

Biometal	Certified value ± SD (ppm)	EDXRFS ± SD (ppm)	EDXRFS % deviation
Mn	18.5 ± 0.2	16.9 ± 1.2	5.6
Fe	205.8 ± 6.8	238 ± 15.6	4.9
Cu	71.6 ± 1.6	66.8 ± 4.6	6.7
Zn	1424 ± 46	1280.5 ± 29.0	6.8
Se	2.06 ± 0.15	3.6 ± 1.1	31.1

complex biomedical sample specimen (Sichangi et al., 2019). The similar comparison between EDXRFS measured and certified concentrations of biometals in Oyster tissue shows the potential for the quantitative determination of trace biometals towards cancer diagnosis. The poor performance by Se could be due to suppressed levels of Se close to its detection limit of 1.6 ppm (Okonda 2015) in Oyster tissue hence the low fluorescence yield (Okonda et al., 2017) thus EDXRFS was inefficient in quantifying low concentrations such as that of Se. This was clear in the performance of the PCA models above which showed Mn, Fe, Cu and Zn as better trace metal cancer biomarkers compared to Se.

3.2.2.1. Concentration of trace biometals in cell cultures. The concentrations of the cancer biomarker metals in the cell cultures were obtained using simulate validated ANN multivariate calibration model (Okonda et al., 2017). These concentrations are shown for both cancerous and normal cell cultures in Table 3 as a result of biological alterations (Carvalho et al., 2007).

The above results depict the subtle alterations of the trace biometals that occurs during the various stages of cancer development. The notably elevated levels of Fe and Zn in cancer cultures (SV) as compared to corresponding normal cultured tissues (SL) may be attributed to increased cell division and proliferation as cancer develops from the early to the malignant stage. Elevated Fe in malignant tissues (Geraki et al., 2004) was found probably due to its regulatory role for angiogenesis and its catalytic role in Fenton reaction which results in an

increased requirement for DNA synthesis (Heath et al., 2013) for tumor angiogenesis.

The concentrations alterations of trace biometals at each stage of cancer proliferation are as shown in Fig. 5.

The clearest distinctions were noted in concentrations of Fe, Zn and Se, particularly due to their role in the Fenton reaction that results in carcinogenesis (Jomova and Valko 2011). The accumulation of Fe and Zn in cancerous cultures (SV) can be attributed to increased cell division and proliferation (Guntupalli et al., 2007). The trace biomarkers (Fe, Cu and Zn) indicated similar patterns for early-stage cancer (SV1-SV4) and advanced-stage cancer (SV5-SV9) which explains the observed clustering for early and late stages of cancer development. Cancerous cultures (SV7, SV8, SV9 and SV10) at a late stage of cancer development had elevated concentrations of Zn compared to corresponding non-cancerous cultured samples which can be attributed to demands for excessive cell division in rapidly growing tumors. Sample SV10 had low concentrations of Fe (101 ± 28 ppm), Zn (59 ± 4 ppm) and Cu (21 ± 1 ppm). This can be explained based on the lag phase stage of cancer development which does not require high amounts of Fe, Cu and Zn due to reduced cell division and proliferation in mature cancer tissues.

The observed differences in trace biometal concentrations in healthy and cancerous cell cultures were evaluated by Student's t-test (Pearson correlation) as shown in Tables 4–6, which reveal latent associations between the trace biometals and their role during tumor growth.

The strong positive correlations between Fe and other trace biometals (Cu, Zn and Se) during the onset of cancer development (early stage) were evident and probably attributed to the exponential phase of cell division and proliferation as a result of increased cellular activity (Raju et al., 2006). It further emphasized the function of Cu as an important co-factor for the metabolism of Fe in tissues. Another distinctly notable strong correlation (0.976) for this early-stage cancer is between Zn and Se for rapid cell division and proliferation. Nonetheless associated with this early stage of cancer development are two very strong negative correlations between Mn and Se (−0.973) and between Mn and Zn (−0.900) due to their antioxidant effects.

During the intermediate stage, the association among the trace biometals is considerably different with only three significant correlations: positive correlations between Fe and Cu and between Cu and Zn; and another strong negative correlation between Se and Zn. The stronger positive correlation between Cu and Zn can be linked to an increase in metabolic activities in malignant tissues (Kubala-Kukuś et al., 2004).

At the advanced stage of cancer development, there are three notable strong positive associations: between Fe and Zn and Se (the very strong association between Fe and Cu that was manifest during early cancer is now absent). Nonetheless, the very strong correlation between Se and Zn that was also noted for early-stage cancer. It should be noted that this correlation was highly negative in the intermediate stage of cancer development.

The above results show varying levels of trace biometals in the cancerous cultures. The trace biometals (Cu, Fe, Zn, Mn and Se) are imperative for the maintenance of physiological homeostasis in organism proteins involved in cellular processes (Gruber et al., 2020) thus their alterations has a potential role in carcinogenesis. The levels of Fe, Cu and Zn in cultured cancer tissues are in agreement with previous studies (Carvalho et al., 2007) as a result of increased cellular activity in cancerous tissues. This study further indicates elevated levels of Zn in cancerous tissue samples for cell proliferation that supports tumor growth (Lee et al., 2003) and cell division. Additionally, high levels of Cu in cancerous samples were also evident due to the increased free radicals in the Fenton reaction (Yaman et al., 2005).

4. Conclusion and prospects

The potential of chemometrics Energy Dispersive X-Ray Fluorescence and Scattering (EDXRFS) spectrometry for direct and rapid determination of trace biometals (Fe, Cu, Mn, Se and Zn) fluorescence

Table 3
ANN model predicted biometal concentration for cultured samples.

Stage	Sample	Mean biometal concentrations ± standard deviation (ppm)				
		Mn	Fe	Cu	Zn	Se
Early-stage	SV1	4.4 ± 0.1	42 ± 9	21 ± 2	44 ± 4	2.4 ± 0.2
	SL1	4.4 ± 0.2	54 ± 11	18.7 ± 0.1	55 ± 8	1.5 ± 0.2
	SV2	4 ± 1	126 ± 14	21.1 ± 0.3	82 ± 7	3 ± 1
	SL2	4 ± 1	76 ± 8	21 ± 2	58 ± 6	2.5 ± 0.2
	SV3	4.5 ± 0.2	99 ± 29	22 ± 3	56 ± 3	3 ± 1
	SL3	4 ± 1	84 ± 33	19 ± 1	66 ± 23	2 ± 1
Intermediate stage	SV4	4.5 ± 0.1	72 ± 1	24 ± 8	50.8 ± 0.3	1.7 ± 0.1
	SL4	4.4 ± 0.2	83 ± 3	19.1 ± 0.3	52 ± 4	2.0 ± 0.1
	SV5	3.6 ± 0.4	160 ± 16	25 ± 2	93 ± 7	4 ± 1
	SL5	4 ± 1	57 ± 16	20 ± 1	47 ± 7	1.7 ± 0.2
	SV6	3.8 ± 0.3	161 ± 11	23.7 ± 0.1	87 ± 14	3.9 ± 0.3
	SL6	4.3 ± 0.2	64 ± 4	19 ± 1	57 ± 2	1.6 ± 0.4
Late-stage	SV7	4 ± 1	148 ± 7	22 ± 3	86 ± 4	3.8 ± 0.4
	SL7	4 ± 1	75 ± 13	18 ± 1	53 ± 2	3 ± 1
	SV8	3.9 ± 0.1	156 ± 14	24 ± 2	87 ± 7	4 ± 1
	SL8	4.6 ± 0.3	59 ± 9	19 ± 1	55 ± 4	1.9 ± 0.1
	SV9	4.2 ± 0.1	168 ± 14	25 ± 3	108 ± 5	3.9 ± 0.2
	SL9	3.4 ± 0.2	40 ± 11	18 ± 2	40 ± 30	1.7 ± 0.3
Late-stage	SV10	4.3 ± 0.1	101 ± 28	24 ± 5	59 ± 6	2.2 ± 0.4
	SL10	4.3 ± 0.1	52 ± 26	16 ± 4	40 ± 18	1.5 ± 0.3

