

United Arab Emirates University

Scholarworks@UAEU

Dissertations

Electronic Theses and Dissertations

1-2020

**ENHANCED LABEL FREE NORMAL AND CANCER CELLS
CLASSIFICATION USING COMBINED PARAMETRIC MODELING
AND OPTICAL TECHNIQUES**

Ayshathul Fouzia Abdul Gani

Follow this and additional works at: https://scholarworks.uaeu.ac.ae/all_dissertations



Part of the [Biological Engineering Commons](#)

United Arab Emirates University

College of Engineering

ENHANCED LABEL FREE NORMAL AND CANCER CELLS
CLASSIFICATION USING COMBINED PARAMETRIC MODELING
AND OPTICAL TECHNIQUES

Ayshathul Fouzia Abdul Gani

This dissertation is submitted in partial fulfilment of the requirements for the degree
of Doctor of Philosophy

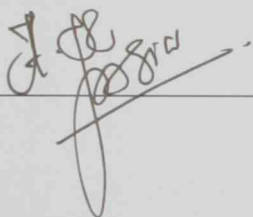
Under the Supervision of Dr. Mahmoud F. Al Ahmad

January 2020

Declaration of Original Work

I, Ayshathul Fouzia Abdul Gani, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this dissertation entitled "*Enhanced label free normal and cancer cells classification using combined parametric modeling and optical techniques*", hereby, solemnly declare that this dissertation is my own original research work that has been done and prepared by me under the supervision of Dr. Mahmoud F. Al Ahmad, in the College of Engineering at UAEU. This work has not previously been presented or published, or formed the basis for the award of any academic degree, diploma or a similar title at this or any other university. Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my dissertation have been properly cited and acknowledged in accordance with appropriate academic conventions. I further declare that there is no potential conflict of interest with respect to the research, data collection, authorship, presentation and/or publication of this dissertation.

Student's Signature: _____



Date: _____

27-FEB-2020

Copyright © 2020 Ayshathul Fouzia Abdul Gani
All Rights Reserved

Approval of the Doctorate Dissertation

This Doctorate Dissertation is approved by the following Examining Committee Members:

- 1) Advisor (Committee Chair): Dr. Mahmoud F. Al Ahmad

Title: Associate Professor

Department of Electrical Engineering

College of Engineering

Signature 

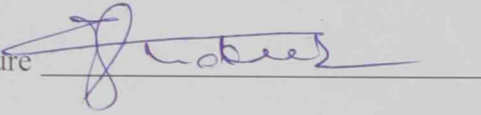
Date 25/02/20

- 2) Member: Dr. Mohammad Shakeel Laghari

Title: Associate Professor

Department of Electrical Engineering

College of Engineering

Signature 

Date 25-02-2020

- 3) Member: Dr. Adel Najar

Title: Associate Professor

Department of Physics

College of Science

Signature 

Date 25/02/2020

- 4) Member (External Examiner): Professor Dominique M. Durand

for Title: EL Lindseth Professor of Biomedical Engineering

Department of Biomedical Engineering, Physiology, Biophysics

Institution: Case Western Reserve University, Ohio, USA.

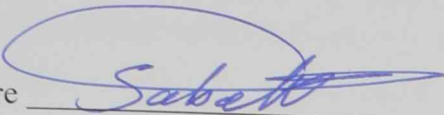
Signature 

Date 25/02/2020

Dr. Tamen El-Maadawy
PhD Coordinator

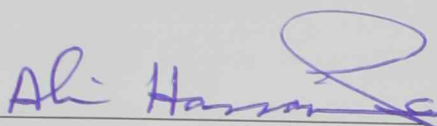
This Doctorate Dissertation is accepted by:

Dean of the College of Engineering: Professor Sabah Alkass

Signature 

Date 27/2/2020

Dean of the College of Graduate Studies: Professor Ali Al-Marzouqi

Signature 

Date 27/2/2020

Advisory Committee

1) Advisor: Dr. Mahmoud F. Al Ahmad

Title: Associate Professor

Department of Electrical Engineering

College of Engineering

2) Member: Dr. Tahir A. Rizvi

Title: Professor

Department of Microbiology

College of Medicine and Health Sciences

3) Member: Dr. Farah Mustafa

Title: Associate Professor

Department of Biochemistry

College of Medicine and Health Sciences

Abstract

Development of label-free methods for cell classifications has been driven by the importance of early detection and identification of cancer disease. The future point-of-care (POC) treatment methods require rapid and real-time cancer screening techniques. As the labelled methods of cell classification are time-consuming process and require large amount of sample preparation along with skilled persons, they do not appear to be suitable for POC treatment methods. This necessitates the importance of such development. The label-free methods incorporate the biophysical properties of cells instead of biomarkers. The optical properties of cells have been frequently utilized for cell classification. This is due to their capability to interact with light. This interaction depends strongly on intrinsic properties and composition of cells.

Cells from different tissues as well as normal and cancerous of same tissue exhibit different optical profiles. Therefore, the objective of this work is to combine the optical techniques with numerical methods to enhance the accuracy in classifying different type of cells. The variation in light interactions with different type of cells is studied and the observations are further analyzed using numerical methods. Prony and autoregressive (AR) techniques are used to extract set of parameters such as poles and coefficients, to enable cell classifications.

For demonstration, six types of cells: lung normal, lung cancer, liver normal, liver cancer, kidney normal, and cervical cancer cells are considered in this work. Their corresponding optical signals have been measured. The measured signals are then estimated and approximated using Prony and AR models. It is shown that the variation in the extracted poles and coefficients for different type of cells form a vital tool in cell classification enhancement. Statistical tool such as analysis of variance (ANOVA) helps in determining the significant AR coefficients.

The results revealed that the poles obtained through the Prony method for different cells differ in their magnitude and location. A figure of merit (FOM) is developed and adapted here which correlates the magnitude and location of poles. It is found that the distribution of FOM in complex z -plane is closer to the center of the unit circle for normal cell lines than for cancer cell lines taken from the same tissue. Furthermore, the AR model of same order for different types of cells exhibit different coefficient

and pole values. To reduce redundancy and to arrive with a concise AR model (order optimization), ANOVA analysis has been used to determine the significance in the AR coefficients. After that, the dominant poles have been determined. With optimizing the order, the differences in the pole values of normal and cancer cell increases, enabling cell classification enhancement. This shows the role of statistical tools is a further enhancement for better accuracy of classification. The findings of this work form the foundation stage in the domain of cell classification for early detection of diseases like cancer.

Keywords: Analysis of variance (ANOVA), Autoregressive (AR), Cancer, Cell Classification, Figure of Merit (FOM), Optical, Prony.

Title and Abstract (in Arabic)

تعزيز تصنيف الخلايا السرطانية الطبيعية والمسمومة بدون تمييز باستخدام النماذج البارامترية والتقنيات البصرية

الملخص

لقد كان الدافع وراء تطوير أساليب خالية من الملصقات لتصنيف الخلايا هو الاكتشاف المبكر وتحديد مرض السرطان. تتطلب طرق علاج نقطة الرعاية المستقبلية POC أساليب الفحص السريع في الوقت الحقيقي. نظرًا لأن الأساليب المحددة لتصنيف الخلايا تستغرق وقتًا طويلاً وتتطلب قدرًا كبيراً من تحضير العينة إلى جانب الأشخاص المهرة، فإنها لا تبدو مناسبة لطرق علاج POC تعتمد الطرق الخالية من الملصقات على الخصائص الفيزيائية الحيوية للخلايا بدلاً من العلاقات الحيوية وكثيراً ما تستخدم لتصنيف الخلايا بسبب قدرتها على التفاعل مع الضوء حيث يعتمد هذا التفاعل على الخصائص الجوهرية لتكوين الخلايا.

تحمل الأنسجة المختلفة للخلايا الطبيعية والسرطانية ملامح بصرية مختلفة. لذلك فإن الهدف من هذا العمل هو الجمع بين التقنيات البصرية والطرق العددية لتعزيز الدقة في تصنيف الأنواع المختلفة من الخلايا. حيث تم دراسة التباين في تفاعلات الضوء من أنواع مختلفة من الخلايا وتحليل الملاحظات بشكل أكبر باستخدام الطرق العددية. تم سيتم استخدام تقنيات Prony و autoregressive لاستخراج مجموعة من المعاملات التي تمكنا من تصنيف الخلايا.

هنالك ستة أنواع من الخلايا تم دراستها في هذا العمل وهي: خلايا الرئة الطبيعية، خلايا سرطان الرئة، خلايا الكبد الطبيعي، خلايا سرطان الكبد، خلايا الكلى الطبيعية وخلايا سرطان عنق الرحم. ثم يتم تقدير الإشارات المقاسة وتقريبها باستخدام نموذج Prony و AR. يتبين أن التباين في الأعمدة المستخرجة ومعاملات الأنواع المختلفة من الخلايا يشكل أداة حيوية في تعزيز تصنيف الخلية. تساعد الأداة الإحصائية مثل تحليل التباين ANOVA في تحديد معاملات AR الهامة.

أظهرت النتائج أن الأقطاب التي تم الحصول عليها من خلال طريقة Prony للخلايا المختلفة تختلف في حجمها وموقعها. تم تطوير وتعيين رقم الجدارة FOM هنا والذي يرتبط بحجم الأقطاب وموقعها. لقد وجد أن توزيع FOM في z-plane أقرب إلى مركز دائرة الوحدة

لخطوط الخلايا الطبيعية منه لخطوط الخلايا السرطانية المأخوذة من نفس الأنسجة إن نموذج AR من نفس الترتيب لأنواع مختلفة من الخلايا يحمل قيم معامل وقطب مختلفة. لتقليل التكرار والوصول إلى نموذج AR موجز، تم استخدام تحليل ANOVA لتحديد الأهمية في معاملات .AR

إن تحسين الترتيب يؤدي إلى زيادة الاختلافات في القيم القطبية للخلية السرطانية مما ينتج تحسين تصنيف الخلايا. تشكل نتائج هذا العمل مرحلة الأساس في مجال تصنيف الخلايا للكشف المبكر عن أمراض مثل السرطان.

كلمات البحث الرئيسية: تحليل التباين (ANOVA)، الانحدار الذات (AR)، السرطان، تصنيف الخلية، رقم الجدارة (FOM)، البصري، الضمني.

Acknowledgements

All praise and gratitude be to the Almighty Allah for showering His abundant mercy on me, blessing me good education and surrounding me with good people throughout my life.

Firstly, I would like to express my genuine appreciation to my research supervisor Dr. Mahmoud F. Al Ahmad for his patience, support, immense knowledge, and the values he imparted in my professional and personal life. Rough sea makes the best sailors. I am indebted throughout my life for his relentless efforts in making his students, the best professionals. Without his help and supervision, this dissertation would not have been possible.

My sincere thanks to the United Arab Emirates University for granting fellowship for my PhD study.

I would like to thank Dr. Tahir Rizvi and Dr. Farah Mustafa for accepting to be my PhD Advisory committee members. I thank Dr. Dominique Durand, Professor, Case Western University, Ohio, USA for the comprehensive review and valuable feedback of my thesis. I would like to register my deepest gratitude to Dr. Mohammed Shakeel Laghari and Dr. Adel Najjar for being my PhD examining committee members and for all their support.

From the bottom of my heart, I genuinely thank Prof. Ali Al-Marzouqi, Dean-CGS, for his constant encouragement and spreading kindness and positivity. I thank Dr. Sabah Alkass, Dean-COE, for providing the congenial environment to carry out my studies. My sincere appreciation to Dr. Imad Barhumi, Chairman, Department of Electrical Engineering, Dr. Mousa Hussein, Dr. Falah Awwad and Dr. Amine El Moutaouakil for their support and help in my teaching assistantship position I held a

while.

I would like to acknowledge honestly the faculties and staff members of the Department of Electrical Engineering of UAEU, especially Dr. Rashad Ramzan, Dr. Atef Abdrabou, Dr. Addy Wahyudie, Eng. Hassan Gafar, my schoolteachers, my university professors in NIT, Nagpur, CEC, Chennai, Mr. M. Sreenivasan, Scientist-D and the other scientists in SAMEER-CEM, Chennai.

I owe my thanks to my friends Maha Shehada, Suha Glal, Aameena Farukh, Thahira, Khawla, Roya, Mus'ab Asad, Moni, Jeya, Fatheen, Huda Musa, Amina, Aasha Abdullah, Sadie khosravi, Sidiqul Akbar, with special thanks to my mentor-friend Dr. Hanan Al-Tous. Their encouraging words and kindness helped me to wade through some difficult times.

I affectionately thank my sisters Zareena, Zeenath, Jahan, binnus Rajesh and Gnaniyar and my dear Amrin for being there for me whenever I needed. I lovingly thank my - nephew Roshan and nieces Mehreen, Reshma and Zafreen – my stress busters.

My warm thanks to my dear mother Fathima for supporting me spiritually, emotionally and for being my strength, forever. Finally, my heartfelt gratitude to my most beloved father, Abdul Gani, watching whose life helped me to learn the values of honesty, hard work and gratefulness. Without him, this journey is never possible.

Dedication

*Dedicated to my IMME and to the
Ever-loving soul of my DAD!*

Table of Contents

Title	i
Declaration of Original Work	ii
Advisory Committee	iv
Approval of the Doctorate Dissertation	v
Abstract	vii
Title and Abstract (in Arabic)	ix
Acknowledgements	xi
Dedication	xiii
Table of Contents	xiv
List of Tables.....	xvii
List of Figures	xviii
List of Abbreviations.....	xx
Chapter 1: Introduction	1
1.1 Overview	1
1.2 Motivation	1
1.3 Statement of the Problem	2
1.4 Thesis objective.....	3
1.5 Thesis outline	4
Chapter 2: State of the Art	6
2.1 Biomarkers for classification of cells.....	6
2.2 Labeled biomarkers – conventional biochemical markers.....	8
2.2.1 Protein biomarkers	9
2.2.2 Enzyme biomarkers.....	9
2.2.3 Deoxyribonucleic acid biomarkers.....	10
2.2.4 Ribonucleic acid biomarkers	11
2.2.5 Drawbacks of labeled biomarkers	12
2.3 Label-free biomarkers for cell classification.....	12
2.3.1 Biomechanical markers	13
2.3.2 Bioelectrical markers.....	15
2.3.3 Bio-optical markers	20
2.3.4 Remarks on label-free-based cell classification	23
2.4 Numerical techniques for classification of cells	24
2.4.1 Biomedical signal processing.....	24
2.4.2 Non-parametric models	25
2.4.3 Parametric models	25

2.5 Relevant work using the Prony technique.....	27
2.5.1 Application of the Prony technique in power systems.....	28
2.5.2 Application of the Prony method in target discrimination.....	29
2.5.3 Application of Prony in biomedical signal processing.....	29
2.6 Statistical tools in the classification of cells	36
2.7 Summary	38
Chapter 3: Fundamental Theory	40
3.1 Introduction.....	40
3.2 Prony – basic theory.....	40
3.3 Prony –Algorithm	41
3.4 Autoregressive modeling – introduction.....	48
3.4.1 AR modeling – basic theory.....	48
3.4.2 AR modeling – algorithm.....	49
3.5 Analysis of Variance – ANOVA.....	52
3.5.1 Introduction	52
3.5.2 Basic concepts of ANOVA	53
3.5.3 Assumptions, notations, and abbreviations in ANOVA	53
3.5.4 Subdivision of variability into components	54
3.5.5 Significance of F-static.....	55
3.6 Conclusion	55
Chapter 4: Experimental Setup	57
4.1 Introduction.....	57
4.2 Cells used in the present work	57
4.3 Cell preparation.....	57
4.4 Cell suspension for culturing the cells	58
4.5 Experimental setup.....	61
4.5.1 Optical setup.....	61
4.5.2 Design of the experiment	64
Chapter 5: Classifications based on Prony models	65
5.1 Results and Discussion.....	65
5.2 Conclusion	75
Chapter 6: Classification based on Autoregression models.....	76
6.1 Results and Discussion.....	76
6.2 Conclusion	85
Chapter 7: Conclusion and Future Outlook	86
7.1 Conclusion	86
7.2 Future outlook.....	87
References	89
List of publications.....	100

Appendix 101

List of Tables

Table 2.1. Electrical properties of four sub-classes of prostate cancer cells.....	18
Table 2.2. Extracted parameters showing the difference in the values of specific membrane capacitance and cytoplasm conductivity for fixed and live cells	20
Table 2.3. Alterations in the scattering properties of normal and cancer cells.	21
Table 2.4. Accuracy, sensitivity, and specificity in classifying healthy, benign, and cancerous breast tissue using a k-mean algorithm on ARMA parameters.....	26
Table 2.5. Classification of 2D-ARMA features vectors (573 vectors) of normal, benign, and cancer breast images	27
Table 2.6. Extracted parameter values for tumors with different conductivity values	32
Table 2.7. Performance of the four different types of classifiers in classifying normal and cancerous cervical tissues.....	38
Table 3.1. Sample ANOVA table	55
Table 4.1. Specifications of the xenon light source	62
Table 4.2. Specifications of the mini-spectrometer (optical sensor).....	63
Table 6.1. Performance measure of a fitted AR model for the transmittance response of different types of cells for order 6.....	78
Table 6.2. Set of extracted AR coefficients for different types of cells.....	78
Table 6.3. Set of poles extracted for different types of cells.....	79

List of Figures

Figure 2.1. Types of biomarkers for classification of cells.....	7
Figure 2.2. Various types of chemical biomarkers conventionally used for cell classification.	8
Figure 2.3. Illustration of HER-2 membrane-bound protein released into blood acting as a breast cancer biomarker.	9
Figure 2.4. Illustration of ELISA technique providing a color change to indicate cancer (HRP is horseradish peroxidase).....	10
Figure 2.5. Structure of DNA and RNA biomarkers.	11
Figure 2.6. Variation in the Young's modulus value obtained for cells from different tissues.....	14
Figure 2.7. Response of normal bladder cells (HCV29) and cancerous bladder cells (HT1376) to the substrate stiffness.....	15
Figure 2.8. Difference in the scatter plot of capacitance of specific membrane versus conductivity of cytoplasm.	17
Figure 2.9. Difference in the IR absorbencies of normal and cancerous gastric epithelial cells	22
Figure 2.10. Poles extracted via the Prony technique to characterize breast tumors	31
Figure 2.11. Z-plane showing the difference in the location of poles extracted using the Prony method on a normal breast and malignant breast.	33
Figure 2.12. Models of breast with and without chest wall	34
Figure 2.13. Z-plane of poles of malignant breast and normal breast for Chest Model I.	35
Figure 2.14. Z-plane of poles of malignant breast and normal breast for Chest Model II.....	35
Figure 3.1. Flowchart showing the Prony algorithm.	47
Figure 4.1. Optical setup for measuring the optical profile of cells.....	61
Figure 5.1. Measured optical transmittance response of HeLa cells 293T fitted with Prony estimations. The measured transmittance is sampled at a uniform sampling interval of $T_s = 2.3$ nm.	67
Figure 5.2. Extracted parameters versus number of exponentials in the fitted model for 293T cell line: amplitude, damping factor, frequency and phase.	68

Figure 5.3. Z-plane plot (unit circle) showing coefficients (in blue) and locations of poles (in red) for HeLa and 293T cell suspensions.....	70
Figure 5.4. Extracted figure of merit (FOM) for 293T (in blue) and HeLa (in red).....	70
Figure 5.5. Extracted figure of merit (FOM) for normal and cancer cell lines from same tissue: lung normal, lung cancer, liver normal and liver cancer.	73
Figure 5.6. Figure of merit distributions: lung normal and cancer cells, and liver normal and cancer cells. The blue dots represent the normal cell lines and the red dots count for the cancer cell lines.....	73
Figure 6.1. Measured optical transmittance response of cancer liver, normal liver, cancerous lung, and normal lung cells. The measured transmittance was sampled at a uniform sampling interval of $W_s = 2.3$ nm.	77
Figure 6.2. Z-plane showing the distribution of the poles of the normal (blue) and cancer (red) lung cells and the normal (blue) and cancer (red) liver cells.	80
Figure 6.3. Z-plane showing the distribution of the reduced poles of the normal (blue) and cancer (red) lung cells and the normal (blue) and cancer (red) liver cells.....	83
Figure 6.4. Z-plane showing the distribution of the Q-factor of the normal (blue) and cancer (red) lung cells and the normal (blue) and cancer (red) liver cells.	84

List of Abbreviations

A/D	Analog to Digital
AFM	Atomic Force Microscopy
ANOVA	Analysis of Variance
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ATCC	American Tissue Culture Collection
BEBM.	Bronchial Epithelial Basal Medium
CAD	Computer Aided Design
df	Degree of Freedom
DFT	Discrete Fourier Transform
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic Acid
EATC	Ehrlich Ascites Tumor Cells
EGF	Epidermal Growth Factor
EIT	Electrical Impedance Tomography
ELISA	Enzyme-linked Immunosorbent Assay
EMD	Empirical Mode Decomposition
FBS	Fetal Bovine Serum
FDTD	Finite Difference Time Domain
FFT	Fast Fourier Transform
FPT	Fast Padé Transformation
GA-PLS-DA	Genetic Algorithm- Partial Least Squares- Discriminant Analysis
HOG	Histogram of Oriented Gradient

IMF	Intrinsic Mode Functions
LBP	Local Binary Pattern
LS	Least Square
LTI	Linear Time-Invariant
MA	Moving Average
MRS	Magnetic Resonance Spectroscopy
mRNA	Messenger RNA
MS	Mean Square
MSE	Mean Square Error
NS&F	Not Stained and Fixed
PAA	Polyacrylamide
PC	Personal Computer
PCA	Principal Component Analysis
POC	Point-of-Care
RNA	Ribonucleic Acid
RPMI	Roswell Park Memorial Institute
S&F	Stained and Fixed
SNR	Signal-to-Noise Ratio
SS	Sum of Squares
SSE	Sum of Squares within Groups
SSG	Sum of Squares between Groups
SST	Sum of Squares Total

Chapter 1: Introduction

1.1 Overview

Classification of normal and abnormal cells plays a major role in early detection of diseases like cancer. Early detection of such diseases helps patients to receive prompt treatment and lead better lives. In the past two decades, diversified research work has been carried out to develop efficient methods for detection and classification of cells. This includes research work in electrical, mechanical, optical, and biochemical fields. Most of the existing labeled techniques used for cell classification are time consuming, require skilled professionals, and require large amount of cell samples, antibodies, antigens, and biomarkers. In addition, repeated biopsies are required if the false positive rate is high. These snags in disease treatment processes necessitate the development of label-free methods for cell classification. Label-free cell classification methods that combine optical, electrical, or mechanical techniques and numerical methods are in currency. The essence of such methods is their efficacy in classifying cells in less time with a reduced amount of cell samples. This work combines optical measurement techniques and numerical methods. The parameters extracted via modeling are analyzed to classify the cells. Statistical tools are also utilized to enhance the classification. The preliminary results obtained are promising and lay a concrete foundation for advanced study in the area of cell classification.

1.2 Motivation

Diagnosing life-threatening diseases like cancer at an early stage when they are not too large and not been spread increases the chances of patients being treated

successfully. Early detection helps patients receive appropriate treatment and increases their survival rate. If diagnosed late, effective treatment becomes difficult, resulting in a low survival rate. Nearly 10 years increase in survival rate of women diagnosed with breast cancer is 90% when the disease is diagnosed at the local stage. In contrast, less than 20% of women survive for 5 years when the disease is diagnosed at the distant stage. Similarly, 93% of patients diagnosed with colon cancer at an early stage have a 5-year increase in survival compared to diagnosis at a late stage [1]. These statistical results emphasize the importance of early detection of diseases such as cancer. This motivates researchers to devise methods of detecting the abnormalities in cells and classifying them at an early stage. Abnormalities in cells alter the intrinsic properties and composition of cells. These alterations can change the electrical, mechanical, optical, physical, or chemical properties of the cells. Early detection of infection is possible if techniques are available to mine these properties. Since the existing methods are invasive and time consuming and require bulky equipment, large amount of sample, and so on, there is a rising demand for a method that simultaneously overcomes these drawbacks and classifies cells efficiently. This forms the motivating factor of the present work. The proposed method is label-free based cell classification and utilizes the alteration in the optical properties of normal and abnormal cells. The measured signals are analyzed using numerical techniques. The parameters extracted from the signal response via modeling are further processed using statistical tools for detection and classification of cells.

