Enhancing Blood Cancer Diognosis with Data Driven Techniques

Balaji J., Sanjai M., Hemadharshini V., Gunavathi R. and Gobisha K.
Department of Information Technology, Nandha College of Technology, Erode, Tamil Nadu, India

Keywords: Blood Cancer Classification, Deep Learning, resnet50 Model, Blood Cancer Detection.

Abstract: Blood is needed by the human body to carry essential nutrients, oxygen and immune cells. Leukemia and lymphoma are two deliverables that can be traced to abnormalities in blood cells. Blood cancers which include a range from leukemia to lymphomas and plasma cell diseases are among the most difficult to diagnose, yet early detection is critical for effective treatment and positive clinical outcomes. In this scenario, the invention of an integrated automated blood cancer detection system is essential. This pipeline classifies the blood sample pictures and classifies between Benign, Early, Pre and Pro cancer types using a pre- trained ResNet50 model. Images are pre-processed for it to comply with its input requirements, and then the model learns crucial aspects to predict accurately. The model produced the type of cancer and a confidence score, which is an indication of how likely the prediction is correct. The system developed using Keras as its deep learning framework and Streamlit as its user interface, provides a reliable and portable tool that effectively automates the picture processing procedure to aid with blood cancer identification. The implementation of these consequences allows doctors in clinics to diagnose faster and more accurately by providing consistent and accurate, classifications, while reducing manual work.

1 INTRODUCTION

Blood fills a major role in the human body, delivering oxygen, nutrients, and immune cells throughout organs and tissues. But in several blood cancers, such as leukemia, lymphoma and myeloma, mutations or uncontrolled growth in blood cells can occur. Early detection of these tumors is needed for effective treatment and improved survival rates. Diagnosing blood cancer, a labor-intensive process that is prone to human error, requires blood samples to be painstakingly reviewed under microscope. The rapid developments in the field of machine learning and image analysis have made it increasingly feasible to combine the automated identification and categorization of blood malignancies.

By reducing the requirement for manual tasks and facilitating better diagnostics, these technologies could help with the fastest, efficient, and accurate blood sample analysis. This method aims to assist medical professionals by providing a more accurate, more automated solution for detecting and classifying blood malignancies, through the use of deep learning models to classify images of blood samples. This will eventually result in better diagnostic outcomes and less costly treatments.

1.1 Blood Cancer Classification

Blood cancer classification is the process of recognizing different types of blood cancer through the study of blood characteristics. Blood malignancies, including leukemia, lymphoma, and myeloma, can have a myriad of presentations and specific pathologic features. Classification is the identification of those tumors which is based on the abnormalities present in blood cells, including size, shape, organization, etc. This step is critical in determining the most appropriate treatment plan, as well as assessing the severity of disease. Prompt and accurate categorization allows healthcare professionals to make informed decisions about the most effective therapies, which greatly increases the chance of well-timed success in any course of therapy. Use of classification also ensures identification of the stage or development of the cancer which is important for long-term care and judgment.

1.2 Deep Learning

Deep Learning is a powerful machine learning technique back up by the architecture of multi-layered

J., B., M., S., V., H., R., G. and K., G.
Enhancing Blood Cancer Diognosis with Data Driven Techniques.
DOI: 10.5220/0013872000004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 1*, pages 738-743
ISBN: 978-989-758-777-1

neural networks to allowing system to learn by itself and improve from data without given explicit instruction. The features can be derived from unprocessed input, for instance, text or pictures, which eliminates the need for human feature selection. Deep learning excels in uncovering intricate interdependencies within volumes of data that would inevitably elude conventional approaches. As an example, for blood cancer diagnosis, deep learning models such as convolutional neural networks (CNNs) are allowed to learn how to analyze images of blood samples and to detect subtle patterns indicative of cancer. These models are trained on high volumes of expert-labeled data and due to this, they evolve over a period of time and can identify and classify blood cancer types with high accuracy with minimal human intervention.

1.3 Resnet50 Model

The ResNet50 Model is a well-known model type of CNN a deep learning framework effective for computer vision tasks, particularly in image classification applications. Each layer of the model has 50 layers to help it learn hierarchies of patterns, allowing it to understand and capture complex components in photographs. One of the major aspects of ResNet50 is its use of residual connections, which allow the network to reduce degradation in performance as it deepens. Due to its ability to maintain great accuracy despite having deep architectures, ResNet50 is extremely useful for tasks that involve large and complex datasets. The ResNet50 model works exceptionally well in classifying Such medical pictures including blood samples by identifying the distinguishing characteristics like size, shape, and texture of blood cells.! Its ability to analyze huge amounts of data and spot complex patterns makes it an ideal candidate for automating blood cancer classification.