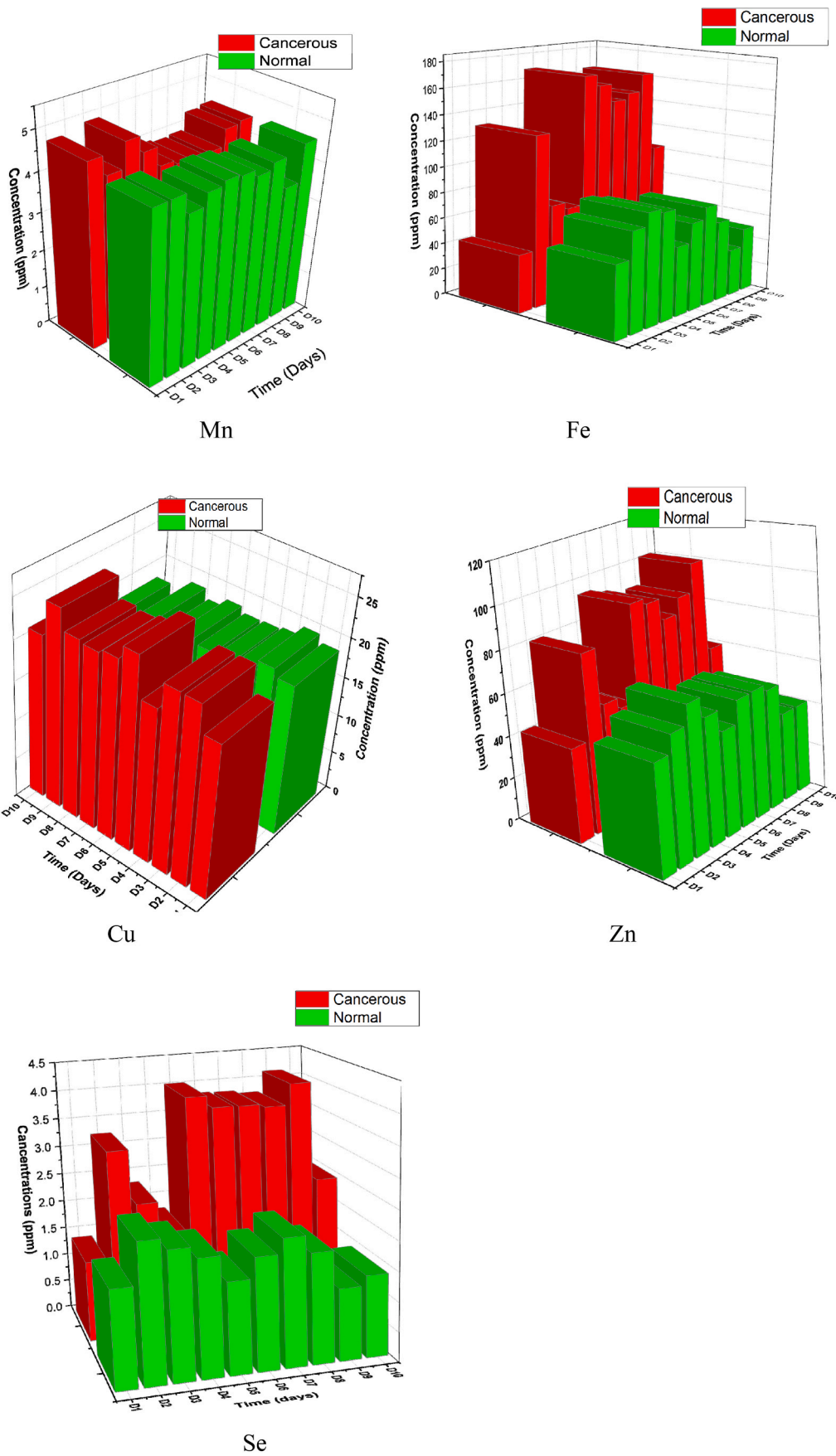


Fig. 5. Comparative concentrations of Cu, Fe, Se, Zn and Mn in cancerous (SV) and normal (SL) cultures.

Table 4
Correlation coefficients of biomarkers in cultures for early-stage cancer.

Element	Mn	Fe	Cu	Zn	Se
Mn	–				
Fe	–0.654	–			
Cu	0.050	0.723	–		
Zn	–0.900	0.918	0.391	–	
Se	–0.973	0.811	0.183	0.976	–

Table 5
Correlation coefficients of biomarkers at an intermediate stage.

Element	Mn	Fe	Cu	Zn	Se
Mn	–				
Fe	0.567	–			
Cu	0.089	0.871	–		
Zn	–0.385	0.542	0.885	–	
Se	0.693	–0.201	–0.656	–0.932	–

Table 6
Correlation coefficients of biomarkers in cultures at an advanced stage.

Element	Mn	Fe	Cu	Zn	Se
Mn	–				
Fe	–0.571	–			
Cu	0.546	0.376	–		
Zn	–1.000	0.964	0.610	–	
Se	–0.811	0.943	0.047	0.820	–

and Compton scatter, their multivariate associations and alterations in model human tissues (cell cultures) as cancer biomarkers has been established. This was demonstrated in the detection and characterization of cancer in model human samples (tissue cultures). The trace biometals were detected and quantified in cancerous and healthy cultures for cancer diagnostics. The trace biometals changes (concentrations and alterations) were detected and characterized as “fingerprints” for cancer diagnostics. The concentrations of Fe, Cu and Zn were elevated in cancerous as compared to normal cell cultures. The strong positive correlation between Fe and Cu (0.871) was evident in cancerous samples as compared to healthy cultures. The results indicated predictable associations of trace biometals fluorescence together with the Compton scatter as cancer biomarkers with the potential to be used for cancer detection and characterization, especially at the local stage of development. ANN characterized cancer into early, intermediate and advanced stages of development by the utility of the trace biometal profiles (fluorescence and corresponding scatter signatures). This work has demonstrated that the change in concentrations, alterations and correlations of trace biometals can be used as biomarkers for cancer diagnostics. However, there is a need for further studies with a wider sample size to better validate and consolidate the present conclusion on the potential of our method for the determination of trace biometals as cancer biomarkers.

Ethical approval

This study was ethically approved (P604/10/2017) by the Kenyatta National Hospital/University of Nairobi Ethics Research Committee.

CRedit authorship contribution statement

J.J. Okonda: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Conceptualization. **H.K. Angeyo:** Supervision, Resources, Project administration, Investigation, Conceptualization. **A. Dehayem-Kamadjeu:** Supervision, Resources, Project administration, Methodology, Investigation. **A.E. Rogena:**

Visualization, Supervision, Resources, Project administration, Methodology, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

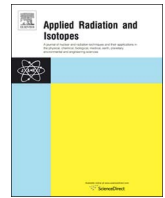
Acknowledgment

The authors wish to appreciate the financial support from Kenya National Research Fund (NRF) to the first author (NRF/PhD/02/146). We also acknowledge Mr. Julius Muchiri of Kenya Medical Research Institute (KEMRI) for his support during the cell culture process and sample preparation stages for this study.

References

- Angeyo, K., Gari, S., Mangala, J., Mustapha, A., 2012. Principal component analysis-assisted energy dispersive X-ray fluorescence spectroscopy for non-invasive quality assurance characterization of complex matrix materials. *X Ray Spectrom.* 41 (5), 321–327.
- Angeyo, K., Patel, J., Mangala, J., Narayana, D., 1998. Radioisotope photon-excited energy dispersive X-ray fluorescence technique for the analysis of organic matrices. *X Ray Spectrom.* 27 (3), 205–213.
- Banaś, A., Kwiatek, W., Zajac, W., 2001. Trace element analysis of tissue section by means of synchrotron radiation: the use of GNUPLOT for SRXRF spectra analysis. *J. Alloys Compd.* 328 (1), 135–138.
- Brereton, R.G., 2003. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant.* John Wiley & Sons.
- Carvalho, M., Brito, J., Barreiros, M., 1998. Study of trace element concentrations in human tissues by EDXRF spectrometry. *X Ray Spectrom.* 27 (3), 198–204.
- Carvalho, M., Magalhães, T., Becker, M., Von Bohlen, A., 2007. Trace elements in human cancerous and healthy tissues: a comparative study by EDXRF, TXRF, synchrotron radiation and PIXE. *Spectrochim. Acta B Atom Spectrosc.* 62 (9), 1004–1011.
- Compton, A.H., Allison, S.K., 1935. *X-Rays in Theory and Experiment.*
- da Silva, M.P., Zucchi, O.L.A.D., Ribeiro-Silva, A., Poletti, M.E., 2009. Discriminant analysis of trace elements in normal, benign and malignant breast tissues measured by total reflection X-ray fluorescence. *Spectrochim. Acta B Atom Spectrosc.* 64 (6), 587–592.
- De Groot, P., Postma, G., Melssen, W., Buydens, L., Deckert, V., Zenobi, R., 2001. Application of principal component analysis to detect outliers and spectral deviations in near-field surface-enhanced Raman spectra. *Anal. Chim. Acta* 446 (1–2), 71–83.
- De Oliveira, L., Antunes, A., Bueno, M., 2010. Direct chromium speciation using X-ray spectrometry and chemometrics. *X Ray Spectrom.* 39 (4), 279–284.
- Einax, J.W., Zwanziger, H.W., Geiss, S., 1997. *Chemometrics in Environmental Analysis.* John Wiley & Sons.
- Farquharson, M., Al-Ebraheem, A., Cornacchi, S., Gohla, G., Lovrics, P., 2013. The use of X-ray interaction data to differentiate malignant from normal breast tissue at surgical margins and biopsy analysis. *X Ray Spectrom.* 42 (5), 349–358.
- Farquharson, M., Geraki, K., 2004. The use of combined trace element XRF and EDXRD data as a histopathology tool using a multivariate analysis approach in characterizing breast tissue. *X Ray Spectrom.: Int. J.* 33 (4), 240–245.
- Ferreira, E.C., Milori, D.M., Ferreira, E.J., Da Silva, R.M., Martin-Neto, L., 2008. Artificial neural network for Cu quantitative determination in soil using a portable laser induced breakdown spectroscopy system. *Spectrochim. Acta B Atom Spectrosc.* 63 (10), 1216–1220.
- Geladi, P., 2003. *Chemometrics in spectroscopy. Part 1. Classical chemometrics.* *Spectrochim. Acta B Atom Spectrosc.* 58 (5), 767–782.
- Geraki, K., Farquharson, M., Bradley, D., Hugtenburg, R., 2004. A synchrotron XRF study on trace elements and potassium in breast tissue. *Nucl. Instrum. Methods Phys. Res. Sect. B Beam Interact. Mater. Atoms* 213, 564–568.
- Goyal, M., Kalwar, A., Vyas, R., Bhati, A., 2006. A study of serum zinc, selenium and copper levels in carcinoma of esophagus patients. *Indian J. Clin. Biochem.* 21 (1), 208–210.
- Gruber, A., Müller, R., Wagner, A., Colucci, S., Spasić, M.V., Leopold, K., 2020. Total reflection X-ray fluorescence spectrometry for trace determination of iron and some additional elements in biological samples. *Anal. Bioanal. Chem.* 412 (24), 6419–6429.
- Guntupalli, J.N.R., Padala, S., Gummuluri, A.R.M., Mukhtineni, R.K., Byreddy, S.R., Sreerama, L., Kedarisetti, P.C., Angalakuduru, D.P., Satti, B.R., Venkatathri, V., 2007. Trace elemental analysis of normal, benign hypertrophic and cancerous tissues