1.3 Statement of the Problem

In recent years, there has been increased use of the optical properties of biological cells to study the activities of cells. Diagnosing diseases at an early stage and gaining

knowledge of the stage of progress of the disease are made possible by studying the activities of cells. Optical properties of biological cells vary with the variation in the intrinsic properties and composition of the cells. This change in the optical properties of different types of cells helps in studying the activities of the cells and hence in their identification and classification. Most of the existing optical techniques in cell classification involve image-processing systems. The drawbacks of such techniques are many including the need for high-contrast images and sophisticated algorithms to process the images. To address these issues, development of a label-free optical technique that classifies cell in an efficient way is required. In the proposed method, the change in the optical properties of normal and cancer cells is used as a biomarker to classify the cell type. Parametric methods such as Prony/AR modeling techniques are utilized to enhance the cell classification.

1.4 Thesis objective

The objective of the present work is to develop an efficient label-free method for classifying cells with improved accuracy. The approach is to measure the optical profile of the cells and model it with numerical techniques. Signal-modeling techniques such as Prony and Autoregressive (AR) methods are utilized. The parameters extracted from the techniques are used as tools for classifying the cells. Optical measurements are carried out to obtain the transmittance profiles of the cells. The transmittance measurements of the cells are sampled to obtain a discrete set of data. The Prony algorithm approximates the sampled data to a set of damping exponential signals. Parameters such as Prony coefficients and poles that are extracted from the fitted model are analyzed for each cell sample, based on which cell classification is done. In autoregressive modeling, the discrete data obtained from the

transmittance profile of each cell sample is modeled with AR technique. The pole information extracted from the AR coefficients of each cell is used as the key for cell classification. To improve the accuracy, a statistical tool called analysis of variance (ANOVA) is used on the AR coefficients to determine the significant coefficients. Poles extracted from the significant AR coefficients are analyzed to perform cell classification. The deviation in the pole location of a cell from the locations obtained previously for a normal cell shows the presence of abnormalities in the cell. This forms the basis for classification of normal and cancer cells.

1.5 Thesis outline

In this section, the structure of the thesis is presented. The thesis details a label-free technique that combines the advancements in optical techniques and numerical methods for enhancing the classification of cells.

In Chapter 2, conventional as well as the state of the art techniques used in cell classification is discussed. This includes a concise discussion on the role of labelled biomarkers in cell classification, in particular normal and cancerous cells, their drawbacks and the advancements of label-free biomarkers. The role of different signal modeling techniques in bioengineering and the utilization of various statistical tools in cell classification used in the past are discussed.

Chapter 3 elaborates the fundamental concepts of Prony and autoregressive signal modeling methods and the algorithms of the methods. This is followed by a discussion on the theory of the statistical tool – analysis of variance.

The validation of the approach proposed in the work starts with the measurements of optical profiles of the six types of cells used in the work. The method of cell culturing, the optical experimental setup used for measuring the optical profiles of the cells and

the procedure of the conduction of the experiment are explained in chapter 4.

In Chapter 5, the classification of cells using Prony modeling technique and the findings of the work are presented.

In Chapter 6, the use of autoregressive technique for modeling the optical profile of the cell samples and the efficacy of the use of analysis of variance -statistical tool- in finding the significant coefficients and hence order optimizations are shown.

In Chapter 7, conclusion and future outlook of the research work is discussed.

Chapter 2: State of the Art

2.1 Biomarkers for classification of cells

Biological cells are by nature highly heterogeneous. A cell primarily consists of two regions: an inner cytoplasmic region enveloped by an outer cellular membrane. Cells exhibit electrical, mechanical, chemical, and optical properties. These properties depend on the composition and type. Any abnormality in a cell caused by inflammation or infection can be examined via the changes in the cell's inherent properties. Examining the properties of the suspected cells will help to diagnose the existence of abnormality at the early stages, especially for diseases with no specific symptoms at the incipient stage.

Extensive research has been carried out in studying the cell property indicators. These indicators present in blood, serum, urine, and stool or in any bodily fluid are called biomarkers [2]. Biomarkers are helpful in classifying cells as normal or abnormal as well as identifying the abnormality of an infected cell. Biomarkers are categorized as biochemical, biomechanical, bioelectrical, and bio-optical markers. Proteins [3], enzymes [4, 5, 6], deoxyribonucleic acid (DNA) [7, 8, 9, 10], and ribonucleic acid (RNA) [11, 12] are some of the biochemical markers conventionally used for cell classification.

A rapid paradigm shift in the field of biosensors has led researchers to focus on label-free biomarkers. In contrast to labeled biomarkers, label-free markers circumvent the need for secondary antibodies. Although labeled biomarkers reduce the rate of false-positive results, the process is time-consuming and therefore inappropriate for rapid and real-time screening [13].

The biomechanical, bioelectrical, and bio-optical properties of the cells are the label-free biomarkers that can be used for cell classification. The elongation of cell structure [14], cell deformation [15], cell stiffness [16, 17], and elasticity [18, 19] are the parameters considered as biomechanical markers. Membrane capacitance and cytoplasm conductivity [20, 21], dielectric properties [22], and cell impedance [23, 24, 25, 26, 27, 28] are considered bioelectrical markers. Similarly, optical properties of the cells provide important information about the state of cells [29]. Properties of light such as absorption [30], refractive index [31, 32], reflectance and transmittance [33], and scattering effects [34, 35] of light are considered optical biomarkers. Figure 2.1 shows the different types of label-based and label-free methods used in classification of cells.

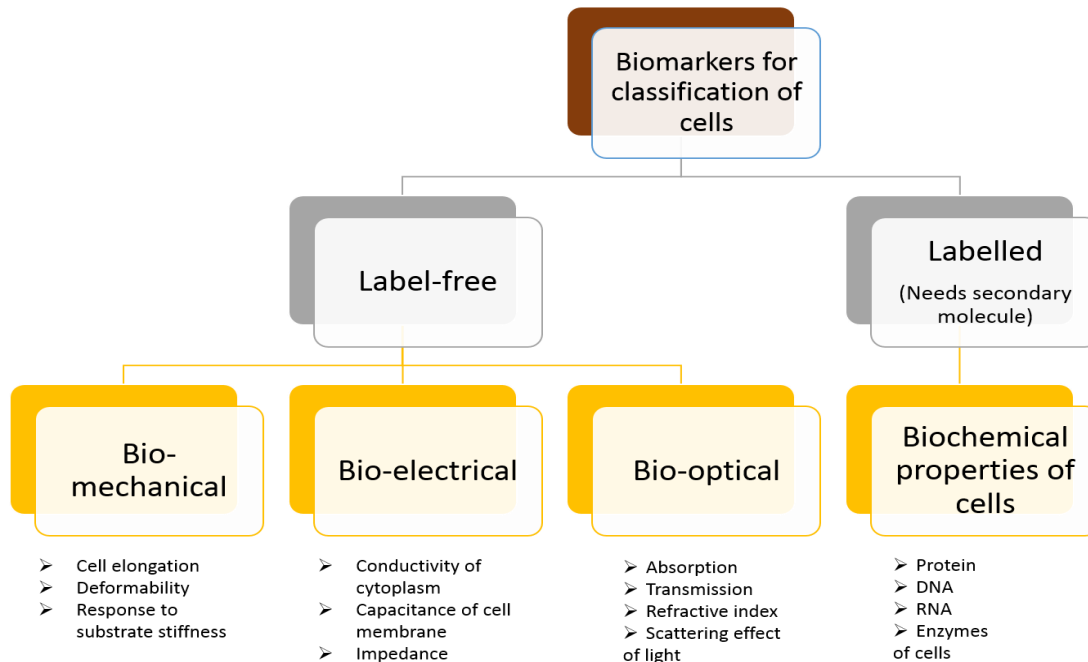


Figure 2.1. Types of biomarkers for classification of cells.

2.2 Labeled biomarkers – conventional biochemical markers

Traditionally, detection and classification of cells were carried out with the aid of biomarkers present in serum or tissues, which involves biopsy. This includes a range of macromolecules such as proteins, messenger RNA (mRNA), DNA, and so on [36]. Most of the biochemical based classification techniques require secondary molecules such as secondary antibodies. In addition, they involve a time-consuming amplification process as seen in polymerase chain reactions (PCRs). The PCR technique amplifies a specific amount of DNA through which cell classification is performed [37]. In the following sections, a brief discussion of various biochemical markers utilized for classification of cells is engaged in. Figure 2.2 shows the various biomarkers conventionally used to classify cells.

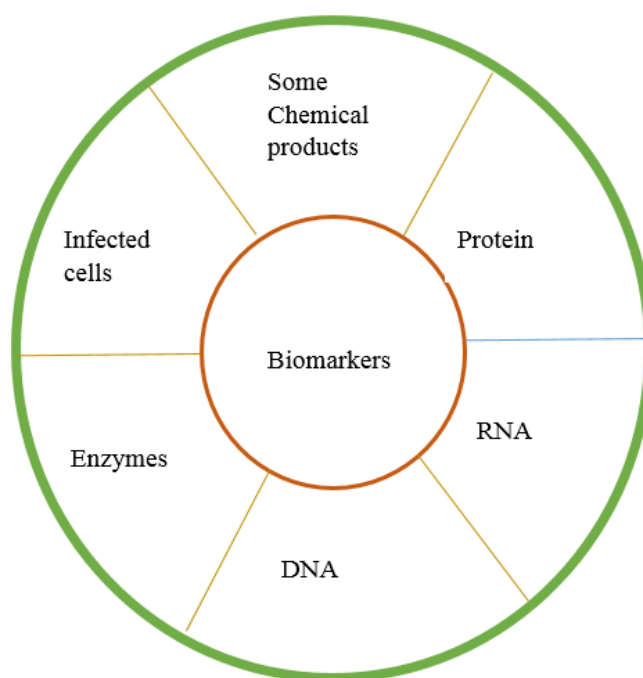


Figure 2.2. Various types of chemical biomarkers conventionally used for cell classification.

2.2.1 Protein biomarkers

Protein biomarkers are present to a larger extent in blood and to certain extent in urine [38]. The protein expression of blood or tissues is considered a potential biomarker for cell classification [38]. Classification of cells is done based on diagnosing the presence of an abnormal protein expression. The main drawback of using protein biomarkers for cell classification is that proteins do not replicate and hence they cannot be amplified for diagnostic study [36]. In addition, proteins are too sensitive to surrounding conditions such as temperature and pH, and this poses a difficulty in detecting abnormal protein patterns in lower concentrations [36]. Lopez has discussed the efficiency of HER-2 oncoprotein functioning as a breast cancer biomarker [39]. The HER-2 protein released in blood is tested to detect the presence of breast cancer. This process is illustrated in Figure 2.3 [1].

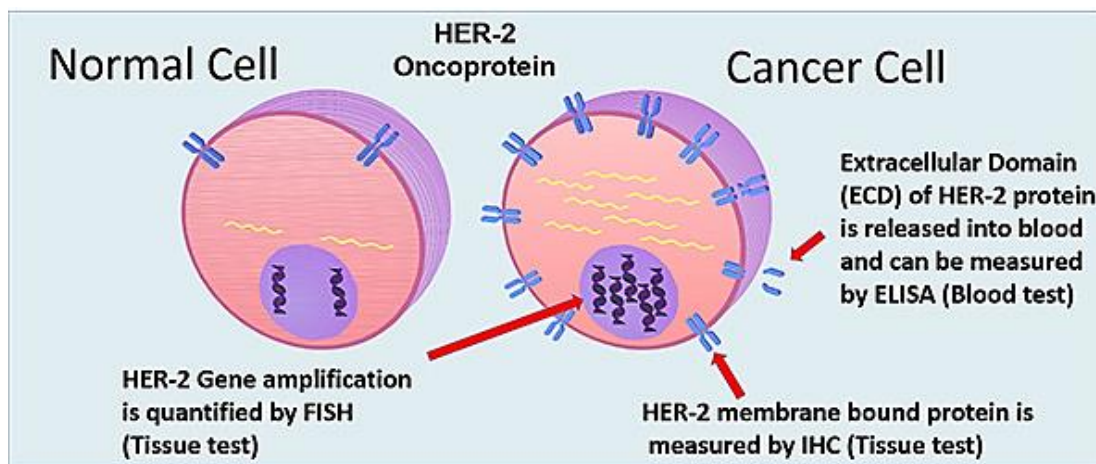


Figure 2.3. Illustration of HER-2 membrane-bound protein released into blood acting as a breast cancer biomarker (Adapted from [39]).

2.2.2 Enzyme biomarkers

A biochemical technique known as enzyme-linked immunosorbent assay (ELISA) is

another traditional technique used for cell classification [6]. The concept of ELISA is shown in the block diagram in Figure 2.4. In this technique, cells are classified by detecting the presence of a specific antibody or antigen in a blood or serum sample. Enzyme-linked immunosorbent assay is a fluorescent or colorimetric immunoassay technique; a visible color change occurs due to the reaction of an enzyme-linked antibody to a substrate. The color change indicates the presence of a particular antibody or antigen in the sample, indicating an infection. Rai et al. have reported problems and limitations in using ELISA such as impure antigens and protein A contamination [40]. If the concentration of infected biomarkers is low in the sample, then achieving distinct color change is challenging [36].

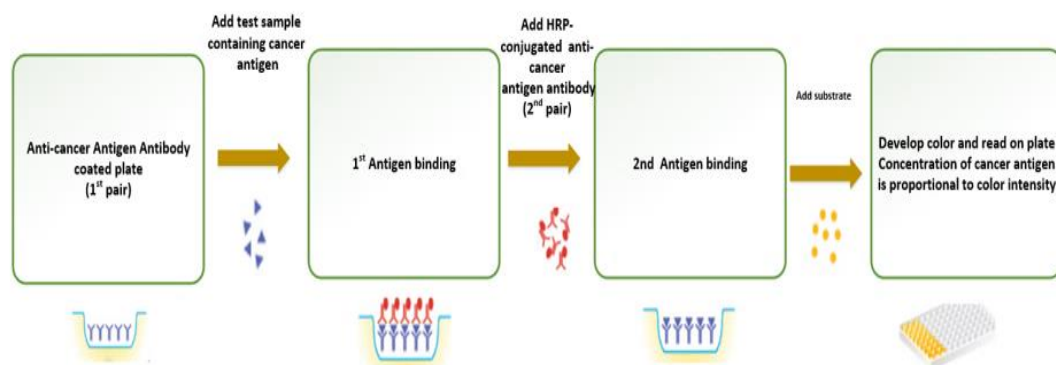


Figure 2.4. Illustration of ELISA technique providing a color change to indicate cancer (HRP is horseradish peroxidase).

2.2.3 Deoxyribonucleic acid biomarkers

Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are molecules that exist in cells [41]. DNA is a double-stranded structure while RNA is single stranded. Figure 2.5 illustrates the structure of DNA and RNA present in a cell.

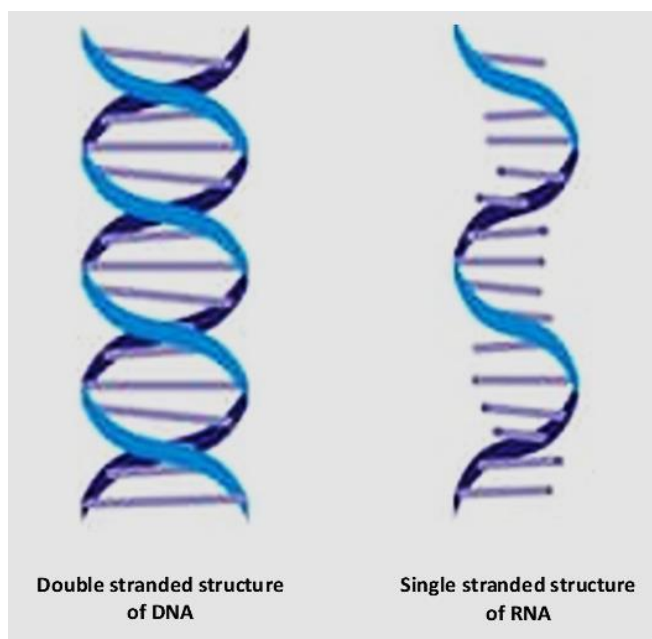


Figure 2.5. Structure of DNA and RNA biomarkers.

An infected or inflamed cell contains DNAs that have altered genes from those of normal cells [10]. An infected cell releases modified DNAs into circulation and detecting the presence of gene-modified DNAs serves as a means of identification of the presence of abnormality [42]. Thakur et al. has reported the role of DNA as a biomarker for cancer cell detection [8]. The results indicate that exosomes of cancer cells contain double-stranded DNA, which therefore acts as a potential biomarker for cell classification.

2.2.4 Ribonucleic acid biomarkers

Like DNA, RNA is a macromolecule in cells that undergoes change when the cells are abnormal. Several studies conducted in the past prove that RNA, particularly mRNA, is a capable biomarker for classifying cells [43]. Analysis of mRNAs connected with tumors and cancer are in currency for detecting cancer cells. Cagir et al. have examined the expression of mRNA in colorectal cells and have shown that it is an excellent

biomarker for diagnosing colorectal cancer [12]. Luparello et al. have differentiated two differently cloned breast cells (BC-3A and BC-61) from a parental cell (8701-BC) by analyzing the hsp- β gene in mRNA [11]. The results support that the hsp- β is an excellent biomarker for classifying the cells.

2.2.5 Drawbacks of labeled biomarkers

Most of the labeled techniques discussed above are invasive techniques. This causes damage to the cells or alters the intrinsic characteristics of the cells. In addition, these techniques necessitate large amounts of samples, antigens, and antibodies; require skilled and trained professionals to make sound decisions; and are time-consuming [13]. Hence, there is an urgent need for the development of label-free biomarkers that help in rapid and point-of-care (POC) detection and classification of cells [44].

2.3 Label-free biomarkers for cell classification

A new era of cell detection and classification methods was initiated with the development of label-free biomarkers. Label-free methods are based on biomechanical, bioelectrical, and bio-optical properties of cells. The salient features of label-free methods of classification are their non-invasiveness and their ability to retain the intactness of the cells during the whole process [45]. This ensures that the composition and inherent properties of cells remain undisturbed. Unlike labeled methods, label-free methods do not require secondary molecules to perform classification. In addition, de-embedding techniques help to remove the contributions of the medium and setup used, providing sample-specific measurements [46]. The following sections discuss some of the label-free methods used in cell type classification.

2.3.1 Biomechanical markers

Mechanical properties exhibited by cells such as stiffness, elongation, deformability, and elasticity aid in analyzing the composition and state of cells. These properties known as biomechanical markers are reportedly used in label-free methods of cell classification [14, 15, 16, 17, 18]. Swaminathan et al. have discussed the variation in the stiffness of normal and cancerous cells [16]. The outcomes of their work prove that the measurement of stiffness is an efficient mechanical phenotype for quickly performing cell classification. Kim et al. have reported the use of a micropipette aspiration technique to analyze the elongation of cells [17]. The elongation of cells is related to Young's modulus. The results indicate that the characterization of cells based on their elongation is a promising method for cell classification. Wang et al. have used the values of the Young's modulus parameter recorded for different cell lines as a mechanical cell type indicator, enabling classification of hundreds of cells [21]. Gossett et al. have examined the deformability property of cells to study inflammation and infections in cells [15]. The results prove that deformability varies with the composition of cells and thus serves as an excellent biomarker for cell classification. Lekka in her in-depth study on characterizing normal and cancer cells based on biomechanical properties using atomic force microscopy (AFM), demonstrated that the deformability of cells is mainly attributed to the changes in the cells due to loss in actin filaments when they become cancerous [19]. The deformability is one order of magnitude larger in cancer cells than in normal cells. Figure 2.6 illustrates the change in the value of the Young's modulus computed for normal and cancer cells from five different tissues: thyroid, breast, prostate, bladder, and kidney. It is evident that except for kidney tissues the cancer tissues exhibit higher

Young's modulus values than their normal counterparts.

Another mechanical biomarker under study for the classification of normal and cancerous cells is the response of cells to substrate stiffness [47, 48] Polyacrylamide (PAA) gels combined with laminin have been used as substrates. The change in the response of the bladder cells to the substrate stiffness has been demonstrated. The HT1376 cancer bladder cells and HCV29 normal bladder cells demonstrated a substantial change in their response to the substrate stiffness as shown in Figure 2.7

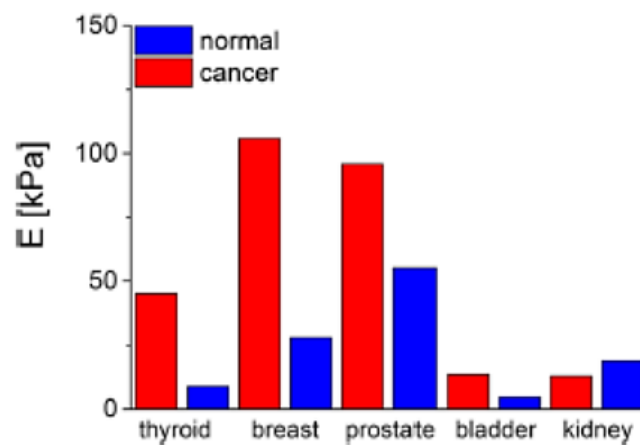


Figure 2.6. Variation in the Young's modulus value obtained for cells from different tissues (Adapted from [19]).

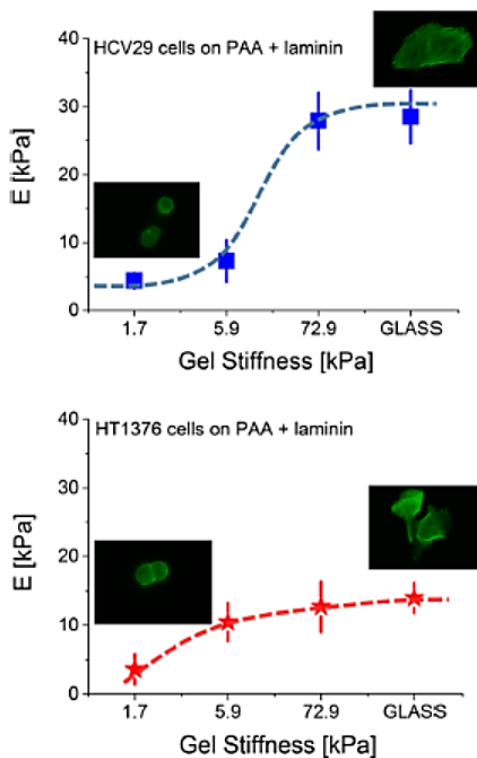


Figure 2.7. Response of normal bladder cells (HCV29) and cancerous bladder cells (HT1376) to the substrate stiffness. The substrate is a polyacrylamide gel with laminin.
(Adapted from source [19]).

2.3.2 Bioelectrical markers

The electric and dielectric properties of cells are linked to the composition and state of the cells. Electrical conductivity and resistivity [21], impedance [23, 24, 25, 26, 27, 28], and capacitance [49] are some of the quantifiable electrical properties exhibited by cells that serve as electrical biomarkers for cell classification. Extensive work carried out using electrical properties of cells as biomarkers has been reported in the literature.

Huang et al. have reported their success in using the conductivity of cytoplasm of cells and membrane capacitance as viable electrical biomarkers for the classification of cells

[50]. The electrical property study was done for four sub-classes of prostate tumor cells in a microfluidic system: untreated, stained, fixed, and fixed and stained, with anti-EpCAM as the stainer. The presence of unique types of proteins in the cell's membrane is exposed by the variation in the values of membrane capacitance obtained for the subtypes of prostate cells. Figure 2.8(a) shows the difference in the scatter plot of capacitance of specific membrane versus conductivity of cytoplasm for not stained and fixed (NS&F) cells and stained and fixed (S&F) cells. The plot reveals a substantial difference in the specific membrane capacitance quantified as $2.16 \pm 0.72 \mu\text{F}/\text{cm}^2$ versus $1.66 \pm 0.46 \mu\text{F}/\text{cm}^2$ of stained and not stained cells, respectively, with no difference in the cytoplasm conductivity: 0.59 ± 0.10 versus $0.59 \pm 0.10 \text{ S/m}$. This shows that antigen staining does not affect the cytoplasm. Similar observations on capacitance of specific membrane and conductivity of cytoplasm have been recorded for not stained and not fixed (NS&NF) cells versus stained and not fixed (S&NF) and in NS&NF versus NS&F cells as shown in Figure 2.8(b) and (c). The observed changes are summarized in Table 2.1.