1.4 Blood Cancer Detection

Blood cancer detection focuses on studying abnormal activity of blood cells in order to diagnosis of various blood cancers. The process involves both traditional diagnostic techniques such as manual inspection of blood samples, as well as modern computer tools such as deep learning for automatic segmentation. The common way to detect blood cancer is through finding deviations from normal blood cell properties, such as atypical cell sizes, shapes or architectures that may indicate the presence of malignancy. Determining the most effective treatment strategies and enhancing patient outcomes depend on the fast and precise identification of blood malignancies. The detection process may be improved and expedited with the use of sophisticated models and algorithms, giving medical practitioners quicker, more precise findings that facilitate wise clinical judgment.

2 LITERATURE REVIEW

According to Md. Aslam Mollah et al, In order for a sensor to be able to measure the salinity of seawater, it has to have high sensitivity, structural simplicity, and durability, according to the study published in theJournal of Molecular Biology. In this work, an ultrahigh sensitivity PCF salinity sensor was proposed based on a sagnac interferometer (SI). The finite element method (FEM) is used to investigate the propagation properties of the suggested PCF. A sensitivity of up to 37,500 nm/RIU and 7.5 nm/% was achieved throughout the salinity of 0% to 100% obtained. The PCF proposed here achieves the highest resolutions of 2.66 × 10−06 RIU and 1.33 × 10−02% as well as good linearities with the length of 2.20 cm is inferior to PCF with other lengths value, up to 0.9924. Due to its remarkable results, this proposed sensor has the ability to detect salinity in seawater. Salt levels matter because they greatly affect activities beneath the sea and ocean species. But the whoifiée indicator of salinity is the electric conductivity of the chlorite ions. But this measurement is affected by the interference from other contamination ions. Research has been attracted due to the various benefits of fiber optic salinity sensors such as their programmable birefringence, small structure, remote sensing, tunable dispersion and immunity to electromagnetic interference (EMI). Here, a PCF salinity sensor based on the SI phenomenon is studied numerically. Every one of the PCF air gaps is believed to be filled with mixed salt concentrations of the seawater (M.R.B.A. Faysal, et., al. 2020).

Most surface plasmon resonance (SPR)- based photonic crystal fiber (PCF) sensors have been reported for detecting the analyte refractive index (RI) values between 1.33 and 1.41 (i.e.: human blood, body fluid), whereas rising state-of the-art detection techniques requires a more improved platform for the accurate detection of HMIs. Here we propose a way to mitigate this effect through a transformation of the structure, which can be achieved using the existing fabrication processes that the sensor architecture is built from. We present our sensor that uses an air-core PCF, and unlike existing PCF-based sensors,

the analyte enter the core of the PCF via vertical side opening channel to measure RI of analytes higher than that of the background of PCF. We employ a chemically stable plasmonic (gold) material, and since the plasmonic material is not directly contacted by the analyte, the interference effect is minimized. For the analyte RI of 1.42, the sensitivity, and resolution of the spectrum have been found to be 11,700 nm/RIU and 8.55 × 10−6 RIU, respectively. However, our proposed sensor still has the potential to catch active samples of these chemical and biological liquids. Various organizations have released several surface plasmon resonance (SPR) sensors in the past few years that detect analytes with a range of 1.33 to 1.41. Micro- fluidic slotted sensors, internal and external metal- coated PCF-based sensors, nanowire-based sensors, and D-shaped configuration-based sensors (A.K. Paul, et., al. 2020) were derived from it, resulting in five types of PCF-based SPR sensors.

Mohammad Al Mahfuz et al. In this work, we have proposed a dual-core photonic crystal fiber (DC-PCF)-based surface plasmon resonance (SPR) biocompatible sensor for the refractive index (RI) sensing of bio-organic molecules and biochemical analytes in the visible to near-infrared (0.5 to 2 μm) region. The sensor construction is easy with two hexagonal ring lattices all with round air aperture. The use of plasmonic material and an analyte detecting layer on the outer surface of the fiber allows practical applications to be made. Gold (Au) with 30 nm of thickness, a noble plasmonic material, is utilized to excite the plasmons on the surface. It is also suggested that a thin layer (~5 nm) of TiO2 (titanium oxide) acts as an interlayer cementing the Au and the silica glass. Based on the mode solver the finite element method (FEM) is utilized to investigate the sensor response. Using both amplitude and wavelength interrogation methods, numerical findings reveal an optimal wavelength sensitivity (WS) of as high as 28000 nm/RIU, an optimal amplitude sensitivity (AS) of 6829 RIU−1, an optimal amplitude resolution (AR) of 5x10−6 RIU, and a wavelength resolution (WR) of 3.57x10−6 RIU for the proposed sensor. Moreover, such a PCF-SPR sensor, having 2800 RIU−1 as the highest FOMpeak value, is currently the starkest sensor (M. Al Mahfuz, et., al 2020).