- of the prostate gland using the particle-induced X-ray emission technique. *Eur. J. Cancer Prev.* 108–115.
- Heath, J.L., Weiss, J.M., Lavau, C.P., Wechsler, D.S., 2013. Iron deprivation in cancer—potential therapeutic implications. *Nutrients* 5 (8), 2836–2859.
- Hinestroza, M.C., Dickersin, K., Klein, P., Mayer, M., Noss, K., Slamon, D., Sledge, G., Visco, F.M., 2007. Shaping the future of biomarker research in breast cancer to ensure clinical relevance. *Nat. Rev. Cancer* 7 (4), 309.
- Jolliffe, I.T., 1982. A note on the use of principal components in regression. *Applied Statistics* 300–303.
- Jomova, K., Valko, M., 2011. Advances in metal-induced oxidative stress and human disease. *Toxicology* 283 (2–3), 65–87.
- Jose, M., Neil, W., 2009. Basic chemometric techniques in analytical spectroscopy. *RSC Analytical Spectroscopy Monographs* 10, 244–263.
- Kaniu, M., Angeyo, K., Mwala, A., Mangala, M., 2012. Direct rapid analysis of trace bioavailable soil macronutrients by chemometrics-assisted energy dispersive X-ray fluorescence and scattering spectrometry. *Anal. Chim. Acta* 729, 21–25.
- Kessler, T., Hoffmann, P., Greve, T., Ortner, H., 2002. Optimization of the identification of chemical compounds by energy-dispersive x-ray fluorescence spectrometry and subsequent multivariate analysis. *X Ray Spectrom.: Int. J.* 31 (5), 383–390.
- Kubala-Kukuś, A., Kuternoga, E., Braziewicz, J., Pajek, M., 2004. Log-stable concentration distributions of trace elements in biomedical samples. *Spectrochim. Acta B Atom Spectrosc.* 59 (10–11), 1711–1716.
- Kunzendorf, H., 1972. Quick determination of the average atomic number Z by X-ray scattering. *Nucl. Instrum. Methods* 99 (3), 611–612.
- Kwiatkiewicz, W., Hanson, A., Paluszkiwicz, C., Galka, M., Gajda, M., Cichocki, T., 2004. Application of SRXRF and XANES to the determination of the oxidation state of iron in prostate tissue sections. *J. Alloys Compd.* 362 (1–2), 83–87.
- Lee, R., Woo, W., Wu, B., Kummer, A., Duminy, H., 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 (6), 689–696.
- Malenica, M., Vukomanović, M., Kurtjak, M., Masciotti, V., Dal Zilio, S., Greco, S., Lazzarino, M., Krusić, V., Perčić, M., Jelovica Badovinac, I., 2021. Perspectives of microscopy methods for morphology characterisation of extracellular vesicles from human biofluids. *Biomedicines* 9 (6), 603.
- McClelland, J.L., Rumelhart, D.E., Group, P.R., 1986. *Parallel Distributed Processing*. MIT press Cambridge, MA.
- Meglen, R.R., 1992. Examining large databases: a chemometric approach using principal component analysis. *Mar. Chem.* 39 (1–3), 217–237.
- Mirji, S., Badiger, N.M., Sanyal, K., Kulkarni, S.S., Misra, N.L., Gai, P.B., 2018. Determination of trace elements in normal and malignant breast tissues of different age group using total reflection X-ray fluorescence spectrometer. *X Ray Spectrom.* 47 (6), 432–440.
- Mulware, S.J., 2013. Trace elements and carcinogenicity: a subject in review. *3 Biotech* 3 (2), 85–96.
- Nagata, N., Peralta-Zamora, P.G., Poppi, R.J., Perez, C.A., Bueno, M.I.M., 2006. Multivariate calibrations for the SR-TXRF determination of trace concentrations of lead and arsenic in the presence of bromine. *X Ray Spectrom.: Int. J.* 35 (1), 79–84.
- Okonda, J., Angeyo, K., Mangala, J., Kisia, S., 2017. A nested multivariate chemometrics based calibration strategy for direct trace biometal analysis in soft tissue utilizing Energy Dispersive X-Ray Fluorescence (EDXRF) and scattering spectrometry. *Appl. Radiat. Isot.* 129, 49–56.
- Okonda, J.J., 2015. *Development of Chemometrics Aided Energy Dispersive X-Ray Fluorescence and Scattering (Edxrf) Method for Rapid Diagnostics of Cancer*. University of Nairobi.
- Podgorczyk, M., Kwiatek, W., Zajac, W., Dulirńska-Litewka, J., Welter, E., Grolimund, D., 2009. Zinc in native tissues and cultured cell lines of human prostate studied by SR-XRF and XANES. *X Ray Spectrom.* 38 (6), 557–562.
- Raju, G.N., Sarita, P., Kumar, M.R., Murty, G.R., Reddy, B.S., Lakshminarayana, S., Vijayan, V., Lakshmi, P.R., Gavarasana, S., Reddy, S.B., 2006. Trace elemental correlation study in malignant and normal breast tissue by PIXE technique. *Nucl. Instrum. Methods Phys. Res. Sect. B Beam Interact. Mater. Atoms* 247 (2), 361–367.
- Sichangi, E.K., Angeyo, H.K., Dehayem-Massop, A., 2019. Trace metal biomarker based Cancer diagnostics in body tissue by energy dispersive X-ray fluorescence and scattering (EDXRF) spectrometry. *Spectrochim. Acta B Atom Spectrosc.* 105635
- Silva, M.P., Soave, D.F., Ribeiro-Silva, A., Poletti, M.E., 2012. Trace elements as tumor biomarkers and prognostic factors in breast cancer: a study through energy dispersive x-ray fluorescence. *BMC Res. Notes* 5 (1), 194.
- Speransky, S.R., 2019. *Identifying Novel Biomarkers of Metastatic Potential in Prostate and Breast Cancer*. University of Miami.
- Tomas, R.C., Sayat, A.J., Atienza, A.N., Danganan, J.L., Ramos, M.R., Fellizar, A., Notarte, K.I., Angeles, L.M., Bangaol, R., Santillan, A., 2022. Detection of breast cancer by ATR-FTIR spectroscopy using artificial neural networks. *PLoS One* 17 (1), e0262489.
- Toplak, M., Read, S.T., Sandt, C., Borondics, F., 2021. Quasar: easy machine learning for biospectroscopy. *Cells* 10 (9), 2300.
- Vatankhah, S., Moosavi, K., Salimi, J., Geranpayeh, L., Peyrovani, H., 2003. A Pixe Analysis for Measuring the Trace Elements Concentration in Breast Tissue of Iranian Women.
- Waddell, R.J., NicDaéid, N., Littlejohn, D., 2004. Classification of ecstasy tablets using trace metal analysis with the application of chemometric procedures and artificial neural network algorithms. *Analyst* 129 (3), 235–240.
- Wentzell, P., Hou, S., 2012. Exploratory data analysis with noisy measurements. *J. Chemometr.* 26 (6), 264–281.
- Yaman, M., Atici, D., Bakirdere, S., Akdeniz, I., 2005. Comparison of trace metal concentrations in malign and benign human prostate. *J. Med. Chem.* 48 (2), 630–634.
- Zaichick, V., Zaichick, S., 2016. Prostatic tissue levels of 43 trace elements in patients with prostate adenocarcinoma. *Cancer Clin. Oncol.* 5 (1), 79.



A nested multivariate chemometrics based calibration strategy for direct trace biometal analysis in soft tissue utilizing Energy Dispersive X-Ray Fluorescence (EDXRF) and scattering spectrometry



J.J. Okonda^{a,*}, K.H. Angeyo^a, J.M. Mangala^b, S.M. Kisia^c

^a Department of Physics, University of Nairobi, P.O. Box 30197-00100, Nairobi, Kenya

^b Institute of Nuclear Science and Technology, University of Nairobi, P.O. Box 30197-00100, Nairobi, Kenya

^c Department of Veterinary Anatomy and Physiology, University of Nairobi, P.O. Box 30197. 0100, Nairobi, Kenya

HIGHLIGHTS

- The X-Ray fluorescence and scatter regions can be utilized to derive comprehensive sample composition information.
- Multivariate calibration techniques enhance direct rapid simultaneous analysis of trace biometals in model soft tissues.
- The nested approach (ANN and PCR) for multivariate calibration is an efficient technique for direct quantitative analysis of specific trace elements in model soft tissue.

