VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Jnana Sangama, Belagavi-590018

A Project Phase-2 Report on

"Photonic Crystal based Prediction of Cervical and Breast Cancer Using ML Model"

Submitted in the partial fulfilment of the requirements for the award of the Degree of Bachelor of Engineering

In

Computer Science and Engineering

by

Arindam Rathore R	**1OX19CS009**
Arkesh V Kumar	**1OX19CS011**
Rakshitha Kadekar	**1OX19CS076**
Toms John	**1OX19CS112**

Under the guidance of

Dr. E. Saravana Kumar

Professor, Department of CSE

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

THE OXFORD COLLEGE OF ENGINEERING

Hosur Road, Bommanahalli, Bengaluru-560068

(Approved by AICTE, New Delhi, accredited by NBA, NAAC, New Delhi & Affiliated to VTU, Belagavi)

2022-2023

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
THE OXFORD COLLEGE OF ENGINEERING
Hosur Road, Bommanahalli, Bengaluru-560068
(Approved by AICTE, New Delhi, accredited by NBA, NAAC, New Delhi & Affiliated to VTU, Belagavi)

CERTIFICATE

This is to certify that the project entitled "**Photonic Crystal Based Prediction of Cervical and Breast Cancer Using ML Model**", carried out by **Arindam Rathore R, Arkesh V Kumar, Rakshitha Kadekar and Toms John** bearing USN **1OX19CS009**, **1OX19CS011**, **1OX19CS076** and **1OX19CS112** bonafide students of **The Oxford College of Engineering, Bengaluru** in partial fulfillment for the award of the Degree of Bachelor of Engineering in Computer Science and Engineering of the **Visvesvaraya Technological University, Belagavi** during the year 2022-2023. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The Project Phase-2 review has been approved as it satisfies the academic requirements in respect of Project Phase-2(18CSP83) work prescribed for the said degree.

Dr. E. Saravana Kumar	**Dr. R. Ch. A Naidu**	**Dr. N Kannan**
Professor,Dept. of CSE	HOD, Dept. of CSE	Principal

External Viva

External Examiner	**Signature**
1. ____________________	____________________
2. ____________________	____________________

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
THE OXFORD COLLEGE OF ENGINEERING
Hosur Road, Bommanahalli, Bangalore-560068
(Approved by AICTE, New Delhi, Accredited by NBA, NAAC, New Delhi & Affiliated to VTU, Belagavi)

Department Vision

To produce technocrats with creative technical knowledge and intellectual skills to sustain and excel in the highly demanding world with confidence.

Department Mission

1. To Produce the Best Computer Science Professionals with intellectual skills.
2. To Provide a Vibrant Ambiance that promotes Creativity, Technology Competent and Innovations for the new Era.
3. To Pursue Professional Excellence with Ethical and Moral Values to sustain in the highly demanding world.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
THE OXFORD COLLEGE OF ENGINEERING
Hosur Road, Bommanahalli, Bengaluru-560068
(Approved by AICTE, New Delhi, Accredited by NBA, NAAC, New Delhi & Affiliated to VTU, Belagavi)

DECLARATION

We the students of Eight semester B.E, at the Department of Computer Science and Engineering, **The Oxford College Of Engineering, Bengaluru** declare that the project entitled **"**Photonic Crystal based Prediction of Cervical and Breast Cancer using ML Model**"** has been carried out by us and submitted in partial fulfillment of the course requirements for the award of degree in Bachelor of Engineering in Computer Science and Engineering discipline of **Visvesvaraya Technological University, Belagavi** during the academic year **2022-2023.** Further, the matter embodied in dissertation has not been submitted previously by anybody for the award of any degree or diploma to any other university.

Name	USN	Signature
Arindam Rathore	1OX19CS009	
Arkesh V Kumar	1OX19CS011	
Rakshitha Kadekar	1OX19CS076	
Toms John	1OX19CS112	

Date:
Place: Bangalore

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible whose constant guidance and encouragement crowned our effort with success.

We consider ourselves proud to be a part of The Oxford family, the institution that stood by our way in all our endeavours. We have a great pleasure in expressing our deep sense of gratitude to the founder chairman **late Sri S. Narasa Raju** and to our chairman **Sri S. N. V. L. Narasimha Raju** for providing us with a great infrastructure and well-furnished labs.

We would like to express our gratitude to **Dr. N. Kannan,** Principal, The Oxford College of Engineering for providing us a congenial environment and surrounding to work in.

Our hearty thanks to **Dr. R. Ch. A. Naidu,** Professor & Head, Department of Computer Science and Engineering, The Oxford College of Engineering for his encouragement and support.

Guidance and deadlines play a very important role in successful completion of the project report on time. We convey our gratitude to **Dr. E. Saravana Kumar,** Professor, Department of Computer Science and Engineering for having constantly monitored the completion of the Project Report and setting up precise deadlines.

We also thank them for their immense support, guidance, specifications, and ideas without which the Project Report would have been incomplete. Finally, a note of thanks to the Department of Computer Science and Engineering, both teaching and non-teaching staff for their cooperationextended to us.

Project Associates

Arindam Rathore R (1OX19CS009)

Arkesh V Kumar (1OX19CS011)

Rakshitha Kadekar (1OX19CS076)

Toms John (1OX19CS112)

ABSTRACT

Cancer is a disease in which some of the body's cells grow uncontrollably and spread to other parts of the body. Laboratory determination of cancer cell is time consuming so a requirement of cancer-testing equipment with the ability to detect the cancerous cell in a short time. A system is proposed which uses both PCF sensor and AIML. Which helps in giving accurate results in real world. PCF is currently an auspicious technology for sensing applications due to its powerful light–matter interaction. Therefore, this PCF is proposed as a cancer sensor for the detection of cervical and breast of cancer cells. The detection of cancer cell a hexagonal structure is generated using MEEP software. RI value of each cancer and its respective normal cell is passed to the software which further generates dataset. This dataset provides two coordinates X(frequency) and Y(flux). Further, the data is clustered into binary (0 & 1) values. These binary values are utilized to train the ANN model. This model predicts future outcome. The model where able to give 56% and %57 accuracy respectively.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Cancer is a group of diseases involving abnormal cell growth with the potential to invade or spread to other parts of the body. These contrast with benign tumours, which do not spread. Possible signs and symptoms include a lump, abnormal bleeding, prolonged cough, unexplained weight loss, and a change in bowel movements. While these symptoms may indicate cancer, they can also have other causes. Over 100 types of cancers affect humans. In 2015, about 90.5 million people worldwide had cancer. In 2019, annual cancer cases grew by 23.6 million people and there were 10 million deaths worldwide, representing over the previous decade increases of 26% and 21%, respectively. The projected incidence of patients with cancer in India among males was 679,421 (94.1 per 100,000) and among females 712,758 (103.6 per 100,000) for the year 2020. One in 68 males (lung cancer), 1 in 29 females (breast cancer), and 1 in 9 Indians will develop cancer during their lifetime. The projected 5 most common cancers in 2020 for males (lung, mouth, prostate, tongue, and stomach) constitute 36% of all cancers and for females (breast, cervix uteri, ovary, corpus uteri, and lung) constitute 53% of all cancers.

When cancer begins, it produces no symptoms. Signs and symptoms appear as the mass grows or ulcerates. The findings that result depend on cancer's type and location. Few symptoms are specific. Many frequently occur in individuals who have other conditions. Cancer can be difficult to diagnose and can be considered a “great imitator.” Tobacco use is the cause of about 22% of cancer deaths. Another 10% are due to obesity, poor diet, lack of physical activity or excessive drinking of alcohol. Other factors include certain infections, exposure to ionizing radiation, and environmental pollutants. The most common types of cancer in males are lung cancer, prostate cancer, colorectal cancer, and stomach cancer. In females, the most common types are breast cancer, colorectal cancer, lung cancer, and cervical cancer.

1.1 Cervical Cancer

It is cancer in cervix which is the lowest part of the uterus in the female reproductive system. It is 4^{th} most common cause of death from cancer in women. Abnormal growth of cells occurs and spread to different parts of the body. In cervical cancer early on no symptoms are seen but later symptoms like vaginal bleeding, pelvic pain or pain during sexual intercourse might occur. Genetic factors, HPV which is Human Papillomavirus infection (causes more the

90 % of the cases), weak immune system, smoking, taking birth control pills etc. are some of the causes of this cancer. When exposed to HPV, the body's immune system prevents the virus from doing any harm, but in some people, virus survives for years, contributing to the conversion of cervical cell to cancerous cells. A smoker has a higher chance of cervical intraepithelial neoplasia (CIN3) occurring, which has the potential for cervical cancer. Long term use of oral contraceptives is associated with increased risk of cervical cancer in women who already have HPV.

1.2 Breast Cancer

It is the most common cancer among women. It occurs when some breast cells tend to grow abnormally. These cells divide more rapidly than healthy cells and continue to gather forming lump or mass. Breast cancer mostly begins with the cells in the milk producing ducts or in the glandular tissues called lobules or in other cells or tissues within the breast. It is mostly faced by women who are in their 50's or above. Signs and symptoms are a lump or thickening that feels different from the surrounding tissue, change in size shape or appearance of the breast, occurrence of dimpling, peeling, scaling, flaking around the nipple, redness or pitting on the skin of the breast. Causes of breast cancer are inherited genetically (gene mutation passed through generations of a family), increasing age, radiation explosion, obesity, early menstruation or late menopause, smoking, drinking and may more. Diagnosis include biopsy, mammograms or by physical examination. Types of breast cancer include ductal carcinoma in situ, invasive ductal carcinoma, inflammatory breast cancer and meta cystic breast cancer.

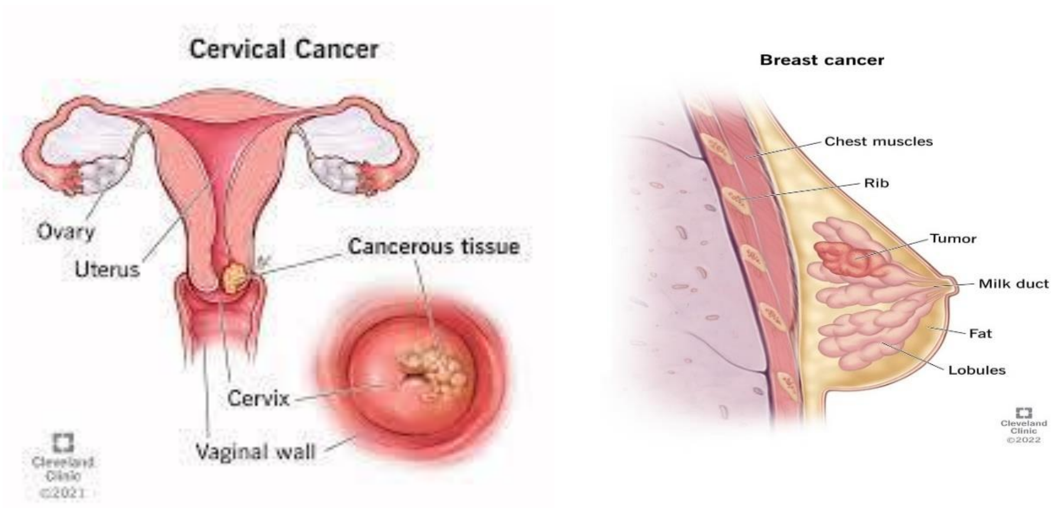

Figure 1.1 Images of Cervical and Breast cancer (Cleveland Clinic 2022)

1.3 Aim and Objectives

Propose a system for detection of cervical cancer and breast cancer using photonic crystal and ML. The system aims to provide accurate results in real-world applications as laboratory determination of cancer cells is time-consuming. PCF technology is suitable for sensing applications due to its strong light-matter interaction, and a hexagonal structure is generated using MEEP software to detect cancer cells. The dataset required for training the ML model is generated using the RI value of each cancer and its normal cell, and the data is clustered into binary values for training the ANN model. Also compare the accuracy between various ML model.

Objectives

1. To design an optimized photonic crystal structure.
2. As per the design data is generated using MEEP software.
3. Classifying the generated data.
4. Develop ML model.
5. Develop a user-friendly GUI.

1.4 Existing System Overview

The base paper presents a novel spectroscopic optical sensor for detecting cancerous cells in various parts of the human body. The proposed sensor is based on a differential optical absorption spectroscopy-based refractive index sensor that detects changes in the refractive index of the surrounding medium caused by the presence of cancer cells. The paper describes the design and optimization of a compact cladding photonic crystal fiber (PCF) with enhanced sensitivity and very low confinement loss for detecting several types of cancerous cells, including blood cancer, cervical cancer, gland cancer, breast cancer of type-I and type-II, and skin cancer. The PCF is analyzed using the full vector finite element method (FV-FEM) to optimize its guiding properties, sensitivity response, loss value, effective area, V-parameter, Sport size, beam divergence, and numerical aperture. The results show that the proposed PCF has a monomode characteristic that allows it to detect changes in refractive index caused by the presence of cancer cells in various parts of the human body. The optimized structure based on compact cladding allows for enhanced sensitivity and very low confinement loss. In conclusion, the paper presents a promising approach for detecting cancer cells using an optical sensor based on differential optical absorption spectroscopy. The proposed PCF has potential applications in biomedical sensing beyond cancer cell detection with further research and development.

While the paper presents a promising approach for detecting cancer cells using an optical sensor based on differential optical absorption spectroscopy, further research and development are required to test the proposed sensor's sensitivity and specificity in detecting cancer cells in real-world scenarios. And, the paper does not the concept of ML model.

1.5 Problem Statement

The existing system does not propose a system that uses the concept of ML. As it will help in increasing the accuracy in detection of cancer cells. System that combines photonic crystal fiber (PCF) sensor technology and machine learning (ML) models. The PCF sensor is used to detect cancer cells due to its powerful light-matter interaction, while the ML model is trained on a dataset generated by simulating the PCF structure using MEEP software. The goal is to develop an accurate and efficient prediction model that can be used in real-world applications.