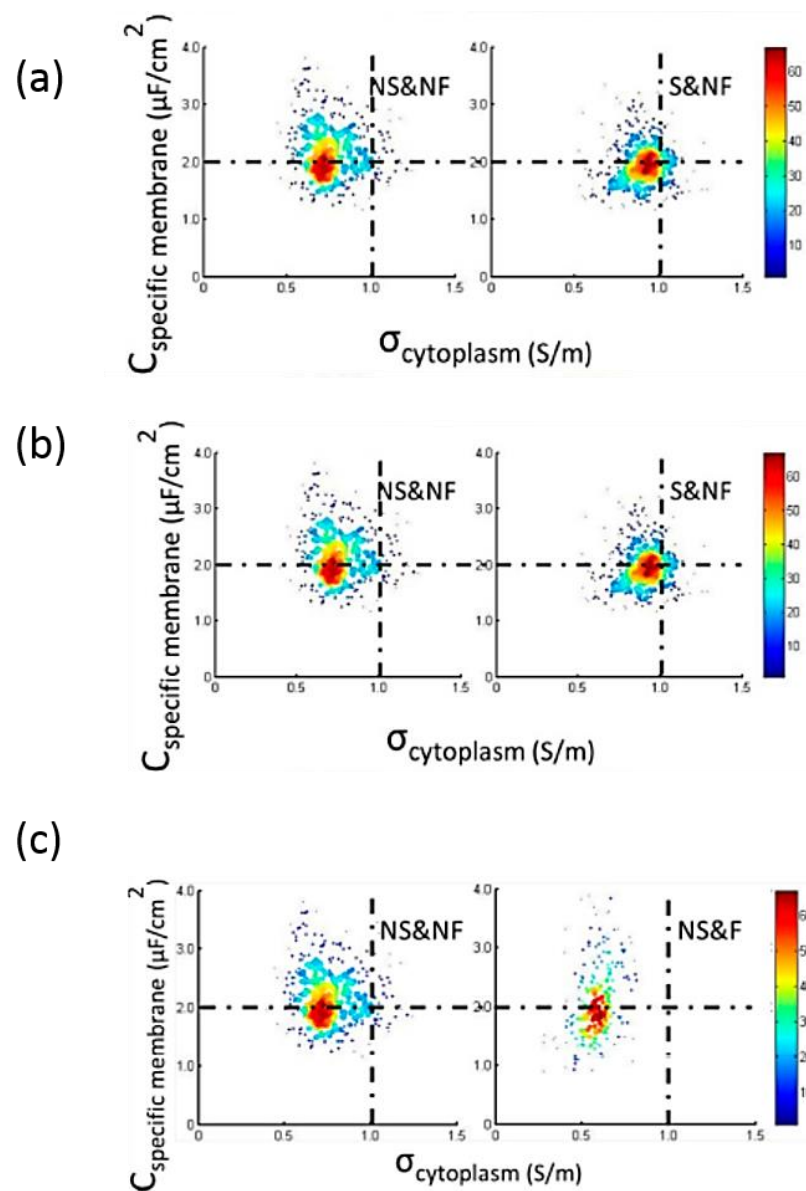


Figure 2.8. Difference in the scatter plot of capacitance of specific membrane versus conductivity of cytoplasm (a) Scatter plot showing the contribution of anti-EpCAM on the specific membrane capacitance of cells with no change in the cytoplasm region. The staining is correlated to the presence of specific membrane protein. (b) Significant difference in both specific membrane capacitance and cytoplasm conductivity illustrating the contribution of the staining agent. (c) Significant change only in the conductivity of cytoplasm due to the fixing process while the membrane conductivity remains almost of the same value for both fixed and not fixed cells. (Source [50]).

Table 2.1. Electrical properties of four sub-classes of prostate cancer cells.
(Adapted from [50]).

Cell sub-type	Electrical property	
	Membrane capacitance $C_{\text{specific membrane}} (\mu\text{F}/\text{cm}^2)$	Cytoplasm conductivity $\sigma_{\text{cytoplasm}} (\text{S}/\text{m})$
NS and F ($n_{\text{cell}} = 208$)	2.16 ± 0.72	0.59 ± 0.10
S and F ($n_{\text{cell}} = 252$)	1.66 ± 0.46	0.59 ± 0.10
NS&NF ($n_{\text{cell}} = 415$)	2.21 ± 0.49	0.77 ± 0.15
S&NF ($n_{\text{cell}} = 417$)	1.97 ± 0.39	0.90 ± 0.13

Aberg et al. have reported a non-invasive technique for classifying benign and malignant skin cells based on the electrical impedance measurements of cells [27]. The electrical impedance of five different types of cells was measured using an electrode system. The results indicate that electrical impedance of cells is a powerful tool for classifying normal and cancer skin cells. Dua et al. have proposed a similar idea of obtaining the electrical impedance of cells and using it as biomarker for classifying normal and cancer cells [28]. The obtained results were further analyzed using numerical techniques. The results show that 100% of the skin lesions were precisely classified for normal cells and 85% correctly classified in the cancerous cells. Zhao et al. have discussed how electrical phenotyping can be helpful in classifying tumor samples [51]. Membrane capacitance and cytoplasm conductivity of the two types of tumor samples (A549 and H1299) from three mice were measured in a microfluidic setup. These two intrinsic electrical parameters are independent of the size of the cells [52]. The results indicate that the cytoplasm conductivity and membrane capacitance

measured for cells of different cell lines (A549 and H1299) display substantial differences substantiating these two parameters as prominent electrical biomarkers that aid in cell classification. Al Ahmad et al. classified cells taken from lung, liver, and breast tissues based on the capacitance value measured. The results indicate that the dielectric property of normal cells is altered when it becomes cancerous; lowering the capacitance values of malignant cells and these findings for different cells help in their classification. Gao et al. in their work on distinguishing normal and cancer lung tissues, have suggested using electrical impedance tomography (EIT) to measure the impedance spectra of the tissues [26]. Two features, namely the cole-cole circle radius and complex modulus, were extracted from the measured impedances based on which cell classification is performed. The results indicate that an average of 85.4% of the samples are classified correctly. Zhou et al. have reported using electrical impedance values of embryonic stem cells of mice to describe the state of a cell [53]. The impedance measurements were taken for both fixed and live cells. During the fixation process the properties of membrane and cytoplasm of cells changed. This alteration in the state of the cells is exemplified in the observed results as shown in Table 2.2, authenticating the prowess of electrical biomarkers in classifying cells.

Table 2.2. Extracted parameters showing the difference in the values of specific membrane capacitance and cytoplasm conductivity for fixed and live cells (Adapted from [53]).

	Membrane-specific capacitance (F/m ²)	Cytoplasm conductivity (S/m)
Fixed cells	0.026 ± 0.004	0.48 ± 0.05
Live cells	0.035 ± 0.006	0.53 ± 0.03

2.3.3 Bio-optical markers

In recent years the study of light interaction with cells has gained popularity since optical properties such as scattering of light, refractive index, and light absorption and extinction are quantifiable for cells. Since these parameters are closely connected with the inherent properties of the cell, they act as bio-optical markers for cell classification. Pentilla et al. have discussed in detail the change in the optical properties of normal and injured cells [29]. In their work, Ehrlich ascites tumor cells (EATCs) were subjected to various conditions of staining and fixation and properties such as scattering, absorption, and extinction of light were analyzed. The results obtained prove that optical properties of cells are remarkable biomarkers for cell type classification. Scattering of light is a medium-dependent optical property. Some media are lightly scattering and some are highly scattering. Mourant et al. have studied the scattering properties of cells for cancer diagnosis [34]. The study was conducted on M1 and MR1 cells. M1 cells are immortalized non-tumorigenic cells taken from normal rat embryos and MR1 is a derivation of M1 that is tumorigenic. The results tabulated in Table 2.3 reveal that the change in scattering measurements in both cells

is attributed to the change in the dimension of scatters in the cells, which is directly linked to the morphological and constituent changes in cancerous and noncancerous cells. M_{44}/M_{11} represents the normalized backscattering of polarized light, with M_{44} representing cells response for polarized light whereas M_{11} is measurements with no polarizer. Similarly, the amount of absorption of light and hence the transmission is dependent on the properties of the medium on which the light is impinging. Salomatina et al. in their work have shown that the absorption and scattering parameters demonstrate differences in the values measured for normal and cancer skin cells in the 1,050 to 1,400 nm spectral range [54].

Table 2.3. Alterations in the scattering properties of normal and cancer cells.

	In exponential phase of cell growth		In plateau phase of cell growth	
	M1	MR1	M1	MR1
M_{44}/M_{11}	-1.053 ± 0.014	-0.088 ± 0.01	-0.065 ± 0.01	-0.098 ± 0.013
Apparent scatter dimension	236 ± 10 nm	261 ± 8 nm	245 ± 8 nm	268 ± 10 nm

M1: non-tumorigenic cells; MR1: tumorigenic cells (Adapted from [34]).

Al Ahmad and his group has demonstrated the efficiency of using optical biomarkers in cell classification [55]. Optical parameters of several types of cells such as BEAS-2B, HeLa, HEK-293T, HCC-87, MCF 10A, THLE2, HepG2, and MDA MB231 were measured. The results indicate that transmittance of normal cells is lower than that of cancerous cells in the spectral range 640 to 1010 nm. The deviation in the optical profiles of the different cells played the role of a biomarker for classifying the cells.

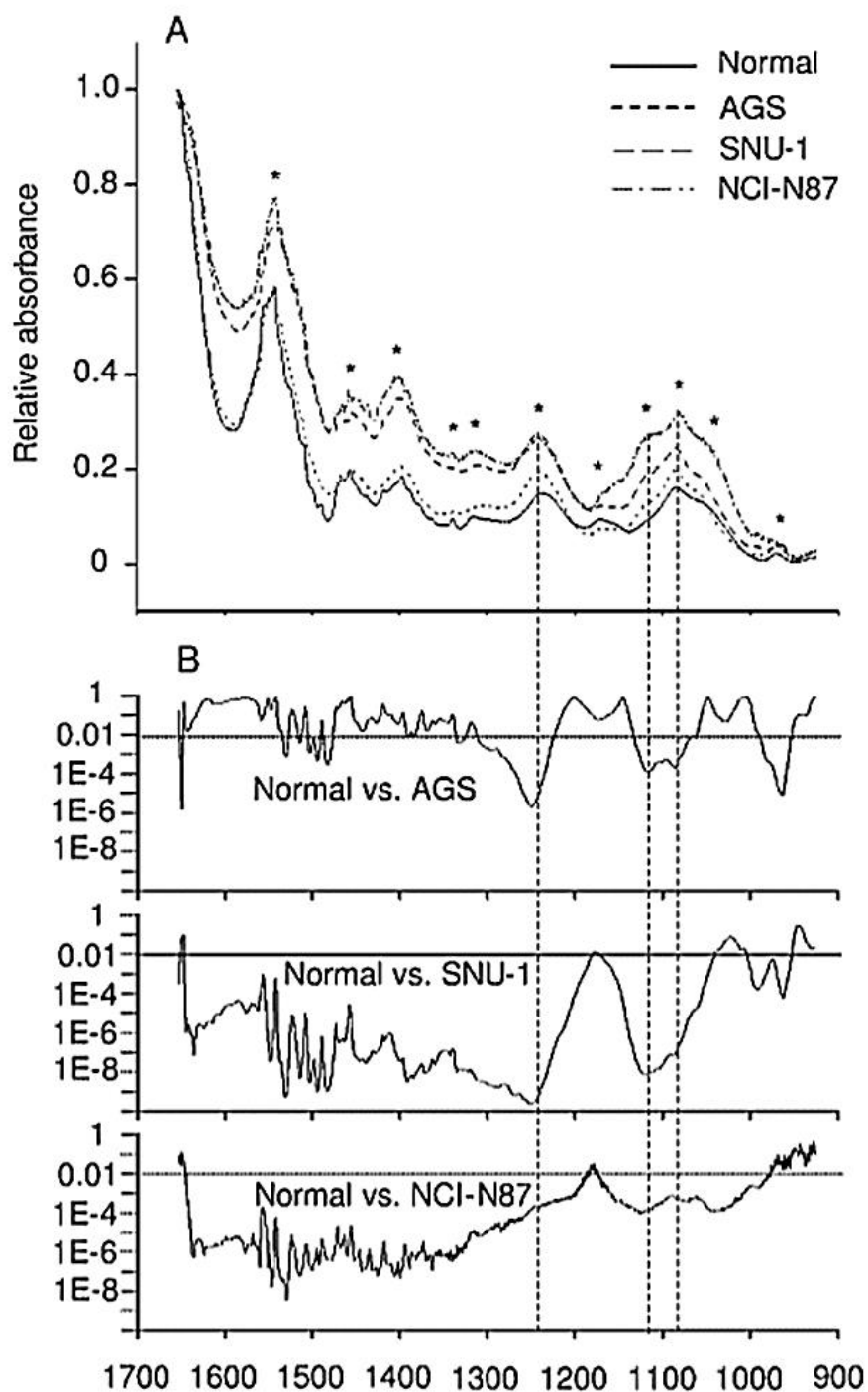


Figure 2.9. Difference in the IR absorbencies of normal and cancerous gastric epithelial cells (a) Relative IR absorbance of normal gastric epithelial cells (mean absorbance of eight observations), AGS, SNU-1, and NCI-87 (mean absorbance of 14, 7, and 5 observations, respectively). (b) t-test results of normal cells versus each type of cell showing the significant difference in the absorbance in the wavenumber around 1240, 1120, and 1080 cm^{-1}

(Source: [30]).

Fujioka et al. carried out a similar work where the absorption property of cells acted as biomarkers [30]. The IR absorption property of three gastric cancer cell lines – AGS, SNU-1, and NCI-N87 – along with normal gastric cells (mucosal epithelial) were studied. It was found that the cells exhibit different absorptions in the spectral range of 1650 to 925 cm^{-1} . The results showing the difference in the IR absorbencies of the four types of cells and the statistical differences in the values obtained using t-tests are shown in Figure 2.9(a) and (b).

2.3.4 Remarks on label-free-based cell classification

At the juncture of POC testing and treatment gaining popularity, time consumption is a major concern with labeled biomarkers. Label-free biomarkers provide an effective solution for the issue. The advancements in engineering and technologies provide methods to mine the mechanical, electrical, and optical properties of the cells and tissues that act as label-free biomarkers. The review of the earlier works in the field of label-free biomarkers demonstrates the different platforms available for improvements in this area. Nevertheless, when the measured electrical, optical or mechanical profile of different cells vary in miniscule for different cells, classifying cells based on the measurements alone becomes a difficult task. Developments in signal modeling techniques and numerical computations open new arenas for improving the accuracy of the results obtained via label-free markers. The following section focuses on signal processing techniques and statistical tools utilized in biomedical signal processing for optimizing cell classification.

2.4 Numerical techniques for classification of cells

With tremendous breakthroughs in diagnosis and treatment on the medical front, the crux lies in analyzing the data/output collected from the sophisticated instrumentation. The principal advantage of such analysis techniques lies in their flexibility of application to responses obtained from mechanical, electrical, or optical experimental set up. Certain features are extracted from the output response and are further assessed by numerical methods. Application of machine learning algorithms [56, 57, 58, 59] such as artificial neural networks (ANNs) [60], support vector machines (SVMs) [61, 62], and genetic algorithms (GAs) [63, 64] in the classification of cells can be found in the literature. To enhance the process of classification, the signal output from the experimental set up is processed by various signal-processing techniques before feature extraction. The processing techniques are broadly classified as non-parametric signal analyzing methods such as discrete Fourier transform (DFT) [65], wavelet transform [66], and parametric modeling such as fast Padé transform [67], autoregressive moving average (ARMA) modeling [68, 69], autoregressive integrated moving average (ARIMA) [70], and Prony modeling. The following section discusses the various signal-processing techniques found in literature with a focus on the Prony signal modeling technique in Section 2.5

2.4.1 Biomedical signal processing

Biomedical signals provide information about the attributes and characteristics of the biological entities that includes cells and tissues. This section reviews the various signal processing techniques used in the analysis of biomedical signals.

2.4.2 Non-parametric models

The non-parametric models such as DFT and wavelet transforms that were used in processing biomedical signals or images and that are reported in the literature do not assume a model. Moradi et al. have demonstrated in their work the application of DFT on ultrasound RF time series signal combined with a neural network for classification of normal and cancerous prostate tissues [65]. The results indicate a mean accuracy of 91% in classification with 92% sensitivity and 90% specificity. In early diagnosis of cancer, appearance of micro-calcifications is a symptom of breast cancer. Detection and segmentation of micro-calcification in images for classifying breast cancer is reported in Strickland [66]. Wavelet transforms is the signal processing technique applied in the mammography for detecting micro-calcifications. The results show that every single micro-calcification is clearly highlighted in the output image, making segmentation much easier.

2.4.3 Parametric models

Conventionally, fast Fourier transform (FFT) was the signal processing technique used to analyze medical images. Since FFT is a linear transform, the noise in the measured data in time domain gets translated to the frequency domain, which worsens the signal-to-noise ratio. Furthermore, FFT only provides the shape of the spectra. Extraction of essential parameters requires further processing. Due to these pitfalls, the conventional FFT is replaced with parametric modeling techniques for processing biomedical signals. Fast Padé transformation (FPT) is a parametric signal processing technique that is used to process biomedical signals. Belkic has explained in detail the drawbacks of using FFT and how FPT is an excellent solution for overcoming the issues with FFT

[67]. The in vitro magnetic resonance spectroscopy (MRS) obtained from breast cancer patients is processed with both FFT and FPT. The results demonstrate the efficiency of FPT over FFT in extracting information from the MRS signals to classify normal and cancerous breast tissues.

When computer-aided diagnosis gained popularity, a new class of parametric discrete-signal-modeling technique called ARMA and its variant ARIMA emerged as useful techniques in biomedical signal processing, particularly in the classification of tissues and cells. The “integrated” part of the ARIMA model comes from the fact that a differencing operation carried out one or more times if the data shows non-stationarity.

Zielinski et al. have applied ARMA modeling on ultrasound images of breasts to classify normal breast and cancer breast cells [69]. With two-dimensional (2D)-ARMA modeling, the Zielinski et al. achieved a classification accuracy of 93.87% in classifying malignant and benign breast tumors. The results are tabulated in Table 2.4.

Table 2.4. Accuracy, sensitivity, and specificity in classifying healthy, benign, and cancerous breast tissue using a k-mean algorithm on ARMA parameters (Adapted from [69]).

Modeling method	Accuracy (%)	Sensitivity (%)	Specificity (%)
2D-ARMA	93.87	92.03	94.14
1D-ARMA	78.51	59.54	79.76

Kumar et al. have done work similar to that of Zielinski et al., but modeled the data obtained for breast images with the ARIMA model [70]. Kumar et al. have achieved classification of healthy, benign, and malignant breast tissues.

Abdulsadda et al. have used a 2D-ARMA model in combination with a multilayer perception (MLP) ANN tool to classify three kind of breast tissues: normal, benign, and malignant [68]. The ARMA (1, 1, 1, 1) order had been used to represent the ultrasound images of the breast. The extracted ARMA features were processed with three algorithms: momentum BPA, delta-bar-delta, and Lvenberg-Macquard. The results shown in Table 2.5 highlight the ability of ARMA modeling to classify normal, benign, and tumor breast images.

Table 2.5. Classification of 2D-ARMA features vectors (573 vectors) of normal, benign, and cancer breast images (Adapted from [68]).

Modeling method	Steady-state MSE	Accuracy (%)	Sensitivity (%)	Specificity (%)
Momentum BPA	0.18	83	96	94
Delta-bar-delta	0.09	89	94	96
Lvenberg-Macquard	0.065	95	98	97

MSE: mean square error

2.5 Relevant work using the Prony technique

In 1795, Gaspard de Prony, a French mathematician, introduced the Prony technique to describe the expansion of gases [71]. Like Fourier analysis, Prony is a signal modeling technique, but unlike Fourier, Prony is also a parametric technique [72]. In

Fourier analysis, any signal is modeled as a sum of sine and cosine signals, meaning that Fourier series fit the measured signal to a set of undamped complex exponentials, whereas the Prony method models a signal to a sum of damped complex exponential signals [73]. Consequently, parameters such as amplitude, frequency, phase, and the damping factor of the damped exponentials are obtained via Prony modeling [74]. These parameters obtained from signals measured for different type of cells greatly assist in cell classification.

In the Prony technique, a uniformly sampled discrete time signal is sampled by fitting the signal to a sum of damped exponential signals. In addition to signal analysis, the Prony technique is used for system identification. For analysis purposes the measured signal can be in the time or frequency domain. This is because Prony analysis can be done for any signal that has transients. For these benefits, the Prony method has found wide application in diversified fields including power systems [75, 76], radar [77], sonar [78], and biomedical engineering [79, 80, 81, 82, 83, 84, 85].

2.5.1 Application of the Prony technique in power systems

The application of Prony in modeling power systems has been reported in literature [75, 76]. The Prony method had been utilized for modal analysis of the response data obtained from a power system in the United States of America. Dynamic modeling of the system had been done based on the signal components extracted using Prony in combination with Fourier and frequency domain techniques. The results of this work demonstrate the success of the application of the Prony technique in modal analysis. In another power system application, the Prony technique had been utilized for online low-frequency oscillation identifications in a power system [76]. A model of a real

grid was generated using a simulator and the order of the model was determined using the Prony algorithm. The results obtained for simulated responses show that Prony is a better technique than FFT, which is the conventional technique used in power systems.

2.5.2 Application of the Prony method in target discrimination

The viability of the Prony technique has found its application in characterizing radar targets [77]. The technique is used to analyze a synthesized signal that represents the backscattered signals from a radar target. The poles obtained through the Prony method represent the natural resonant frequencies of the targets. The results show the efficacy of the Prony technique in identifying and distinguishing targets. The time-consuming computations in numerical search techniques have been overcome via the Prony method. It has been proven that the Prony method is more efficient than the numerical search methods. In sound navigation and ranging, to distinguish underwater objects of interest from other interfering objects, the signals from these objects are modeled using the Prony technique [78]. The targets were detected and discriminated by analyzing the signals observed, both by FFT and Prony techniques. The results ascertain that Prony analysis has better detection capabilities than the FFT technique. The Prony technique retains the transients of interest, while these transients are lost in the FFT method of analysis.

2.5.3 Application of Prony in biomedical signal processing

An extensive application of the Prony technique is found in biomedical signal processing such as characterizing tumors [79], detecting and identifying cancers [80],

studying DNA sequences [85] , and so on. Huo et al. have reported using the Prony technique in tumor detection [79]. The tumor was treated as a hidden target and was subjected to electromagnetic pulse. Similar to detecting military targets, the tumors were detected by analyzing the backscattered signals from the tumors using Prony technique. The resonant frequencies (poles) of the backscattered signals extracted using the Prony technique characterize the morphological and electrical properties of the targets. These properties vary for normal and tumor breast tissues, and these variations were exemplified in the resonant frequencies as shown in Figure 2.10. The tumors are simulated as an ellipse with different dimensions, conductivity (σ), and dielectric constants (ϵ_r). As can be seen in Figure 2.10, the poles are different for different cases, proving that the change in the electrical properties of the tissues resulted in changes in pole values, namely the complex frequency of the scattered signal from the tissue.