ARTICLE INFO

Keywords:

Chemometrics
Calibration strategy
Biometals
Model soft tissue
Oyster tissue

ABSTRACT

Compton scatter-modulated fluorescence and multivariate chemometric (artificial neural network (ANN) and principal component regression (PCR)) calibration strategy was explored for direct rapid trace biometals (Mn, Fe, Cu, Zn, Se) analysis in “complex” matrices (model soft tissues). This involved spectral feature selection (multiple fluorescence signatures) normalized to or in conjunction with Compton scatter. ANN model resulted in more accurate trace biometal determination ($R^2 > 0.9$) compared to PCR. Hybrid nested (ANN and PCR) approach led to optimized accurate biometals’ concentrations in Oyster tissue ($\leq \pm 10\%$).

1. Introduction

Energy dispersive X-ray fluorescence (EDXRF) spectroscopy (variant of X-ray fluorescence spectrometry) is an analytical technique for identification and quantification of elements in materials by means of measurement of the elements’ specific characteristic fluorescence energies and intensities (Silva et al., 2012). However, its utility in direct rapid analysis of elements is faced by challenges arising from the enhanced spectral background due to high mass absorption coefficients of especially “dark matrices” (low-Z dominated), extreme matrix effects and spectral overlaps (Angeyo et al., 1998). In order to minimize some of the above challenges, Total Reflection X-ray Fluorescence (TXRF) is used as its sensitivity can be improved by the fact that, at very low incidence glancing angles, the high background due to scatter from the sample support is minimized (Eppis et al., 2001). TXRF spectrometric quantification through internal standardization for small sample mass leads to low detection limits due to improvement in the peak to

background ratio (Eppis et al., 2001; Markowicz et al., 1993). However, TXRF is limited by tedious and invasive sample preparation (digestion process) which compromises the integrity of the sample (Kallithrakas-Kontos et al., 2009) and may also lead to loss of vital information especially volatile elements. This study exploits conventional EDXRF instrumentation to address some of the above limitations by developing chemometric multivariate calibration strategies for direct and rapid analysis of selected trace biometals (Mn, Fe, Cu, Zn and Se) in model soft tissues. In this domain, chemometrics aids in modeling multivariate spectroscopic data (Adams, 1995). EDXRF and scattering spectrometry (Angeyo et al., 2012; Kaniu et al., 2012); exploits not only fluorescence peaks, but also scattering profiles for quantitative and qualitative modeling (Compton and Allison, 1935). Previously, Compton scattering has been used for correction of matrix effects through normalization of the fluorescence line (Feather and Willis, 1976; Nielson, 1977).

Biometals are regarded as nutritionally essential for various

* Correspondence to. Department of Physics, University of Nairobi, P.O. Box 30197, Nairobi, Kenya.
E-mail address: okondajustus@uonbi.ac.ke (J.J. Okonda).

biochemical processes in humans and are present in the order of a few micrograms per gram (Prasad, 2013) in proteins, enzymes and cellular membranes (Carvalho et al., 1998; Rose, 2016). They act as activators or inhibitors of enzymatic reactions, enhances competition between elements and metal proteins for binding positions and modification in permeability of cellular membranes (Carvalho et al., 1998; Yaman et al., 2007). The levels of trace biometals in tissues are altered by pathogenesis and metabolic processes. Thus knowledge of the changes in concentrations of certain trace elements in body tissues can be used as biomarkers for disease diagnostics (Kubala-Kukuś et al., 2004). Synchrotron Radiation X-ray Fluorescence (SR-XRF) with high sensitivity has been utilized for determination trace biometals' levels in tissues (Kwiatkiewicz et al., 2004; Silva et al., 2009). Though highly sensitive, SR-XRF analysis of trace biometals is expensive with laborious sample preparations.

In this study, we exploit chemometric enabled energy dispersive X-ray fluorescence (EDXRF) and scattering spectrometry for developing a robust hybrid multivariate calibration strategy for direct rapid and non-invasive quantification of trace biometals in "complex" matrices (soft model tissues) prepared as paraffin wax-embedded thin sections on a Mylar based film. Chemometrics enables direct relation of spectral intensities to analyte concentration through multivariate calibration in which peak overlaps, weak signals and matrix effects are mathematically resolved and compensated (Eriksson et al., 2014). In this study we have utilized Artificial Neural Networks (ANNs) and Principal Component Regression (PCR) for developing the multivariate calibration strategies. ANNs consists of a large number of processing units (neurons) connected together to mimic the biological structure of the brain and nervous system (Liu et al., 1993). On the other hand, PCR performs the regression analysis of concentration based on selected principal components (with greatest variance) of spectral data (Martens and Naes, 1992).

2. Materials and methods

2.1. Sample preparation

Model soft body tissues were prepared from high purity (analytical grade) paraffin wax as the tissue equivalent material (Bethesda, 1989; Ferreira et al., 2010). For each element of interest, the corresponding compound of a given mass was dissolved in 15 ml of ethanol to obtain a stock solution of pre-determined concentrations. Appropriate volumes of stock solutions corresponding to the analyte concentrations spanned for preparation of simulate samples were pipetted and homogeneously mixed in 1.8 ml vials which was then topped to 1.8 ml using ethanol that resulted to a dilution factor of 2 for the initial concentrations. This was done in triplicate (constituting 30 samples) at different concentration within the range in which the biometals occur in native soft body tissues (Banaś et al., 2001; Silva et al., 2012) as Cu: 5–32.5 ppm, Zn: 5.0–67.5 ppm, Fe: 11.5–120 ppm, Mn: 0.8–9.0 ppm and Se: 0.5–6 ppm

The multi-element solutions of predetermined concentration were pipetted onto steel moulds and appropriate amounts of acetone (universal secondary fixative with low toxicity and freely miscible with water and organic solvents) added together with 0.2 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) and heated to a temperature of 56°C in a tissue cassette (for acetone to evaporate) thus leaving the elements of interest embedded in paraffin wax matrix which mimics the soft tissue. The mixture was further stirred for about 3–5 min to increase homogeneity and then cooled on a cold plate of the tissue cassette. The solid simulate samples were dried on steel moulds at room temperature until solidified and then kept in sealed labeled plastic containers. For each sample, 10 μm thick sections were made using a microtome and mounted on a 2 μm Mylar foil with same sample geometry maintained for EDXRF

analysis.

In order to evaluate the accuracy our method, a certified reference material (NIST Oyster tissue1566b) was dried at room temperature in a bio-safety cabinet to a constant mass and subjected to similar sample preparation for EDXRF and scattering analysis. The Oyster tissue samples were also analyzed by conventional classical EDXRF method and analyzed by AXIL-QXAS algorithm (Bernasconi, 1993) in the fundamental parameters approach for comparison with the results obtained by our method. The analysis conditions were; irradiation time of 50 s and applied voltage of 5–50 kV in the X-ray tube current of automated current in the range 0.001–1 mA.

2.2. Multivariate calibration

This study is mainly focused on the complementary (nested) utility of the two chemometric approaches namely: ANN and PCR, for multivariate calibration utilizing EDXRF and scattering spectrometry. Multivariate calibration involves development of a mathematical model that furnishes specific sample properties (De Oliveira et al., 2010) describing the relation between the EDXRF (and scatter) spectral data and the analyte concentration. It consists of modeling, which establishes a mathematical relation between X (spectral data) and Y (concentration) by determining the regression coefficient which can be used to reproduce Y given X , and validation which involves optimizing the model for prediction of the same elements in unknown samples of similar matrix composition (Nagata et al., 2006)

The neural network was constituted by three node layers; input, hidden and output with a large number of parallel connected neurons (Kallithrakas-Kontos et al., 2009). The multi-input/output relationships mimic human cognitive processes using Back propagation (BP) ANN training function with Levenberg-Marquardt (LM) algorithm for training the feed forward networks for function approximation (non-linear regression). As compared to other algorithms, BP-ANN was found appropriate enhances lower mean square errors and faster convergence (Beale et al., 2012).

The EDXRF and scatter trace biometal spectral responses in the training data set were input into the first layer in the network to yield outputs X_i . The net input Net_j is given by (Luo et al., 1997);

$$Net_j = \sum (x_i W_{ij}) + \theta_{ij} \quad (1)$$

where W_{ij} and θ_{ij} are weight and node threshold parameters respectively.

After successive transformation by nonlinear sigmoid function f to the node k in the last layer, the output (OUT_j) of the node j is given by;

$$OUT_k = f(Net_k) \quad (2)$$

The output (OUT_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 (Liu et al., 1993)

$$Y_k = F(OUT_k) + \varepsilon \quad (3)$$

where F is the network training function and ε the error of calibration as a result of the residuals.

