Epigenetically Reprogrammed Intelligent Cancer Biomarker for Cancer Detection

Dr. Bai V. Bhagirathi ♣, Tejas Deepak Hundre ♣, Magdum Sunad Sanmati ♣, Aditya Ganapati Nimbalkar ♣, M. Rakesh ♣
Mechatronics Engineering Acharya Institute of Technology Bangalore, India

ABSTRACT:
This paper presents a novel approach to leukemia detection in blood samples using advanced technologies such as medical imaging and machine learning. Leveraging Convolutional Neural Networks (CNNs) and the MobileNetV2 architecture, the study develops an integrated predictive model trained on a diverse dataset of blood smear images. Hardware integration, particularly utilizing Raspberry Pi, facilitates efficient image processing and analysis. Rigorous testing ensures the system's accuracy and reliability, offering advantages in enhanced diagnostic accuracy, speed, and user-friendly interface. Despite challenges such as data variability and ethical considerations, the proposed approach demonstrates promise for improving patient outcomes and healthcare efficiency.

Keywords: Cancer detection, Artificial Intelligence (AI), Machine learning (ML), Mobile.netV2, Intelligent system, Image segmentation, Conventional neural network (CNN).

INTRODUCTION
Cancer, particularly blood cancer, poses a significant global health concern, affecting individuals of all ages, especially children [1]. Leukaemia, lymphoma, and myeloma are common forms of blood cancer, each presenting unique challenges. [2] For instance, in leukemia, excessive production of white blood cells by the bone marrow compromises the immune system and decreases red blood cell and platelet levels. [1] There are various types of leukemia, including Acute Myelogenous Leukaemia (AML), Acute Lymphocytic Leukaemia (ALL), Chronic Myelogenous Leukaemia (CML), and Chronic Lymphocytic Leukaemia (CLL), each requiring distinct treatments. [1] Medical imaging techniques such as MRI, CT-Scan, and blood smear analysis play a pivotal role in diagnosing and classifying different tissues to aid in accurate treatment. [4] Automatic diagnosis systems utilizing machine learning (ML) techniques have emerged to expedite diagnoes and improve treatment outcomes. [8] Recent advancements in ML-based medical image analysis (MIA) have focused on leveraging techniques like Support Vector Machines (SVM), Artificial Neural Networks (ANN), and deep learning (DL) to classify diseases and assist healthcare professionals in identifying abnormalities.

[5] The application of microscopy image processing and computer-assisted pathology in cancer diagnosis has gained traction, particularly due to its ability to reduce errors and processing time. [6] Image segmentation methods play a crucial role in analysing blood cells to aid in the identification of cancer patterns, such as in cases of Leukaemia, characterized by abnormal proliferation of white blood cells in the bone marrow, lymph nodes, and blood. [6-7].
LITERATURE SURVEY

The research in medical imaging revolves around the segmentation and classification of images to extract diseases and areas of interest. Convolutional Neural Networks (CNNs) have emerged as a popular choice due to their robust classification capabilities. For instance, the utilization of CNN-based models for breast cancer diagnosis [4], also reported superior performance of CNNs over handcrafted feature-based methods in lung sound classification [9]. Similarly, the proposal of an optimal CNN for early detection of skin cancer. In the context of classifying cancerous and non-cancerous lymphocyte cells, studies often involve a two-step process: feature extraction followed by classification. [13] For example, the development of a computer-based method for acute lymphoblastic leukemia detection using image pre-processing, feature extraction, and SVM classification. [14,12] Similarly, a proposed quantitative microscopic approach for ALL screening, achieves high accuracy with multiple classifiers. [18]

Recent advancements in medical image processing include transfer learning and efficient CNN frameworks to address issues like small databases and low-dimensional data. The employment transfers learning-based models with ResNet50 architecture, while also introducing depth-wise separable convolutions and linear bottleneck architecture for improved classification accuracy in low-dimensional datasets. [16,8] Additionally, there are also proposed machine learning-based approaches to enhance blood sample images, achieving high accuracy through adaptive un-
sharpening and image normalization techniques. [19] Further transfers of histopathological transfer learning for improved ALL detection in limited dimensional databases.[20]

Prior research underscores the importance of diverse datasets for improving model generalization. Comparative studies have explored the performance of classification models using various deep learning architectures like ResNet and DenseNet. Additionally, there is emerging interest in extending model applicability to other blood cancer types beyond acute lymphoblastic leukemia (ALL) and multiple myeloma (MM).

The proposed approach, achieving up to 96.66% classification accuracy, leverages MobileNet V2, a CNN architecture, which contributes to its uniqueness among CNN algorithms. Its high accuracy and precision are further enhanced by training on a diverse dataset, ensuring robust detection capabilities.