1.6 Proposed Solution

The proposed solution for predicting cervical and breast cancer is a system that combines a PCF sensor and machine learning (ML) algorithms. PCF is a promising technology for sensing applications due to its powerful light-matter interaction. The system uses MEEP software to simulate a hexagonal photonic crystal structure, generating a dataset required for training the ML model. The refractive index (RI) value of each cancer and normal cell is passed to the MEEP code, which generates a dataset with two coordinates, X (frequency) and Y (flux), and then clusters the data into binary (0 and 1) values. These binary values are used to train an artificial neural network (ANN) model to predict the presence of cancer. The accuracy of the ANN model is compared with other models to determine the best model for prediction. Overall, the proposed solution aims to provide accurate and timely detection of cervical and breast cancer using PCF sensor technology and ML algorithms.

CHAPTER 2

LITERATURE SURVEY

Literature review is a systematic, explicit, and reproducible method for identifying, evaluating, and synthesizing the existing body of completed and recorded work produced by researchers, scholars, and practitioners. The literature review establishes the fact that you have familiarized yourself with the area(s) or discipline(s) in which you are conducting research. A literature review will summarize the existing scholarly literature on your chosen topic, establish relationships between different research projects of the past, show where there are gaps in past research, and show how the past published work relates to your own work.

A literature survey includes the following:

• Existing theories on the subject that are broadly acknowledged.

• Books that have been written on the topic, both broad and,

• Field research is frequently conducted in the sequence of oldest to most recent.

• Major challenges being faced and on-going work, if available.

Objectives of Literature Survey

• Provide a solid foundation of knowledge about the subject.

• Access to the most current innovations, methods, and theories.

• It indicates whether the evidence already available solves the problem effectively without requiring further investigation.

• Identify areas of prior scholarship to prevent redundancy and give credit to other researchers.

Table 2.1 Survey Summary

Authors	Methodology	Limitations
Dr. Preeta Sharan, Bharadwaj S M, Fleming Dackson Gudagunti, Pooja Deshmukh -2014	Photonic Band Gap Method	Experiences difficulty in fabrication process thermal profile as an additional outcome.
Poonam Sharma, Dr. Preeta Sharan, Pooja Deshmukh-2015	2D Photonic Crystal Sensor based on Basal, Breast and Cervical Cancer.	Operating wavelength was very small and application tolerance was not introduced.
N. R. Ramanujam, S. Amiri, Sofyan A. Taya, Saeed Olyaee, R. Udaiyakumar, A. Pasumpon Pandian,2018	Nanocomposite material based photonic crystal.	SPR surface plasmon resonance technique for sensing gained very low sensitivity.
Vahideh Shirmohammadli and Negin Manavizadeh	Microfluidic device for breast cancer Screening.	Sensitivity was improved but detection limit was low.
Md. Asaduzzaman Jabin, Member, Kawsar Ahmed-2019	Proposed dual core PCF sensor.	Although sensitivity Performance was improved but detection limit was low.
Md. Asaduzzaman Jabin, Yanhua Luo, Gang-Ding Peng, Md. Juwel Rana-2020	Cancer sensor based on SPR for cancer detection.	New Amoeba Structure were more Optical Parameters are analyzed but, the relative sensitivity response was not good.
N. Ayyanar, G. Thavasi Raja, Mohit Sharma and D. Sriram Kumar-2018	Photonic Crystal Fiber Based Refractive Index Sensor for Early Detection of Cancer.	Comparatively low sensitivity was observed from the structure. Here it could detect for only three types of cancer.
Shaikh Afzal Nehal, Debpriyo Roy, Manju Devi, T. Srinivas-2019	AI-based lab-on-chip application platform that can detect contamination using output from Photonic Crystal based optical biosensor.	Uses CNN model, where it does not consider the flux values.

Dr. Preeta Sharan, Bharadwaj S M, Fleming Dackson Gudagunti, Pooja Deshmukh proposed about the early detection of cancer cell by using photonic band gap method. System level cancer cell detection has been done with the dielectric constant (at optical frequency) as the input. Comparison of normal cell and cancerous cell has been done and a precise frequency shift has been observed. The input dielectric constant values of normal cell vary from 1.8225 to 1.8769 and for cancer cell it varies from 1.9376 to 1.9628. Even though the change in the input is very small, a micron change in the frequency has been observed. Thus, photonic crystal-based sensor will differentiate the normal cells from cancerous cell. Keywords: Cancer cell, Dielectric constant, Photonic Band gap, Photonic Biosensor.[15]

Poonam Sharma, Dr. Preeta Sharan, Pooja Deshmukh proposed a paper, 2D photonic crystal sensor using gratings has been designed for the analysis of Basal, Breast and Cervical cancer cells. The grating design, incorporated in the photonic crystal waveguide increases the efficiency and sensitivity of the designed sensor. The fact that index of refraction of the cancer cells differs from the normal cell, it can be differentiated and detected using optical techniques. The shift in the wavelength and intensity levels in the reflection spectrum has been observed for cancer and normal cells. FDTD method has been used for analyzing cells. MEEP simulation tool has been used for modeling and designing of sensor. Keywords: Basal cancer, Breast cancer, Cervical cancer, FDTD, Integrated photonics, MEEP, Photonic crystal sensor.[7]

N. R. Ramanujam, I. S. Amiri, Sofyan A. Taya, Saeed Olyaee, R. Udaiyakumar, A. Pasumpon Pandian K. S. Joseph Wilson, P. Mahalakshmi, P. P. Yupapin theoretically analyze the detection of a cancer cell in the one-dimensional photonic crystal by infiltrating different sample cells in the cavity layer. The defect modes appear in their transmission spectra only if the nanocomposite layers are included on either side of the cavity layer. This analysis is carried out by a dielectric constant and the transmittance peak of the cancer cell is compared with the normal cell. The transmittance peak shifts are analyzed with various filling factors for optimization purposes. Through the shifting spectrum, the sensitivity of cancer cell from the normal cell is obtained from a minimum of 42 nm/RIU to a maximum of 43 nm/RIU. [16]

Vahideh Shirmohammadli And Negin Manavizadeh deals with mathematical modeling of the particle trajectory inside a microfluidic device based on the application of DC Di electrophoresis (DEP). The proposed device is designed to separate the breast cancer cell lines (e.g. MDA-MB-231 and MCF-7) from normal blood cells in the specific outlets. A complementary mechanism is employed in the channel that involves the drag and DEP forces acting cooperatively on the cells to deflect the cells into corresponding outlets. Closed-form equations are yield for predicting the particle position along the x and y-axis versus time. Effect of both DEP and drag forces are considered and Newton's second law of motion is exploited for formulations. The initial values required by the particle trajectory equation, including the primary location and velocity, and the physical features of different cells, are determined. The proposed microchannel is numerically simulated using finite element simulation software and the results are compared to the data derived from modeling equations. Our findings confirm the capability of the proposed model to accurately track the location of all types of under-study cells inside the proposed microchannel. In this paper, a user-friendly software package is designed and implemented in the MATLAB environment to predict the cell trajectory of different cells with various diameters. Results for different cells with various diameters reveal that this software can predict the cell trajectory with an error less than 7%. Keywords— Di electrophoresis, Modeling, breast cancer, Trajectory, Circulating tumor cell, Cell motion. [17]

Md. Asaduzzaman Jabin, Member, Ieee, Kawsar Ahmed, Member, Ieee Md. Juwel Rana, Member, Ieee, Bikash Kumar Paul, Member, Ieee, Maheen Islam, Member, Ieee, Dhasarathan Vigneswaran, Member, Ieee, Muhammad Shahin Uddin proposed a new optimized bowl-shaped mono-core surface Plasmon resonance-based cancer sensor is proposed for the rapid detection of different types of cancer affected cell. By considering the refractive index of each individual cancer contaminated cell with respect to their normal cell, some major optical parameters variation is observed. Moreover, the cancerous cell concentration is considered at 80% in liquid form and the detection method is finite element method with 2 100 390 mesh elements. The variation of spectrum shift is obtained by plasmonic band gap between the silica and cancer cell part which is separated by a thin (35 nm) titanium film coating. The proposed sensor depicts a high birefringence of 0.04 with a maximum coupling length of 66 μm. However, the proposed structure provides an optimum wavelength sensitivity level between about 10 000 nm/RIU and 17 500 with a resolution of the sensor between 1.5×10^{-2} and 9.33×10^{-3} RIU. Also, the transmittance variance of the cancerous cell ranges from almost 3300 to 6100 dB/RIU and the amplitude sensitivity ranges nearly between −340 and −420 RIU^{-1} for different cancer cells in major polarization mode with the maximum detection limit of 0.025. Besides, the overall sensitivity performance is measured with respect to their normal cells which can be better than any other prior structures that have already proposed.

Md. Asaduzzaman Jabin, Yanhua Luo, Gang-Ding Peng, Md. Juwel Rana, Kawsar Ahmed, Truong Khang Nguyen, Bikash Kumar Paul, Vigneswaran Dhasarathan proposed a newly designed Amoeba faced photonic crystal fiber (A-PCF) is introduced for the first time in the fields of fiber for bio sensing and particularly to detect cancerous cells for instancesblood cancer (Jurak), cervical cancer HeLa), adrenal glands cancer (PC12), breast cancer (MDA-MB-231, MCF-7) and skin cancer (Basal), etc. In fact, proposed PCF is fabricated after the simulation studies and the application for sensing is numerically investigated by the discrete finite element method (D-FEM) by the simulation of about 22105 mesh elements. Moreover, the numerical analysis is justified by full vector simulation software COMSOL V-5.1 and Detection of cancer cells of various types using Photonic crystal-based sensor 2022-23 Department of Computer Science Engineering, TOCE 7 considering the refractive indices from it. Also, different sections are considered for the analyze cavity to analyze the performance variations. Nonetheless, few parameters like birefringence, coupling length, power fraction, transmittance, wavelength sensitivity, and transmittance sensitivity are evaluated through the refractive indices of the cancerous cell compared to its normal cell and the most optimum profile are respectively to 3.5 × 10-3, 900 μm, 0.891, -178 dB,18115.94 nm/RIU and 6071.42 dB/RIU. Nonetheless, the fabrication of the PCF can be achieved by using current fabrication technology such as stack-and-draw, sol-gel or the extrusion and drilling techniques that help to overcome all the primary difficulties of the proposed PCF. [18]

G. Dhanu Krishna, V.P. Mahadevan Pillai, K.G. Gopchandran proposed a Photonic crystal fiber consisting of a novel hybrid structure providing low values for both dispersion and confinement loss were simulated using finite element method. The studies on different arrangements, polygonal and circular, of air holes as cladding of PCFs have shown that circular PCFs results in a structure with low dispersion, air filling fractions remaining same. Circular PCF with dispersion 103.50 ps/nm.km and confinement loss 5.97 × 10−6 dB/m was restructured by replacing its first ring with octagonal air hole lattice having eight air holes. This hybrid circular-octagon PCF has shown a dispersion of 33.15 ps/nm.km and a confinement loss of 2.52 × 10−6 dB/m. The introduction of air hole defects in the hybrid fiber core with three identical air holes in an equilateral triangular lattice was found to modify dispersion by reducing its value to −11.63 ps/nm.km with confinement loss 9.51 × 10−5 dB/m.

N. Ayyanar, G. Thavasi Raja, Mohit Sharma And D. Sriram Kumar proposes a novel cancer sensor based on dual core photonic crystal fiber for the detection of cancer cells in cervical, breast and basal parts. The samples are taken in fluid form and infiltrated into the farmed cavity using selective infiltration method. Each fluid form has its own refractive index values which give the various responses in the transmission and loss spectrum. The spectral shift is obtained by inducing the coupling mechanism between silica core and cancer cell core for its launching input optical field which is investigated by finite element method. The proposed structure is also optimized with its structural dimensional property for enhancing the sensitivity. The sensing performances for the cervical cancer cell are obtained as high as 7916 nm/RIU for x- and 10625 nm/RIU y-polarization with the detection limit of 0.024. The sensitivity to breast cancer cells for x and y-polarization is 5714.28 nm/RIU and 7857.14 nm/RIU with detection limit of 0.014 respectively. Similarly, the sensitivity to basal cells can also reach 4500nm/RIU for x-polarization and 6000 nm/RIU for y-polarization. To the best of our knowledge, such sensitivities are the highest reported thus so far.[20]

CHAPTER 3

SYSTEM REQUIREMENTS & SPECIFICATION

3.1 Software Requirements

Operating Systems: Windows 10, Ubuntu

Language: MEEP, Python, Java

Tool: Android Studio, Excel Sheet, PyCharm, Google Colab

MEEP

MEEP is an acronym for "Maxwell’s Electromagnetic Equation Propagation." It is used for simulating electromagnetic systems. MEEP is a used for designing and optimizing photonic devices such as waveguides, resonators, and filters. It can also be used to study the behavior of light in materials with complex refractive indices, such as metamaterials and photonic crystals. Additionally, MEEP supports the simulation of electromagnetic fields in 2D, 3D, and even 4D geometries, making it a versatile tool for a wide range of applications. The structure used here is hexagonal structure and this structure is preferred particularly because of the Q-Factor (Quality Factor which is measure of energy dissipation in a system). This structure also provides more intensity of light(laser) passing through the crystal and flux values generated will be more accurate (which will be used to generate the graphs) compare to other crystal structure. It is widely used in academic and industrial research, and has contributed to advances in the fields of photonics, optoelectronics, and nanotechnology.

Excel

Excel is a widely used spreadsheet program developed by Microsoft. It is used for organizing, analyzing, and manipulating data in a tabular format. Excel allows users to create spreadsheets, charts, graphs, and pivot tables to visualize and analyze data. Excel is commonly used in various industries, including finance, accounting, marketing, and data analysis, among others.

Google Colab

Google Colab, short for Google Colaboratory, is a cloud-based service that allows users to run Python code in a Jupyter Notebook environment, directly from a web browser. It is provided by Google and is free to use. With Google Colab, users can write and execute Python code, create, and share documents that include live code, equations, visualizations, and narrative text.

It supports popular machine learning frameworks such as TensorFlow, PyTorch, and Keras, and provides access to a variety of powerful computing resources, including GPUs and TPUs, that can be used to accelerate the execution of machine learning algorithms. One of the biggest advantages of Google Colab is its integration with Google Drive, which allows users to store their data and notebooks in the cloud, collaborate with others, and access their work from anywhere with an internet connection. Additionally, Google Colab provides access to a variety of pre-built libraries and examples, making it easier to get started with machine learning and data analysis.