In this work a novel technique named as 2-D photonic crystal waveguid (PCW) based cell detection has been proposed by Abinash Panda et al. for identifying the nature of either normal or malignant cells. The proposed metamaterial is designed with 5 × 5 silicon-based rods on a square lattice with a central defect and air as the background. To correctly sense, we classify two sets of live cells: Group I: Malignant Cell (YD-10B); Group II: Normal Cell (INOK). Properly adjusting the plane wave expansion (PWE) method, the electric field distribution and the peak reflected wavelength have been achieved in the designed PCW structure. Accurate identification of normal and malignant cells, a large number of structural parameters, including but not limited to lattice spacing, circular rod diameter, and backdrop material type, are needed. This MATLAB simulation indicates that yellow color (i.e. reflected wavelength) belongs to cancerous cells and orange color (i.e. reflected wavelength) corresponds to healthy cells. Similarly, the negative dispersion coefficient, scattering loss, and the nonlinear coefficient of the proposed structure are precisely evaluated on normal and malignant cells separately. Moreover, this proposed sensor has a high sensitivity of 2360.12 nm/RIU, low resolution of 1.78 × 10−6, and high-quality factor (as high as 99.765) when differentiating normal and malignant cells (P.P. Devi and A. Panda, 2020).

Chunlian Cen et al. Here, we propose to use critical coupling and impedance matching theory to computationally simulate the perfect absorption of monolayer graphene. We studied a perfect single-band absorption of the monolayer graphene by using the important coupling effect and impedance matching. Errors from the data fittings remained within 10%, leading to a high quality factor (Q-factor = 664.2) absorption spectrum with the absorbance ~100% in the near-infrared range. The position of the absorption spectrum can be adjusted by changing the ratio of the air hole radii of the elliptic cylinder to structural period. The attained S = 342.7 nm/RIU (refractive index unit) and FOM = 199.2 (figure of merit) could be achieved by an absorber, which shows great potential for biosensor technology development. We hope that our research can serve as an interesting application for graphene photonics and optoelectronics. Plasmon metamaterials have recently been a hot topic due to their unique EM (electromagnetic) control capability. As such, it is currently the most studied material due to its optical and physical properties. The scope of electromagnetic metamaterials has gone from being limited to microwave frequency range to terahertz, infrared, and almost the entire visible light electromagnetic spectrum as illustrated by the study (Jiang L et., al.2020) Metamaterial absorbers, on the other hand, have been shown to improve absorption in solar, microwave, infrared, and optical systems.

3 EXISTING SYSTEM

To realize early detection of blood cancer, the current paper proposes twin-core photonic crystal fiber (TC-PCF), which involves the refractive index (RI) of healthy and cancerous blood cells. Normal and cancerous cells are thought to be 30 to 70 percent liquid, according to what is known. Because the middle air hole is longer than the other two, these samples should be infiltrated at that spot. Fig. 1: Schematic setup of the known TC-PCF for the normal and malignant cells, using a two-dimensional finite element method (FEM) to monitor the variations of the transmitted spectrum and coupling length while considering the refractive index (RI). As per the transmitted spectrum shift, the proposed sensor may show a high sensitivity (8571.43 nm/RIU). Due to the straightforward detection strategy, the suggested TC-PCF sensor can potentially be applied to detection of blood cancer in a convenient and cost-effective way.

4 PROPOSED SYSTEM

This technique combines Convolutional Neural Networks (CNN) with ResNet architecture to enhance the analysis of blood cell images for improved accuracy in identifying anomalies, including blood cancers. The integration of residual learning into ResNet ensures efficient information flow by using shortcut connections to avoid problems such as disappearing gradients. The classification is made through multiple convolutional layers that help extract features such as size, form and texture that help distinguish between healthy blood cells and malignant ones. Using pooling and fully connected layers, those features are then merged to obtain a proper classification. Honed for processing efficiency, the system handles large, diverse datasets and is flexible enough to perform a variety of analytical tasks. Providing a consistent, automated way to analyse medical images, it has the potential to be used in multiple disease settings because of its scalability. Automated blood smears analysis has improved significantly with this state-of-the-art technique, increasing diagnostic fidelity and enabling accurate analysis, diagnosis, and prediction.