The same trace biometal spectral responses were also analyzed by PCR in a two-step process. PCR considers the most descriptive principal components (PC) obtained by Principal Component Analysis (PCA) as independent variables instead of adopting original variables (Pires et al., 2008). The PCA analysis decomposes the raw data matrix, X into two matrices, T and V , (Buono et al., 2005).

$$X = TV^T + \varepsilon \quad (4)$$

where T is the score matrix which represents the position of the samples in the new coordinate system and contains elemental concentration information while V is the loadings matrix and describes the hidden profiles that are common to all measured data – i.e. in the case of EDXRF and scatter spectra, the loadings contain qualitative information

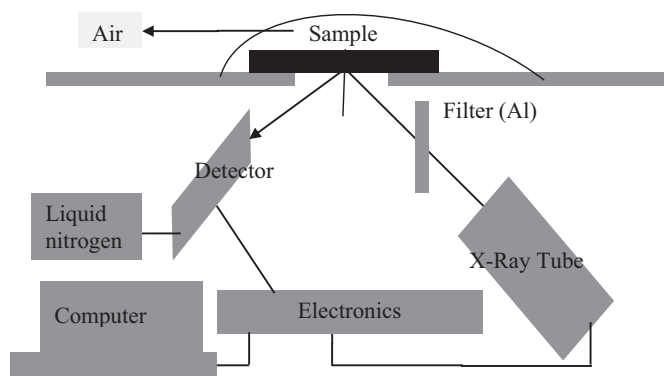


Fig. 1. Schematic presentation of EDXRF instrumentation.

Table 1
Selected spectral region (feature selection) for multivariate calibration.

Element	Peak	Energy (KeV)	Channel numbers (spectral region of interest)
Mn	K_{α}	5.89	287–296
	K_{β}	6.49	320–329
Fe	K_{α}	6.40	315–324
	K_{β}	7.06	347–356
Cu	K_{α}	8.04	397–406
	K_{β}	8.90	440–449
Zn	K_{α}	8.63	428–437
	K_{β}	9.57	474–483
Se	K_{α}	11.21	561–568
	K_{β}	12.49	618–627

Table 2
Detection limits (ppm) for the direct analysis of indicated analytes in model soft tissue.

Element	K_{α} energy (keV)	Atomic number, Z	Detection limit (ppm)
Mn	5.89	25	3.2
Fe	6.40	26	10.0
Cu	8.04	29	4.5
Zn	8.63	30	4.6
Se	11.21	34	1.6

such as elements giving rise to the spectra, and ϵ the residual. The unique scores from each dataset were used to regress back against the known concentration matrix for quantitative prediction given by

$$\hat{y}_i = x_i \hat{V} \quad (5)$$

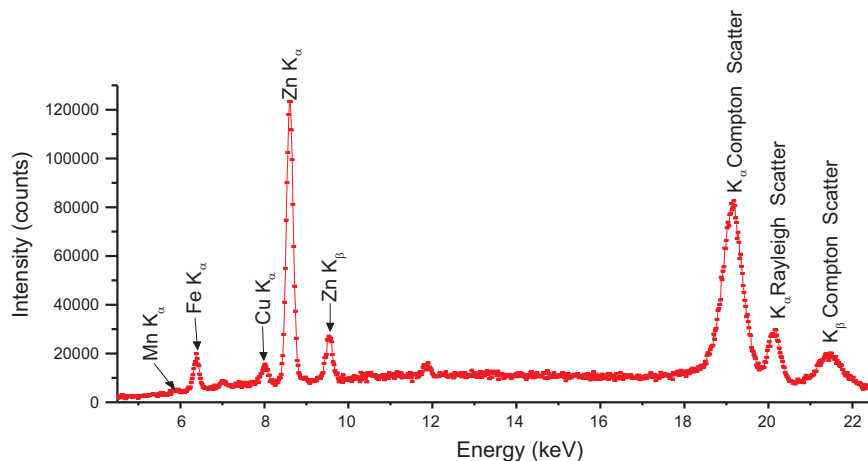


Fig. 2. Typical spectrum of certified reference material (NIST 1566b, Oyster tissue) at 50 s.

where \hat{y}_i is the predicted concentration of sample i , and \hat{V} the loadings of the selected PCs of X .

The root mean squared error of prediction (RMSEP) was also determined for the expected average error when constituent concentrations of samples with similar matrix composition are to be predicted (Facchin et al., 1999).

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_{pred,i} - y_i)^2}{n}} \quad (6)$$

where $y_{pred,i}$ is the predicted concentration and y_i the known concentration for sample i , and n the number of samples.

2.3. Spectral analysis

Fig. 1 shows a typical schematic diagram of Shimadzu EDX-800HS model CE (212–23701-36) spectrometer that was used in this study.

The spectrometer is equipped with a Rhodium (Rh) X-ray tube and aluminum (Al) filter and collimation diameter of 10 mm spectrometer. In order to minimize attenuation effects, each sample was placed at about 45° to the incident radiation (Tertian and Claisse, 1982), and irradiated at 50 s (time optimized for rapid analysis and reasonable signal to noise ratio of analytes). The spectral measurements were made in the range 0–40 keV using a multi-channel analyzer in air at dead time of < 25% with current and voltage were adjusted automatic during spectral acquisition. Both the characteristic fluorescence and scattered radiation from the sample were detected by the Si (Li) detector at about 90° (for strong linear polarization of radiation beam for enhancing signal to noise ratio) to the incident beam with 140 eV full width at half maximum (FWHM) for $Mn K_{\alpha}$ at 5.9 KeV. For each spectrum, 2047 channels were acquired under same irradiation conditions. Standard reference material (NIST Oyster tissue1566b) samples were also analyzed in triplicate under same irradiation conditions as the model tissues.

Peak identification and spectrum deconvolution was done on a representative simulate sample using the AXIL (Analysis of X-ray spectra by Iterative Least-square fitting) for feature selection of the fluorescence spectral signature positions corresponding to the selected trace biometals (Beckhoff et al., 2007). The independent variables were constructed in such a way that the rows corresponded to samples and columns to the spectral signature intensities corresponding to selected spectral regions of each trace element of interest as shown in Table 1.

Additionally, K_{α} Compton scatter region corresponding to channel numbers 940–980 and energy 19.15keV of Rh K_{α} was selected. Biomedical samples are mostly composed of the “dark matrix” (low-Z matrix due to carbon, hydrogen and oxygen (C-H-O) and low-Z elements ($Z < 10$) which dominates Compton scatter (Knoll, 2010; Verbi et al., 2005). The Compton peak being much broader with non-

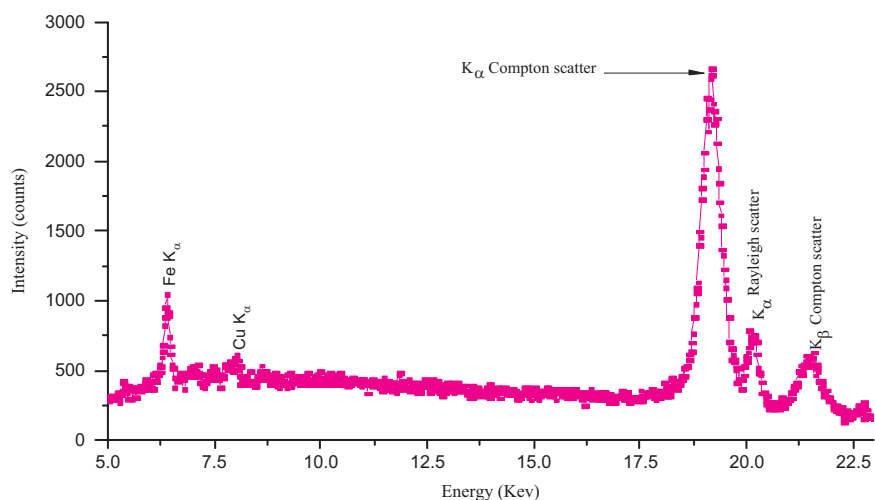


Fig. 3. Typical EDXRF spectra of domestic dog tissue obtained at 50 s irradiation time.

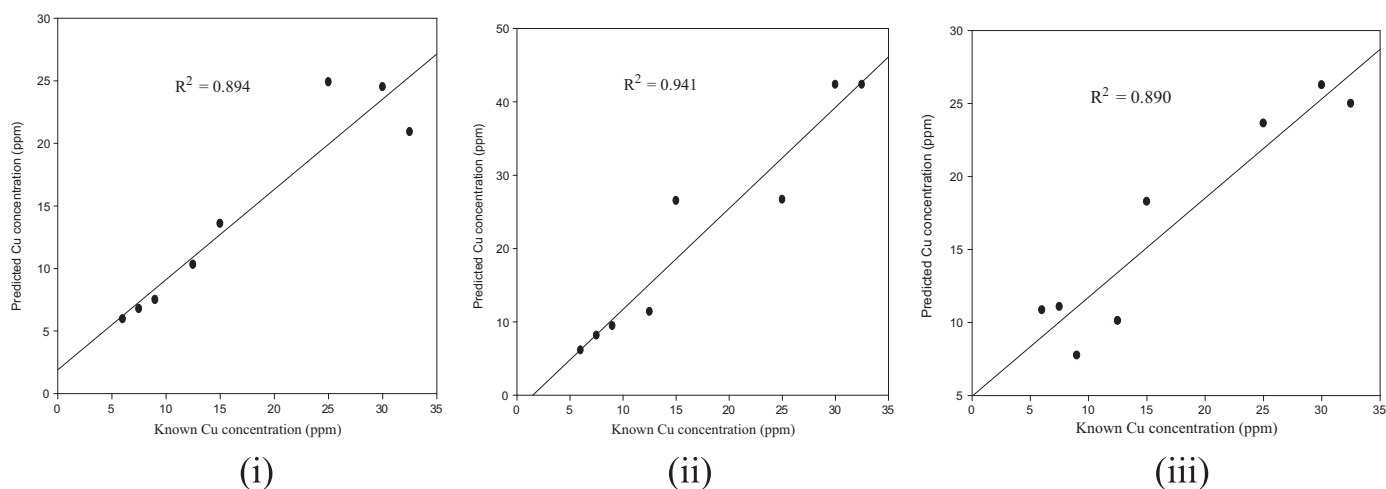


Fig. 4. ANN regression plots for predicted versus known Cu utilizing (i) multiple fluorescence peaks, (ii) multiple fluorescence peaks including Compton scatter and (iii) multiple fluorescence peaks normalized to the Compton scatter.