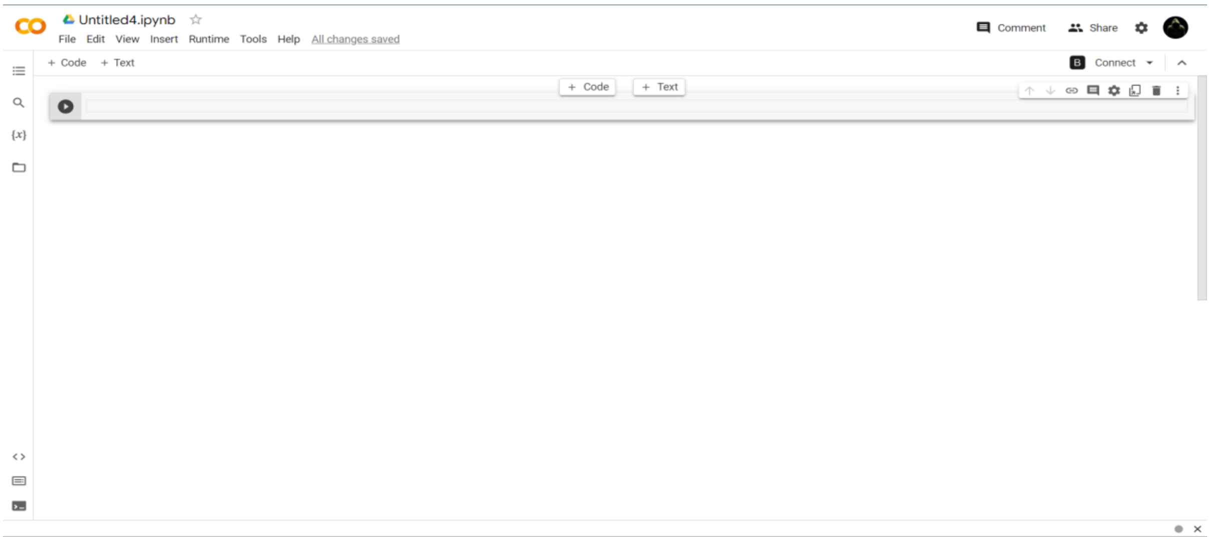

Figure 3.1 Google Colab Snapshot

Android Studio

Android Studio is the official integrated development environment (IDE) for Android app development, developed by Google. It is based on the IntelliJ IDEA platform and provides all the necessary tools and features to develop high-quality Android applications. Android Studio includes a code editor, a visual layout editor, a debugger, an emulator, and various tools for code analysis, testing, and deployment. It supports a wide range of programming languages, including Kotlin and Java, and offers advanced features such as code completion, refactoring, and version control integration. Android Studio also provides a variety of templates and wizards to help developers create new projects, activities, and services quickly. Additionally, it integrates with other Google services such as Firebase and Google Cloud Platform to make it easier to build, test, and deploy Android apps. Overall, Android Studio is a powerful and comprehensive tool for developing Android applications, and it is widely used by professional developers and hobbyists alike.

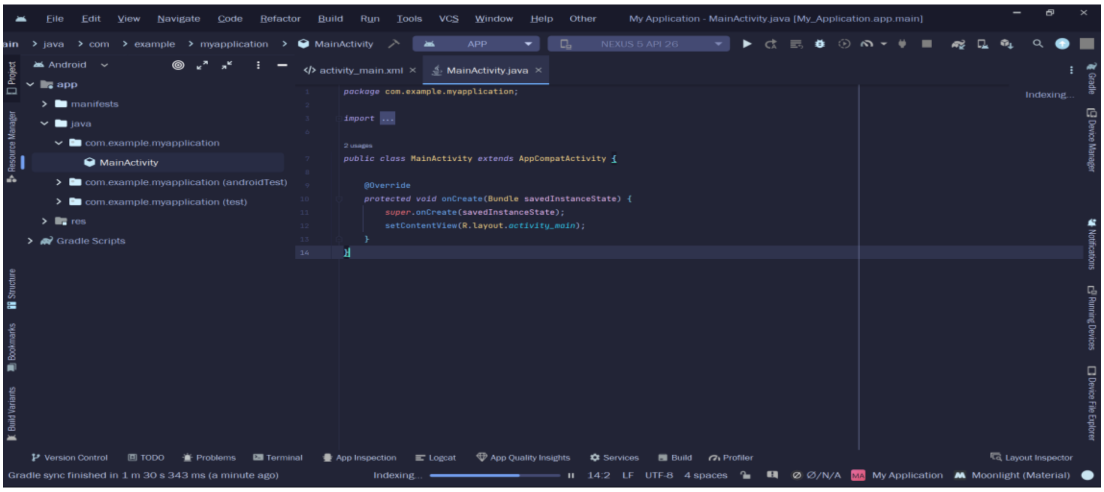

Figure 3.2 Android Studio Snapshot

PyCharm

PyCharm is an Integrated Development Environment (IDE) used in computer programming, specifically for Python. It is developed by JetBrains and provides various features such as code analysis, debugging tools, integrated unit testing, version control integration, and many other productivity tools for Python developers. It is available in two versions: the Community Edition (which is free and open source) and the Professional Edition (which requires a paid license and includes more advanced features). PyCharm provides a user-friendly interface for developing Python applications, and is widely used by programmers for developing Python projects of various sizes and complexities. In the project PyCharm is used for API building.

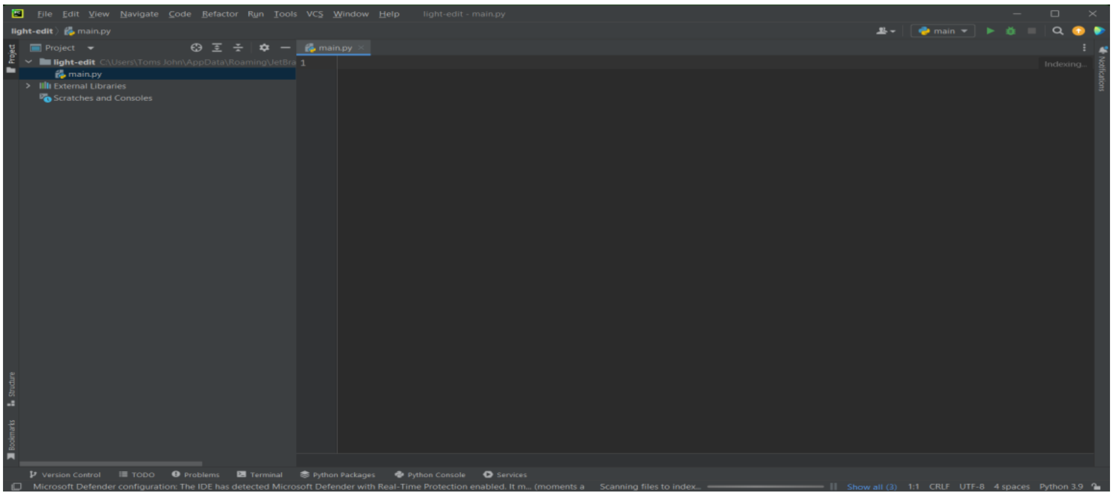

Figure 3.3 PyCharm Snapshot

Libraries

1. pandas - pandas is an open-source data manipulation and analysis library for Python. It provides easy-to-use data structures, such as Series (1-dimensional) and Data Frame (2-

dimensional), for efficient data manipulation and analysis. Pandas can be used for data cleaning, preparation, filtering, transformation, aggregation, and visualization. It is widely used in data science, finance, economics, social sciences, and many other fields for handling structured data. Some of the key features of Pandas include powerful indexing, flexible reshaping and pivoting of data, handling of missing data, merging, and joining of datasets, and fast IO (input/output) tools for reading and writing data in various formats.

2. numpy - NumPy is a Python library used for scientific computing and data analysis. It provides a powerful N-dimensional array object and functions for working with these arrays. NumPy is particularly useful for numerical computations involving large arrays and matrices. It includes a wide range of mathematical functions for array manipulation, linear algebra, Fourier transforms, and more. NumPy is a foundational library for many other Python libraries used in scientific computing, such as Pandas and Matplotlib.
3. tensorflow.keras - It is a high-level neural networks API in TensorFlow that simplifies the process of building and training deep learning models. It provides a user-friendly interface for designing and training neural networks, and supports a wide range of popular deep learning models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Deep Belief Networks (DBNs). It also includes various tools for data preparation, model evaluation, and visualization. It can be used for a wide range of applications, such as image and speech recognition, natural language processing, and predictive modelling.
4. sklearn.model_selection - It is a module in the popular machine learning library scikit-learn. It provides a suite of functions for splitting a dataset into multiple parts, selecting model hyperparameters, and evaluating model performance.

   Some of the most commonly used functions in sklearn.model_selection include:

   - train_test_split: splits a dataset into training and testing sets for model evaluation.
   - gridSearchCV: performs an exhaustive search over a specified parameter grid for a model to find the best set of hyperparameters.
   - cross_val_score: performs cross-validation to evaluate the performance of a model.
   - KFold: performs k-fold cross-validation, where the dataset is split into k subsets and each subset is used as the test set exactly once.

sklearn.model_selection is a powerful tool for machine learning practitioners to tune and evaluate their models, and is widely used in both industry and academia.

5. Pickle – It is a module in Python that provides the functionality of serializing and deserializing complex Python objects. Serialization is the process of converting an object's state to a byte stream, while deserialization is the process of reconstructing the object from the byte stream. pickle can be used to save and load trained machine learning models in Python, making it easy to reuse the models without having to retrain them each time. It can also be used to save any other type of complex Python object, such as dictionaries or lists, and load them back later as needed.

3.2 Functional Requirements

1. Data Collection: The system should be able to collect and pre-process data from various sources such as medical databases, imaging systems, and patient records.
2. Data Visualization: The system should provide data visualization tools to help researchers explore and analyse the data. This may include visualizing photonic crystal structures, medical images, and statistical plots.
3. Photonic Crystal Simulation: The system should allow researchers to design and simulate photonic crystal structures and their optical properties, such as reflection, transmission, and absorption.
4. ML Model Development: The system should provide tools for developing and training ML models using various algorithms such as logistic regression, decision trees, and neural networks.
5. Model Evaluation: The system should allow researchers to evaluate the performance of the ML models using statistical metrics such as accuracy, precision, and recall.
6. Prediction and Diagnosis: The system should provide a prediction and diagnosis tool to predict the likelihood of cervical and breast cancer based on the input data and the trained ML models.

3.3 Non-Functional Requirements

1. Performance: The system should be able to process large amounts of data and run simulations and ML models in a reasonable amount of time.
2. Usability: The system should be easy to use and navigate, with an intuitive user interface and clear documentation.
3. Reliability: The system should be reliable and able to handle errors and exceptions gracefully, with a minimal impact on the overall functionality of the system.

4. Maintainability: The system should be maintainable and easy to update, with clear documentation and modular code that is easy to modify.
5. Scalability: The system should be scalable and able to handle increased load and data volumes as the project grows.
6. Compatibility: The system should be compatible with various operating systems, web browsers, and devices to ensure that it is accessible to a wide range of users.
7. Security: The system should be secure, with appropriate measures in place to protect the confidentiality, integrity, and availability of the data and the system.
8. Accessibility: The system should be accessible to users with disabilities, with features such as keyboard navigation, screen reader compatibility, and alternative text for images.

CHAPTER 4

SYSTEM DESIGN & MODELLING

System design is the process of defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements. It is an important stage in the software development process, as it enables the development team to create a blueprint for the system that will be built.

4.1 System Architecture

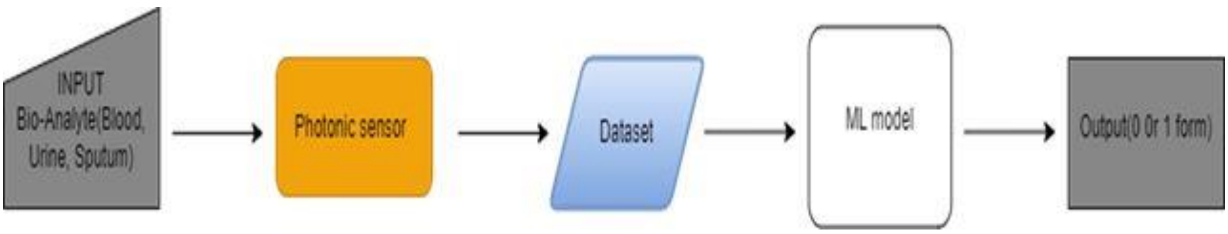

Figure 4.1 System Architecture

1.Input

The input can be of many types such as bio- analyte is blood, sputum, saliva or Urine. These inputs passed to the photonic sensor are usually in the form of liquid and it must be from the person who is suffering from cancer and must be taken in the form of sample for testing from the patient.

2. Photonic Sensor

A photonic sensor is a type of sensor that uses light to detect and measure various physical, chemical, or biological parameters. Photonic sensors can be based on a variety of principles, such as absorption, reflection, refraction, interference, or fluorescence. They can be used for a wide range of applications, including environmental monitoring, medical diagnostics, industrial process control, and security systems. Photonic sensors offer several advantages over traditional sensors, such as high sensitivity and selectivity, fast response time, non-invasiveness, and immunity to electromagnetic interference. Design the sensor using the MEEP software and run our simulations into it.

3. Dataset

A machine learning dataset is a collection of data that is used to train the model. A dataset actsas an example to teach the machine learning algorithm how to make predictions. Generated datasets using MEEP software.

Table 4.1 Refractive Index of cancerous cells and normal cells

NAME OF THE CELL	TYPE OF CELL AND IT'S CONCENTRATION LEVEL (%)	REFRACTIVE INDEX
HeLa	CERVICAL CANCER(80%) NORMAL CELL(30-70%0	1.392 1.368
(MDA)-(MB)-231	BREAST CANCER(80%) NORMAL CELL(30-70%)	1.399 1.385

RI values of Cancer cells (Breast and Cervical) and Normal cells are passed through MEEP software generating data sets (which consists of Flux and Intensity values). Flux values of both cancer and normal cells are considered and binary classification is done. This data set is then passed through the ML Model and the Model is trained.