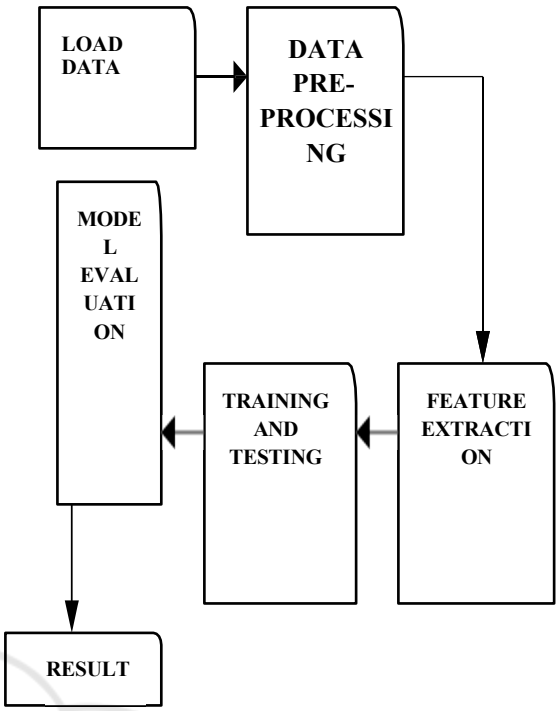

Figure 1: System Flow Diagram.

4.1 Load Data

The Load Data Module is the first step and it is used to collect and prepare the dataset for further operations. This module is responsible for recovering images, which are tagged with the type of blood cancer or whether it is non-cancerous. To facilitate organizing and retrieving the pictures, they are organized systematically This module ensures that the inputs are ready in such a fashion and structure, that the system can use it appropriately in the subsequent workflows, which is training the data and testing the data.

4.2 Data Preprocessing

The data preprocessing module's objective is to clean and transform the raw data into a form that could be used as input in the model. This means that all the pictures have to be resized to a standard size, their pixels have to be transformed to the same range, as well as performing all the other required manipulations to get the data compatible with the model. The objective of this module is to obtain better quality data so that noise and inconsistencies are discarded, and high-level structured data is clear and organized for the best possible outcome during the analysis and learning stages.

4.3 Feature Extraction

This means that it is the job of the Feature Extraction Module to find and select features of the blood sample photos that are relevant to classification. This module utilizes the various textures, patterns, and shapes found in the images to recognize and classify distinct types of blood cancer. These important features can be automatically extracted by deep learning technique types of computations including convolutional neural networks (CNNs). This module not only simplifies the data but focuses well on the important portions of the pictures, while also helping the model to figure, and, classify the pictures accurately.

4.4 Training and Testing

The Module for Training and Testing incorporates the learning process of the model. During the training phase, the model learns to recognize patterns in the photos based on labeled data. It adjusts its internal parameters to minimize errors and improve predictions. An independent set of data is used to evaluate the model's accuracy and generalizability after training. This ensures that the model can accurately classify new and unseen images, predicting how well it will perform when used on other datasets in the future.

4.5 Model Evaluation

The fine-tuning phase of the Model Evaluation Module determines the performance of a trained model and tests its accuracy. After training/testing metrics like F1-score, accuracy, precision, recall are used to determine the model performance. Such metrics assess the models ability to classify the benign and malignant samples and the types of blood cancer. Make sure that the final model is trustworthy and gives consistent quality predictions by taking a look at the output of the model and to find out if improvements can be made.

5 RESULT ANALYSIS

The system's ability to detect and classify different types of blood cancer using picture data is demonstrated in the analysis of the project's outcome. Performance evaluation exhibits how the ability of the model to accurately predict benign vs malignant samples and classify them under certain categories as early, pre or advanced phases. Metrics of the classification process, such as accuracy, precision, recall, and F1-score, help ensure that the feature extraction and model training modules have successfully identified key patterns that help in the human defects. The outcomes, which demonstrate pleasingly high degrees of accuracy and consistency in predictions, confirm the approach of analyzing blood sample images through deep-learning algorithms. This review illustrates how one such initiative measured up against its goals and shows how it might aid in diagnosis of early and precise blood cancer type. Accuracy for the existing system and proposed system are tabulated in table 1 and illustrated in figure 2.

Table 1: Comparison Table.