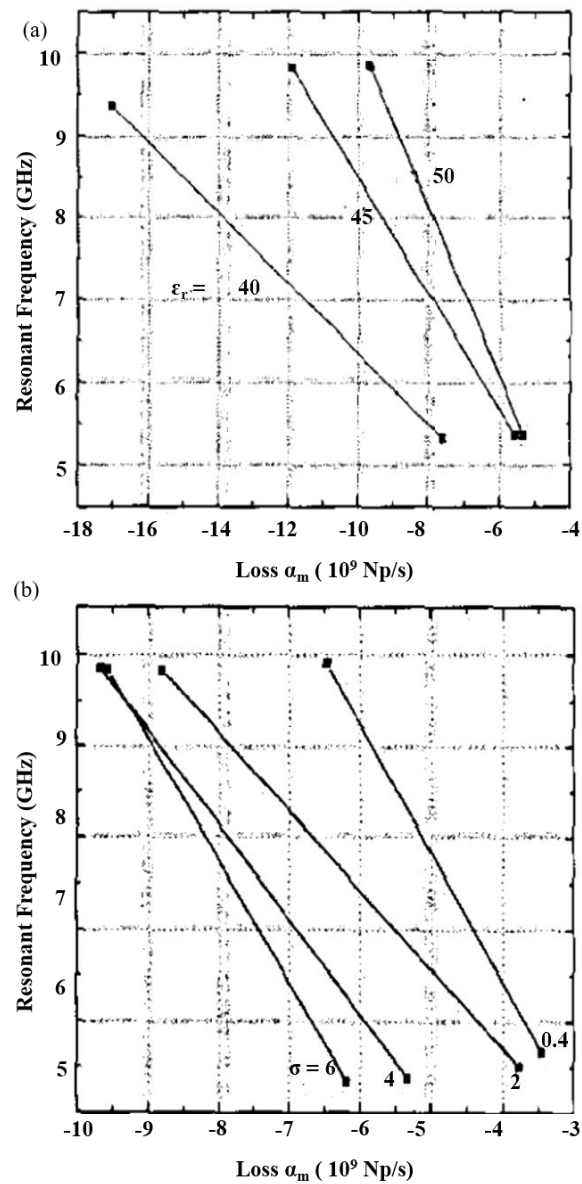


Figure 2.10. Poles extracted via the Prony technique to characterize breast tumors (a) with conductivity $\sigma = 4$ S/m and differing ϵ_r and (b) with dielectric constant $\epsilon_r = 50$ and varying σ (Source [79]).

Li et al. have conducted a similar study on breast tumor detection where the tumor was simulated using the finite difference time domain (FDTD) method [80]. The time domain response of the simulated model of the tumor was analyzed with the Prony modeling method. The breast tumor was simulated using the FDTD method for varying dimensions and electrical properties. The extracted parameters such as

amplitude A_m , damping factor α_m , and frequency f_m of the m -poles are different for tumors of different morphological and electrical properties. Table 2.6 shows the values of the extracted parameters. As can be seen, the extracted parameters are viable tools in classifying tissues as normal and cancerous. The results prove that Prony analysis is a promising method for characterizing tumors.

Table 2.6. Extracted parameter values for tumors with different conductivity values (Diameter = 20 mm) (Adapted from [80]).

Tumor #	Conductivity σ (S/m)	Pole #	α_m (10^9 Np/s)	f_m (GHz)	A_m (relative magnitude)
1	0.6	1	-2.6699	8.2807	0.0163
		2	-2.5744	6.5848	0.0559
		3	-1.6542	3.2364	0.0087
2	1.05	1	-2.8243	8.2554	0.0163
		2	-2.7039	6.5656	0.0545
		3	-1.7881	3.2209	0.0086
3	1.5	1	-2.8670	8.2045	0.0137
		2	-2.9507	6.5231	0.0469
		3	-1.8629	3.1777	0.0068

An added advantage of using the Prony technique for signal processing lies in its efficiency in modeling a signal even in the presence of noise [81]. Wang et al. conducted experiment on three different groups of tumors simulated with varying conductivity values, dielectric constants, and radii. The results obtained indicate that even for signal-to-noise ratio (SNR) levels above 25 dB, the real and imaginary part of errors remain at a low level.

Bannis et al. have discussed the role of Prony analysis in the detection and identification of breast tumors [82]. The scattered electromagnetic signal from a breast tumor model was analyzed using the Prony method. The poles that were extracted via the Prony method were used to detect and identify breast tumor. The complex z-plane plot of the poles shows that the poles of the normal and malignant breast tumors differ in their locations as shown in Figure 2.11(a) and (b).

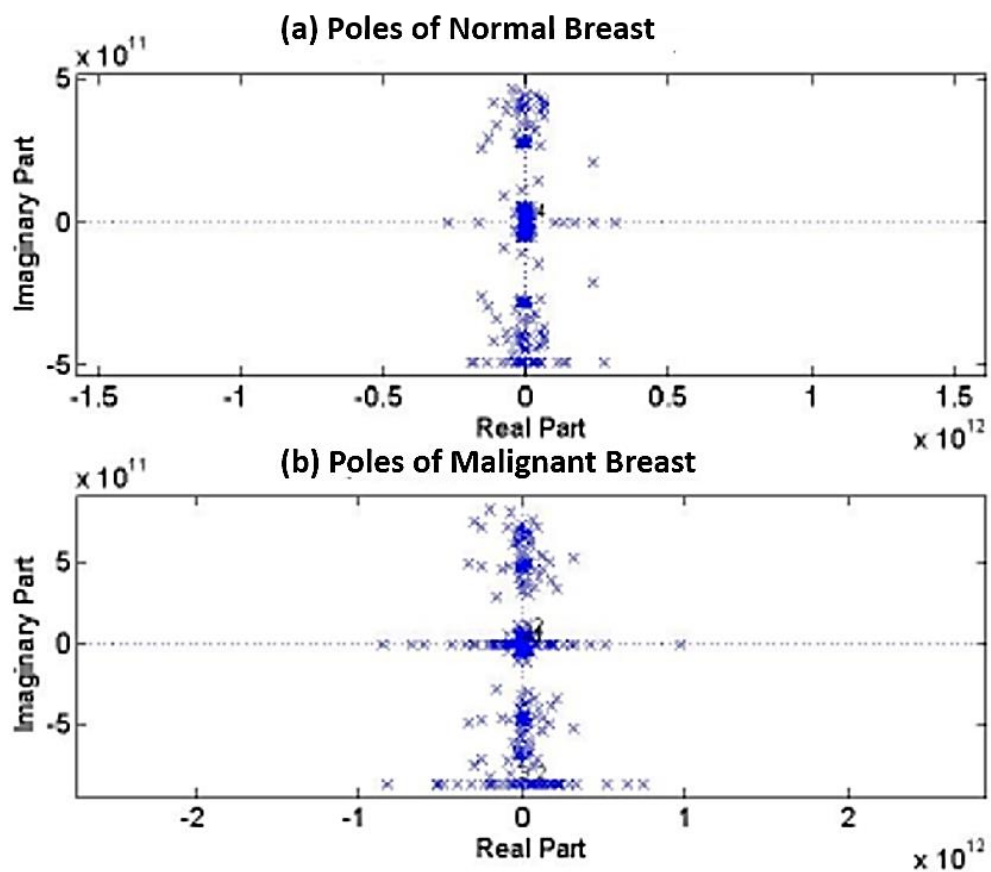


Figure 2.11. Z-plane showing the difference in the location of poles extracted using the Prony method on a (a) normal breast and (b) malignant breast (Source [82]).

In a study of the role of the chest wall in breast cancer detection, Bannis et al. modeled the chest wall (see Model I and Model II) and breast tumor with different dielectric

properties [83]. The results indicate that even in the presence of the chest wall the poles obtained through the Prony method for normal breast tissue are different from those of malignant breasts with tumors of radii 5 mm. The two different chest models are shown in Figure 2.12 (a) and (b) and z-plane plot of poles obtained for the models are shown in Figures 2.13 (a) and (b) and 2.14 (a) and (b).

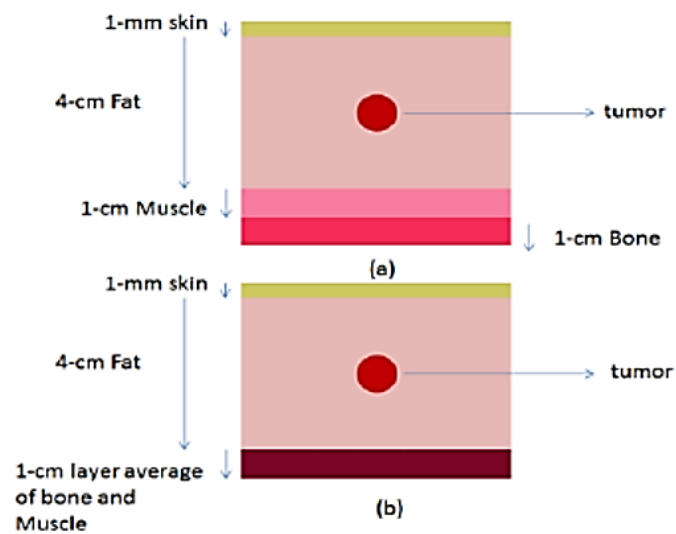


Figure 2.12. Models of breast with and without chest wall

(a) Chest Model I (b) Chest Model II (Source [83]).

In the medical field, discriminating coding and non-coding sections of DNA sequence is a promising method for disease detection. Prony technique was utilized for distinguishing the sequences. The work reported by Roy et al. discusses the application of the Prony technique in estimating the power spectral density of a DNA sequence and the viability of the technique for differentiating the coded and non-coded region of a DNA sequence [85]. The DNA sequences extracted from Celegan Cosmid F56F11.4A, D13156.1, and T12B5.A were utilized and their power spectrum

estimation had been conducted via the Prony technique. The results indicate that the Prony technique in distinguishing coded and non-coded sequences of DNA is superior to the conventional period gram methods.

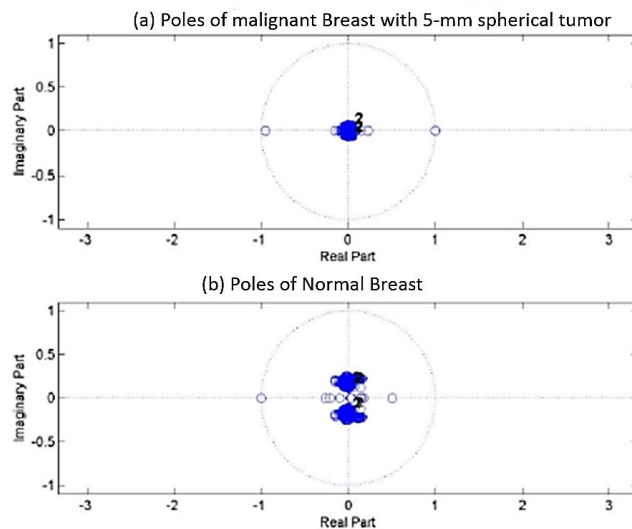


Figure 2.13. Z-plane of poles of (a) malignant breast and (b) normal breast for Chest Model I. (Source [83]).

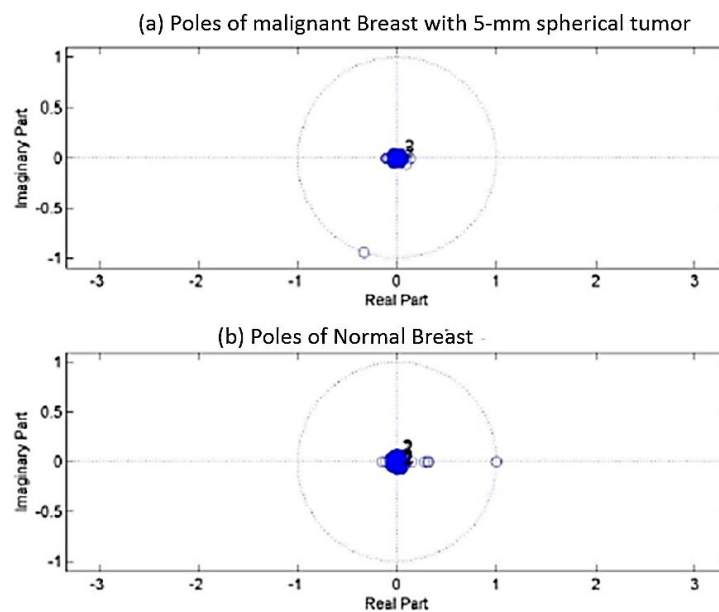


Figure 2.14. Z-plane of poles of (a) malignant breast and (b) normal breast for Chest Model II. (Source [83]).

2.6 Statistical tools in the classification of cells

In order to enhance the accuracy in classification of cells, the electrical, mechanical, or the optical profiles observed for the cells are analyzed with statistical tools. This is typically useful when the observed profiles overlap or differ in miniscule scale, especially at the early stages of infection or inflammation. Statistical analysis is helpful in observing a trend in data; the trend makes it possible to look for the gradual changes in the observations made for different subjects. For instance, the parameters obtained through processing the output response of different kinds of cells are analyzed with the aid of statistical tools such as GAs and machine learning algorithms such as SVMs, ANNs, and so on. Utilization of such statistical tools enhances the precision and reliability of the classification of cells.

The spectra obtained from Raman spectroscopy is a very promising tool for characterizing normal and cancerous tissues and cells from different parts of the body such as lungs, livers, breasts, prostates, and so on [86]. If the spectra of different tissues or cells obtained through this technique show subtle variation, statistical tools such as the GA can provide support in extracting prominent features for classifying with improved accuracy. Li et al. have combined the Raman spectra with the GA (with linear discriminant analysis) to identify significant Raman bands for classifying nasopharyngeal normal and cancer tissues [87]. The results demonstrate that the classification done with this combined technique leads to improved outcomes with sensitivity of 69.2% and specificity of 100% compared to the results obtained by applying principal component analysis (PCA) to the same dataset. In the PCA method, 63.3% sensitivity and 94.6% specificity were obtained. Li et al. used the same

combination of techniques to classify normal and cancerous bladder serum and obtained a sensitivity of 90.9% and a specificity of 100% in classification [64]. Duraipandian et al. have demonstrated the application of a variation of the GA called genetic algorithm partial least squares discriminant analysis (GA-PLS-DA) combined with Raman spectra of different tissues obtained using Raman spectroscopy for classifying cervical normal and pre cancer tissues [63]. The algorithm was successful in identifying seven significant Raman bands and yielded an accuracy of 82.9% with 72.5% sensitivity and 89.2% specificity. In a similar work on classifying cervical normal and cancerous tissues, the dataset obtained from optical spectroscopy was processed by an empirical mode decomposition (EMD) signal processing technique by which the measured signal is decomposed into intrinsic mode functions (IMFs) [61]. A machine-learning algorithm called the SVM was deployed as a classifier with different kernels such as linear, quadratic, polynomial, and RBF. The areas of the IMFs act as feature vectors for the classifier, providing a best classification accuracy of 95.65% with 87.5% specificity and 100% sensitivity with linear kernel. Table 2.7 shows the performance of the four different types of classifiers in classifying normal and cancerous cervical tissues.

Table 2.7. Performance of the four different types of classifiers in classifying normal and cancerous cervical tissues (Adapted from [61]).

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
SVM (linear kernel)	95.65	100	87.5	93.75
SVM (quadratic kernel)	86.95	93	75	87.5
SVM (polynomial kernel)	82.60	80	87.5	92.3
SVM (RBF kernel)	91.30	100	75	88.23

Ramasamy et al. have developed a multiclass classifier based on the SVM algorithm to classify 16,063 gene expressions from 218 tumor samples [62]. The results demonstrate an excellent classification accuracy of 78% while the accuracy obtained with random classification is much lower of 9%. Adetiba et al. have demonstrated the potency of an ANN ensemble in predicting lung cancer [60]. Using the Voss DNA encoding method, a genomic sequence was generated from normal and mutated genomes of nucleotides. Feature extraction was done via histogram of oriented gradient (HOG) and local binary pattern (LBP), which were used to train the classifiers. The results reveal a best fit of the data in this study providing an accuracy of 95.9%, validating the efficiency of the statistical tool in detection and prediction.

2.7 Summary

In summary Prony technique has proven to be powerful in the signal modeling and approximation. Prony parametric modeling technique provides parameters such as

amplitude, frequency, phase, and damping factor of the complex exponential signals in the model. The parameters extracted from Prony fitting are used to characterize the signal source as well as for system identification. These features make Prony an exemplary technique for signal modeling, finding its application in various fields.

Chapter 3: Fundamental Theory

3.1 Introduction

Cell identification and classification is closely linked with source characterization and system identification based on signal processing techniques. In general, an appropriate model is fitted to measured dataset and certain parameters are extracted. The extracted parameters play a key role in analyzing the signals and in identification of the source. This chapter discusses the basic theory of two techniques used in signal modeling, namely Prony analysis and AR modeling, and the algorithm of those techniques. Signal modeling is broadly classified into two categories: non-parametric and parametric modeling. Fourier methods such as DFT come under non-parametric modeling, in which no particular model is assumed that generated the data. In contrast, parametric modeling such as Prony and AR modeling is based on the use of models such as a linear time-invariant (LTI) model for the data. The fitted model is used to extract parameters such as amplitude, frequency, phase, damping factor of the exponential signals in the fitted model and corresponding coefficients in AR models. The extracted set of parameters and coefficients are then used in classification process.

3.2 Prony – basic theory

The concept of Prony modeling is fascinating and important in its countless applications in diversified fields. Prony modeling, a parametric time-series modeling technique, is an extension of the non-parametric Fourier methods [75]. The Fourier method models a discrete set of data in time domain as a sum of sines and cosines; in other words, it fits the data to a set of undamped exponentials [88]. In practice, fitting a set of data to a sum of complex exponentials is indeed fitting it to exponentially

decaying signals [89]. Consequently, Prony modeling fits a set of uniformly sampled data to a sum of damped exponentials; parameters such as amplitude, frequency, phase, and damping factor of the decaying complex exponentials are extracted from the Prony fitting.

The essence of this technique is that it converts a non-linear approximation into linear by solving a linear equation set and finding solutions to the set of equations [74]. As such, Prony is a robust technique for extracting exponential parameters associated with the measured data and determining their complex poles, which are called Prony roots. In the classic Prony method, a uniformly spaced data sequence of N samples with $N = 2p$ are approximated to a sum of p complex exponential decaying functions. The p complex exponentials have p different amplitudes, frequencies, phases, and damping factors. These extracted parameters and the poles play a key role in characterizing the source that generated the data and in system identification. Thus, Prony is a signal modeling as well as system identification technique [72]. The next section details the algorithm of the Prony modeling technique.

3.3 Prony –Algorithm

The Prony algorithm begins with fitting the observed discrete set of data to a sum of decaying complex exponential signals. Let $y[n]$ be the measured response that can be fitted with a sum of decaying exponential signals as given in Equation 3.3.1 [90, 91]:

$$y[n] = \sum_{i=1}^p A_i e^{j\theta_i} \cdot e^{(\alpha_i + j2\pi f_i)T_s(n-1)}, \quad n = 1, 2, 3, \dots, N \quad (3.3.1)$$

where N is the total number of data samples, T_s is the sampling period, and p is the

total number of decaying complex exponential signals in the response signal (same as the order of the model). The i^{th} complex exponential component has amplitude A_i (same unit as the response signal), frequency f_i in Hz, phase θ_i in radians, and damping factor α_i in s^{-1}

Equation 3.3.1 can be written in time-dependent and time-independent terms as in Equation 3.3.2:

$$y[n] = \sum_{i=1}^p h_i z_i^{n-1} \quad (3.3.2)$$

where the time-independent term h_i is the complex amplitudes. The time-dependent term z_i is the complex exponentials and the poles of the system. This representation provides a tractable way to solve linear, constant co-efficient difference equations. These parameters can be written as given in Equations 3.3.3 and 3.3.4:

$$h_i = A_i e^{j\theta_i} \quad (3.3.3)$$

and

$$z_i = e^{(\alpha_i + j2\pi f_i)T_s} \quad (3.3.4)$$

The algorithm of Prony modeling is a three-step procedure as stated below:

- Step 1: Solve an autoregressive (linear prediction) model for the observed data.
- Step 2: Solve for the roots of the characteristic or Prony polynomial; determine frequency and damping factor.
- Step 3: Solve the set of linear equations for the estimation of amplitude and phase.

Systematic explanation of the Prony algorithm

➤ **Step 1: Solve an AR model (linear prediction) for the observed data.**

The observed data $y[n]$ is written as a linear prediction AR model as given in Equation 3.3.5 as

$$y[n] = a_1 y[n-1] + a_2 y[n-2] + a_3 y[n-3] \dots \dots a_p y[n-p], n = 1, 2, \dots, N \quad (3.3.5)$$

Writing the autoregressive model in matrix (square Toeplitz) form as below:

$$\begin{bmatrix} y[p] \\ y[p+1] \\ \vdots \\ \vdots \\ y[N] \end{bmatrix} = \begin{bmatrix} y[p-1] & y[p-2] & \dots & y[0] \\ y[p] & y[p-1] & \dots & y[1] \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ y[N-1] & y[N-2] & \dots & y[N-p] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_p \end{bmatrix} \quad (3.3.6)$$

or

$$\mathbf{y} = \mathbf{Y} \cdot \mathbf{a} \quad (3.3.7)$$

where \mathbf{y} , the forward linear prediction matrix, \mathbf{Y} , the observation vector, and \mathbf{a} , the linear prediction coefficients vector, given as:

$$\mathbf{y} = \begin{bmatrix} y[p] \\ y[p+1] \\ \vdots \\ \vdots \\ y[N] \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} y[p-1] & y[p-2] & \dots & y[0] \\ y[p] & y[p-1] & \dots & y[1] \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ y[N-1] & y[N-2] & \dots & y[N-p] \end{bmatrix}, \text{ and } \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_p \end{bmatrix}.$$

Solving the linear system of equations gives the linear prediction coefficients vector \mathbf{a} , which can be computed from Equation 7 as $\mathbf{a} = \mathbf{Y} \setminus \mathbf{y}$ in MATLAB using a backslash operator.

- **Step 2: Solve for the roots of the characteristic or Prony polynomial; determine frequency and damping factor.**

A characteristic polynomial, also called a Prony polynomial, is formed from the vector of linear coefficients obtained in Step 1. The polynomial is written as follows:

$$T(z) = \prod_{i=1}^p (z - z_i) = \sum_{i=0}^p a[i]z^{p-i}; \quad a[0] = 1 \quad (3.3.8)$$

Since the linear prediction coefficient vector \mathbf{a} is known from Step 1, the characteristic polynomial can be readily solved for z_i , which gives the poles of the system.

In MATLAB, it is implemented using the MATLAB instruction [92]

```
t = transpose ([1; a]);
```

```
m = roots (t).
```

The linear prediction vector is a column vector \mathbf{a} ; the input vector to the **roots** syntax must be a row vector. Hence, the transpose of \mathbf{a} is computed and given as input to compute the roots of the polynomial formed from \mathbf{a} .

The damping factors α_i and frequencies f_i are determined as follows:

$$\alpha_i = \frac{\ln|z_i|}{T_s} \quad (3.3.9)$$

and

$$f_i = \frac{\tan^{-1} \left[\frac{\text{Im}(z_i)}{\text{Re}(z_i)} \right]}{2\pi T_s} \quad (3.3.10)$$

➤ **Step 3: Solve the set of linear equations for the estimation of amplitude and phase.**

The roots of the system of linear equations defined in Equation 3.3.2 are determined based on which exponential amplitude and phase are determined. Equation 3.3.2 can be written in matrix form as follows:

$$\begin{bmatrix} y[1] \\ y[2] \\ \cdot \\ \cdot \\ y[p] \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_1^1 & z_2^1 & \dots & z_p^1 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ z_1^{p-1} & z_2^{p-1} & \dots & z_p^{p-1} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \cdot \\ \cdot \\ h_p \end{bmatrix} \quad (3.3.11)$$

or

$$\mathbf{y} = \mathbf{z} \cdot \mathbf{h}, \quad (3.3.12)$$

where

$$\mathbf{y} = \begin{bmatrix} y[1] \\ y[2] \\ \cdot \\ \cdot \\ y[p] \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_1^1 & z_2^1 & \dots & z_p^1 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ z_1^{p-1} & z_2^{p-1} & \dots & z_p^{p-1} \end{bmatrix} \quad \text{and} \quad \mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \\ \cdot \\ \cdot \\ h_p \end{bmatrix}.$$

Solving the linear system of equations provides the values of complex amplitude vector \mathbf{h} . Since the values of complex amplitudes h_i and poles z_i are known, the exponential amplitude and phase can be determined as follows:

$$A_i = |h_i| \quad (3.3.13)$$

$$\theta_i = \tan^{-1} \left[\frac{\text{Im}(h_i)}{\text{Re}(h_i)} \right] \quad (3.3.14)$$

Thus, the parameters – amplitude, frequency, phase, and damping factor of the exponentials in the Prony fitting are extracted. These poles z_i along with their complex amplitude h_i and the parameters extracted aid in analyzing the signals and characterizing the system that generated the signals. Characterization of the system based on the poles is a domain-independent operation. The properties of the system that generated the signals can be studied based on the pole values and its distribution in the z-plane. The Prony algorithm for signal modeling and system characterization is shown in the flowchart in Figure 3.1.