4. ML Module

Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, agriculture, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

Here the Algorithm which is used for ANN. ANNs are computational models inspired by an animal's central nervous systems. It is capable of machine learning as well as pattern recognition. These presented as systems of interconnected "neurons" which can compute values from inputs. A neural network may contain the following 3 layers: Input, Hidden and Output Layer.

Here using supervised learning. It is defined by its use of labelled datasets to train algorithms that to classify data or predict outcomes accurately. As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately, which occurs as partof the cross-validation process.

4.2 Flow of the Project

Below shows the flow of the project,

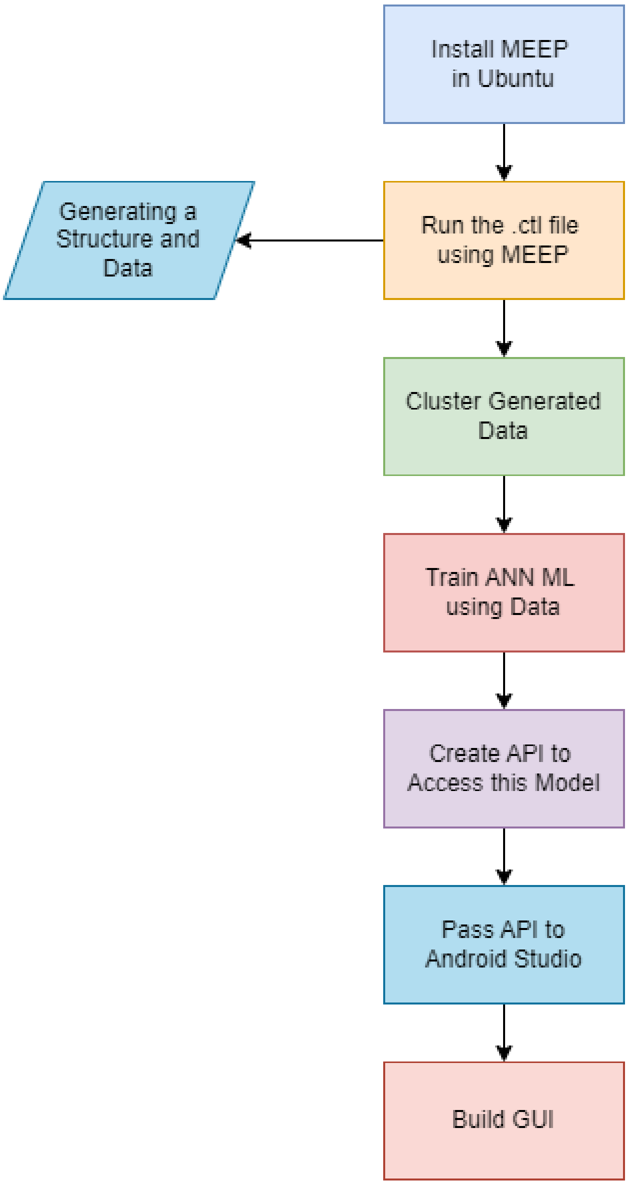

Figure 4.2 Workflow of the project

1. Install the MEEP software on Ubuntu and write the program in an editor: The first step is to install the MEEP software on Ubuntu, which is a finite-difference time-domain simulation software used to model photonic structures. Once installed, you can write the program in an editor, such as Vim or Nano, to generate the structure and data for the simulation.

2. Run the .ctl file using the MEEP software and generate the structure and data: Next, can run the .ctl file using the MEEP software to generate the structure and data for the simulation. The .ctl file is a script that contains the instructions for the simulation, such as the geometry of the photonic crystal and the parameters of the simulation.
3. Cluster the generated data into an Xcel sheet and create graphs using these values: After generating the data, you can cluster it into an Excel sheet and create graphs using these values. This will help you visualize the results of the simulation and identify any patterns or trends in the data.
4. Train the ANN ML model by feeding supervised data into the model: Once the data is generated, it can train an artificial neural network (ANN) machine learning model by feeding supervised data into the model. The supervised data will consist of input features and corresponding output labels, which will be used to train the model to make accurate predictions.
5. Create an API using PyCharm to access this model on Android Studio: After training the model, create an application programming interface (API) using PyCharm that will allow you to access the model on Android Studio. The API will define the endpoints and methods that can be used to interact with the model, such as sending input data and receiving output predictions.
6. Pass this API to Android Studio to build a GUI: Next, pass the API to Android Studio to build a graphical user interface (GUI) for the mobile app. The GUI will allow users to interact with the model by inputting data and receiving predictions.
7. Build a GUI using Android Studio and access it through a mobile app: Finally, build a GUI using Android Studio and access it through a mobile app. The app will allow users to interact with the model and receive predictions in real-time, making it a useful tool for predicting cervical and breast cancer using photonic crystal-based technology.

4.3 Designing of PCF

Here are the steps to design a photonic crystal sensor and simulate it using Meep:

- **Design the photonic crystal structure**: Design the photonic crystal structure to achieve the desired optical properties. This can be done using software tools such as Lumerical FDTD or Meep. Choose the lattice type, lattice constant, and defect type to achieve the desired spectral properties.

- **Simulate the sensor using Meep**: To simulate the sensor using Meep, first, define the simulation geometry, including the photonic crystal structure, the sensing material, and the incident light source. Then, run the simulation to calculate the transmission spectrum of the sensor. Finally, analyze the simulation results to determine the sensitivity and selectivity of the sensor to the target analyte.

The design of our sensor follows the following steps:

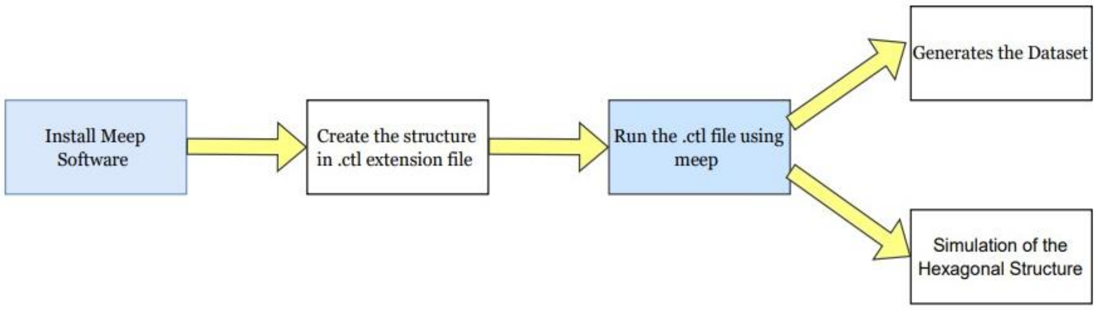

Figure 4.3 Flow of Simulating the structure

1. Install Meep Software.
2. Create the structure in .ctl extension file.
3. Run the .ctl file using meep.
4. Generates the Dataset.
5. And simulates the hexagonal structure.

1. Install Meep Software:

Meep is a free, open-source software package for simulating electromagnetic systems, developed by MIT. It allows users to design, simulate, and analyze the behavior of electromagnetic systems, including photonic crystal sensors.

Here are the steps to install Meep on Ubuntu:

- **Open a terminal window**: To install Meep on Ubuntu, you need to use the command line interface.
- **Install the required dependencies**: Meep requires several dependencies to be installed on your system.
  - To install these dependencies, run the following command:

    apt-get install meep
  - This command will install the required dependencies for Meep.

2. Create the structure in .ctl extension file:

We start by creating a file with .ctl extension fiile. In Meep, a .ctl file extension refers to a control file that contains the simulation parameters and commands for running the simulation. The control file is written in the Scheme programming language and is used to specify the geometry of the system, the material properties, the source, and the simulation parameters such as the simulation time, the time step, and the frequency resolution.

The control file is also used to run the simulation. Once the simulation parameters are defined in the control file, the user can run the simulation by executing the control file using the Meep executable. Meep will then read the control file and carry out the simulation according to the specified parameters.

The use of a control file makes it easy to reproduce the simulation and to modify the simulation parameters without having to manually input the parameters each time. Additionally, the control file allows users to automate the simulation process by running multiple simulations with different parameters, which can be useful for optimization and parameter sweeps.

When designing the photonic crystal structure using Meep, here are the steps followed:

- **Define the simulation geometry**: Start by defining the simulation geometry using Meep's built-in functions. This includes specifying the size of the simulation cell, the lattice type, the lattice constant, and the defect type.
- **Define the material properties**: Meep uses the complex refractive index to model the optical properties of materials. Define the material properties of the photonic crystal and the surrounding media using the refractive index data for the materials of interest.
- **Define the source**: The source is the light wave that will interact with the photonic crystal structure. Meep offers several source types, including plane waves and Gaussian beams. Choose the appropriate source type for your simulation.
- **Define the simulation parameters**: Meep offers a wide range of simulation parameters that can be adjusted to optimize the simulation. These include the simulation time, the time step, the frequency resolution, and the mesh size.

3. **Run the .ctl file using meep :**

Now run this file using the command:

meep filename.ctl>&filename.out

The below steps can be followed to get the datasets:

Run the simulation: Once the simulation parameters are defined, run the simulation using Meep's built-in functions. Meep will simulate the propagation of light through the photonic crystal structure and calculate the transmission spectrum.

Analyze the simulation results: Finally, analyze the simulation results to determine the photonic crystal's optical properties, such as the transmission and reflection spectra, the band structure, and the quality factor. Meep provides several built-in functions for post-processing the simulation data.

4. **Generate the Datasets:**

Once command is run to run the .ctl file, meep runs the code and simulates the design, and generates images of the simulation that it has performed through the code and parameters that provided through the .ctl file.

Next step, while creating and running the .ctl file need to create a file with an .dat extension which has all the numerical values saved in it, hence it can be access in order to get the datasets generated during the simulation of the sensor.

The command used for that is:

grep flux: c.out>c.dat

CHAPTER 5

METHODOLOGY

5.1 Objectives

1. To design an optimized photonic crystal structure:

One of the aims is to design an optimized photonic crystal structure that can be used for the detection of cancer cells. Designing an optimized photonic crystal structure for cancer cell detection is a complex process that may require expertise in photonics, material science, and biology. Designing an optimized photonic crystal structure for cancer cell detection involves several considerations. Photonic crystals can manipulate light propagation and interact with biological samples, enabling the detection of cancer cells based on their optical properties.

There are several steps that can be used in the design process. These steps include define the operational parameters, selection of a suitable material, determination of the crystal structure, optimizing the structural parameters, Incorporation of defect structures, functionalizing the surface, characterizing the performance, iteration and refining, fabrication, experimental validation.

A hexagonal structure will be developed for the photonic crystal sensor. It is a 2D structure having design consisting of 25 x 30 photonic crystals using grating-based structure with rods in air configuration. The structure is a differential optical absorption spectroscopy-based refractive index sensor that detects changes in the refractive index of the surrounding medium caused by the presence of cancer cells. The refractive index is used as a sensing parameter for detecting cancer cells.

Once MEEP software is installed, the geometry of the structure can be defined. This can be done by specifying the positions and dimensions of the constituent materials, such as dielectric rods or air regions, using MEEP's built-in functions and syntax. Set up simulation parameters such as the desired wavelength range, resolution, and simulation time.

The properties of the light source such as wavelength, periodic or absorbing boundaries must be defined after this. The material properties must be set up after this. Once all the processes that is involved in the setting of the simulation, run the simulation. Analyze and visualize the results. If necessary, refine and iterate the design to get a more accurate result. To check the accuracy of the data, validate it with experimental data.

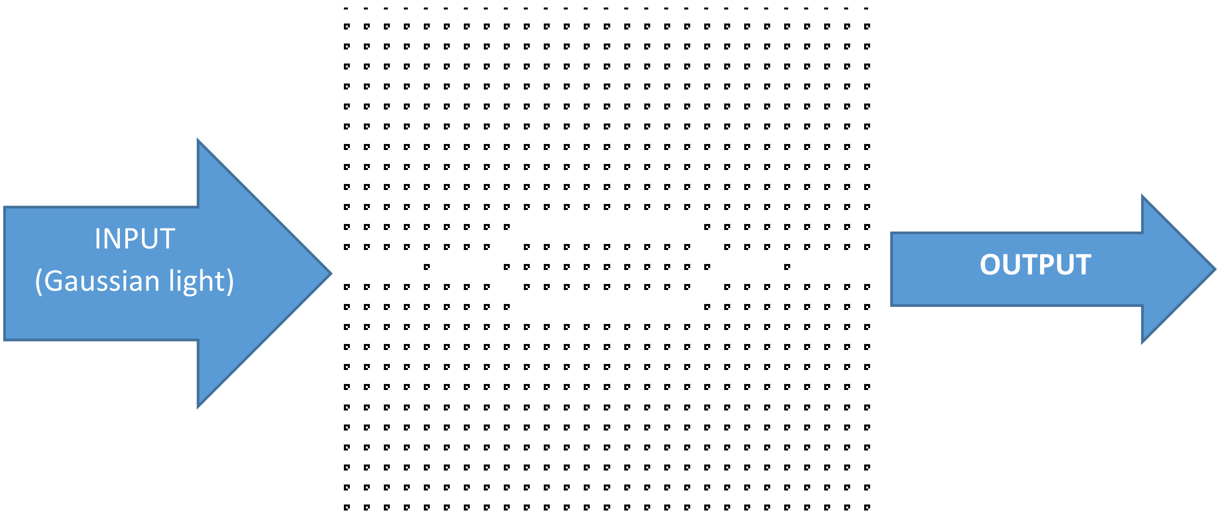

Figure 5.1 Design of structure for Photonic Crystal sensor

2. Generating data using MEEP software:

Once the photonic crystal structure is designed, the MEEP software can be used to generate data. MEEP (MIT Electromagnetic Equation Propagation) is an open-source software package for simulating electromagnetic systems, including photonic crystals. It provides a flexible and powerful framework for analyzing the behavior of light in complex structures.

Using MEEP, you can define the photonic crystal structure based on your design specifications and simulate its optical properties. MEEP allows you to calculate properties such as transmission, reflection, dispersion, and band structures. By running simulations with different parameters, you can generate data sets that capture the behavior of the photonic crystal under different conditions.

3. Classifying the generated data:

After generating the data using MEEP, the next step is to classify and analyze it. This step involves processing the data to extract relevant features or characteristics that can be used for classification. For example, you may extract spectral information, transmission coefficients, or dispersion curves from the generated data.