Algorithm	Accuracy
Existing system	77
Proposed system	85

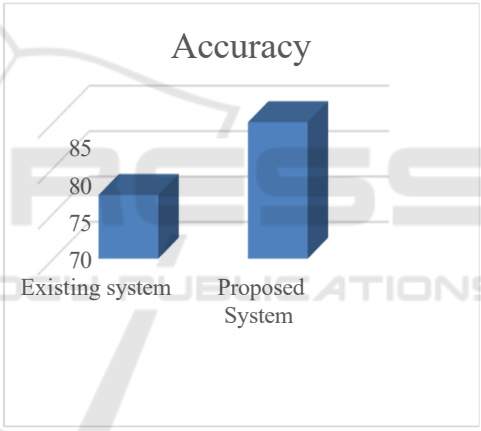

Figure 2: Graph Diagram.

6 CONCLUSIONS

In conclusion, this study provides an approach to employ state-of-the-art machine learning algorithms to diagnose and classify blood cancer using image data. By processing the data in a systematic manner, they were able to extract relevant features, train the model and classifier such that the blood samples were able to identify the visual patterns in the images. Model evaluation ensures accuracy and reliability of the model. This method shows how successful machine learning will be impactful in the medical industry by providing a helpful tool to assist the diagnosis and detection of blood-related disorders. The method is a significant advance in medical

analysis as it automates the identification process and produces reliable results.

Future Work.

To increase the generalizability of the model, this study can be further improved in the future by including more heterogeneous datasets including samples from other demographics and a wider spectrum of blood cancer types. Adding in data from additional sources, such as genetic markers or clinical test results, could potentially improve the predictive power of the system. Further studies could potentially focus on enhancing feature extraction techniques; perhaps by utilizing highly advanced deep learning architectures ormethods such as transfer learning to bolster model performance. Also, we hope to reduce the computation complexity and to tune the model for faster training steps to scale up the system for more applications. Another potential future direction is to automate the process itself for various blood-related diseases. This would make the system more useful than just cancer detection and ultimately facilitate more accurate and efficient diagnosis in clinical environments.

REFERENCES

A.K. Paul, M.S. Habib, N.H. Hai, S.A. Razzak, Air-core photonic crystal fiber-based plasmonic sensor for high refractive index sensing, Opt. Commun. 464 (2020) 125556.

Cen C, Chen X, Zhang Y, Yang H, Yi Z, Yao W, et al. A two-band metamaterial absorber for surface plasmon resonance at terahertz in graphene. 2020; Physica E. 10, 117:113840.

Dai, W., Chen, F., Luo, H., Xiong, Y., Wang, X., Cheng, Y., and Gong, R. are the authors of this work. For enhanced microwave absorption capabilities, carbonyl iron@void@ nitrogen-doped carbon with a yolk-shell structure is created. J. Alloy Compd. 2020, 812, 152083.

Jiang L, Xu D, Yi Z, Chen Z, Cen C, and others. Graphene is a high quality factor, high sensitivity metamaterial that is the perfect absorber due to critical coupling theory and impedance matching. Nanomaterials, 2020, 10:95

M. Al Mahfuz, M.A. Hossain, E. Haque, N.H. Hai, Y. Namihira, and Ahmed A plasmonic RI sensor with a dual-core photonic crystal fiber that operates in the visible to near-infrared spectrum, In IEEE Sensors Journal (2020), https://doi.org/10.1109/JSEN.2020.2980327

M.R.B.A. Faysal, M.R. Hasan, M.B. Sagnac interferometer-based photonic crystal fiber salinity sensor with high sensitivity, Results Phys. 16 (2020) 103022,

Miao, C.; Fang, R.; Mou, H.Y.; and Xiao, W. Sandwich-like Si@TiO2@rGO composites are easily synthesized for application as high-performance anodes in lithium ion batteries. J. Alloy, Compd. C 2020, 818, 152884. [Source]

P.P. Devi and A. Panda. Refractive index-based photonic crystal biosensor for the identification of cancerous cells, Opt. Fiber Technol. 54 (2020) 102123,

Z. Fan, Z. Guo, X. Kong, and Z. Meng, Wide- range refractive index sensor based on photonic crystal fiber with phase matching between metal defect mode and core mode, Opt. Commun. (2020) 125233.

Zhang, C.F.; Xu, D.Y.; Wu, P.H.; Chen, Z.Q.; Jian, R.H. Graphene-Based Unpatterned Narrow Dual-Band Monolayer Perfect Absorber with Near-Infrared Critical Coupling. Micromachines 11, 58. 8, 2020.