The Prony technique is a generalization of the Fourier methods, but the Prony technique provides a better frequency resolution than the Fourier counterpart. This is because the Prony technique is based on AR modeling [74]. The downside of Prony modeling is that the technique becomes unstable when the number of exponentials is high. An effective solution for this is pre-filtering the signal before Prony fitting [74].

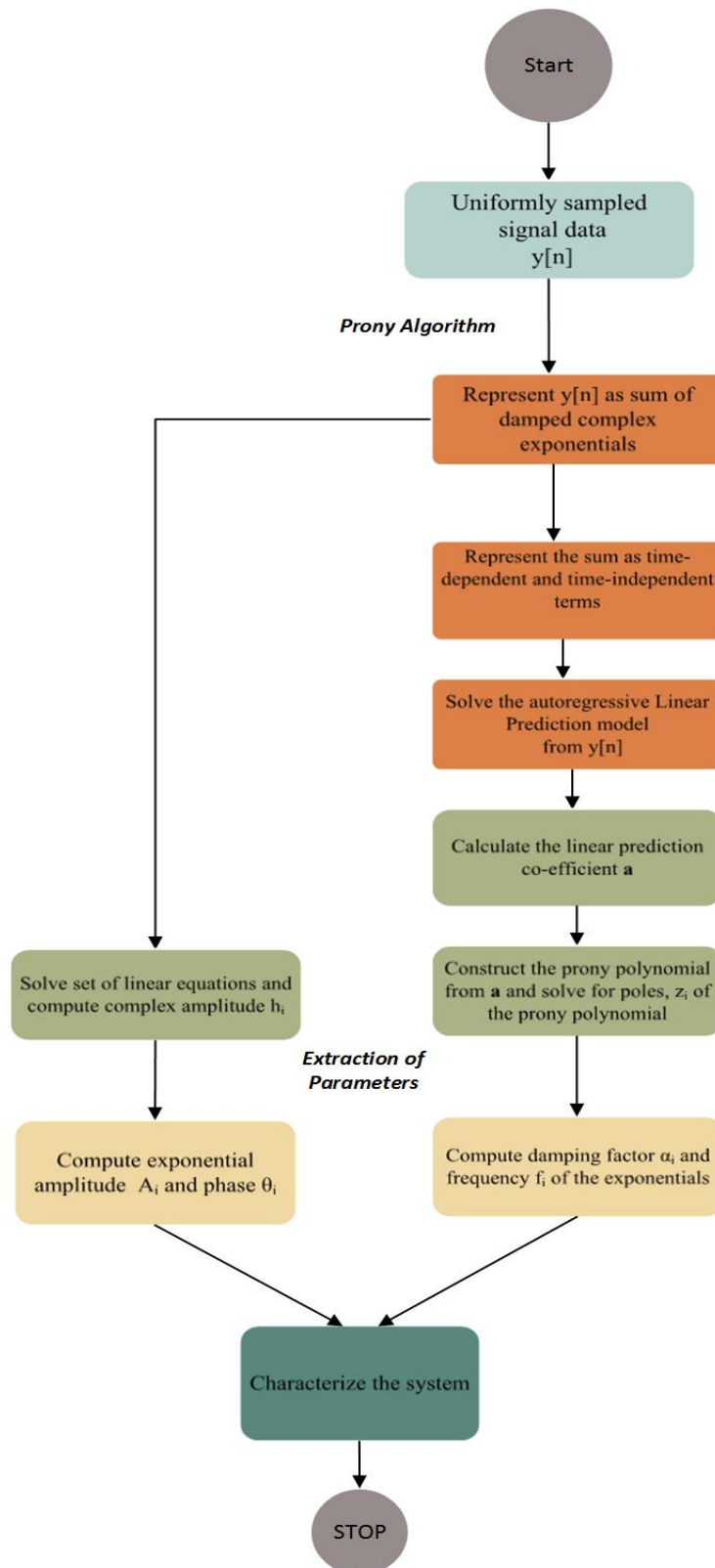


Figure 3.1. Flowchart showing the Prony algorithm.

3.4 Autoregressive modeling – introduction

The previous section discussed the fundamentals of Prony modeling and its algorithm. This section discusses a stochastic time series model that is commonly used in signal modeling, namely the AR modeling technique and its application in bioengineering [93]. This is one of the most well-known and frequently used univariate modeling methods. This model falls under the parametric modeling category and parameters such as AR coefficients and the poles extracted from the model are used as tools for characterizing the source. The basic theory and the algorithm of AR modeling are discussed in the following sections.

3.4.1 AR modeling – basic theory

The basic idea behind AR modeling is using linear regression to estimate the present value of the observed data as a linear combination of past observations and a random shock [94, 95]. Certain assumptions such as stationarity, linearity of the data series, and Gaussian distribution of the shocks are made for the implementation of the model. Here stationarity means that the roots of the characteristic equation of the AR model must lie outside the unit circle [96]. A data series is linear if the present value of the series is a linear combination of past observations. The random shock or the random noise is assumed as white noise following Gaussian distribution with zero mean and constant variance [97]. Nevertheless, most of the real-time data series are non-linear; linear models draw attention because they are simple to understand and easy to implement. The AR process is defined by a set of linear difference equations with constant coefficients called AR coefficients. The extracted poles from the AR coefficients represent the intrinsic characteristics of the cells under study; enabling

their classifications.

The crux of the AR modeling technique is the efficacy of the technique for modeling a discrete set of data. Though the model is popularly used for estimation and prediction of time series, it can be readily used for analyzing any set of discrete data. Like in Prony modeling, the poles extracted via AR modeling technique characterize the source, but unlike the Prony technique, the poles obtained via the AR method can be further processed for system identification. The next section outlines the algorithm of the AR modeling technique.

3.4.2 AR modeling – algorithm

In AR modeling, because the present output is modeled as a linear combination of previous outputs, it is also called all-pole modeling since the model only has poles by definition [98]. The order of the model is based on the number of past outputs that are linearly combined to represent the present output.

The algorithm of AR modeling is based on the following steps:

Step 1: Construct the AR model from the measured data.

Step 2 Compute the AR coefficients.

Step 3: Extract poles from the AR coefficients.

➤ **Step 1: Construct the AR model from the measured data.**

Let $y[n]$ be a discrete data response. The AR model of the data of order p , AR (p) can be written as follows:

$$y[n] = a_1y[n - 1] + a_2y[n - 2] + \dots + a_p y[n - p] + e[n]; \quad n = 1, 2, \dots, N \quad (3.4.1)$$

where N represents the total number of data samples and a_k is the AR coefficients. The random shock or random noise is assumed to be white Gaussian noise: $WN(0, \sigma^2)$.

The all-pole model can be represented in z domain as follows:

$$A(z) = (1 + a_1z^{-1} + a_2z^{-2} + \dots + a_pz^{-p})^{-1} \quad (3.4.2)$$

The model thus has p unknown coefficients, unknown σ^2 , a total of $p + 1$ unknown parameters that characterize the model.

➤ **Step 2: Computation of AR coefficients**

Some efficient methods available for computing the AR coefficients are as follows [99, 100]:

- Least square (LS) method,
- Maximum likelihood method, and
- Yule-Walker equation solved by the Levinson recursion algorithm.

The LS method of determination of AR coefficients is discussed below.

- ***Least square method:***

As the name indicates, the ordinary LS technique is used for estimation of the AR coefficients. This is possible since the random shock $e[n]$ are uncorrelated with the past values of y ($y[n - j]$, $j = 1, 2, \dots, p$).

The output $y[n]$ can be written in a p -dimensional vector form as follows:

$$\mathbf{Y}[\mathbf{n}] = (y[n] \ y[n - 1] \ \dots \ y[n - p + 1])' \quad (3.4.3)$$

and the random error $e[n]$ is written in vector form as follows:

$$\mathbf{u}[\mathbf{n}] = (e[n] \ 0 \ \dots \ 0)' \quad (3.4.4)$$

The prime denotes the transpose of the matrix.

A matrix \mathbf{A} is defined as follows:

$$\mathbf{A} = \begin{bmatrix} a_1 \dots & a_{p-1} & a_p \\ I_{p-1} & & 0_{(p-1) \times 1} \end{bmatrix} \quad (3.4.5)$$

The AR model given in Equation 3.2.1 can now be written as follows:

$$\mathbf{Y}[\mathbf{n}] = \mathbf{A} \mathbf{Y}_{\mathbf{n}-1} + \mathbf{u}_{\mathbf{n}} \quad (3.4.6)$$

$\mathbf{U}_{\mathbf{n}}$ are the errors, the square of which has to be minimized to obtain the LS estimators.

The LS estimators for determining the AR coefficients that are in the first row of the \mathbf{A} matrix are as follows:

$$\hat{\mathbf{a}} = (\sum_{n=2}^N \mathbf{Y}_{\mathbf{n}-1} \cdot \mathbf{Y}_{\mathbf{n}-1}')^{-1} \cdot \sum_{n=2}^N \mathbf{Y}_{\mathbf{n}-1} \cdot Y_{\mathbf{n}} \quad (3.4.7)$$

Here, X_t is a scalar quantity. The first element alone is considered because the

objective is to determine the AR coefficients that are present in the first row of the matrix \mathbf{A} . After obtaining the AR coefficients, the variance σ^2 is determined as follows:

$$\widehat{\sigma^2} = \frac{1}{N-p-1} \cdot \sum_{n=p+1}^N (y_n - \widehat{a}_1 y_{n-1} - \dots - \widehat{a}_p y_{n-p})^2 \quad (3.4.8)$$

giving $p + 1$ unknowns determined by the LS method.

➤ **Step 3: Extract poles from the AR coefficients.**

After computing the p - AR coefficients, the all-pole model given in Equation 3.4.2 is solved to determine the poles.

In MATLAB, the syntax *arx* performs the LS algorithm to construct the AR polynomial, solving which gives the AR coefficients [101]. The poles of the model are determined from the AR coefficients [99] and metrics such as percentage of prediction accuracy, mean square error (MSE) , and forward prediction error (FPE) [102] are obtained to measure the performance of the fitted model.

3.5 Analysis of Variance – ANOVA

3.5.1 Introduction

With the advent of computer-aided design (CAD), the use of statistical tools in analyzing the output responses obtained from an experimental setup for different subjects/conditions is in currency. These tools help in quantifying the significant differences between the measured responses by which means the most significant parameters are obtained [103]. In statistical terms, the conditions are called treatments. The statistical tools such as t-test, chi square estimator, ANOVA, and so on estimate

the population mean/variance or differences in the population mean/variance; the choice of an appropriate tool depends on the sample size. For instance, if population means are to be compared for a set of two, a simple t-test is used. If the sample size expands (> 2), sophisticated tools such as ANOVA are used to study the variance within and between different populations [104].

3.5.2 Basic concepts of ANOVA

Analysis of variance is an efficient statistical tool that helps to incorporate the effect of a subpopulation on the variability of total population. It can be used to measure and compare more than two means and determine the effects of factors. The underlying idea behind ANOVA is to subdivide the total variation in a dataset into two or more components; each component has its own source of variation. Analysis of variance quantifies the contributions of each source to the total variation. In other words, ANOVA allows us to determine the contributions of different factors to the variability in the total dataset. In general, a quantitative response variable is connected to one or more explanatory variable. In such a scenario where it is necessary to quantify the response variable, ANOVA is the most fitting statistical technique.

3.5.3 Assumptions, notations, and abbreviations in ANOVA

- Certain assumptions are made in the ANOVA technique in order to get accurate results from the calculations [105].

Assumption 1: Samples chosen are simple random samples.

Assumption 2: Each population is normally distributed.

Assumption 3: The variance is the same for all the populations.

➤ The following notations are used in the upcoming sections:

m – number of populations (groups)

n_i – size of sample from population i

Z_{ij} – j^{th} response sampled from i^{th} population

$\bar{z}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} z_{ij}$ – sample mean of responses from i^{th} group

$s_i = \frac{1}{n_i-1} \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_i)^2$ – sample standard deviation from the i^{th} group

$n = \sum_{i=1}^m n_i$ and $\bar{z} = \frac{1}{n} \sum_{ij} z_{ij}$ – total sample and mean of all responses, respectively

(irrespective of groups)

➤ Abbreviations [106]:

SST – Sum of squares total,

SSG – Sum of squares between groups, and

SSE - Sum of squares within groups

3.5.4 Subdivision of variability into components

Considering the populations as a single sample and not as “ m ” samples from each population, the total amount of variability is given as follows:

$$SST = \sum_{i=1}^m \sum_{j=1}^{n_i} (z_{ij} - \bar{z})^2$$

(3.5.1)

Eq. 3.5.1 is split into two parts: $SST = SSG + SSE$

where $SSG = \sum_{i=1}^m n_i (\bar{z}_i - \bar{z})^2$ is the variability between group means around \bar{z}

and $SSE = \sum_{i=1}^m \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_i)^2$ is the variability within group means.

3.5.5 Significance of F-static

The F-static is the ratio between the measures of the variability between treatments and the variability within treatments [107]. If the F-static is large, between-treatments variability is more than within treatments. This indicates that the means of the populations from which the samples are taken are different, causing the null hypothesis to be rejected (the means of populations is the same for all groups). The opposite of the null hypothesis is the alternate hypothesis. If the F value is small, variability within treatments is more than that of between treatments and the null hypothesis holds. A sample of ANOVA table ([115, 116]) is shown in Table 3.1.

Table 3.1. Sample ANOVA table

Source of Variation	SS	df	MS	F-static
Between samples	SSG	m-1	$MSG = \frac{SSG}{m-1}$	$\frac{MSG}{MSE}$
Within samples	SSE	n-m	$MSE = \frac{SSE}{n-k}$	
Total	SST	n-1		

df: degree of freedom; SS: sum of squares; MS: mean square.

3.6 Conclusion

In conclusion, Prony is a viable technique for analyzing experimental data in discrete-time domain. The poles extracted from the time domain data are employed for the

generation of transfer function and for system identification. On the other hand, if the measured response, irrespective of the domain, has transients then the poles and other parameters extracted can be utilized for characterizing the source, but for system identification. A technique analogous to the Prony decomposition which is used for frequency domain data is reported by Kumaresan [108] validates this. Similar to Prony, AR modeling is a parametric signal modeling method; parameters such as AR coefficients and the poles extracted from the coefficients play a significant role in signal modeling and system identification. The salient feature of the AR technique is its flexibility for use with any set of discrete data. Analysis of variance is an efficient statistical tool that gives the significant coefficient, thereby reducing redundancy and simplifying the generated model. Thus, numerical methods applied to biological signals are promising techniques for cell sorting.

Chapter 4: Experimental Setup

4.1 Introduction

Classification of cells based on label-free-based optical biomarkers combined with numerical techniques is the objective of this work. To demonstrate the effectiveness of the optical techniques combined with numerical methods such as Prony and AR modeling in classification of cells, normal and cancerous cells from two different cell lines were utilized. The cells used in the work were obtained according to the norms of the American Tissue Culture Collection (ATCC), Manassas, VA, USA and were cultured in specific media. The details of different types of cells used in the proposed method to demonstrate cell classification and their specific culture media are discussed in the following sections.

4.2 Cells used in the present work

The cell lines used in this study were taken from human lung and liver tissues. The cells from each cell line were in turn normal and cancerous. Thus, cells from the four different cell lines used in the experiment were normal lung cells (BEAS 2B), cancer lung cells (CC-827), normal liver cells (THLE2), and cancer liver cells (HEPG2). To show the effectiveness of the proposed method for other cell types, two more cell lines were included: normal kidney cells (293T) and cancer cervical cells (HeLa). Thus, a total count of six cell types is used for classification by the proposed method.

4.3 Cell preparation

All the cells used in the experimental work were procured from ATCC. They are

cultured and trypsinized in accordance with the recommendations framed by ATCC. Trypsinization is the process of cell detachment using a proteolytic enzyme called trypsin [109]. The addition of trypsin digests the proteins to detach the cells from the vessel in which the cells are cultured and are passed through a new container. At the end of trypsinization, the cells are in their suspension.

4.4 Cell suspension for culturing the cells

The suspensions used for cell culturing can be homogenous or non-homogenous. In homogenous suspension, the suspension contains cells of a single type, whereas a non-homogenous suspension has cells from different cell lines. The suspensions used in this work were homogenous as they contained cells of single cell lines. The homogeneity of the suspension was crosschecked with a confocal fluorescence microscope. The number of cells in a suspension is also counted using the microscope. With this hemocytometer, the cell population was adjusted to be 10^7 cells per mL for each type of cell with a mean error of 5%.

The *in vitro* nourishment requirement of cells for their survival and proliferation differs for different types of cells. Indeed, the nutrition needs for normal and cancer cells from the same cell line are different [110, 111]. The following section elaborates the culture medium for the six type of cells used in this study. A humidified air ambience with 5% carbon dioxide (CO₂) at 37°C was maintained for all the cells.

(a) BEAS 2B – normal lung cells

As per the ATCC guidelines, the culture plates on which the cells were cultured were

pre-coated with a pre-coating mixture. The mixture used for BEAS 2B cells contains fibronectin (0.01 mg/mL), bovine collagen (0.03 mg/mL) and bovine serum albumin (0.01 mg/mL) diluted in bronchial epithelial basal medium (BEBM). The reagents used in the work were procured from Sigma-Aldrich. The BEGM bullet kit (Lonza™ Clonetics™) that includes the essential additives (gentamycin/amphotericin was discarded) for primary culture was used for BEAS 2B cells. Supplements such as penicillin (100 units/mL) and streptomycin (100 mg/mL) were added to the medium. For trypsinization, an EDTA solution (0.53 mM) with 0.5% polyvinylpyrrolidone (PVP) was used.

(b) CC-827 – lung cancer cells

The ATCC-recommended medium suitable for culturing CC-827 lung cancer cells is the Roswell Park Memorial Institute (RPMI) 1640 medium. The RPMI-1640 is a product of Hyclone™, US. The medium is suitable for culturing a variety of mammalian leukemic cells. The medium had a 10% heat-inactivated fetal bovine serum (FBS) supplement as base. The trypsinization of the cells was done with 0.25% trypsin (a 0.53 mM EDTA solution).

(c) THLE2 – normal liver cells

A mixture consisting of 2.9 mg/mL of collagen I, 1 mg/mL of fibronectin, and 1 mg/mL of bovine serum albumin in BEBM was used as a pre-coating mixture coated on the culturing plates. The reagents were procured from Sigma-Aldrich. Discarding the gentamycin/amphotericin and epinephrine, the Lonza™ Clonetics™ BEGM bullet kit with a base of epidermal growth factor (EGF) (5 ng/mL), phosphoethanolamine (70

ng/mL), and other additives in the kit were used as growth medium for the THLE2 cells. The supplements for the media were heat-inactivated FBS (Hyclone™, US – 10%) and penicillin-streptomycin (Gibco – 1%). Trypsinization was carried out with 0.5% trypsin (0.53 mM EDTA solution).

(d) HEPG2 – liver cancer cells

The HEPG2 cancer cells from liver tissue were grown in Dulbecco's modified Eagle's medium (DMEM – Hyclone™) in culture plates. Ten percent of FBS (Hyclone™, US) and 1% of penicillin-streptomycin (Gibco) were supplements for the medium. As per ATCC guidelines, trypsinization for these cells was done using 0.5% trypsin (0.53 mM EDTA solution).

(e) 293T – normal kidney cells

These normal cells from kidney tissue were cultured in DMEM (Hyclone™) base. The medium was supplemented with 10% FBS and antibiotics like penicillin-streptomycin and gentamicin.

(f) HeLa – cervical cancer cells

According to the ATCC standard, the HeLa cells were cultured in DMEM (Hyclone™) with 7% fetal calf serum (FCS) and the antibiotics PenStrep and gentamicin as supplements.

4.5 Experimental setup

4.5.1 Optical setup

Optical properties such as absorption and transmittance of light by the cells act as label-free optical biomarkers for classification of cells [55]. These optical profiles of the six types of cells discussed above were measured using an optical experimental set up as shown in Figure 4.1.

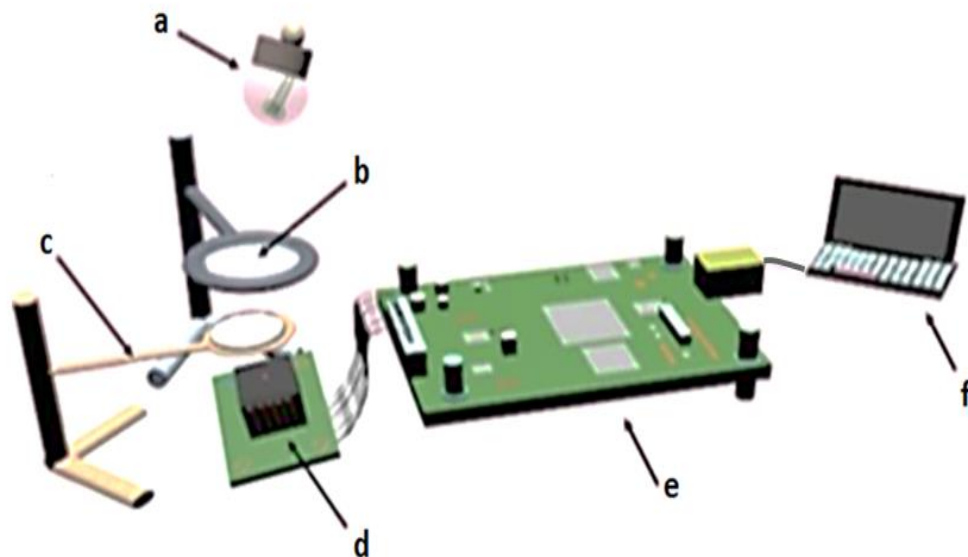


Figure 4.1. Optical setup for measuring the optical profile of cells (a) Xenon light source (b) convex lens holder (c) sample holder (d) Optical sensor (C11708MA - Mini spectrometer with sensor board - C113451-02) (e) Evaluation board (C113451-01) for interfacing (f) laptop equipped with HMS evaluation software for data acquisition (source [54]).

The major components of the setup are as follows:

(a) Xenon light source

(b) Optical sensor

- (c) Interfacing board
- (d) Data acquisition system (personal computer)

The details and description of the components of the setup are outlined below.

(a) Xenon light source

Spectrometers can sense light from two types of sources, namely optical fiber and monochromatic source. An example of a monochromatic light source is a xenon lamp. Monochromatic sources were preferred over optical fiber sources due to their simplicity and the suitability of the specifications for the experimental requirements. The xenon light source (450–1,000 W research arc lamp) from Newport used in the experiment has the specifications listed in Table 4.1 [112]. The advantages of monochromatic light sources are that they have long lifespans and are capable of producing a wide range of light from ultraviolet (UV) to near-infrared (NIR) with ripples kept at a minimum.

Table 4.1. Specifications of the xenon light source

Parameter	Range
Current	(18.0–20.0) A
Output voltage	4.2 V
Power	(450–501) W

(b) Optical sensor – mini-spectrometer

The C11708MA mini-spectrometers (optical sensors) [113] with a sensor board C113451-02 used in this work is from Hamamatsu, Japan. The optical sensor forms the heart of the experimental setup. It is very important to have a compatibility between the wavelength of the light source and that of the sensor. The spectrum of wavelength utilized for the present work is in the range of 640–1,010 nm. The sensor senses the light impinging on it and produces an electrical output signal in accordance with the intensity of the light. The sensor acts as an analog-to-digital (A/D) convertor, and the signal output from the sensor is an A/D count. The specifications of the C11708MA mini-spectrometers used in this study are listed in Table 4.2.

Table 4.2. Specifications of the mini-spectrometer (optical sensor)

Feature	Specification
Thumb size	27.6 x 16.8 x 13 mm
Weight	9 g
Spectral response range	640–1,010 nm
Spectral resolution	20 m

(c) Interfacing evaluation board

The data from the mini-spectrometer was transferred to a personal computer (PC) via an evaluation board. The evaluation board used in the work was the C113451-01, which interfaces to the PC through a C113451-02 sensor board (Hamamatsu, Japan).

(d) Data acquisition system (PC)

The data acquired from the mini spectrometer was processed in a PC with an analyzer software called HMS evaluation software installed. The software converted the A/D count from the sensor to a corresponding light level.

4.5.2 Design of the experiment

This section elaborates on the measurement of the optical biomarkers such as absorption and transmittance properties of the six cells used in the present work. Each cell sample in its suspension in a cuvette was placed in the sample holder and the holder was positioned beneath a lens holder. The convex lens focused the light from the xenon lamp to pass through the sample. The mini-spectrometer fixed below the sample received the light transmitted through the sample, translated the light beam to an electrical signal, and output an A/D count. The evaluation board with the sensor board acted as an interface for the mini-spectrometer and the PC. The A/D count output was transferred to the PC with the evaluation software. The software processed the count and converted the count to a light level.