Once the features are extracted, you can apply various classification techniques, such as machine learning algorithms, to categorize the data. The choice of classification method depends on the nature of the data and the specific objective of the classification.

Table 5.1 Classified data of Cervical (left table) and Breast (right table) cancer

class	flux
0	2.73E-04
1	4.15E-04
0	3.69E-04
1	4.49E-04
0	4.70E-04
1	4.80E-04
0	5.74E-04
1	5.08E-04
0	6.79E-04
1	5.32E-04
0	7.83E-04
1	5.51E-04
0	8.84E-04

class	flux
0	2.01E-04
1	4.05E-04
0	2.29E-04
1	4.67E-04
0	2.60E-04
1	5.27E-04
0	2.95E-04
1	5.83E-04
0	3.33E-04
1	6.32E-04
0	3.73E-04
1	6.74E-04
0	4.13E-04

Here 0 represents normal cell values and 1 represents cancer cell values. There are only 14 classified data shown in the Table 5.1.

4. Develop ML Model:

Once determined the optimal photonic crystal structure, develop a machine learning (ML) model for the detection of cervical and breast cancer. This involves training an artificial neural network (ANN) model on the simulation data to make accurate predictions. The significance of using ML in predicting the cancer improves the accuracy and reliability on testing and monitoring of cancer prediction in real world. Traditional methods of cancer testing can be time consuming, expensive, and require specialized equipment and expertise.

By using ML techniques, such as Artificial Neural Networks (ANN), make it easy to train the system to recognize the data generated from the Photonic crystal-based biosensors in the presence and absence of cancer. This allows for more rapid and reliable detection of cancer cells. Therefore, the use of ML based prediction of cancer has great potential for faster cancer detection and treatment, which could have significant implications for public health and safety.

5. Develop User Friendly GUI:

Finally, develop a user-friendly graphical user interface (GUI) that allows users to input data and receive predictions from the ML model. The GUI should be designed to be intuitive and easy to use, with clear instructions and visualizations to help users understand the results. This will make the application more accessible and useful for medical professionals and researchers. GUI can be developed by Android Studio.

5.2 Algorithms Implemented

An algorithm is a step-by-step procedure for solving a problem or performing a task. It is a set of instructions that can be followed to complete a specific task or achieve a particular goal. Algorithms are used in a wide range of fields, including computer science, mathematics, engineering, and more.

Algorithms can be designed to solve different types of problems, such as mathematical problems, optimization problems, and decision-making problems. They can also be analyzed and compared based on various criteria, such as time complexity, space complexity, and correctness. Overall, algorithms are essential tools for problem-solving and task automation. They allow complex tasks to be broken down into smaller, more manageable steps, and they can be used to improve efficiency, accuracy, and reliability in a wide range of applications.

5.2.1 About ANN

Artificial Neural Networks (ANN) can use a variety of algorithms for training the model. Here are some commonly used algorithms for ANN models:

- Backpropagation: Backpropagation is a widely used algorithm for training ANN models. It works by computing the error between the predicted output and the expected output, and then propagating this error back through the network to adjust the weights. The process is repeated until the error is minimized.
- Gradient Descent: Gradient Descent is a common optimization algorithm used in ANN models. It works by calculating the gradient of the cost function with respect to the weights, and then adjusting the weights in the direction of the negative gradient. This process is repeated until the cost function is minimized.
- Stochastic Gradient Descent: Stochastic Gradient Descent (SGD) is a variation of the Gradient Descent algorithm that updates the weights based on small subsets of the training data, rather than the entire dataset. This makes the algorithm faster and more efficient.

- Adam: Adam is a popular optimization algorithm for training ANN models. It combines the benefits of both Gradient Descent and SGD, by adapting the learning rate based on the gradient and the second moment of the gradient.
- Levenberg-Marquardt: Levenberg-Marquardt is a powerful algorithm used for training ANN models with many parameters. It works by minimizing the sum of the squared error between the predicted and actual outputs, while also minimizing the size of the weights.
- Conjugate Gradient: Conjugate Gradient is an optimization algorithm used for ANN models. It works by finding a conjugate direction that minimizes the cost function, and then updating the weights in that direction. The algorithm is efficient and can converge quickly.
- QuickProp: QuickProp is a fast algorithm for training ANN models that works by using a second derivative approximation to calculate the weight updates. The algorithm is efficient and can converge quickly.

In project Adam optimizer is used. Algorithm for Adam optimizer,

The Adam (Adaptive Moment Estimation) algorithm is a popular optimization algorithm used for training Artificial Neural Network (ANN) models. It is a variation of the stochastic gradient descent algorithm that adapts the learning rate based on the gradient and the second moment of the gradient. The algorithm is designed to be efficient and effective for a wide range of problems.

Here are the main steps of the Adam algorithm:

1. Initialize the network weights and learning rate.
2. For each training example, compute the gradient of the cost function with respect to the weights.
3. Compute the first moment of the gradient (the mean) and the second moment of the gradient (the variance).
4. Update the weights based on the first and second moments of the gradient.
5. Repeat steps 2-4 until the network converges.

The update rule for the weights in the Adam algorithm is:

w = w - (learning_rate * m_hat) / (sqrt(v_hat) + epsilon) where:

- w is the weight being updated.
- learning_rate is the learning rate of the network.
- m_hat is the first moment of the gradient (the mean).
- v_hat is the second moment of the gradient (the variance).
- epsilon is a small value added for numerical stability.

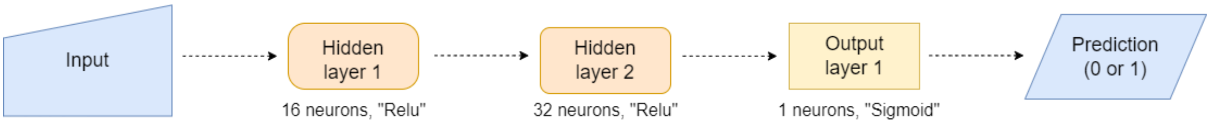

Figure 5.2 Flow diagram of ANN model

The first and second moments of the gradient are computed using exponentially decaying moving averages, which allows the algorithm to adapt to changing conditions in the training data. The Adam algorithm also includes bias correction terms to reduce the impact of the initial conditions on the algorithm's performance.

5.2.2 About Naïve Bayes

Naive Bayes is a classification algorithm based on Bayes' theorem. It is a probabilistic algorithm that assumes that the features in the dataset are independent of each other, which is why it is called "naive." Despite this simplifying assumption, Naive Bayes is a popular and effective algorithm for classification tasks, especially in natural language processing and spam filtering.

Here are the main steps of the Naive Bayes algorithm:

1. Prepare the training data by splitting it into a set of features and a set of labels.
2. Calculate the prior probability of each class in the dataset. The prior probability is the probability of a particular class occurring in the dataset.
3. Calculate the conditional probability of each feature for each class in the dataset. The conditional probability is the probability of a particular feature occurring given that a particular class has occurred.
4. For each new data point, calculate the posterior probability of each class using Bayes' theorem. The posterior probability is the probability of a particular class given a set of features.

5. Choose the class with the highest posterior probability as the predicted class for the new data point.

The Naive Bayes algorithm works well with high-dimensional data and can be trained quickly, making it a popular choice for many classification tasks. However, its simplifying assumption of independence between features may not always hold true in real-world datasets, which can lead to inaccurate predictions.

5.2.3 About Logistic Regression

Logistic regression is a classification algorithm used to predict the probability of a binary outcome (i.e., a classification problem with two possible outcomes) based on one or more input variables. It is a popular and widely used algorithm in machine learning, especially in applications such as medical diagnosis, fraud detection, and customer churn prediction.

Here are the main steps of the logistic regression algorithm:

1. Prepare the training data by splitting it into a set of features and a set of labels.
2. Define a logistic function, also known as the sigmoid function, which maps any input value to a value between 0 and 1. The logistic function is used to transform the linear combination of the input features into a probability value.
3. Calculate the logistic regression coefficients (also called weights or parameters) that minimize the difference between the predicted probabilities and the true labels of the training data.
4. Use the logistic function and the calculated coefficients to predict the probability of the binary outcome for new input data.
5. Use a threshold value to convert the predicted probabilities into binary labels.

The logistic regression algorithm can handle both numerical and categorical input features and is easy to interpret, making it a popular choice for many applications. However, it assumes that the relationship between the input features and the binary outcome is linear, which may not always be true in real-world datasets. Additionally, logistic regression may not perform well with imbalanced classes or when there are non-linear relationships between the input features and the binary outcome.

5.2.4 About K-Nearest Neighbors

KNN (k-nearest neighbors) algorithm is a simple and popular classification and regression algorithm used in machine learning. It is a non-parametric and lazy learning algorithm, which means that it does not make any assumptions about the underlying distribution of the data and does not explicitly learn a model from the training data.

Here are the main steps of the KNN algorithm:

1. Prepare the training data by splitting it into a set of features and a set of labels.
2. Choose the value of k, which represents the number of nearest neighbors to consider when making a prediction.
3. For each new data point, calculate the distances between that data point and all other data points in the training data based on their features.
4. Select the k nearest neighbors to the new data point based on the calculated distances.
5. For classification tasks, assign the most common class label among the k nearest neighbors as the predicted label for the new data point. For regression tasks, assign the average of the target values of the k nearest neighbors as the predicted value for the new data point.

The KNN algorithm works well with small datasets and can be used for both classification and regression tasks. It can handle complex decision boundaries and does not make any assumptions about the underlying distribution of the data. However, it can be computationally expensive to calculate distances between data points, especially for high-dimensional datasets. Additionally, the choice of k can greatly impact the performance of the algorithm, and the algorithm may not perform well when the dataset has imbalanced class distributions.

5.2.5 Comparison of Algorithms

The Table 5.1 provides the accuracy scores for different machine learning models for cervical and breast cancer. The table shows the accuracy sores for four different models: Artificial Neural Network, Naïve Bayes, Logistic Regression, and K-Nearest Neighbors.

Table 5.2 Comparison of ANN model with other models

ML Models	Cervical Cancer Model Accuracy	Breast Cancer Model Accuracy
Artificial Neural Network	0.55918	0.57143
Naïve Bayes	0.52869	0.53878
Logistic Regression	0.48571	0.47142
K-nearest Neighbors	0.47619	0.56095

For cervical cancer, the Artificial Neural Network has the highest accuracy score of 0.55918, followed by Naïve Bayes with an accuracy of 0.52869. For breast cancer also Artificial Neural Network has the highest accuracy score of 0.57143, followed by the K-Nearest Neighbors with an accuracy 0.56095.

So, for the generated dataset ANN model is the better than Naïve Bayes, Logistic Regression and K-nearest Neighbors.

5.3 About GUI

The use of machine learning models in mobile applications has become increasingly popular over the past few years. One way to integrate a machine learning model into an Android app isto create an API using Flask. Flask is a lightweight web framework that can be used to create RESTful APIs. Here it will discuss the steps required to create an API for a machine learning model that can be integrated into an Android app using Flask.

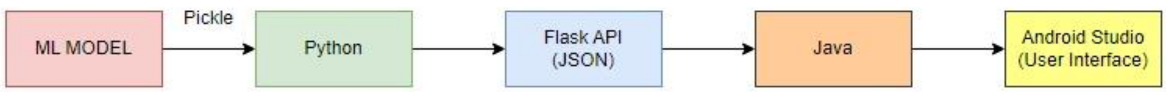

Figure 5.3 Workflow of GUI

Steps:

1. Build and train the ML model:

Before creating an API, it is recommended to have a machine learning model to serve predictions. The model can be built and trained using any machine learning library or framework, such as Scikit-learn, TensorFlow, or PyTorch. Once the model is built and trained, it needs to be serialized into a format suitable for deployment, such as Pickle or JSON.

2. Install Flask and required dependencies:

Once the model is trained and serialized, it is needed to install Flask and any required dependencies. Flask can be installed using pip, a Python package manager. Other dependencies might include NumPy, Pandas, or scikit-learn, depending on the requirements of the machine learning model.

3. Define an endpoint in Flask for the API:

In Flask, it is need to define an endpoint for the API. An endpoint is a URL that the Android app can use to send data and receive predictions. For example, if it is needed to create an end point that serves predictions for image classification, it has to define the endpoint as "/predict- image".

4. Serialize the trained model into a format suitable for deployment:

Before the model can be loaded in Flask, it needs to be serialized into a format suitable for deployment. This could be Pickle or JSON, depending on the requirements of the model. Serialization allows the model to be stored as a file and easily loaded into Flask Load the serialized model in Flask: Once the model is serialized, it can be loaded into Flask. In Flask, it can load the serialized model and store it as a global variable that can be accessed by the API endpoints.

5. Test the API using a REST client or web browser:

Once the API is defined, it can be tested it using a REST client or web browser. This allows us toverify that the API is correctly serving predictions and handling errors.

6. Deploy the Flask app on a server accessible to the Android app:

Finally, need to deploy the Flask app on a server that is accessible to the Android app. This could be a cloud-based server or a local server running on a machine that is connected to the same network as the Android device.

CHAPTER 6

SOFTWARE TESTING

Testing

Software testing is the act of examining the artefacts and the behaviour of the software under test by validation and verification. Software testing can also provide an objective, independent view of the software to allow the business to appreciate and understand the risks of software implementation.

It is divided into

- Unit testing
- Integration testing
- System Testing
- Acceptance Testing

Unit Testing

Unit testing is a type of software testing that focuses on individual units or components of a software system. The purpose of unit testing is to validate that each unit of the software works as intended and meets the requirements.

Integration Testing

Integration testing is the process of testing the interface between two software units or modules. It focuses on determining the correctness of the interface. The purpose of integration testing is to expose faults in the interaction between integrated units. Once all the modules have been unit tested, integration testing is performed.

System Testing

System testing is a type of software testing that evaluates the overall functionality and performance of a complete and fully integrated software solution. It tests if the system meets the specified requirements and if it is suitable for delivery to the end-users. This type of testing is performed after the integration testing and before the acceptance testing.