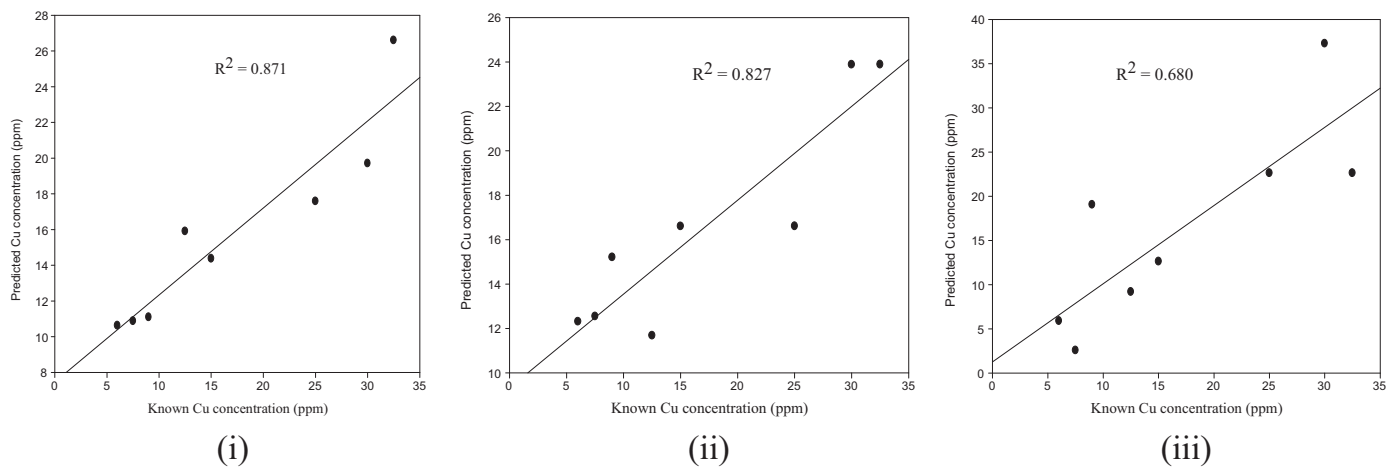


Fig. 5. PCR regression plots for predicted versus known Cu utilizing (i) multiple fluorescence peaks, (ii) multiple fluorescence peaks including Compton scatter and (iii) multiple fluorescence peaks normalized to the Compton scatter.

Gaussian shape makes it difficult for determination of intensity by spectrum evaluation programs (Markowicz et al., 1993) which relies on Gaussian peak shapes. The Rayleigh peak is fairly described by a Gaussian width which corresponds to the spectrometer resolution. However, owing to its partial overlap with the Compton region, the

Rayleigh peak is also difficult to model (Van Gysel et al., 2003). The inclusion of Compton scatter was necessary for compensation of the strong matrix effects through Compton normalized fluorescence peaks (Kaniu et al., 2012) and direct analytical exploitation of spectral data by chemometrics-enabled energy dispersive X-ray fluorescence and

Table 3
Regression coefficients of ANN and PCR for model soft tissue.

Element	Regression coefficients (R^2 values)							
	Fluorescence				Multiple fluorescence + Compton scatter		Multiple fluorescence normalized to the Compton scatter	
	ANN		PCR		ANN	PCR	ANN	PCR
	K_α	$K_\alpha + K_\beta$	K_α	$K_\alpha + K_\beta$				
Mn	0.816	0.923	0.752	0.906	0.679	0.946	0.886	0.998
Fe	0.931	0.880	0.550	0.850	0.916	0.869	0.921	0.936
Cu	0.789	0.894	0.859	0.871	0.941	0.827	0.890	0.680
Zn	0.777	0.876	0.301	0.830	0.995	0.688	0.718	0.772
Se	0.621	0.814	0.366	0.875	0.613	0.318	0.784	0.635

scattering spectrometry method.

As opposed to classical XRF analysis where only a single fluorescence line (normally the one with the highest fluorescence yield, good signal-to-noise ratio and free from spectral overlaps) is used, in this approach multiple fluorescence peaks regardless of spectral overlaps (Kaniu et al., 2011) are exploited in multivariate domain. The selected spectral set (fluorescence and/or the Compton scatter peaks) were pre-processed though normalization and mean centering to minimize noise and enable similar sample contribution to the calibration strategy.

ANN model training involved four stages: (i) assembling of training data, (ii) creation of the network, (iii) training the network, and (iv) simulating the network response to new inputs (Angeyo et al., 2012) of the similar matrices. Back propagation (BP-ANN) training function with Levenberg-Marquardt (LM) algorithm was used for feed forward networks training with an aim of function approximation (Banaś et al., 2001). The input data was divided into three sets; 60% for network training, 20% for validating and 20% for testing how well the network will perform on new data sets of similar matrices. The predictive ability of ANN multivariate calibration was achieved through optimization of the neural network architecture as the models were trained, tested and validated. This was aimed at achieving the smallest training error as a function of the added neurons to the intermediate layer. The best model was initialized as that corresponding to high R value, low mean-squared error and 4 neurons in the hidden layer.

A complementary PCR calibration strategy was also developed using the same spectral data and was validated by using leave-one-out cross validation method (Full Cross validation method) in the PCA. The regression of Y (concentrations) was based on selected principal components of PCA (PC1 and PC2) instead of the original X-variables (Jolliffe, 1982) for relating X-variables to the dependent Y-variable(s). The regression coefficients were obtained by regressing the 2 Pcs against the simulated concentrations to predict the concentrations of the sample constituents in the test data set. Spectra modeling by multivariate

Table 4
Comparison of RMSEP for ANN and PCR for simulate tissue.

Element	Root mean square error of prediction (RMSEP)							
	Multiple fluorescence				Multiple fluorescence + Compton scatter		Multiple fluorescence normalized to the Compton scatter	
	ANN		PCR		ANN	PCR	ANN	PCR
	K_α	$K_\alpha + K_\beta$	K_α	$K_\alpha + K_\beta$				
Mn	1.259	1.257	0.457	1.509	1.259	0.456	2.051	0.253
Fe	15.039	3.795	17.748	3.844	15.039	17.714	10.782	20.859
Cu	3.944	0.987	6.121	4.302	3.943	4.601	2.318	6.348
Zn	1.456	5.576	11.362	9.368	2.061	11.363	9.244	20.646
Se	0.599	0.938	0.961	1.135	0.847	1.245	0.69	0.837

Table 5
Regression coefficients for determination of trace biometals in soft tissue.

Element	Best regression coefficients (R^2)	RMSEP (ppm)	Method	Spectral region utilized
Mn	0.998	0.253	PCR	Normalized fluorescence
Fe	0.880	3.795	ANN	Multiple fluorescence
Cu	0.924	0.987	ANN	Fluorescence + Compton
Zn	0.995	2.061	ANN	Fluorescence + Compton
Se	0.814	0.938	ANN	Multiple fluorescence

chemometric techniques (ANN and PCR) realized detection and quantification of low signal to noise ratio (SNR) analytes in complex matrices through direct transformation of EDXRF and scatter spectra to trace biometal concentrations. The normally strong matrix effects and overlapping spectral lines were mathematically resolved.

The applicability of the chemometric aided EDXRF and scatter method for analysis of trace biometals in model soft tissues was ascertained by calculating the detection limits DL (Table 2), of the biometals in the Standard Reference Material (NIST 1566b, oyster tissue) using the following equation (Debertin and Helmer, 1988; Markowicz et al., 1993);

$$DL = \frac{3}{S} \sqrt{\frac{I_b}{T_b}} \quad (7)$$

where I_b is the background counts under the characteristic peaks, T_b the irradiation time and S is the elemental sensitivity.

It was assumed that the concentrations of trace biometals in any typical soft tissue which are below the detection limits in Table 2 above are impossible to quantify directly by classical EDXRF analysis (perhaps also by the chemometrics-enabled energy dispersive X-ray fluorescence and scattering spectrometry method).

2.4. Results and discussion

Spectral deconvolution enabled feature selection of observable peak positions corresponding to the biometals in Oyster tissue as shown in Fig. 2.

The predominant features evident in the above spectrum are the X-ray fluorescence peaks (K_α and K_β) and enhanced Compton scatter peak. The enhanced background due to the ‘dark matrices’ in Oyster tissue is a limitation for accurate and direct quantitative analysis of low concentration analytes in the absorbing and interfering matrix.

Fig. 3 shows a typical spectrum of domestic dog tissue sample with similar matrices to the model soft tissue.

The immediate apparent features in the above domestic dog spectrum (for a typical soft body tissue) is the enhanced Fe peak and barely discernible fluorescence peaks of Zn, Cu, Mn and Se. The prominent Fe

Table 6
Biometal concentrations in CRM Oyster tissue (NIST1566b) utilizing EDXRF and scattering approach and classical EDXRF fundamental parameter method.