Chapter 5: Classifications based on Prony models

Label-free methods neither cause cell damage nor contribute to any change in cell composition and intrinsic characteristics. Indeed, there is much interest in the scientific community to learn more from existing methods and to develop new label-free based methods for detection and classification of cells. Cell classification using optical measurements has been frequently utilized. When cells interact with light, due to differences in the composition of different types of cells, changes in the optical absorption and transmission response result. This work combined the advancement in optical measurements and Prony techniques to enhance the classification of cells based on their measured optical profiles. In this work, six types of cells, HeLa, 293T, lung-cancer and normal, and liver- cancer and normal, were suspended in their corresponding medium and their transmission characteristics were assessed. After media de-embedding, the transmission profiles were fitted with a sum of exponentially decaying signals using the Prony algorithm. After that, the optical response of each cell was modeled with a set of extracted parameters: amplitude, frequency, phase, and damping factor. The four parameters extracted via the Prony method are related to the coefficients and locations of the poles for each fitted model. A figure of merit (FOM) has been introduced, whose distribution in the complex z -plane plays a major role in the classification of cell type. The changes in the values of FOM are due to the changes in cell composition and intrinsic characteristics of different cells.

5.1 Results and Discussion

Six types of cells were utilized in this study. These cells were utilized to carry out the proposed current approach in terms of detection capabilities. Normal and cancer cells

of lung and liver were used to demonstrate cell identification using the current approach. Using a hemocytometer, the cell concentration in each suspension was adjusted to 10^7 cells per mL with 5% mean error. After that, each type of cell suspension was loaded in the experimental setup and the optical transmittance of the cells was measured over the wavelength of 640-1050 nm with a wavelength reproducibility between -0.5 to 0.5 nm and maximum of 20 nm FWHM spectra, under constant light conditions. The de-embedding of the medium and holder contributions are then performed by subtracting the suspension responses directly from the filled control medial response.

Figure 5.1 shows the signal intensities varying with wavelength. As the measured signal exhibits transient behavior, a wavelength modified Prony algorithm can be applied. Figures 5.1(a) and (b) depict the measured optical responses superimposed with the Prony estimated signal for the HeLa and 293T cell lines, respectively. The least number of exponentials that gives the best fitted model is considered the optimum order of the model. The optimum order (p) was found to be 40, which is the minimum required order that provides the excellent fitting. A higher order (higher p values) will result in redundancy and require further processing resources. The responses were collected using the experimental setup reported in [55]. It is recommended to apply the same order to both the 293T and HeLa cell suspensions for fair comparison. Parameters such as amplitude, frequency, phase and damping factor of the exponentials are extracted from the fitted response of each type of cell. Further information about the parameter estimation for exponential sums approximated by the Prony method has been detailed by Yan et al. in [114]. Furthermore, a description of the extractions of the corresponding transient parameters, such as the order of the

signal model, the data window length, sampling interval and parameters such as the attenuation factor has been explicitly described in [91].

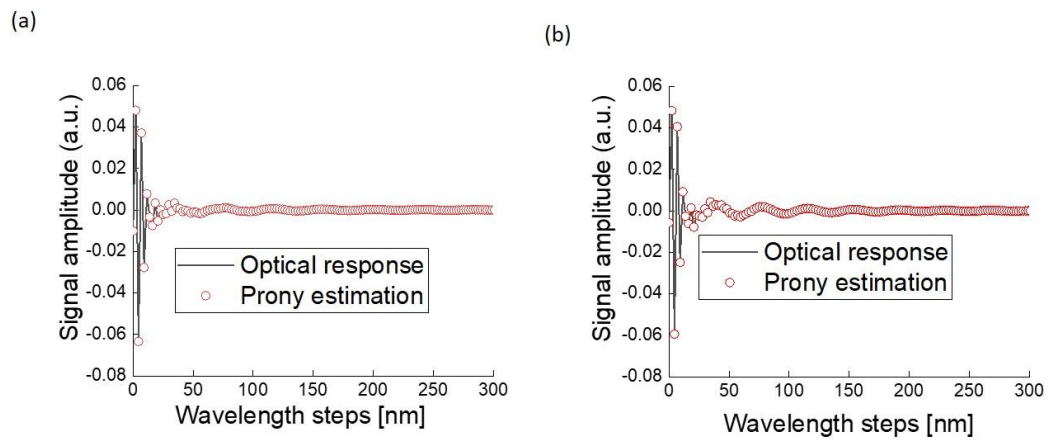


Figure 5.1. Measured optical transmittance response of (a) HeLa cells, (b) 293T fitted with Prony estimations.

The measured transmittance is sampled at a uniform sampling interval of $T_s = 2.3$ nm. This results in a number of samples of $N = 256$.

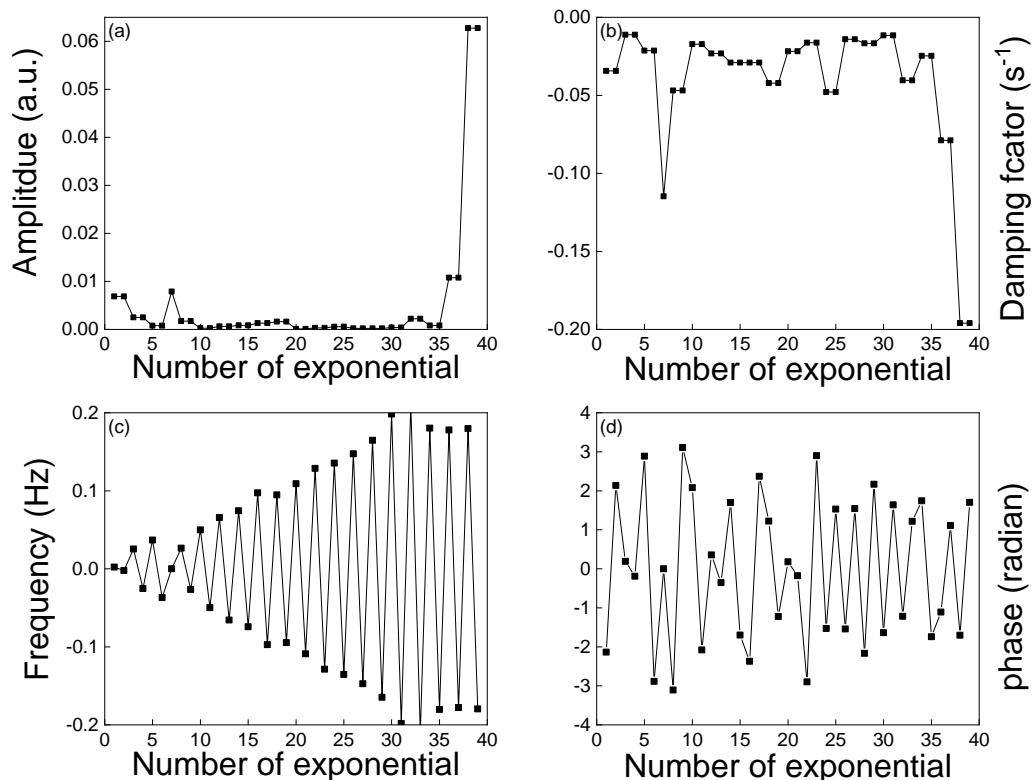


Figure 5.2. Extracted parameters versus number of exponentials in the fitted model for 293T cell line: (a) amplitude (b) damping factor (c) frequency and (d) phase.

Figures 5.2(a), (b), (c) and (d) show the plots of amplitude, damping factor, frequency and phase, respectively, obtained for 293T with a fitting order of 40. The measured data were smoothed using the Savitzky–Golay method [115]. This has been used here to increase the data precision without distorting the signal tendency. The extracted parameters are further processed to extract the corresponding coefficients and pole locations. The coefficients and locations of poles were computed using (3.3.3) and (3.3.4). The extracted coefficients and location of poles for the HeLa and 293T cell suspension are illustrated in Figure 5.2(a) and (b), respectively.

Rodríguez et al. have conducted a review of Prony's method regarding the signal approximation using MATLAB code and have implemented the classical methods to test both performance and Prony approximation [74]. The complete theoretical bases of Prony's method and their piece-by-piece implementation in MATLAB have been presented. Rodríguez's algorithms and codes are adopted in this work. As illustrated in Figure 5.3(a) and (b), the extracted poles are located within the unit circle of the z-plane. The y-axis represents the imaginary part and the x-axis represents the real part. The coefficients of HeLa are focused around the origin point when compared to 293T in the z-plane. The distribution of the coefficients and poles locations is not helpful to be used for cell identification. Therefore, a figure of merit (FOM) is introduced for better identification accuracy. The FOM is defined as follows:

$$\text{FOM}(p) = C(p) / L(p) \quad (5.1.1)$$

where $L(p)$ and $C(p)$ represent the location and coefficients of the poles, respectively. The computed FOM is then normalized for each type of cell with its corresponding maximum value. Although the Prony algorithm was developed for modeling signals in the time domain, it can be applied for responses obtained in frequency domains as well [108]. In his paper, Kumaresan has extracted the poles directly from the frequency response using a technique that is analogous to Prony [108].

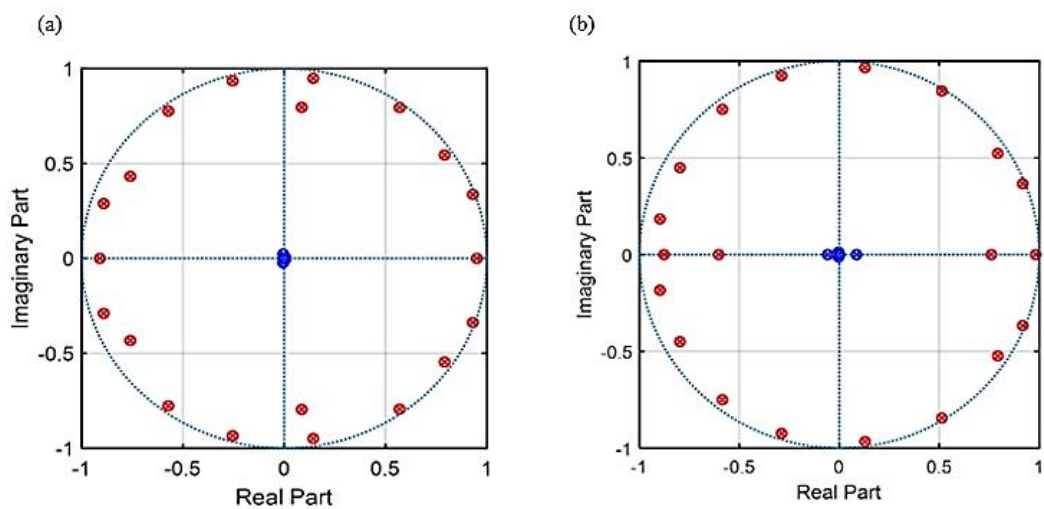


Figure 5.3. Z-plane plot (unit circle) showing coefficients (in blue) and locations of poles (in red) for (a) HeLa and (b) 293T cell suspensions.

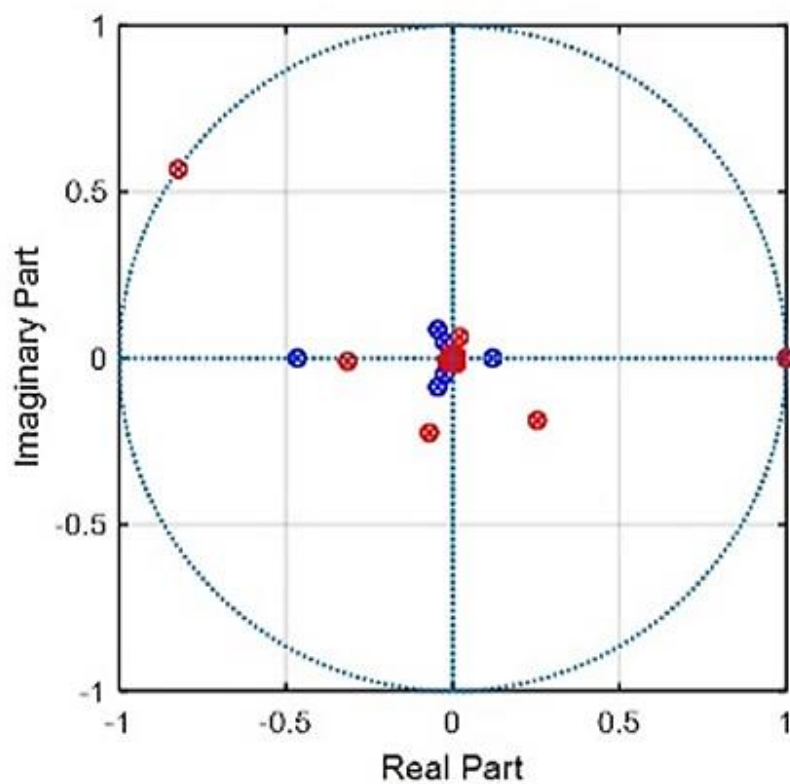


Figure 5.4. Extracted figure of merit (FOM) for 293T (in blue) and HeLa (in red).

Figure 5.4 shows the extracted FOM for HeLa and 293T cells. The FOM distribution for the HeLa is very close to the center of the unit circle. Significant differences in cell composition for normal and cancer cells have been reported. Their interaction with light will cause a change in the optical absorption and transmission response. Due to differences in the composition of the different type of cells, the light interaction with the cells causes an alteration in their absorption and transmission responses. The modifications of the optical responses from normal to cancer were explained mainly by morphological changes, modification of its physiological and biochemical properties that affect the refractive index and allow them to be differentiated from each other. The pole locations and coefficients will be affected accordingly. Empirically, the cancer cells exhibit higher transmittance intensity when compared to normal ones from the same tissue type.

The FOM is inversely proportional to $L(p)$; therefore, for corresponding high locations of poles, lower FOM values are obtained. On the other hand; the complex poles are defined as $\sigma \pm j\omega$, where σ is the damping coefficient and ω is the resonant pulsation. The damping and resonant pulsation are higher in cancer cells compared to normal cells. Therefore, the FOM becomes smaller for cancer cells than for normal cells.

Based on these results, it is evident that the coefficients and poles locations vary with composition and cell morphology. Undeniably, the main difference between normal and cancer cells of the same tissue is in terms of composition and morphology. Hence, the proposed FOM is a distinctive parameter that can be used to explore the detection and identification of normal and cancer cells. This is possible when the technique is used only for fitting the response in the frequency domain to the sum of the damped

exponential and for parameter extraction. The objective here is to make inferences from the obtained parameters and for further processing. These frequency domain measurements cannot be utilized for the generation of a representative equivalent circuit. Hence, this work claims the validity of using the Prony technique to model a frequency domain signal, as the extracted parameters are used for making inferences for cell identification. It is worth mentioning that the focus of this work is to classify normal and cancerous cells for the same tissue. Therefore, the FOM for lung and liver normal and cancerous suspensions were extracted per the introduced procedure and are depicted in Figure 5.5. Figures 5.5(a) and (b) show the FOM of cell lines for lung normal and cancer cells, respectively. Figures 5.5(c) and (d) show the FOM of cell lines for liver normal and cancer cells, respectively.

The distribution of the FOM of cancer cells is closer to the origin of z-plane when compared with that of the normal cells. Each plotted measurement represents the average of 15 measurements. The multiple measurements were conducted on different aliquots taken from the same sample suspension in the same region spot. The error bars in the subfigures of Figure 5.5 represent the average values along with maximum and minimum values. The bar corresponding to the x-axis represents the average in the FOM real part, while the endpoints represent its maximum and minimum values. The bar corresponding to the y-axis represents the average in the FOM imaginary part, while the endpoints represent its maximum and minimum values.

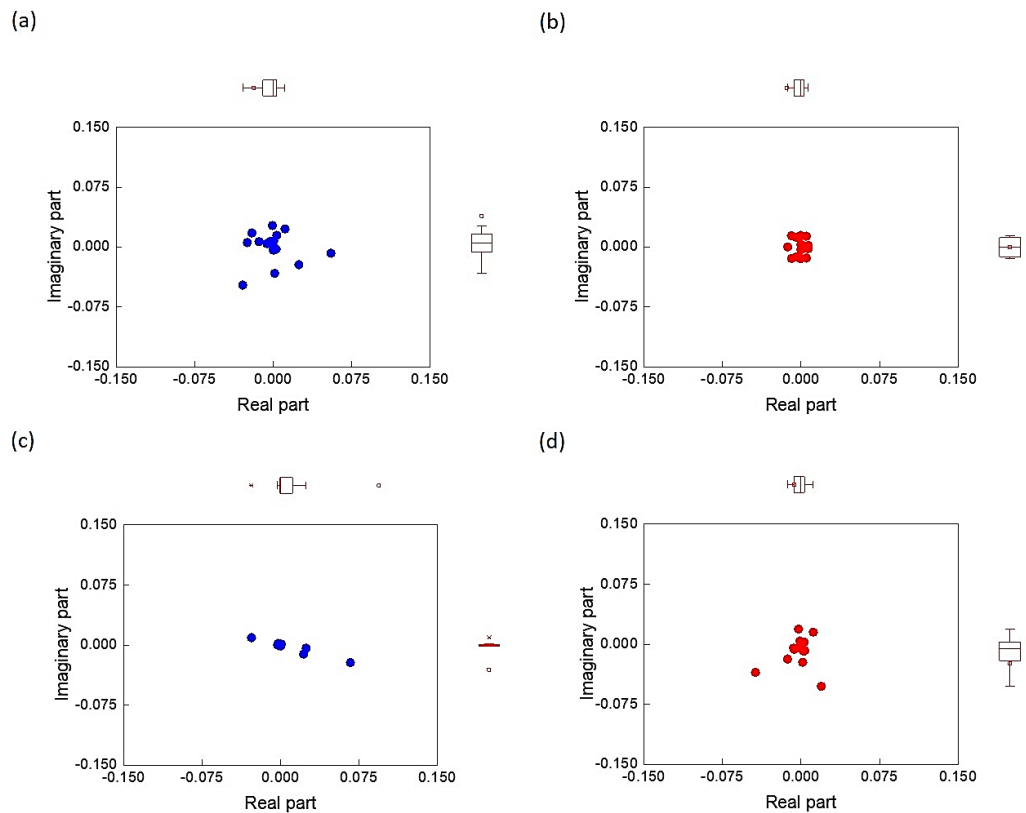


Figure 5.5. Extracted figure of merit (FOM) for normal and cancer cell lines from same tissue: (a) lung normal, (b) lung cancer (c) liver normal and (d) liver cancer.

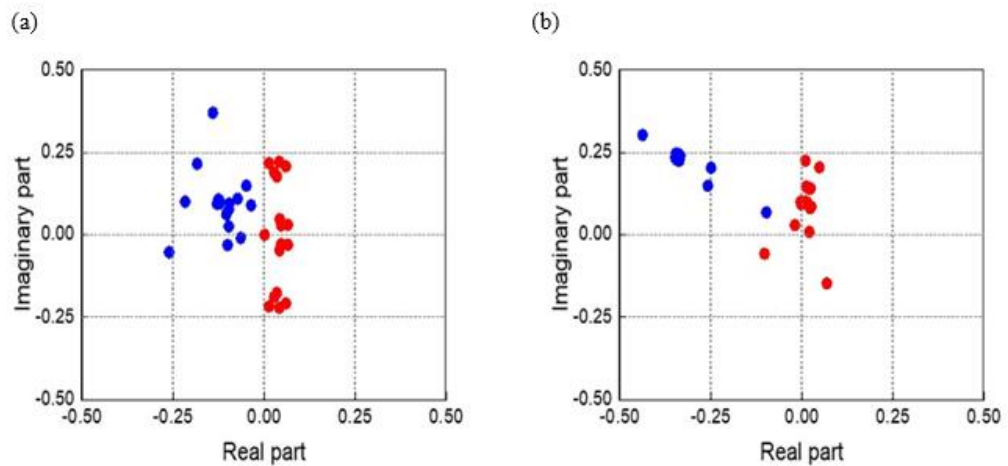


Figure 5.6. Figure of merit distributions: (a) lung normal and cancer cells, and (b) liver normal and cancer cells. The blue dots represent the normal cell lines and the red dots count for the cancer cell lines.

For further investigations, the distribution of the FOM for normal and cancer cells has been superimposed on each other, as depicted in Figure 5.6. Figure 5.6(a)

superimposes the lung normal and cancer corresponding FOMs. Figure 5.6(b) superimposes the FOMs for the liver normal and cancer cell lines. The majority of the real part of the FOMs corresponding to cancer cells have positive real part (located in the right hand side) of the z-plane; the majority of the real part of the FOMs corresponding to normal cells have negative real part (located in the left hand side) of the plot.

The figure of merit (FOM) which is introduced for the first time relates the location of the poles ($L(p)$) and the coefficient of poles $C(p)$). Scientifically: significant differences in cell composition for normal and cancer cells have been reported [55]. Their interaction with light will cause a change in the optical absorption and transmission response. Due to differences in the composition of the different type of cells, the light interaction with the cells causes an alteration in their absorption and transmission responses. The modifications of the optical response from normal to cancer state were explained mainly by morphological changes, modification of its physiological and biochemical properties that affect the refractive index and allowing them to be differentiated from each other. The poles location and coefficients will be affected accordingly. Therefore, it is suggested that within the range -0.5 to $+0.5$ in the z-plane, if 85% of the FOM values have negative real part (located in the left hand side) then the cell lines under study is considered to be normal; else it is cancer cells. There is a clear discrimination strategy: by performing optical measurements on the different in vitro normal and cancer cell line models, the developed data processing procedure based on the Prony method to achieve a label-free discrimination between cancer and healthy cells from the same tissue type works very well.

5.2 Conclusion

In summary, this work addressed the classification and discrimination between normal and cancer cells from the same tissues. A label-free method combining the Prony estimation theory and optical transmittance measurements was introduced and proven to be a powerful technique. The proposed approach has been examined using six types of different cell lines. The measured optical responses of the six types of cells have been reconstructed using the Prony algorithm with same fitting order of 40. Based on the observations, a normalized figure of merit has been introduced for identification. Based on this merit, the distribution of the FOMs around the center of the unit circle of the cancer cell lines was closer than the normal cell lines from same tissues (in the case of lung and liver cells). These preliminary findings can be considered the foundation stage for cell identification using optical measurements combined with the Prony estimation theory.

Chapter 6: Classification based on Autoregression models

Label free based methods received huge interest in the field of bio cell characterizations because they do not cause any cell damage nor contribute any change in its compositions. This work takes a close outlook of cancerous cells discrimination from normal cells utilizing parametric modeling approach. Autoregressive (AR) modeling techniques is used to fit a measured optical transmittance of both cancer and normal cells profiles. Analysis of variance (ANOVA) statistical approach is incorporated in this work to determine the significant AR coefficients. The transmitted light intensity passes through the cells get affected by their intercellular compositions and membrane properties. In this study; four types of cells lung- cancerous and normal, liver- cancerous and normal cells were suspended in their corresponding medium and their transmission characteristics were collected and processed. The AR coefficients of each type of cell were analyzed with the statistical technique ANOVA, which provided the significant coefficients. The poles extracted from the significant coefficients provided an improved demarcation for normal and cancerous cells. These outcomes can be further utilized for cell classification using statistical tools.

6.1 Results and Discussion

The cell count in the suspension of each type of cell was conducted using a hemocytometer with a 5% mean error. The suspension was then loaded in the experimental setup and the optical profile of the cells in the suspension was recorded. The measured transmission profile of the benign and malignant cells of the lung and liver tissues are shown in Figure 6.1(a), (b), (c), and (d), respectively.