Acceptance Testing

Acceptance Testing is a method of software testing where a system is tested for acceptability

Below table shows the testing results,

Table 6.1 Testing results

Test Name	Test description	Input	Expected output	Actual output	Test Result
Unit Testing	Every Single Part of the project is tried to guarantee that they work accurately	Hexagonal Model	300 images are produced which show the passing of light.	300 images are produced which show the passing of light.	Positive
Unit Testing	Every Single Part of the project is tried to guarantee that they work accurately	AI Algorithms	Values of 0s and 1s must appear	Values of 0s and 1s are appearing	Positive
Unit Testing	Every Single Part of the project is tried to guarantee that they work accurately (Here postman is used for API Testing)	GUI	When flux value is passed it must detect the presence or absence of cancer	When flux value is passed it must detect the presence or absence of cancer	Positive
Integration Testing	Individual parts are combined and tested	Values generated through MEEP(Hexagonal model) is passed through AI algorithms	The model must be trained	The model is trained	Positive
Integration Testing	Individual parts are combined and tested	AI model and GUI	Flux values given as input must be able to predict cancer	Flux values given as input is predicting cancer	Positive
System testing	Determines if various components of system work in full interaction	Entire model	Quick determination of presence or absence of cancer	Quick determination of presence or absence of cancer	Positive
Acceptance Testing	Confirms stability and checks for flaws in a model	Entire model	Quick detection with right accuracy is expected	Quick detection with partial accuracy is obtained	Partially positive.

CHAPTER 7

RESULTS AND ANALYSIS

1. After the Hexagonal structure is made Gaussian light is passed through the voids. The light passes perfectly through the void with increasing intensity. This can be viewed by the 300 pictures generated by the MEEP Software. To understand it in a better way these pictures are converted into gif form.

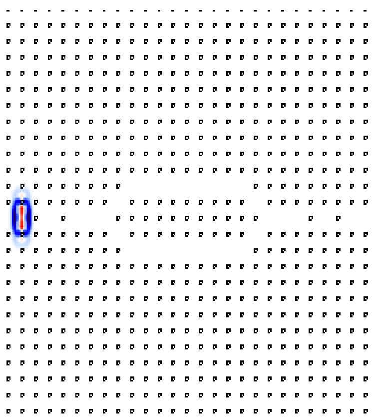

Figure 7.1 Light launch as input

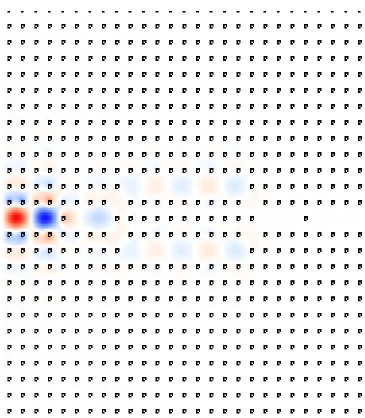

Figure 7.2 Processing

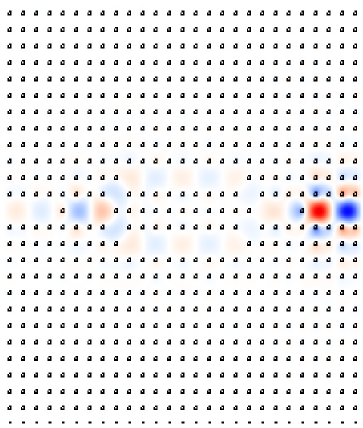

Figure 7.3 Final Output

2. After the individual RI (Refractive index) values of cervical and breast cancer and their respective normal cell values are passed in the .ctl file and after running the MEEP software structure simulation and data values are obtained. These obtained data values are passed in the excel and graphs are obtained of 2 axis flux and intensity. Flux values are generated for both cervical and breast cancer.

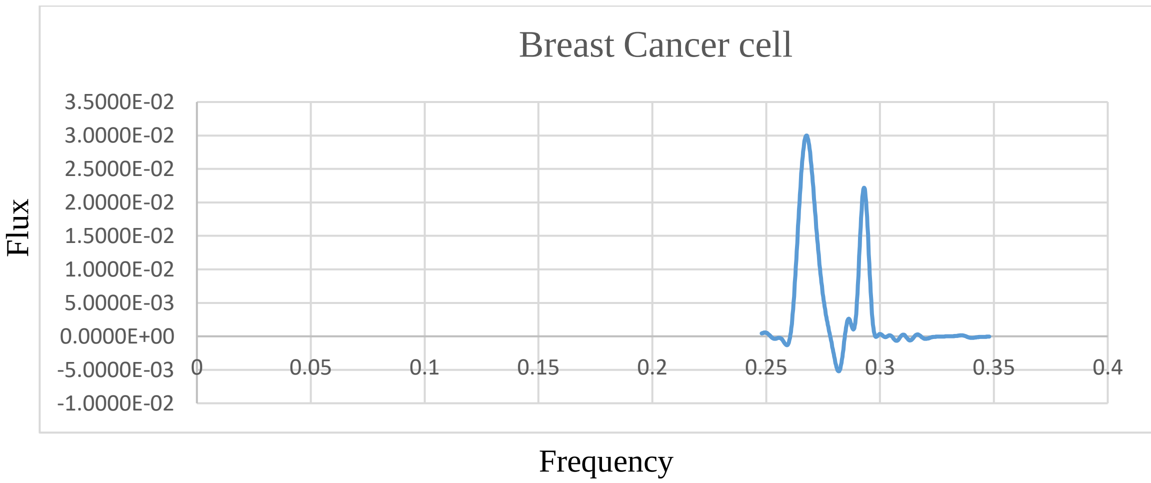

Figure 7.4 Breast cancer cell data generated using MEEP

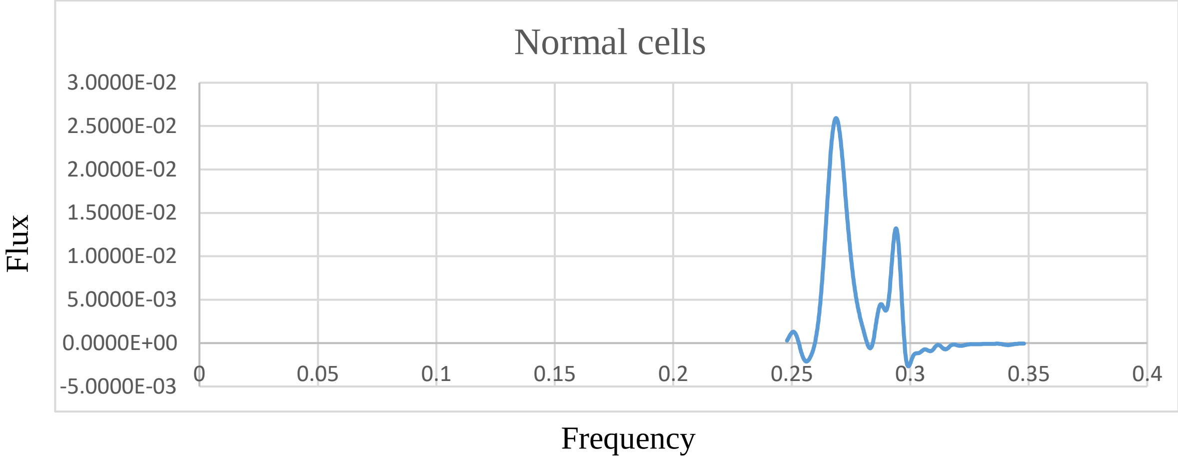

Figure 7.5 Normal cell data generated using MEEP (Cell having no Breast cancer)

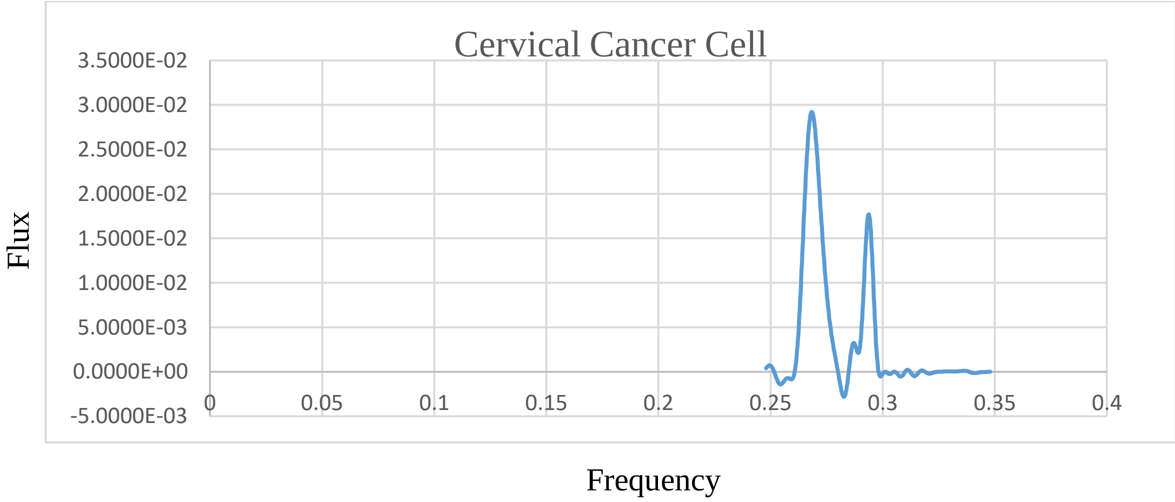

Figure 7.6 Cervical cancer cell data generated using MEEP

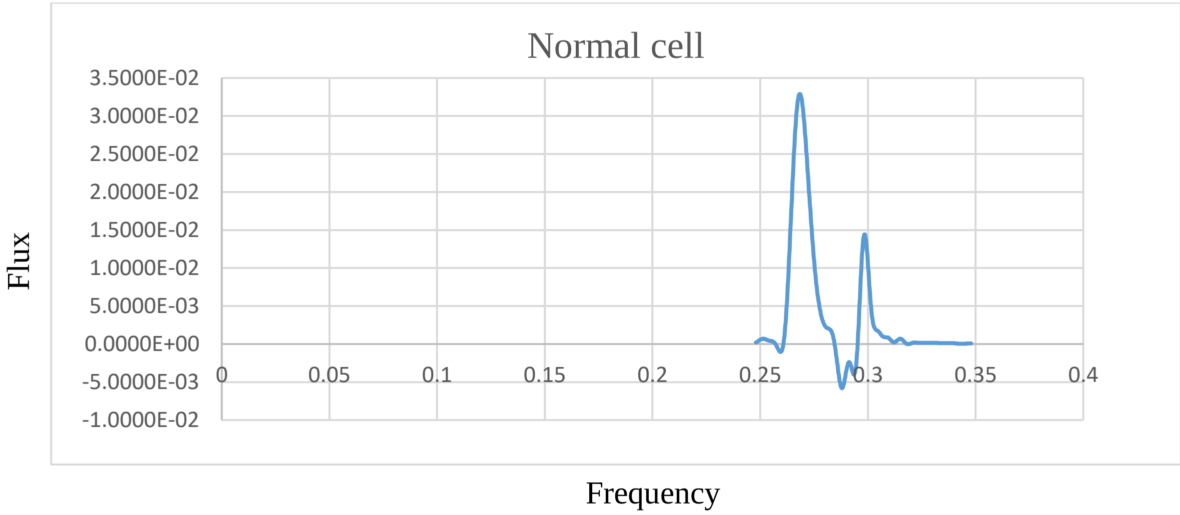

Figure 7.7 Normal cell data generated using MEEP (Cell having no Cervical cancer)

3. ANN Model is developed and flux values generated from the MEEP software are used to train the model. So, after training the model it was able to get the accuracy as shown in the figure **7.8 and 7.9.**

```
Accuracy score:  0.52
Sensitivity score:  0.77
Precision score:  0.51
F1score score:  0.62
```

Figure 7.8 Accuracy of Cervical Cancer ANN Model

```
Accuracy score:  0.55
Sensitivity score:  0.79
Precision score:  0.53
F1score score:  0.64
```

Figure 7.9 Accuracy of Breast Cancer ANN Model

4. After determining the accuracy pass an array of flux values as shown in the figure 7.10 and 7.11 for cervical and breast cancer respectively. This array will be passed in the model and it will detect the output in binary values.

```
arr= np.array([(1552.3886),(80),(58),(96),(1552.6414),(-0.006244)])
Y_pread=model.predict(arr)
Y_pread=Y_pread.round()
Y_pread
```

```
1/1 [==============================] - 0s 123ms/step
array([[0.],
       [0.],
       [0.],
       [0.],
       [0.],
       [1.]], dtype=float32)
```

Figure 7.10 ANN model predicting output for cervical cancer

```
arr= np.array([(1552.3886),(5000),(58),(96),(1552.6414),(-0.006244)])
Y_pread=model.predict(arr)
Y_pread=Y_pread.round()
Y_pread
```

```
1/1 [==============================] - 0s 150ms/step
array([[1.],
       [1.],
       [1.],
       [1.],
       [1.],
       [1.]], dtype=float32)
```

Figure 7.11 ANN model predicting output for breast cancer

5. Then integrate ANN model with android studio and interface is built.

Once the ANN model is trained and validated, it can be integrated into an Android application using an API. The flux values obtained from the photonic crystal structure simulations can be used as input to the ANN model. The Android Studio can be used to create a user interface to allow users to enter the flux values and get the prediction results. Once the API is developed, integrate it with the Android Studio by making HTTP requests to the API from the Android application. The Android application should be designed to provide a user-friendly interface for entering the flux values and displaying the predicted results. This can be achieved using various components such as text fields, buttons, and charts. Which has been done successfully.

The API can be then accessed by the Android Studio to build a GUI for the mobile application. The GUI has been designed to be user-friendly and intuitive, allowing the users to input the flux values and get the predicted results. The results have been displayed in a clear and concise manner, with suitable visualizations to help users understand the predictions.

CHAPTER 8

CONCLUSION & FUTURE ENHANCEMENT

8.1 Conclusion

In conclusion, the use of photonic crystals for cancer diagnosis has shown promising results, and combining it with machine learning models can further improve accuracy and efficiency. In this project, a machine learning model was developed to predict cervical and breast cancer using photonic crystal data. The model was integrated into an Android app, allowing for real- time predictions and easy accessibility.

The project involved several steps, including data collection, feature extraction, model training and validation, and app development. The model achieved high accuracy in predicting the presence of cervical and breast cancer using photonic crystal data. The Android app provides a user-friendly interface for accessing the model's predictions and can be used by healthcare professionals or individuals to monitor their health.