Element	Certified value ± standard deviation (ppm)	EDXRF-FP value ± Standard deviation (ppm)	EDXRF and scattering measured value ± standard deviation (ppm)	EDXRF and scattering % deviation from certified values
Mn	18.5 ± 0.2	44.7 ± 7.3	16.9 ± 1.2	5.6
Fe	205.8 ± 6.8	1583.5 ± 91.8	238 ± 15.6	4.9
Cu	71.6 ± 1.6	51.4 ± 4.2	66.8 ± 4.6	6.7
Zn	1424 ± 46	1333.8 ± 68.8	1280.5 ± 29.0	6.8
Se	2.06 ± 0.15	18.6 ± 2.4	3.6 ± 1.1	31.1

peak can be attributed to its high concentration in tissues based on its significant role in blood circulation in body tissues (Silva et al., 2009). In such cases, the trace elements' fluorescence peaks of interest are hardly discernible against the high background. The Compton peak is observed to be intense which underscores its potential utility in our EDXRF and scattering spectrometric modeling approach in soft body tissues. The major challenge is direct and rapid determination with good accuracy the trace biometals of interest in soft tissues whose spectral features resemble those shown in Fig. 3 above. The ratio of characteristic and Compton scatter intensities is less sensitive to variation in sample composition and thus it can be used to mathematically eliminate strong matrix effects (Markowicz, 2011) thus a linear function of the concentration of the analytes. Additionally, the characteristic fluorescence and the Compton scatter peaks can be simultaneously used to determine the concentration of the analytes in soft body tissues.

The calibration strategy curves of ANN and PCR models of Cu in model tissue samples are shown in Figs. 4 and 5 respectively. They indicate how well (close correlation between measured and predicted trace elements at low detection limits) the models are likely to perform for future prediction evaluation of Cu in tissue samples of similar matrix composition in multivariate domain.

From the above calibration curves using the validation set of model tissue samples, it is clear that ANN model prediction for Cu in model soft tissue is better compared to PCR model. The utility of multivariate fluorescence peaks and Compton scatter peaks in PCR leads to poorer model performance as opposed to the use of multivariate fluorescence peaks. This can probably be attributed to the linearity of PCR model which may not be appropriate for modeling the non-linear relations in multivariate space between the trace biometal concentration and the spectral signatures (fluorescence and Compton scatter) combined as model input data. However, ANN model training with purlin linear function and sigmoid non-linear function used in the hidden layer of neural network resolved both linear and non-linear relations between several responses (sample spectral data matrix) and vector of properties (concentrations of analytes) in the tissue samples (Markowicz, 2011). It is noted that ANN is more robust in reducing matrix effects for direct rapid analysis of Cu and other analytes in soft tissues. The ANN and PCR regression coefficients for all the trace elements of interest (Fe, Mn, Cu, Zn and Se) are summarized in Table 3.

The ANN and PCR multivariate calibration results of multiple fluorescence spectra utilizing K_{α} and K_{β} fluorescence peaks are better (high regression coefficients) than the utility of K_{α} fluorescence. Quantification accuracy is therefore independent of signal-to noise ratio of the analyte fluorescence signatures. The results show enhanced prediction accuracy for both fluorescence and Compton scatter in model development as opposed to K_{α} fluorescence peaks (see Fe, Cu, Zn in the ANN model, and Mn in the PCR model). Compton fluorescence normalized spectral data increased the accuracy even more. The regression coefficients of trace biometals in Table 3 above gives guidance on the choice of the best quantitative model for a specified trace biometal in model soft tissue in the nested model approach. The performance not only provides potential for the simultaneous utilization of both fluorescence and Compton scatter, but also normalization of the fluorescence to the Compton in direct rapid quantitative analysis of trace elements in complex biological matrices. For instance, the quantitative determination of Mn in model tissues can be best achieved using PCR model incorporating use of Compton scatter ($R^2 = 0.998$) and Zn using ANN model ($R^2 = 0.995$) utilizing Compton scatter normalized fluorescence and selected fluorescence (K_{α} and K_{β}) peaks.

The quality of multivariate analytical modeling is not only given by regression coefficients (R^2) but also the root mean square error of prediction (RMSEP) as shown in Table 4. The RMSEP describes how ANN and PCR may be used for prediction of the trace biometals in unknown tissues of matrix composition closely matching the model soft tissues.

In general, the results in Table 5 indicate however that, the overall

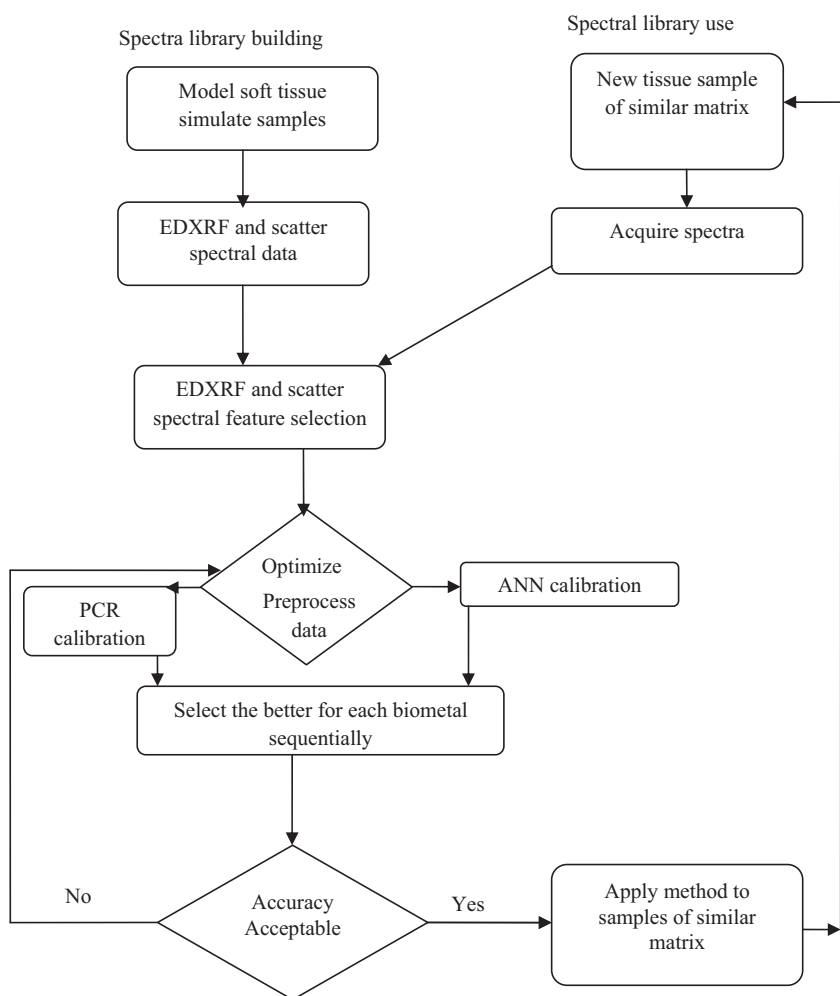


Fig. 6. Nested multivariate calibration model for model soft tissue samples.

ANN model predictions of the concentrations of Mn, Fe, Cu, Zn and Se are better (more accurate for most elements of interest with $R^2 > 0.900$) than the corresponding values obtained through PCR model. This can be attributed to the fact that ANN has the capability to model both linear and non-linear relations (concentration in relation to fluorescence peaks and Compton scatter spectral signatures). The R^2 values were selected based on the low RMSEP in the validation where both the information in Tables 3, 4 were used to make a compromised hybrid selection of the optimal calibration model (Table 5) for direct rapid analysis of each biometal of interest in soft tissue.

The robustness and accuracy of the developed calibration strategy was evaluated by analysis of the trace metals in the standard reference material (Oyster tissue NIST 1566b) using fundamental parameters (FP) method based classical EDXRF for comparison with our method (Table 6).

The EDXRF and scattering measured concentrations of most of the trace biometals in Oyster tissue in Table 6 are in agreement with the certified values to within <10% except for Se. The poor results for Se is probably due to low concentration of Se in the Oyster tissue close to its detection limit (1.6 ppm) thus low fluorescence yield and so the developed analytical model may not be robust enough for such cases. The inapplicability of the fundamental parameter method EDXRF analysis to such tasks (direct rapid simultaneous determination of trace biometals) in complex matrix samples (model soft tissues) is apparent in Table 6 as the concentrations of Mn, Fe, Cu and Se are far from the certified values. This is due to the fact that FP-EDXRF assumes X-ray fluorescence peak intensity (of a well resolved single fluorescence line with appreciable signal-to-noise ratio) as a simple linear function of

concentration (He and Van Espen, 1990). This is not easily realized in direct trace element analysis of biological samples. The results demonstrate that EDXRF and scattering spectrometry coupled with chemometrics is superior alternative to classical method for direct rapid analysis of trace biometals in model soft tissues.

Fig. 6 shows a nested multivariate calibration model for quantitative analysis of trace biometals in model soft tissues. It may be noted that the combined utility of multiple fluorescence and Compton scatter peaks, in EDXRF and scattering spectrometry in conjunction with multivariate chemometric techniques, offers a versatile tool for direct rapid trace elements analysis in soft body tissues. The method is non-invasive (avoids the sample digestion process), affordable and is capable of mining additional information (in Compton scatter) to elemental profiles in the tissues.