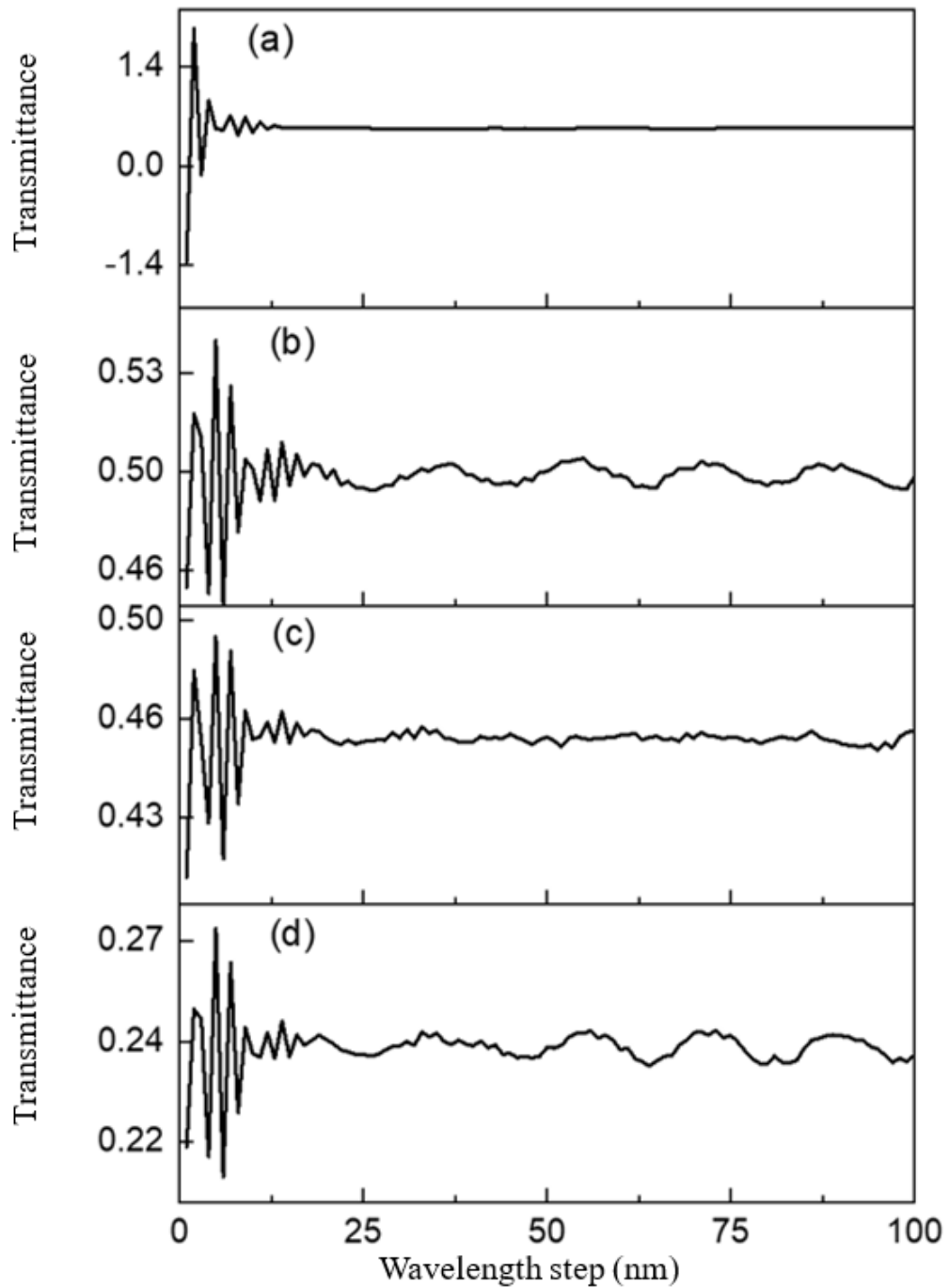


Figure 6.1. Measured optical transmittance response of (a) cancer liver (b) normal liver (c) cancerous lung and (d) normal lung cells. The measured transmittance was sampled at a uniform sampling interval of $W_s = 2.3$ nm.

Table 6.1 summarizes the metrics such as prediction accuracy, MSE, and FPE of the fitted AR model of each type of cell for an order 6. The complexity of the model

increases with the order of the model.

Table 6.1. Performance measure of a fitted AR model for the transmittance response of different types of cells for order 6

Type of cell	Prediction accuracy	MSE	FPE
Normal lung	91.61%	5.72e-08	6.00e-08
Cancerous lung	90.75%	1.13e-07	1.18e-07
Normal liver	93.48%	6.53e-08	6.84e-08
Cancerous liver	99.76%	6.59e-08	6.91e-08

The AR model coefficients of order 6 obtained for the four types of cells are shown in Table 6.2. As shown, the coefficients are of different values for different types of cells. This reflects the alteration in the composition and intrinsic properties of the different cell types. Moreover, the normal and cancer cells from the same tissue have different coefficient values, implying the variation in their composition, morphology, and intrinsic properties.

Table 6.2. Set of extracted AR coefficients for different types of cells

AR coefficients	Normal lung	Cancerous lung	Normal liver	Cancerous liver
a1	- 0.16	+ 0.06	- 0.13	- 0.68
a2	- 0.90	- 0.94	- 0.96	- 0.95
a3	+ 0.03	- 0.23	- 0.08	+ 0.43
a4	+ 0.42	+ 0.36	+ 0.46	+ 0.36
a5	+ 0.05	+ 0.12	+ 0.13	- 0.09
a6	+ 0.01	+ 0.02	+ 0.06	- 0.01

Table 6.3 shows the poles extracted for the normal and cancer cells of the lung and liver tissue used in this work. The poles can be real valued or complex conjugate pairs. For instance, the poles P1 and P2 of liver cancer cells are real and distinct, while their remaining poles (P3–P6) occur as complex conjugate pairs. All the extracted poles of lung (normal and cancer) cells occur as complex conjugate pairs.

The distribution of the poles of the normal and cancer cells of the lung and liver in the z-plane is shown in Figure 6.2(a) and (b), respectively. The poles of the normal cells are illustrated with blue dots whereas the red dots represent the poles of the cancer cells. As shown in Figure 6.2, the poles of the different cells have different distributions in the z-plane. In addition, the distribution of the poles of the normal and cancer cells of the same tissue are also different. Any deviation from the pole values of the normal cells shows the presence of abnormalities.

Table 6.3. Set of poles extracted for different types of cells

Type of cell	P1, P2	P3, P4	P5, P6	P1, P2
Normal lung	$-0.687 \pm 0.305i$	$-0.05 \pm 0.16i$	$0.820 \pm 0.32i$	$-0.687 \pm 0.305i$
Cancerous lung	$-0.688 \pm 0.309i$	$-0.15 \pm 0.21i$	$0.800 \pm 0.24i$	$-0.688 \pm 0.309i$
Normal liver	$-0.71 \pm 0.33i$	$-0.10 \pm 0.32i$	$0.88 \pm 0.32i$	$-0.71 \pm 0.33i$
Cancerous liver	$-0.1 + 0i, 0.35 + 0i$	$-0.65 \pm 0.23i$	$0.87 \pm 0.11i$	$-0.1 + 0i, 0.35 + 0i$

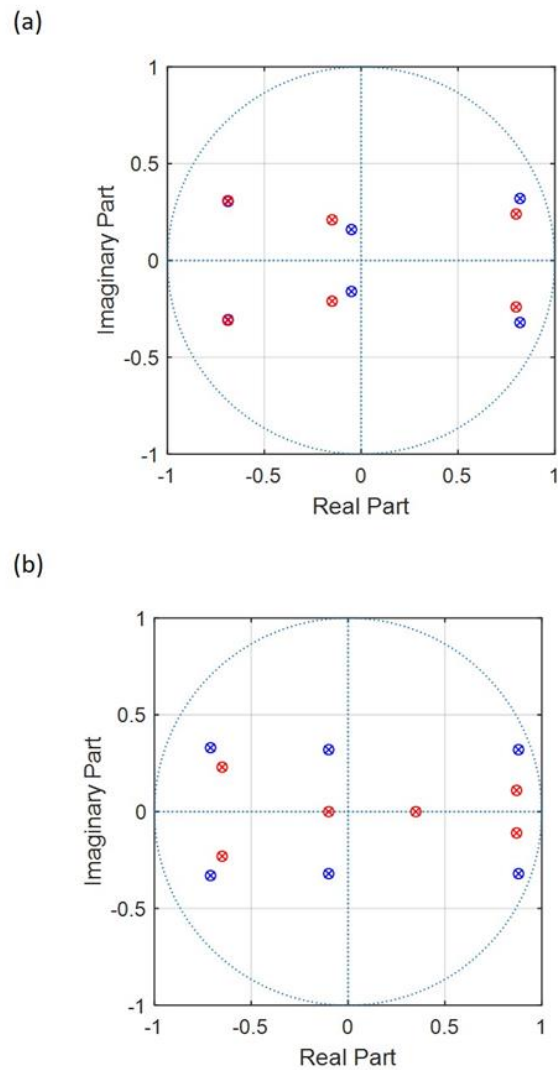


Figure 6.2. Z-plane showing the distribution of the poles of the (a) normal (blue) and cancer (red) lung cells and the (b) normal (blue) and cancer (red) liver cells.

The modifications of the optical response from normal to cancer state were explained mainly by morphological changes, modification of its physiological and biochemical properties that affect the refractive index and allowing them to be differentiated from each other. The poles location and coefficients will be affected accordingly.

It is worth to mention that the measurements were conducted on cells suspended in their corresponding media. The applied optical measurements conditions did not cause

any harsh to the cells. The applied light did not intracellular temperature of the cells. The O_2 is dissolved in the media which will help the cells to survive. The pH has been maintained; has not been affected by light. The pH has been measured before and after. The temperature of the suspension has been measured before and after, almost the same. The measurements have been conducted at room temperature. The cells are suspended in media that is rich with nutrient, to keep them alive. Cells have been subjected to light for less than 5 minutes. This is not significant time to make them die; mainly the cells during measurements were suspended inside the media. Cell viability test, the most common test using the try-band loop staining has been used to check the suspension before and after the optical measurements; before the light the percentage of living cell was above 90%, after the optical measurements; the percentage of living cell was above 85%.

To reduce redundancy and arrive at a concise AR model, statistical tools such as the N-way ANOVA technique were applied. The ANOVA revealed the significance of the AR coefficients. The ANOVA technique was applied to the AR coefficients rather than the poles since the poles were extracted from the coefficients. The coefficient that gives the highest value of the mean square is the significant AR coefficient. Three coefficients – a_1 , a_3 , and a_5 – out of the six coefficients shown in Table 6-2 were found to be significant. Hence the order of the AR model is reduced by one degree. This will reduce the complexity of the system. The new set of reduced poles distribution in the Z-plane are plotted in Figure 6.3.

The complex poles are defined as $\sigma \pm j\omega$, where σ is the damping coefficient (real part of the pole) and ω is the resonant pulsation (imaginary part of the pole). The poles

damping and resonant can be used to identify the quality factor of the pole [116], as follow:

$$Q=\omega/2\sigma \quad (6.1.1)$$

The pole quality (Q) can then be used to discriminate between cancer and normal cells. The pole quality can be considered as a figure of merit (FOM) which is introduced for the first time in the work. It correlates the real part with the imaginary part to develop a discrimination procedure. The location of the poles strongly affected by the differences in cell composition for normal and cancer cells. The modifications of the optical response from normal to cancer state were explained mainly by morphological changes, modification of its physiological and biochemical properties that affect the refractive index and allowing them to be differentiated from each other; as previously indicated. The poles quality factors will be affected accordingly. Empirically, the cancer cells exhibit higher transmittance intensity when compared to normal ones from the same tissue type.

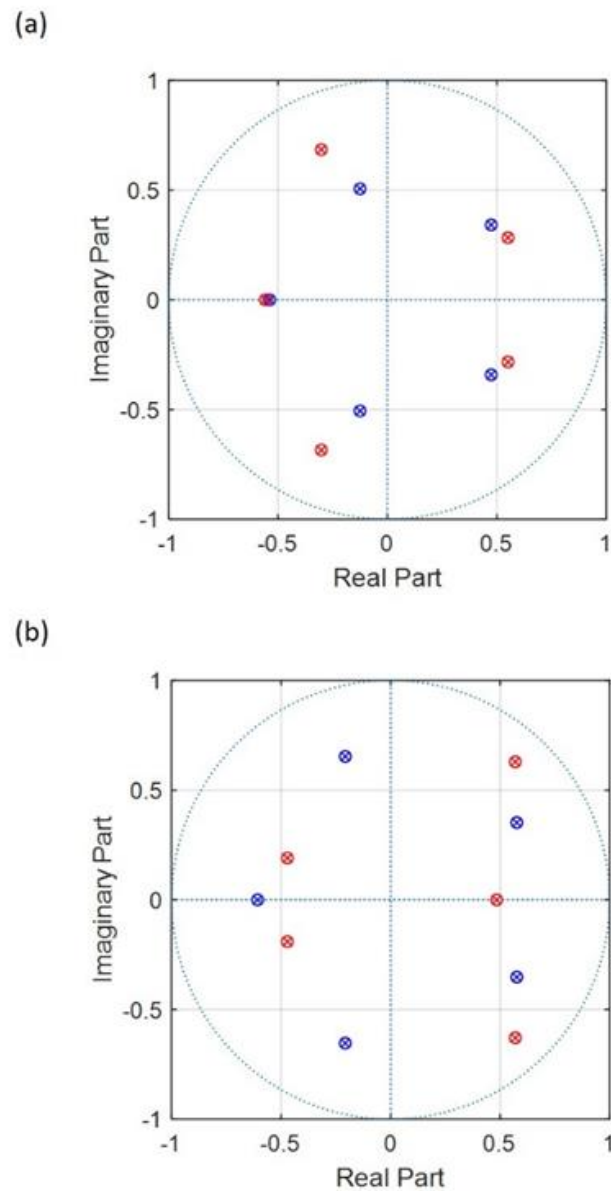


Figure 6.3. Z-plane showing the distribution of the reduced poles of the (a) normal (blue) and cancer (red) lung cells and the (b) normal (blue) and cancer (red) liver cells.

The corresponding computed poles quality factors for the four type of cells under study are shown in Figure 6.4. Figure 6.4(a) and (b) represents the distributions of Q-factor for lung (normal (blue), cancer (red)) and liver cells lung (normal (blue), cancer (red)), respectively. Figure 6.4 revealed that the magnitude of the pole quality factor for cancer cells the cancer cells is higher than normal.

Therefore, the proposed approach offers a very straight forward and clear discrimination strategy: by performing optical measurements on the different in vitro cell normal and cancer cell line models, the developed data processing procedure based on the AR method to achieve a label-free discrimination between cancer and healthy cells from the same tissue type works very well. Furthermore; the proposed approach can be coupled or integrated with existing techniques and methods to enhance the discrimination between cancer and normal cells from same tissues. The current method utilized statistical methods that are less expensive than machine learning based methods.

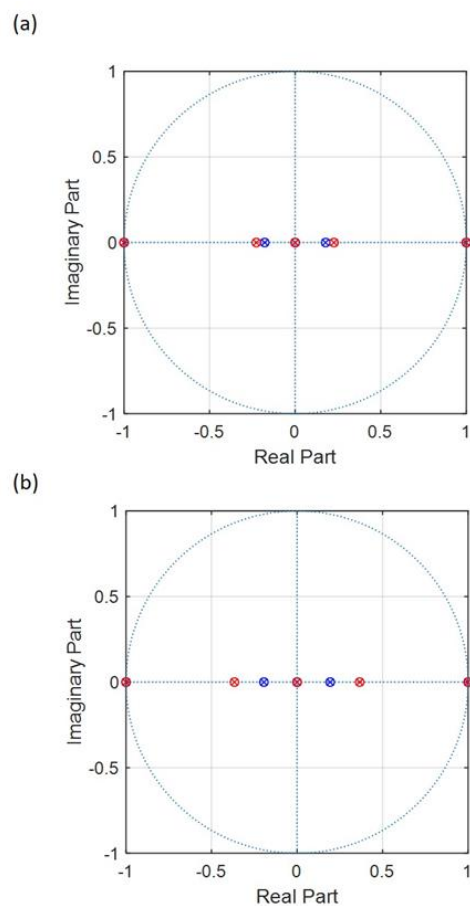


Figure 6.4. Z-plane showing the distribution of the Q-factor of the (a) normal (blue) and cancer (red) lung cells and the (b) normal (blue) and cancer (red) liver cells.

6.2 Conclusion

This work utilized optical techniques combined with an AR modeling method and statistical techniques for the classification of normal and cancer cells. The approach used in the present work was applied to normal and cancer cells from lung and liver tissues. The AR coefficients of each type of cell were analyzed with the statistical technique ANOVA, which provided the significant coefficients. The poles extracted from the significant coefficients provided an improved demarcation for normal and cancer cells. These preliminary outcomes can be further utilized for cell classification using statistical tools.

Chapter 7: Conclusion and Future Outlook

7.1 Conclusion

In this thesis, the downsides of labelled techniques of classification of cells such as invasiveness, need of skilled operators, requirement of large amount of antibodies and antigens are addressed by utilizing a label-free method of cell classification. The rapid and real time label-free based cell classification method presented in the work combines the advancements in optical techniques and numerical methods.

The analysis of the measured optical profiles of different type of cells carried out using Prony modeling shows that the interaction of light with cells largely depend on the intrinsic properties and compositions of the cells. Therefore, any alteration in the composition or intrinsic properties due to abnormality or disease infection is revealed in the difference in the different parameters such as amplitude, phase, frequency and damping factor extracted from the Prony modeling. In addition to these parameters, the location of poles and the coefficient values obtained for cells of different tissues as well as normal and cancerous cells of same tissue exhibited different values. The main contribution of this work is introduction of a figure of merit (FOM) that is determined from the location of poles and the coefficient values. The distribution of the FOM in the complex z-plane showed prominent difference for normal and cancer cells of same tissue and of different tissues as well.

In addition to the Prony modeling, the measured optical profiles of the cells are analyzed using another numerical technique – Autoregressive (AR) modeling. The difference in the AR coefficient values as well as in the AR pole values extracted from

the coefficients substantiate the change in the properties of normal cells due to some abnormality and thus serving as an efficient tool for cell classification.

The focus of the thesis is development of label-free methods for early detection of diseases like cancer. During in the inception stage of diseases like cancer, the change in the intrinsic properties and the composition of the cells is meagre. Hence, the measured optical profile and the extracted parameters are in close proximity that results in difficulty in the classification of cells. The contribution of this work to overcome this is by utilizing statistical tool called Analysis of variance (ANOVA) to determine the significance of the AR coefficients. The AR poles extracted based on the reduced set of significant AR coefficients showed higher difference in their values for normal and cancerous cells.

These analyses were performed on normal and cancerous cells of same tissues (lung and liver) and from different tissues (kidney and cervix). Preliminary results show that the approach adopted in this work serves as an efficient tool for classification of normal and cancerous cells.

7.2 Future outlook

The findings of the work open wide arena for future developments in the domain of cell classification and early detection of diseases like cancer. The analysis conducted on six types of cells form the foundation stage in cell classification. As a future work, the analysis can be conducted on large number of cells – normal cells of different tissues as well as their cancerous counterpart. A look-up table can be built with the coefficient values and pole values of large variety of normal and cancer cells which can be used as a quick reference in the first stage of classification of cells.

The essence of Prony technique is its applicability to analyze transient signals irrespective of the domain of measurement. As a future work, Prony analysis can be conducted on transient signals measured in time domain for different types of cells and then the poles obtained through the modeling technique can be further utilized for development of transfer function followed by the synthesis of electrical equivalent circuits. The coefficients obtained through the modeling techniques can be further analyzed using other statistical tools such principal component analysis (PCA) which aids in the prediction of the state of cells.

The ultimate future of the work will be in the development of portable medical devices for point-of-care treatment of diseases like cancer. In addition to cancer detection, the analysis technique discussed in the work can be applied to neuron cells to define the state of brain and help in taking precaution measures for stroke – prone patients.

References

- [1] V. T Broadbridge, C. S. Karapetis, C. Beeke, R. J. Woodman, R. Padbury, G. Maddern, S. W. Kim, D. Roder, P. Hakendorf, and T. J. Price., "Do metastatic colorectal cancer patients who present with late relapse after curative surgery have a better survival?," *British Journal of Cancer*, vol. 109, no. 5, pp. 1338-1343, 2013.
- [2] K. J. Martin, M. V. Fournier, G. P. V. Reddy, & A. B. Pardee, "A need for basic research on fluid-based early detection biomarkers," *Cancer Research*, vol. 70, no. 13, pp. 5203-5206, 2010.
- [3] P. Kuppusamy, N. Govindan, M. M. Yusoff, & S. J. Ichwan, "Proteins are potent biomarkers to detect colon cancer progression," *Saudi Journal of Biological Sciences*, vol. 24, no. 6, pp. 1212-1221, 2017.
- [4] A. Voller, A. Bartlett, & D.E. Bidwell, "Enzyme immunoassays with special reference to ELISA techniques," *Journal of Clinical Pathology*, vol. 31, no. 6, pp. 507-520, 1978.
- [5] T. Porstmann, & S. T. Kiessig, "Enzyme immunoassay techniques an overview," *Journal of Immunological Methods*, vol. 150, no. 1-2, pp. 5-21, 1992.
- [6] N. Konishi, K. Matsumoto, Y. Hiasa, Y. Kitahori, I. Hayashi, & H. Matsuda, "Tissue and serum pepsinogen I and II in gastric cancer identified using immunohistochemistry and rapid ELISA," *Journal of Clinical Pathology*, vol. 48, no. 4, pp. 364-367, 1995.
- [7] A. Ziegler, A. Koch, K. Krockenberger, & A. Großhennig, "Personalized medicine using DNA biomarkers: a review," *Human Genetics*, vol. 131, no. 10, pp. 1627-1638, 2012.
- [8] B. K. Thakur, H. Zhang, A. Becker, I. Matei, Y. Huang, B. Costa-Silva, & C. Williams, "Double-stranded DNA in exosomes: a novel biomarker in cancer detection," *Cell Research*, vol. 24, no. 6, pp. 766-769, 2014.
- [9] H. E. Mulcahy, J. Lyautey, C. Lederrey, X. Q. Chen, F. Lefort, V. Vasioukhin, & M. Stroun, "Plasma DNA K-ras mutations in patients with gastrointestinal malignancies," *Annals of the New York Academy of Sciences*, vol. 906, no. 1, pp. 25-28, 2000.
- [10] P. Anker, F. R. A. N. C. O. I. S. Lefort, V. A. L. E. R. I. Vasioukhin, J. A. C. Q. U. E. L. I. N. E. Lyautey, C. H. R. I. S. T. I. N. E. Lederrey, X. Q. Chen, & M. J. Farthing, "K-ras mutations are found in DNA extracted from the plasma of patients with colorectal cancer," *Gastroenterology*, vol. 112, no. 4, pp. 1114-1120, 1997.

- [11] C. Luparello, A. Noël, & I. D. A. Pucci-Minafra, "Intratatumoral heterogeneity for hsp90 β mRNA levels in a breast cancer cell line," *DNA and Cell Biology*, vol. 16, no. 10, pp. 1231-1236, 1997.
- [12] B. Cagir, A. Gelmann, J. Park, T. Fava, A. Tankelevitch, E. W. Bittner ... & S. A. Waldman, "Guanylyl cyclase C messenger RNA is a biomarker for recurrent stage II colorectal cancer," *Ann Intern Med*, vol. 131, no. 11, pp. 805-812, 1999.
- [13] S. Mehrabani, A. Maker, & A. Armani, "Hybrid integrated label-free chemical and biological sensors," *Sensors*, vol. 14, no. 4, pp. 5890-5928, 2014.
- [14] J. Chen, Y. Zheng, Q. Tan, E. Shojaei-Baghini, Y. L. Zhang, J. Li, ... & Y. Sun, "Classification of cell types using a microfluidic device for mechanical and electrical measurement on single cells," *Lab on a Chip*, vol. 11, no. 18, pp. 3174-3181, 2011.
- [15] D. R. Gossett, T. K. Henry, S. A. Lee, Y. Ying, A. G. Lindgren, O. O. Yang, & D. Di Carlo, "Hydrodynamic stretching of single cells for large population mechanical phenotyping," in *Proceedings of the National Academy of Sciences*, vol. 109, no. 20, pp. 7630-7635, 2012.
- [16] V. Swaminathan, K. Mythreye, E. T. O'Brien, A. Berchuck, G. C. Blobe, & R. Superfine, "Mechanical stiffness grades metastatic potential in patient tumor cells and in cancer cell lines," *Cancer Research*, vol. 71, no. 15, pp. 5075-5080, 2011.
- [17] W. Kim, & A. Han, "A micro-aspirator chip using vacuum expanded microchannels for high-throughput mechanical characterization of biological cells," in *14th International Conference on Miniaturized Systems for Chemistry and Life Sciences*, Groningen, The Netherlands, 2010.
- [18] M. M. Brandao, A. Fontes, M. L. Barjas-Castro, L. C. Barbosa, F. F. Costa, C. L. Cesar, & S. T. O. Saad, "Optical tweezers for measuring red blood cell elasticity: application to the study of drug response in sickle cell disease.," *European Journal of Haematology*, vol. 70, no. 4, pp. 207-211, 2003.
- [19] M. Lekka, "Discrimination between normal and cancerous cells using AFM," *Bionanoscience*, vol. 6, no. 1, pp. 65-80, 2016.
- [20] G. Qiao, W. Wang, W. Duan, F. Zheng, A. J. Sinclair, & C. R. Chatwin, "Bioimpedance analysis for the characterization of breast cancer cells in suspension.," in *IEEE Transactions on Biomedical Engineering*, 2012.
- [21] K. Wang, Y. Zhao, D. Chen, B. Fan, Y. Lu, L. Chen, ... & J. Chen, "Specific membrane capacitance, cytoplasm conductivity and instantaneous Young's modulus of single tumour cells," *Scientific Data*, vol. 4, pp. 1-8, 2017.