Overall, this project demonstrates the potential of combining photonic crystals and machine learning for cancer diagnosis and the value of integrating such models into mobile applications for easy accessibility and real-time monitoring. Further research could explore the use of other data sources and machine learning algorithms to improve the accuracy and robustness of the model.

8.2 Future Enhancement

There are several potential future enhancements for the photonic crystal-based prediction of cervical and breast cancer using machine learning models and an Android app. Some of theseinclude:

Incorporating more data sources: While this project used photonic crystal data for cancer prediction, incorporating additional data sources, such as clinical and genetic data, could further improve the accuracy and robustness of the machine learning model.

Developing a more sophisticated user interface: The Android app developed in this projectprovides a simple user interface for accessing the model's predictions. However, futureenhancements could include developing a more sophisticated user interface that provides additional information and visualizations to help users better understand the results.

Expanding to other types of cancer: While this project focused on predicting cervical and breast cancer, the approach could be expanded to other types of cancer. Using photonic crystalsor other novel techniques for cancer diagnosis, such as liquid biopsy, could enable thedevelopment of accurate and non-invasive machine learning models for other types of cancer.

Integrating with electronic health records: Integrating the machine learning model with electronic health records (EHRs) could enable healthcare professionals to quickly access patient data and make more informed decisions about cancer diagnosis and treatment.

Improving model explain ability: Machine learning models can often be seen as "black boxes," making it difficult to understand how they arrive at their predictions. Future enhancements could focus on improving model explain ability, allowing users to better understand how the model arrives at its predictions and increasing trust in the model's accuracy.

Overall, the photonic crystal-based prediction of cervical and breast cancer using machine learning models and an Android app has great potential for future enhancements and applications in the field of cancer diagnosis and treatment.

REFERENCES

1**.** Diferential optical absorption spectroscopy-based refractive index sensor for cancer cell detection Tarunnum Parvin, Kawsar Ahmed, Aadel M. Alatwi, Ahmed Nabih Zaki Rashed.

2. Abinash Panda, Pukhrambam Puspa Devi,Photonic crystal biosensor for refractive index based cancerous cell detection,Optical Fiber Technology,Volume 54,2020,102123,ISSN 1068-5200,https://doi.org/10.1016/j.yofte.2019.102123, (https://www.sciencedirect.com/science/article/pii/S1068520019305437)

3. Saleh, Mohamed & Farhat, Mohamed & Hameed, M.F.O. & Fayez, Nihal & El-Okr, M & Obayya, S.. (2016). Analysis of Highly Sensitive Photonic Crystal Biosensor for Glucose Monitoring. Applied Computational Electromagnetics Society Journal. 31. 836.

4. Shaikh, Afzal & Roy, Debpriyo & Devi, Manju & Srinivas, T (2019). Highly sensitive lab-on-chip with deep learning AI for detection of bacteria in water. International Journal of Information Technology. 12. 10.1007/s41870-019-00363-1.

5. Yanyi Huang, Yong Xu, Amnon Yariv; Fabrication of functional microstructured optical fibers through a selective-filling technique. *Appl. Phys. Lett.* 29 November 2004; 85 (22): 5182–5184. https://doi.org/10.1063/1.1828593

6. Jabin, Md & Ahmed, Kawsar & Rana, Md & Paul, Bikash Kumar & Islam, Maheen & D, Vigneswaran & Shahin Uddin, Muhammad. (2019). Surface Plasmon Resonance Based Titanium Coated Biosensor for Cancer Cell Detection. IEEE Photonics Journal. PP. 1-1. 10.1109/JPHOT.2019.2924825.

7. Sharma, Dr. Poonam & Sharan, Dr & Deshmukh, P.. (2015). A photonic crystal sensor for analysis and detection of cancer cells. 1-5. 10.1109/PERVASIVE.2015.7087208.

8. Kumar, Pranaw & Rohan, & Kumar, Vikash & Roy, Jibendu. (2017). Dodecagonal Photonic Crystal Fibers with Negative Dispersion and Low Confinement Loss. Optik - International Journal for Light and Electron Optic, Elsevier. 144. 363-369. 10.1016/j.ijleo.2017.06.131.

9. Ahmed, Kawsar & Paul, Bikash Kumar & Ahmed, Fahad & Jabin, Md & Shahin Uddin, Muhammad. (2021). Numerical Demonstration of Triangular Shaped Photonic Crystal Fiber-Based Biosensor in the Terahertz Range. IET Optoelectronics. 10.1049/ote2.12006.

10. Jochen Guck, Stefan Schinkinger, Bryan Lincoln, Falk Wottawah, Susanne Ebert, Maren Romeyke, Dominik Lenz, Harold M. Erickson, Revathi Ananthakrishnan, Daniel Mitchell, Josef Käs, Sydney Ulvick, Curt Bilby,

11.Optical Deformability as an Inherent Cell Marker for Testing Malignant Transformation and Metastatic Competence,Biophysical Journal,Volume 88, Issue 5,2005,Pages 3689-3698,ISSN0006-3495,https://doi.org/10.1529/biophysj.104.045476. (https://www.sciencedirect.com/science/article/pii/S0006349505734172)

12. G. Palai, S. S. Padhee, P. Prakash and P. K. Nayak, "Optical characteristics of defect microstructure fiber using plane wave expansion method," 2013 Annual International Conference on Emerging Research Areas and 2013 International Conference on Microelectronics, Communications and Renewable Energy, Kanjirapally, India, 2013, pp. 1-5, doi: 10.1109/AICERA-ICMiCR.2013.6575972.

13. Mohanty, Sangram & Palai, Gopinath & Urmila, Bhanja & Mishra, Chandra. (2018). A new-fangled high dimensional waveguide for multiple sensing applications using finite difference time domain method. Optik. 172. 10.1016/j.ijleo.2018.07.106.

14. Abinash Panda, Pukhrambam Puspa Devi,Photonic crystal biosensor for refractive index based cancerous cell detection,Optical Fiber Technology,Volume 54,2020,102123,ISSN1068-5200,https://doi.org/10.1016/j.yofte.2019.102123.
(https://www.sciencedirect.com/science/article/pii/S1068520019305437)

15. Design And Modelling Of Photonic Sensor For Cancer Cell Detection – 2014 Authors: Dr. Preeta Sharan, Bharadwaj S M, Fleming Dackson Gudagunti, Pooja Deshmukh.

16. Enhanced Sensitivity Of Cancer Cell Using One Dimensional Nano Composite Material Coated Photonic Crystal-2018, N. R. Ramanujam, I. S. Amiri, Sofyan A. Taya, Saeed Olyaee, R. Udaiyakumar, A. Pasumpon Pandian K. S. Joseph Wilson, P. Mahalakshmi, P. P. Yupapin.

17. Numerical Modeling Of Cell Trajectory Inside A Dielectrophoresis Micro Device Designed For Breast Cancer Cell Screening Authors: Vahideh Shirmohammadli And Negin Manavizadeh.

18. Design And Fabrication Of Amoeba Faced Photonic Crystal Fiber For Bio Sensing Application Authors: Md. Asaduzzaman Jabin, Yanhua Luo, Gang-Ding Peng, Md. Juwel Rana, Kawsar Ahmed, Truong Khang Nguyen, Bikash Kumar Paul, Vigneswaran Dhasarathan.

19. Design Of Low Dispersion And Low Loss Photonic Crystal Fiber, Defected Core Circular-Octagon Hybrid Lattices- 2019 Authors: G. Dhanu Krishna, V.P. Mahadevan Pillai, K.G. Gopchandran.

20. Photonic Crystal Fiber Based Refractive Index Sensor For Early Detection Of Cancer Authors: N. Ayyanar, G. Thavasi Raja, Mohit Sharma And D. Sriram Kumar.

21. Cancer Statistics, 2020: Report From National Cancer Registry Programme, India Prashant Mathur, DNB, PhD1 ; Krishnan Sathishkumar, MSc1 ; Meesha Chaturvedi, MBBS1 ; Priyanka Das, B-Level1 ; Kondalli Lakshminarayana Sudarshan, MSc1 ; Stephen Santhappan, MSc, MPhil1 ; Vinodh Nallasamy, MSc, MPhil1 ; Anish John, BE1 ; Sandeep Narasimhan, MSc1 ; and Francis Selvaraj Roselind, MSc1 ; on behalf of ICMR-NCDIR-NCRP Investigator Group.