2.5. Conclusion

This study was undertaken to develop and evaluate the potential of energy dispersive X-ray fluorescence and scattering (EDXRFS) spectrometry in conjunction with multivariate chemometric techniques, namely ANN and PCR for direct rapid simultaneous analysis of trace biometals in model soft tissues. The method is rapid (involves minimum sample preparation with no sample pre-treatments, 50 s analysis time), cost-effective (as it can also be easily done using radioisotope sources), direct and is applicable to simultaneous determination of biometals at trace levels in complex matrices. The results clearly demonstrate the nested approach (ANN and PCR) multivariate calibration as an efficient method for quantitative analysis of trace elements in model soft tissues.

The study provides proof of concept for multivariate calibration that needs further studies with more model tissue samples in order to consolidate the present conclusions.

Acknowledgement

The authors wish to thank the University of Nairobi for financial support accorded to this project in the form of an M.Sc fellowship (to the first author) that facilitated this study. They also express their gratitude to Mr. Charles Magiri and Mrs. Martha Maina both of Histopathology Laboratory of the Kenya Medical Research Institute (KEMRI), and Mr. Simon Bartilol of the Institute of Nuclear Science and Technology, University of Nairobi, for their assistance during sample preparation and the technical aspects of EDXRF analysis respectively.

References

- Adams, M.J., 1995. *Chemometrics in Analytical Spectroscopy*. The Royal Society of Chemistry.
- Angeyo, K., Gari, S., Mustapha, A., Mangala, J., 2012. Feasibility for direct rapid energy dispersive X-ray fluorescence (EDXRF) and scattering analysis of complex matrix liquids by partial least squares. *Appl. Radiat. Isot.* 70 (11), 2596–2601.
- Angeyo, K., Patel, J., Mangala, J., Narayana, D., 1998. Radioisotope photon-excited energy dispersive X-ray fluorescence technique for the analysis of organic matrices. *X-Ray Spectrom.* 27 (3), 205–213.
- Banaś, A., Kwiatek, W., Zajac, W., 2001. Trace element analysis of tissue section by means of synchrotron radiation: the use of GNUMPLOT for SRXRF spectra analysis. *J. Alloy. Compd.* 328 (1), 135–138.
- Beale, M.H., Hagan, M.T., Demuth, H.B., 2012. *Neural network toolbox™ user's guide*. Paper presented at the R2012a. The MathWorks, Inc., 3 Apple Hill Drive Natick, MA 01760-02098, <www.mathworks.com>.
- Beckhoff, B., Kannigieser, B., Langhoff, N., Wedell, R., Wolff, H., 2007. *Handbook of Practical X-ray Fluorescence Analysis*. Springer Science & Business Media.
- Bernasconi, G., 1993. IAEA AXIL-QXAS Version Reference Manual. IAEA, Vienna.
- Bethesda, M., 1989. Tissue substitutes in radiation dosimetry and measurement. In: *Proceedings of the International Commission on Radiation Units and Measurements (ICRU)*.
- Bueno, M.I.M.S., Castro, M.T., de Souza, A.M., de Oliveira, E.B.S., Teixeira, A.P., 2005. X-ray scattering processes and chemometrics for differentiating complex samples using conventional EDXRF equipment. *Chemom. Intell. Lab. Syst.* 78 (1), 96–102.
- Carvalho, M., Brito, J., Barreiros, M., 1998. Study of trace element concentrations in human tissues by EDXRF spectrometry. *X-Ray Spectrom.* 27 (3), 198–204.
- Compton, A.H., Allison, S.K., 1935. *X-rays in Theory and Experiment*.
- De Oliveira, L., Antunes, A., Bueno, M., 2010. Direct chromium speciation using X-ray spectrometry and chemometrics. *X-Ray Spectrom.* 39 (4), 279–284.
- Debertin, K., Helmer, R., 1988. *Gamma and X-ray Spectrometry with Semiconductor Detectors*. Elsevier Science Publishers BV, Amsterdam, The Netherlands.
- Eppis, R., Padelfano, R., Lopez, C., Verboso, C., Bianchi, L., Gandia, S., 2001. Some application of TXRF in food, medicinal products and nuclear material. In: *Proceeding of the 3rd Paper Presented at the International Symposium on Nuclear and Related Techniques*. Havana, Cuba.
- Eriksson, L., Trygg, J., Wold, S., 2014. A chemometrics toolbox based on projections and latent variables. *J. Chemom.* 28 (5), 332–346.
- Facchin, I., Mello, C., Bueno, M., Poppi, R., 1999. Simultaneous determination of lead and sulfur by energy-dispersive x-ray spectrometry. Comparison between artificial neural networks and other multivariate calibration methods. *X-Ray Spectrom.* 28 (3), 173–177.
- Feather, C., Willis, J., 1976. A simple method for background and matrix correction of spectral peaks in trace element determination by X-ray fluorescence spectrometry. *X-Ray Spectrom.* 5 (1), 41–48.
- Ferreira, C.C., Ximenes, R., Garcia, C.A.B., Vieira, J.W., Maia, A.F., 2010. Total mass attenuation coefficient evaluation of ten materials commonly used to simulate human tissue. In: *Proceedings of the Paper presented at the Journal of Physics: Conference Series*.
- He, F., Van Espen, P., 1990. An integrated system for quantitative EDXRF analysis based on fundamental parameters. *Nucl. Instrum. Methods Phys. Res. Sect. A: Accel. Spectrom. Detect. Assoc. Equip.* 299 (1–3), 580–583.
- Jolliffe, I.T., 1982. A note on the use of principal components in regression. *Appl. Stat.* 300–303.
- Kallithrakas-Kontos, N., Koulouridakis, P., Hatzistavros, V., Aretaki, I., 2009. Chromium speciation by TXRF analysis. *X-Ray Spectrom.* 38 (2), 152–156.
- Kaniu, M., Angeyo, K., Mangala, M., Mwala, A., Bartilol, S., 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., Angeyo, K., Mwala, A., Mangala, M., 2012. Direct rapid analysis of trace bioavailable soil macronutrients by chemometrics-assisted energy dispersive X-ray fluorescence and scattering spectrometry. *Anal. Chim. Acta* 729, 21–25.
- Knoll, G.F., 2010. *Radiation Detection and Measurement*. John Wiley & Sons.
- Kubala-Kukuś, A., Kuternoga, E., Braziewicz, J., Pajek, M., 2004. Log-stable concentration distributions of trace elements in biomedical samples. *Spectrochim. Acta Part B: At. Spectrosc.* 59 (10), 1711–1716.
- Kwiatek, W., Hanson, A., Paluszkiwicz, C., Gałka, M., Gajda, M., Cichocki, T., 2004. Application of SRXRF and XANES to the determination of the oxidation state of iron in prostate tissue sections. *J. Alloy. Compd.* 362 (1), 83–87.
- Liu, Y., Upadhyaya, B.R., Naghedolfeizi, M., 1993. Chemometric data analysis using artificial neural networks. *Appl. Spectrosc.* 47 (1), 12–23.
- Luo, L., Guo, C., Ma, G., Ji, A., 1997. Choice of optimum model parameters in artificial neural networks and application to X-ray fluorescence analysis. *X-Ray Spectrom.* 26 (1), 15–22.
- Markowicz, A., 2011. An overview of quantification methods in energy-dispersive X-ray fluorescence analysis. *Pramana J. Phys.* 76 (2), 321–329.
- Markowicz, A.A., Van Grieken, R., Markowicz, A., 1993. *Handbook of X-ray Spectrometry*. Marcel Dekker, New York.
- Martens, H., Naes, T., 1992. *Multivariate Calibration*. John Wiley & Sons.
- Nagata, N., Peralta-Zamora, P.G., Poppi, R.J., Perez, C.A., Bueno, M.I., 2006. Multivariate calibrations for the SR-TXRF determination of trace concentrations of lead and arsenic in the presence of bromine. *X-Ray Spectrom.* 35 (1), 79–84.
- 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.
- Pires, J., Martins, F., Sousa, S., Alvim-Ferraz, M., Pereira, M., 2008. Selection and validation of parameters in multiple linear and principal component regressions. *Environ. Model. Softw.* 23 (1), 50–55.
- Prasad, A.S., 2013. *Essential and Toxic Element: Trace Elements in Human Health and Disease*. Elsevier.
- Rose, J., 2016. *Trace Elements in Health: a Review of Current Issues*. Butterworth-Heinemann.
- Silva, M., Tomal, A., Perez, C., Ribeiro-Silva, A., Poletti, M., 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, M.P., Soave, D.F., Ribeiro-Silva, A., Poletti, M.E., 2012. Trace elements as tumor biomarkers and prognostic factors in breast cancer: a study through energy dispersive X-ray fluorescence. *BMC Res. Notes* 5 (1), 194.
- Tertian, R., Claisse, F., 1982. *Principles of Quantitative X-ray Fluorescence Analysis*.
- Van Gysel, M., Lemberge, P., Van Espen, P., 2003. Description of Compton peaks in energy-dispersive X-ray fluorescence spectra. *X-ray Spectrom.* 32 (2), 139–147.
- Verbi, F.M., Pereira-Filho, E.R., Bueno, M.I.M., 2005. Use of X-ray scattering for studies with organic compounds: a case study using paints. *Microchim. Acta* 150 (2), 131–136.
- Yaman, M., Kaya, G., Yekeler, H., 2007. Distribution of trace metal concentrations in paired cancerous and non-cancerous human stomach tissues. *World J. Gastroenterol.* 13 (4), 612.