- [22] L. A. Flanagan, J. Lu, L. Wang, S. A. Marchenko, N. L. Jeon, A. P. Lee, & E. S. Monuki, "Unique dielectric properties distinguish stem cells and their differentiated progeny," *Stem Cells*, vol. 26, no. 3, pp. 656-665, 2008.
- [23] K. Heileman, J. Daoud, & M. Tabrizian, "Dielectric spectroscopy as a viable biosensing tool for cell and tissue characterization and analysis," *Biosensors and Bioelectronics*, vol. 49, pp. 348-359, 2013.
- [24] A. Han, L. Yang, & A. B. Frazier, "Quantification of the heterogeneity in breast cancer cell lines using whole-cell impedance spectroscopy.," *Clinical Cancer Research*, vol. 13, no. 1, pp. 139-143, 2007.
- [25] J. E. Da Silva, J. M. De Sá, & J. Jossinet, "Classification of breast tissue by electrical impedance spectroscopy," *Medical and Biological Engineering and Computing*, vol. 38, no. 1, pp. 26-30, 2000.
- [26] J. Gao, S. Yue, J. Chen, & H. Wang, "Classification of normal and cancerous lung tissues by electrical impedance tomography.," *Biomedical Materials and Engineering*, vol. 24, no. 6, pp. 2229-2241, 2014.
- [27] P. Aberg, I. Nicander, J. Hansson, P. Geladi, U. Holmgren, & S. Ollmar, "Skin cancer identification using multifrequency electrical impedance-a potential screening tool," in *IEEE Transactions on Biomedical Engineering*, 2004.
- [28] R. Dua, D. G. Beetner, W. V. Stoecker, & D. C. Wunsch, "Detection of basal cell carcinoma using electrical impedance and neural networks," in *IEEE Transactions on Biomedical Engineering*, 2004.
- [29] A. N. T. T. I. Penttila, E. M. McDOWELL, & B. J. Trump, "Optical properties of normal and injured cells, Application of cytophographic analysis to cell viability and volume studies," *Journal of Histochemistry & Cytochemistry*, vol. 25, no. 1, pp. 9-20, 1977.
- [30] N. Fujioka, Y. Morimoto, T. Arai, K. Takeuchi, M. Yoshioka, & M. Kikuchi, "Differences between infrared spectra of normal and neoplastic human gastric cells," *Journal of Spectroscopy*, vol. 18, no. 1, pp. 59-66, 2004.
- [31] P. Y. Liu, L. K. Chin, W. Ser, H. F. Chen, C. M. Hsieh, C. H. Lee, & K. Wang, "Cell refractive index for cell biology and disease diagnosis: past, present and future," *Lab on a Chip*, vol. 16, no. 4, pp. 634-644, 2016.
- [32] Z. Wang, G. Popescu, K.V. Tangella, & A. Balla, "Tissue refractive index as marker of disease," *Journal of Biomedical Optics*, vol. 16, no. 11, pp. 1-8, 2011.
- [33] C. R. Simpson, M. Kohl, M. Essenpreis, & M. Cope, "Near-infrared optical properties of ex vivo human skin and subcutaneous tissues measured using the

- Monte Carlo inversion technique," *Physics in Medicine & Biology*, vol. 43, no. 9, pp. 2465-2478, 1998.
- [34] J. R. Mourant, A. H. Hielscher, A. A. Eick, T. M. Johnson, & J. P. Freyer, "Evidence of intrinsic differences in the light scattering properties of tumorigenic and nontumorigenic cells.," *Cancer Cytopathology: Interdisciplinary International Journal of the American Cancer Society*, vol. 84, no. 6, pp. 366-374, 1998.
- [35] X. Huang, P. K. Jain, I. H. El-Sayed, & M. A. El-Sayed, "Gold nanoparticles: interesting optical properties and recent applications in cancer diagnostics and therapy," *Nanomedicine*, vol. 2, no. 5, pp. 681-693, 2007.
- [36] L. Wu, & X. Qu, "Cancer biomarker detection: recent achievements and challenges," *Chemical Society Reviews*, vol. 44, no. 10, pp. 2963-2997, 2015.
- [37] A. M. Wang, M.V. Doyle, & D.F. Mark, "Quantitation of mRNA by the polymerase chain reaction," *Proceedings of the National Academy of Sciences*, vol. 86, no. 24, pp. 9717-9721, 1989.
- [38] S. M. Hanash, S. J. Pitteri, & V. M. Faca, "Mining the plasma proteome for cancer biomarkers," *Nature*, vol. 452, no. 7187, pp. 571-579, 2008.
- [39] M. F. Lopez, "www.genengnews.com," November 2015. [Online]. Available: <https://www.genengnews.com/insights/circulating-protein-biomarkers-a-boon-to-breast-cancer-management/>. [Accessed 5 September 2019].
- [40] G. P. Rai, & K. S. Venkateswaran, "Limitations and Practical Problems in Enzyme Linked Immunosorbent Assays," *Defence Science Journal*, vol. 42, no. 2, pp. 71-84, 1992.
- [41] A. L. Lehninger, D. L. Nelson, & M. M. Cox, *Lehninger principles of biochemistry*, New York: Worth Publishers, 2000.
- [42] P. Anker, H. Mulcahy, X. Q. Chen, & M. Stroun, "Detection of circulating tumour DNA in the blood (plasma/serum) of cancer patients," *Cancer and Metastasis Reviews*, vol. 18, no. 1, pp. 65-73, 1999.
- [43] K. Cuschieri, & N. Wentzensen, "Human papillomavirus mRNA and p16 detection as biomarkers for the improved diagnosis of cervical neoplasia," *Cancer Epidemiology and Prevention Biomarkers*, vol. 17, no. 10, pp. 2536-2545, 2008.
- [44] E. Stern, A. Vacic, N. K. Rajan, J. M. Criscione, J. Park, B. R. Ilic, & T. M. Fahmy, "Label-free biomarker detection from whole blood," *Nature Nanotechnology*, vol. 5, no. 2, pp. 138 -142, 2010.

- [45] K. H. Han, A. Han, & A. B. Frazier, "Microsystems for isolation and electrophysiological analysis of breast cancer cells from blood," *Biosensors and Bioelectronics*, vol. 21, no. 10, pp. 1907-1914, 2006.
- [46] M. Al Ahmad, F. Mustafa, L. M. Ali, J. V. Karakkat, & T. A. Rizvi, "Label-free capacitance-based identification of viruses," *Scientific Reports*, vol. 5, pp. 1-6, 2015.
- [47] L. Glass-Marmor, R. Beitner, "Detachment of glycolytic enzymes from cytoskeleton of melanoma cells induced by calmodulin antagonists," *European Journal of Pharmacology*, vol. 328, no. 2 - 3, pp. 241-248, 1997.
- [48] E. L. Baker, R. T. Bonnecaze, & M. H. Zaman, "Extracellular Matrix Stiffness and Architecture Govern Intracellular," *Biophysical Journal*, vol. 97, no. 4, pp. 1013-1021, 2009.
- [49] X. Liang, K. A. Graham, A. C. Johannessen, D. E. Costea, & F. H. Labeed, "Human oral cancer cells with increasing tumorigenic abilities exhibit higher effective membrane capacitance," *Integrative Biology*, vol. 6, no. 5, pp. 545-554, 2014.
- [50] S. B. Huang, Y. Zhao, D. Chen, S. L. Liu, Y. Luo, T. K. Chiu, & M. H. Wu, "Classification of cells with membrane staining and/or fixation based on cellular specific membrane capacitance and cytoplasm conductivity," *Micromachines*, vol. 6, no. 2, pp. 163-171, 2015.
- [51] Y. Zhao, M. Jiang, D. Chen, X. Zhao, C. Xue, R. Hao, & J. Chen, "Single-cell electrical phenotyping enabling the classification of mouse tumor samples," *Scientific Reports*, vol. 6, pp. 1-8, 2016.
- [52] Y. Zhao, D. Chen, Y. Luo, H. Li, B. Deng, S.B. Huang, ... & X. Zhao, "A microfluidic system for cell type classification based on cellular size-independent electrical properties," *Lab on a Chip*, vol. 13, no. 12, pp. 2272-2277, 2013.
- [53] Y. Zhou, S. Basu, E. Laue, & A.A. Seshia, "Single cell studies of mouse embryonic stem cell (mESC) differentiation by electrical impedance measurements in a microfluidic device," *Biosensors and Bioelectronics*, vol. 81, pp. 249-258, 2016.
- [54] E. V. Salomatina, B. Jiang, J. Novak, & A. N. Yaroslavsky, "Optical properties of normal and cancerous human skin in the visible and near-infrared spectral range," *Journal of Biomedical Optics*, vol. 11, no. 6, pp. 1-9, 2006.
- [55] M. Al Ahmad, A. Najar, A. El Moutaouakil, N. Nasir, M. Hussein, S. Raji, & A. Hilal-Alnaqbi, "Label-Free Cancer Cells Detection Using Optical Sensors.," *IEEE Access*, vol. 6, pp. 55807-55814, 2018.

- [56] T. M. Jørgensen, A. Tycho, M. Mogensen, P. Bjerring, & G. B. Jemec, "Machine-learning classification of non-melanoma skin cancers from image features obtained by optical coherence tomography.," *Skin Research and Technology*, vol. 14, no. 3, pp. 364-369, 2008.
- [57] S. Tarek, R. A. Elwahab & M. Shoman, "Gene expression based cancer classification.," *Egyptian Informatics Journal*, vol. 18, no. 3, pp. 151-159, 2017.
- [58] S. B. Cho, & H. H. Won, "Machine learning in DNA microarray analysis for cancer classification," in *First Asia-Pacific bioinformatics conference on Bioinformatics*, Adelaide, 2003.
- [59] M. Hilario, A. Kalousis, M. Müller, & C. Pellegrini, "Machine learning approaches to lung cancer prediction from mass spectra," *Proteomics*, vol. 3, no. 9, pp. 1716-1719, 2003.
- [60] E. Adetiba, & O.O Olugbara, "Lung cancer prediction using neural network ensemble with histogram of oriented gradient genomic features," *The Scientific World Journal*, vol. 2015, pp. 1-17, 2015.
- [61] S. Mukhopadhyay, I. Kurmi, R. Dey, N. K. Das, S. Pradhan, A. Pradhan, ... & S. Mohanty, "Optical diagnosis of colon and cervical cancer by support vector machine. In Biophotonics: Photonic Solutions for Better Health Care V," *International Society for Optics and Photonics*, vol. 9887, pp. 1-5, 2016.
- [62] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C. H. Yeang, M. Angelo, & T. Poggio, "Multiclass cancer diagnosis using tumor gene expression signatures," in *National Academy of Sciences of the United States of America*, 2011.
- [63] S. Duraipandian, W. Zheng, J. Ng, J.J. Low, A. Ilancheran, & Z. Huang, "In vivo diagnosis of cervical precancer using Raman spectroscopy and genetic algorithm techniques.," *Analyst*, vol. 136, no. 20, pp. 4328-4336, 2011.
- [64] S. Li, L. Li, Q. Zeng, Y. Zhang, Z. Guo, Z. Liu, & S. Liu, "Characterization and noninvasive diagnosis of bladder cancer with serum surface enhanced Raman spectroscopy and genetic algorithms," *Scientific Reports*, vol. 5, pp. 1-7, 2015.
- [65] M. Moradi, P. Mousavi, D. R. Siemens, E. E. Sauerbrei, P. Isotalo, A. Boag, & P. Abolmaesumi, "Discrete Fourier analysis of ultrasound RF time series for detection of prostate cancer," in *IEEE Engineering in Medicine and Biology Society*, Lyon, 2007.
- [66] R. N. Strickland, & H. I. Hahn, "Wavelet transforms for detecting microcalcifications in mammograms," *IEEE Transactions on Medical Imaging*, vol. 15, no. 2, pp. 218-229, 1996.

- [67] K. Belkic,, "Current dilemmas and future perspectives for breast cancer screening with a focus upon optimization of magnetic resonance spectroscopic imaging by advances in signal processing," *Isr. Med. Assoc. J.*, vol. 6, pp. 610-618, 2004.
- [68] A. Abdulsadda, N. Bouaynaya, & K. Iqbal, "Stability analysis and breast tumor classification from 2D ARMA models of ultrasound images," in *IEEE Engineering in Medicine and Biology Society*, Minnesota, 2009.
- [69] J. Zielinski, N. Bouaynaya, & D. Schonfeld, "Two-dimensional ARMA modeling for breast cancer detection and classification," in *Signal Processing and Communications (SPCOM)*, Bangalore, India, 2010.
- [70] N. Kumar, P. Kumari, P. Ranjan, & A. Vaish, "ARIMA model based breast cancer detection and classification through image processing," in *Students Conference on Engineering and Systems*, Allahabad, 2014.
- [71] R. Prony, "Essai experimental," *École Polytechnique Floréal et Plairial*, vol. 2, pp. 24-76, 1795.
- [72] S. Singh, "Application of Prony analysis to characterize pulsed corona reactor measurements," Doctoral dissertation, University of Wyoming, 2003.
- [73] M. S. Mazzola, N. H. Younan, R. Soundararajan, & S. E. Sadow, "Application of the singular value decomposition–Prony method for analyzing deep-level transient spectroscopy capacitance transients," *Review of Scientific Instruments*, vol. 69, no. 6, pp. 2459-2463, 1998.
- [74] A. F. Rodríguez, L. de Santiago Rodrigo, E. L. Guillén, J. R. Ascariz, J. M. Jiménez, & L. Boquete, "Coding Prony's method in MATLAB and applying it to biomedical signal filtering," *BMC Bioinformatics*, vol. 19, no. 1, pp. 1-14, 2018.
- [75] J. F. Hauer, "Application of Prony analysis to the determination of modal content and equivalent models for measured power system response," *IEEE Transactions on Power Systems*, vol. 6, no. 3, pp. 1062-1068, 1991.
- [76] J. Xiao, X. Xie, Z. Hu, & Y. Han, "Improved prony method for online identification of low-frequency oscillations in power systems," *Journal-Tsinghua University*, vol. 44, no. 7, pp. 883-887, 2004.
- [77] C. W. Chuang & D. L. Moffatt, "Natural resonances of radar targets via Prony's method and target discrimination," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 5, pp. 583-589, 1976.
- [78] L. Marple, & T. Brotherton, "Detection and classification of short duration underwater acoustic signals by Prony's method," in *International Conference on Acoustics, Speech, and Signal Processing*, 1991.

- [79] Y. Huo, R. Bansal, & Q. Zhu, "Breast tumor characterization via complex natural resonances," in *IEEE MTT-S International Microwave Symposium Digest*, 2003.
- [80] D. Li, B. Guo, & G. Wang, "Application of sem technique to characterization of breast tumors," in *International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications*, 2007.
- [81] L. Wang, X. Xiao, & T. Kikkawa, "Ultra-wide bandwidth based tumor poles extraction in practical noise situation," in *IEEE International Workshop on Electromagnetics: Applications and Student Innovation Competition*, 2016.
- [82] M. H. Bannis, F. M. El-Hefnawi, H. M. A. El Kader, K. ElMahgoub, & A. Z. Elsherbeni, "Breast cancer detection and identification using prony's method," in *IEEE Antennas and Propagation Society International Symposium*, 2014.
- [83] M. H. Bannis, F. M. Elhefnawi, H. M. A. El Kader, K. ElMahgoub, & A. Z. Elsherbeni, "Effect of chest wall on breast tumor detection using Prony's method," in *IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*, 2015.
- [84] P. K. Gale, & J. W. Pierre, "Prony analysis based parameter estimation of an NMR signal of blood plasma for cancer detection.," in *In 1995 International Conference on Acoustics, Speech, and Signal Processing*, 1995.
- [85] M. Roy, & S. Barman, "Novel approach to power spectrum estimation of DNA sequence by Prony's method," in *International Conference on Communications, Devices and Intelligent Systems*, 2012.
- [86] Z. Movasaghi, S. Rehman, & I. U. Rehman, "Raman spectroscopy of biological tissues," *Applied Spectroscopy Reviews*, vol. 42, no. 5, pp. 493-541, 2007.
- [87] S. Li, Q. Y. Chen, Y. J. Zhang, Z. Liu, H. Xiong, Z. Guo, ... & S. Liu, "Detection of nasopharyngeal cancer using confocal Raman spectroscopy and genetic algorithm technique," *Journal of Biomedical Optics*, vol. 17, no. 12, pp. 1-7, 2012.
- [88] S. K. Mitra, & J. F. Kaiser, *Handbook for digital signal processing*, John Wiley & Sons, Inc., 1993.
- [89] S. L. Marple Jr, & W. M. Carey, "Least Squares Prony Method," in *Digital spectral analysis with applications*, Prentice-Hall, 1989, pp. 303-349.
- [90] D. J. Trudnowski, J. M. Johnson, & J. F. Hauer, "Making Prony analysis more accurate using multiple signals," *IEEE Transactions on power systems*, vol. 14, no. 1, pp. 226-231, 1999.

- [91] D. Potts, & M. Tasche, "Parameter estimation for exponential sums by approximate Prony method," *Signal Processing*, vol. 90, no. 5, pp. 1631-1642, 2010.
- [92] "Mathworks Documentation," MATLAB Inc., 2006. [Online]. Available: <https://www.mathworks.com/help/matlab/ref/roots.html>. [Accessed 5 September 2019].
- [93] E. D. Übeyli & I. Güler, "Spectral analysis of internal carotid arterial Doppler signals using FFT, AR, MA, and ARMA methods," *Computers in Biology and Medicine*, vol. 34, no. 4, pp. 293-306, 2004.
- [94] R. Adhikari, & R.K. Agrawal,, "An introductory study on time series modeling and forecasting,," 2013. [Online]. Available: <https://arxiv.org/ftp/arxiv/papers>. [Accessed 5 September 2019].
- [95] K. Berndtsson, *ARIMA Modeling and Simulation of Currency Pairs*, Gothenburg, Sweden: Master Thesis, 2014.
- [96] L. Magee, "Unit Roots and Characteristic Roots," 2008. [Online]. Available: https://socialsciences.mcmaster.ca/magee/761_762/other%20material/unit%20and%20char%20roots.pdf. [Accessed 5 September 2019].
- [97] H. Akaike, "Fitting autoregressive models for prediction," *Annals of the Institute of Statistical Mathematics*, vol. 21, no. 1, pp. 243-247, 1969.
- [98] J. Ayres, & M. L. Bushnell, , "Analog Circuit Testing Using Auto Regressive Moving Average Models," in *20th International Conference on VLSI Design held jointly with 6th International Conference on Embedded Systems*, Bangalore, 2007.
- [99] J. Huang, J. Zhao, , & Y. Xie, "Source classification using pole method of AR model," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1997.
- [100] K.A. Wear, R.F. Wagner, & B.S. Garra, "A comparison of autoregressive spectral estimation algorithms and order determination methods in ultrasonic tissue characterization," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency control*, vol. 42, no. 4, pp. 709-716, 1995.
- [101] MATLAB Inc, "Estimate parameters of ARX, AR, or ARI model," 2019. [Online]. Available: <https://www.mathworks.com>. [Accessed 5 September 2019].
- [102] M. S. Braasch, "Current developments in signal modeling of the precision distance measuring equipment (Doctoral dissertation)," Ohio University, Athens, 1989.

- [103] H. Scheffe, *The analysis of variance*, vol. 72, John Wiley & Sons, 1999.
- [104] L. S. Huang, & J. Chen, "Analysis of variance, coefficient of determination and F-test for local polynomial regression," *The Annals of Statistics*, vol. 36, no. 5, pp. 2085-2109, 2008.
- [105] G. H. Lunney, "Using analysis of variance with dichotomous dependent variable: An empirical study," *Journal of Educational Measurement*, vol. 7, no. 4, pp. 263-269, 1970.
- [106] M. H. Subhashini, & M. Arumugam, "Analysis of variance (Anova)," *CMFRI Special Publication*, vol. 7, pp. 169-170, 1981.
- [107] L. Fox, W. R. Butler, R. W. Everett, & R.P. Natzke, "Effect of adrenocorticotropin on milk and plasma cortisol and prolactin concentrations," *Journal of Dairy Science*, vol. 64, no. 9, pp. 1794-1803, 1981.
- [108] R. Kumaresan, "On a frequency domain analog of Prony's method," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 1, pp. 168-170, 1990.
- [109] G. B. Gori, "Trypsinization of animal tissues for cell culture: Theoretical considerations and automatic apparatus," *Appl. Environ. Microbiol*, vol. 12, no. 2, pp. 115-121, 1964.
- [110] D. Barnes, & G. Sato, "Serum-free cell culture: a unifying approach.," *Cell*, vol. 22, no. 3, pp. 649-655, 1980.
- [111] M. Arora, "Cell culture media: A review," *Mater Methods*, vol. 3, no. 175, pp. 1-29, 2013.
- [112] "Arc Lamp Source," Arc Lamp, 2018. [Online]. Available: <https://www.newport.com/p/66924-450XV-R1>. [Accessed 5 September 2019].
- [113] "c10988ma-01_etc_kacc1169e.pdf," Hamamatsu, Photonics, June 2019. [Online]. Available: https://www.hamamatsu.com/resources/pdf/ssd/c10988ma-01_etc_kacc1169e.pdf. [Accessed 5 September 2019].
- [114] J. Yan, Y.W. Lan, W. Fang, X. Hou, & J.C. Mei, "Prony algorithm's application on extracting the parameters of the power system transient fault.," *Applied Mechanics and Materials*, vol. 602, pp. 3877-3880, 2014.
- [115] W. Dai, I. Selesnick, J. R. Rizzo, J. Rucker, & T. Hudson, "A nonlinear generalization of the Savitzky-Golay filter and the quantitative analysis of saccades," *Journal of vision*, vol. 17, no. 9, pp. 1-15, 2017.

- [116] J. W. Brown, Complex variables and applications, New York: McGraw-Hill, 2009.

List of publications

Aysha F. Abdul Gani. and Mahmoud Al Ahmad., (2020). Label-free normal and cancer cells classification combining Prony's method and optical techniques. IEEE Access. DOI: 10.1109/ACCESS.2020.2973468, IEEE Access.

Appendix

Aliasing and Calibration

The system has been equipped for anti-aliasing to restrict the bandwidth of the applied optical signal over the measured band of interest. Most of the used optical detectors possess a finite width slot or pixels arranged in an array. Each pixel records the received intensity integrated over its own area. The pixels are usually separated by an array pitch. It is well known that coarse sampling increases the random noise errors in the wavelength. The full width at half maximum (FWHM) as a resolution measure is a very important parameter and indicates the minimum acceptable for signal/noise ratio in the optical detectors. In optical detector, 2 pixels per FWHM represents Nyquist sampling rate, which is the minimum requirement for valid measurements. For the optical sensor used in the work, the wavelength reproducibility measured under constant light input conditions ranges from -0.5 to +0.5 nm. The wavelength temperature dependence ranges from -0.05 to +0.05 nm per one Celsius degree. The reported measurements have been conducted at room temperature and under constant light source.

The optical mini spectrometer system is equipped with a built-in calibration system. The calibration is carried out automatically using the corresponding software. The system performs a reference measurement which is used as baseline for the dark or monitor modes measurements.