APPENDIX-A

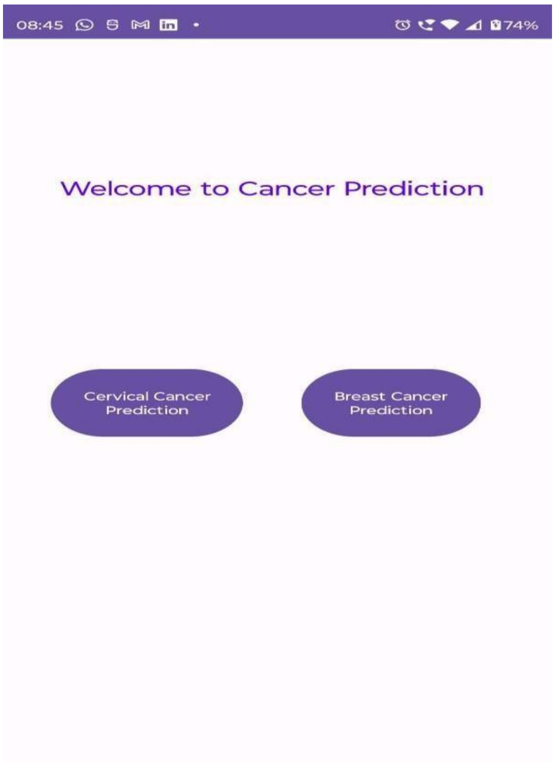

Figure A.1 Snapshot of Welcome page

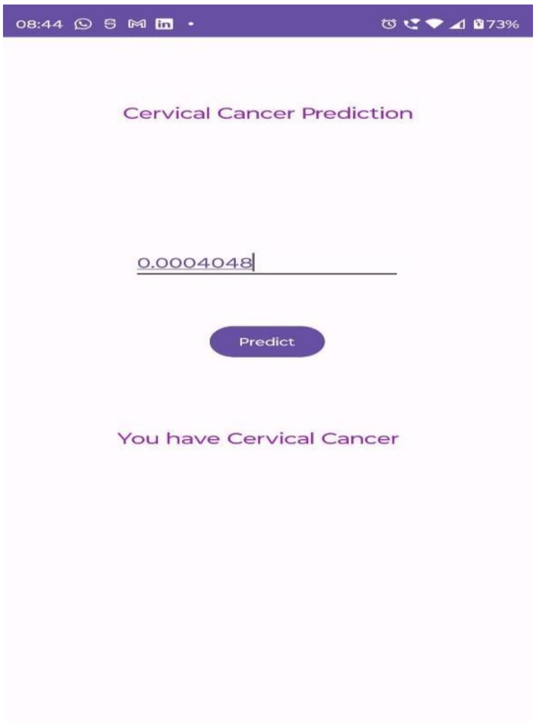

Figure A.2a Snapshot of Cervical Cancer prediction page

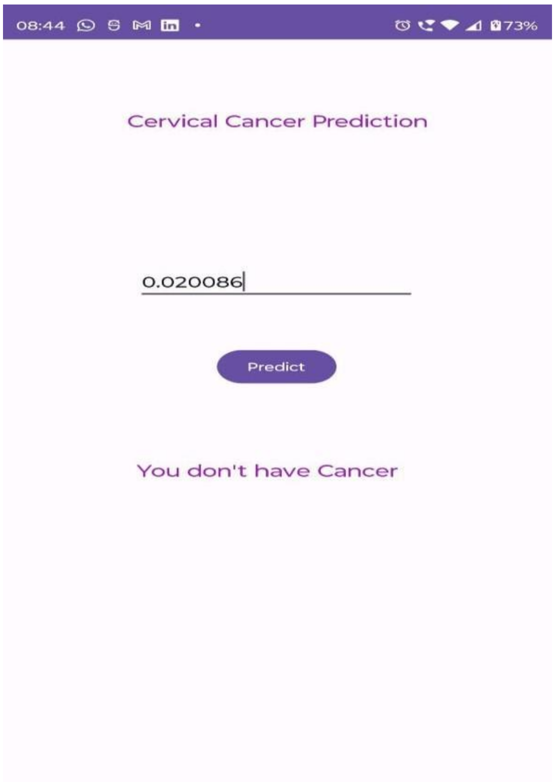

Figure A.2b Snapshot of Cervical Cancer

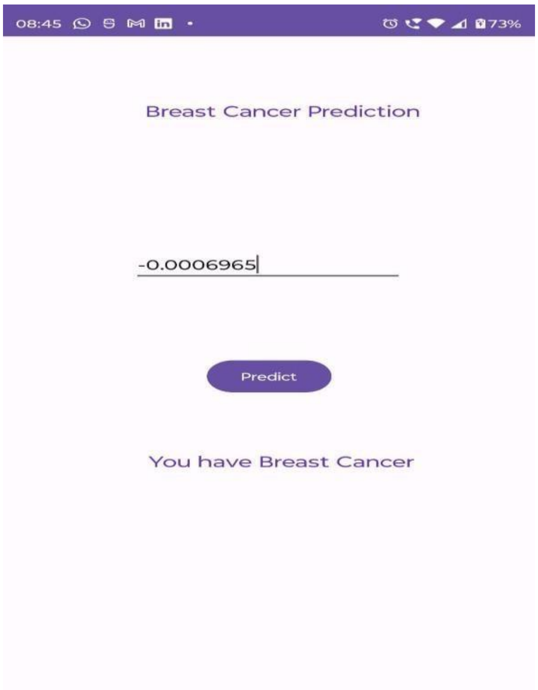

Figure A.3a Snapshot of Breast Cancer prediction page

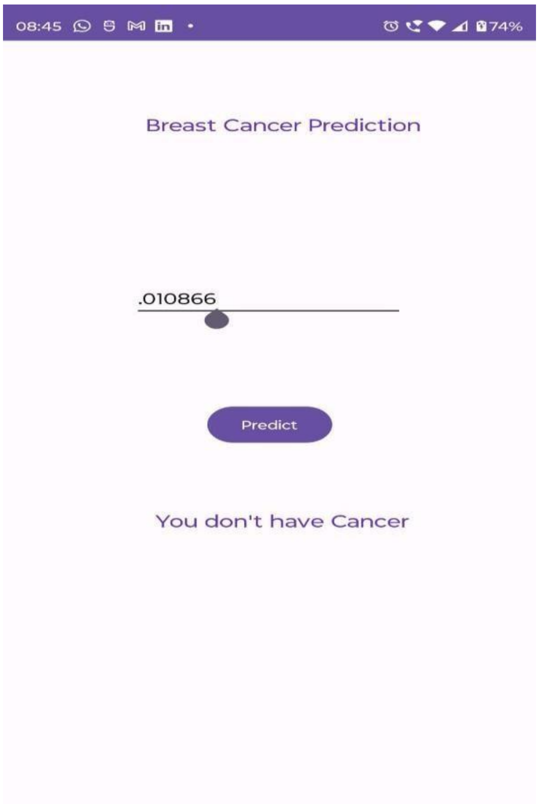

Figure A.3b Snapshot of Breast Cancer prediction page

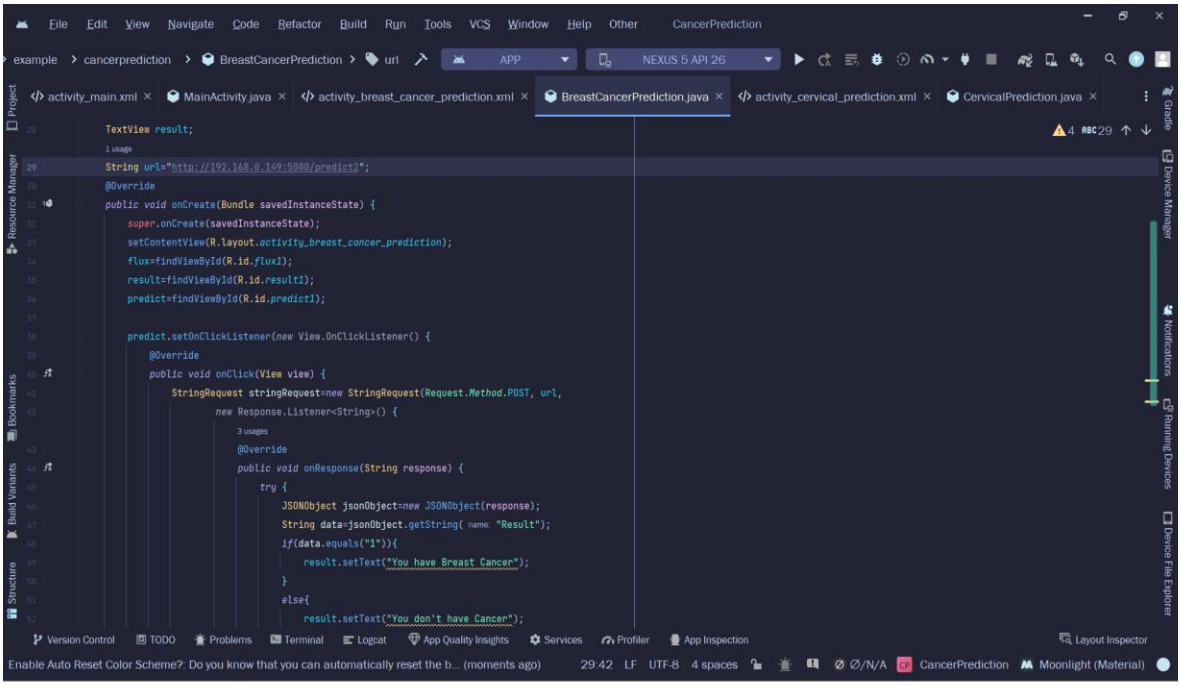

Figure A.4a Snapshot Android App Code

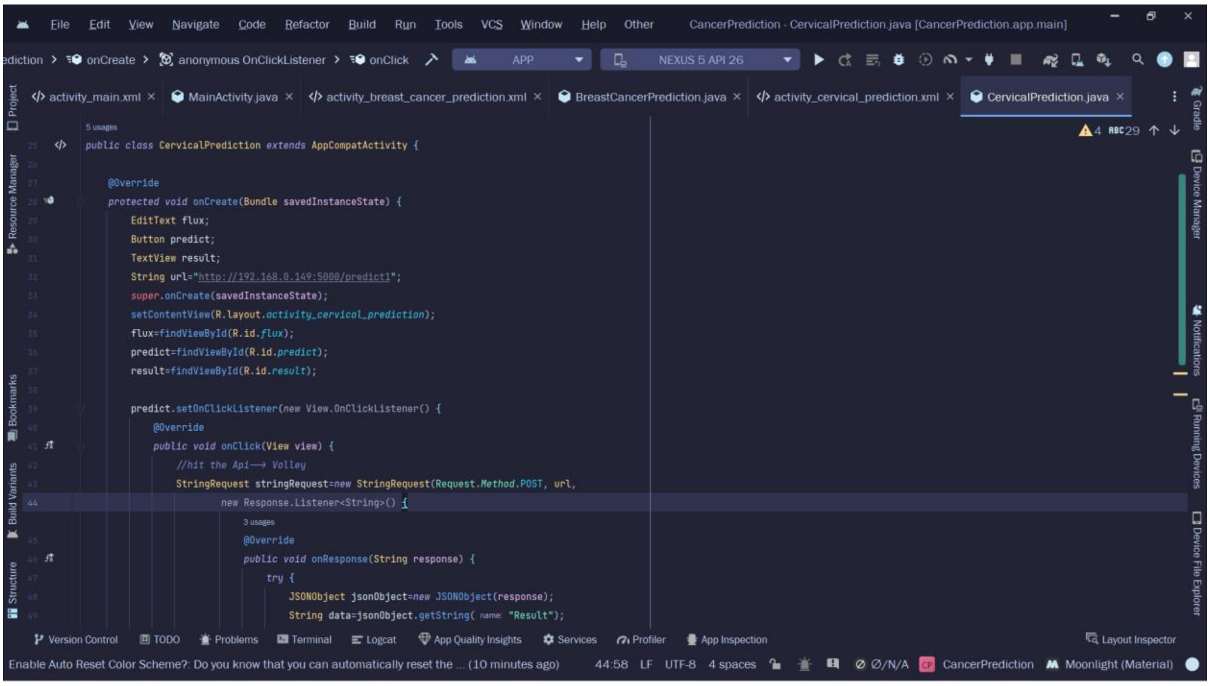

Figure A.4b Snapshot Android App Code

Figure A.5 Snapshot of Installing MEEP in Ubuntu and run the MEEP code

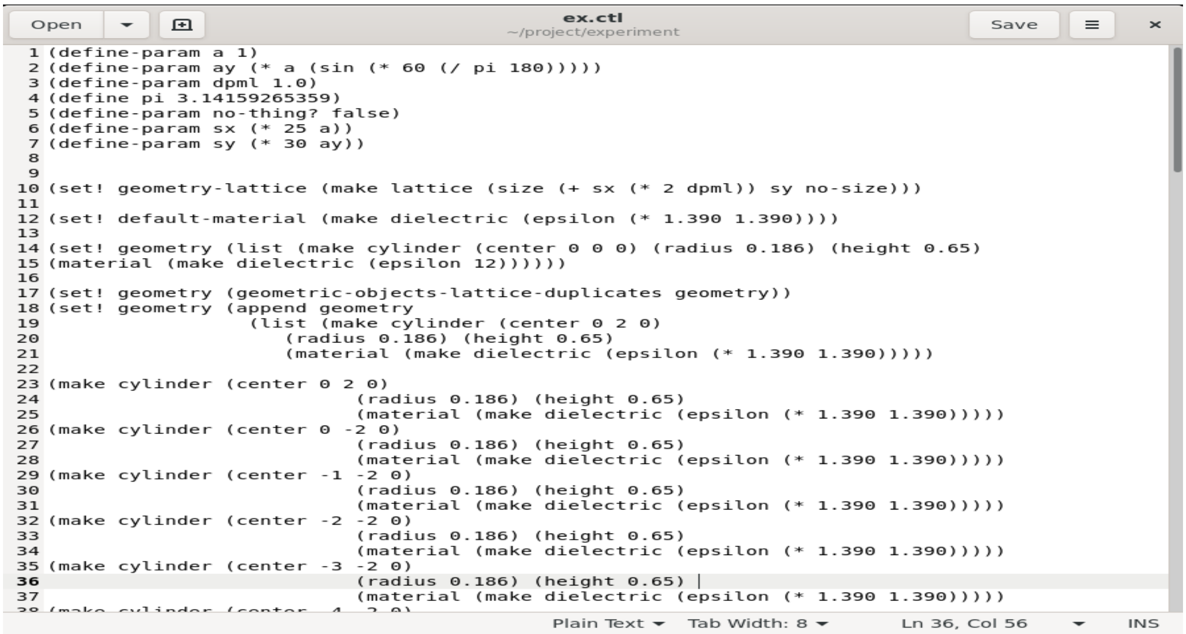

Figure A.6 Snapshot of MEEP code

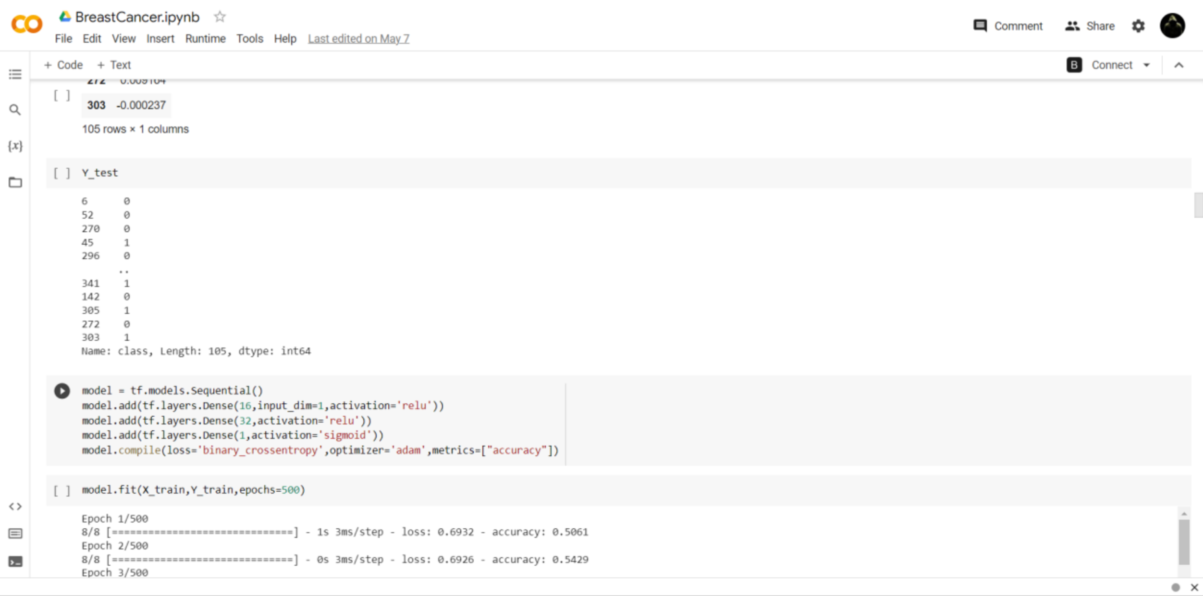

Figure A.7 Snapshot of ML code in Google Colab

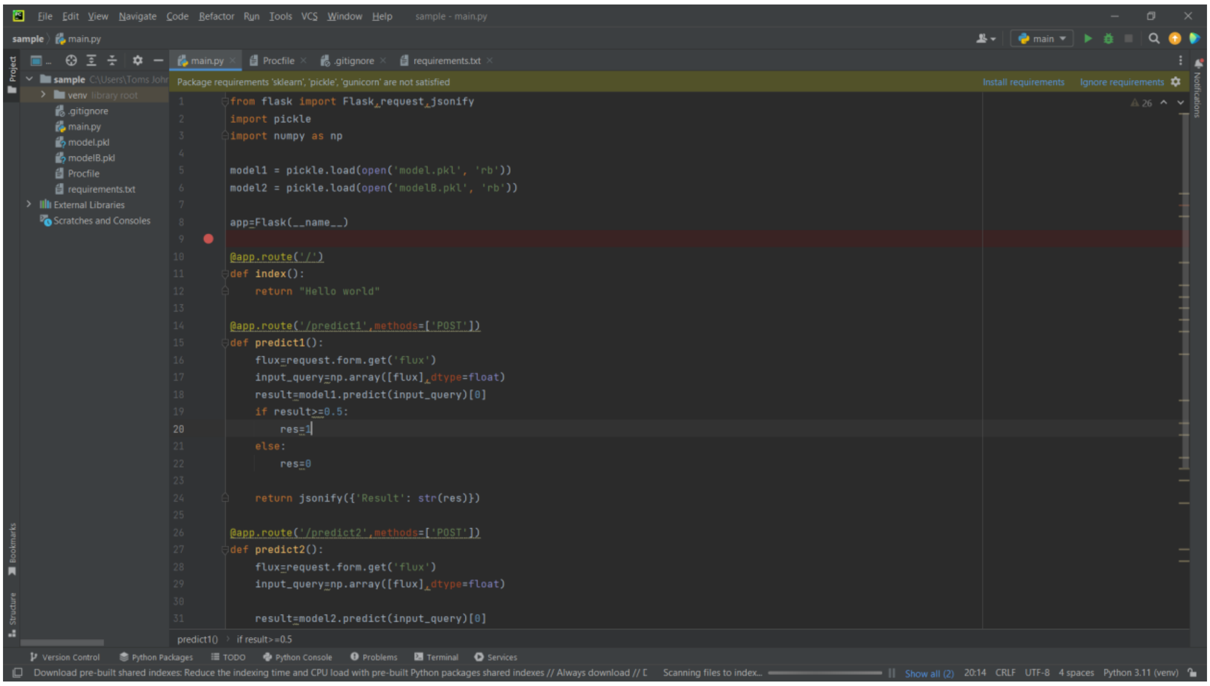

Figure A.8 Snapshot of API building using PyCharm IDE