OBJECTIVES

Certain objectives are as follows

- To Design and Develop a Hardware Intelligent Predictive Kit for Cancer Detection.
- To Train and Test the Cancer Detection Kit for Accurate Results.
- To Integrate Hardware and Intelligent Software for Easy and Fast Cancer Detection.
- To Develop A User-Friendly Interface for Easy and Fast Cancer Detection.

METHODOLOGY

This approach outlines the proposed method for detecting leukemia in blood images using Raspberry Pi as an image processor. It is divided into two main sections: software design and hardware design. This software includes steps such as preprocessing the images, extracting relevant features, training the machine learning model, and evaluating its performance. In the hardware design section, there are three main components: a controlling system, an imaging system, and a fully-fledged model ready for implementation. These components work together to capture and process blood images using the Raspberry Pi platform.

Training the MobileNetV2 Model

The project trained a MobileNetV2 Convolutional Neural Network (CNN) model using a meticulously curated dataset of approximately 10,000 images, encompassing both malignant and non-cancerous cases. This dataset was carefully annotated and pre-processed to ensure quality and consistency. Leveraging the MobileNetV2 architecture's suitability for resource-constrained devices like the Raspberry Pi, the trained model aimed to accurately distinguish between malignant and non-cancerous cases in medical images, facilitating cancer diagnosis and screening, particularly in settings with limited access to specialized expertise.

Developing the Integrated Predictive Model

Following training, the project integrated the MobileNetV2 architecture with the acquired dataset to develop an integrated predictive model. Through fine-tuning and transfer learning techniques, the pre-trained MobileNetV2 model was optimized to discern features indicative of cancerous cell morphology. Rigorous validation procedures confirmed the model's capability to reliably identify cancerous cells within blood smear images. This integrated predictive model emerged as a robust tool for blood cancer diagnosis and screening, offering valuable insights for informed decision-making in patient care.
Designing the User Interface

The project transitioned to designing a frontend user interface (UI) using HTML and CSS, aiming for an intuitive and user-friendly experience. The UI featured a welcoming pop-up display and a seamlessly integrated form for inputting patient details, streamlining access to essential information for analysis and decision-making.

Integrating Frontend and Backend Components

Integration of frontend and backend components was pivotal for cohesive system functioning. This involved synchronizing the MobileNetV2-based blood cancer detection model with the HTML and CSS-based frontend interface, ensuring effective communication and usability.

Electronic Assembly

The electronic assembly methodology involved selecting and integrating cutting-edge hardware components with advanced image processing techniques to facilitate efficient leukemia detection in
blood samples. From hardware selection to software integration, rigorous testing and validation ensured the functionality, accuracy, and reliability of the proposed diagnostic system, offering an efficient solution for leukemia detection in blood samples.

**WORKING PRINCIPLE**

**Acquisition of digitized peripheral blood swarm images**

Standard microscopy techniques are employed to acquire digitized peripheral blood smear images.

**Preprocessing of Images**

The acquired images undergo preprocessing to enhance quality, including noise reduction, contrast enhancement, and image normalization.

**Features Extraction**

Advanced image processing algorithms are utilized to extract features from the preprocessed images, capturing information about cellular morphology, texture, and spatial arrangement crucial for leukemia detection.

**Machine Learning Model**

Features extracted from the images are fed into a machine learning model, leveraging transfer learning with the MobileNetV2 pre-trained model. The model is trained on a diverse dataset of labeled blood smear images, encompassing both normal and leukemia-affected samples.

**Hardware Integration**

A Raspberry Pi controller and touchscreen display are integrated into the system, serving as the computational unit and user interface, respectively. The Raspberry Pi executes image processing algorithms and runs the machine learning model, while the touch screen display facilitates interaction and result visualization.

**Result Presentation**

Predictions generated by the model are displayed on the touch screen display in a clear format, including details about the leukemia stage, the percentage of affected cells, and any abnormal cell morphology detected.

**Testing and Validation**

Rigorous testing and validation procedures are conducted to ensure the system's functionality, accuracy, and efficiency.

**Overall Impact**

The proposed system offers a robust and efficient solution for leukemia detection in blood samples, with the potential to significantly improve diagnostic capabilities and patient outcomes.

**Table 1. Comparison Table**

<table>
<thead>
<tr>
<th>SL.NO</th>
<th>METHODS</th>
<th>DATA SET</th>
<th>ACCURACY IN %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CNN [23]</td>
<td>All-IDB</td>
<td>88.25</td>
</tr>
<tr>
<td>2</td>
<td>SVM with cell energy feature [24]</td>
<td>All-IDB</td>
<td>94</td>
</tr>
<tr>
<td>3</td>
<td>Gao-based method [25]</td>
<td>All-IDB</td>
<td>93.84</td>
</tr>
<tr>
<td>4</td>
<td>Squeeze based CNN [26]</td>
<td>All-IDB</td>
<td>98.3</td>
</tr>
<tr>
<td>5</td>
<td>Mobile-Net V2 (proposed)</td>
<td>All-IDB</td>
<td>96.66</td>
</tr>
</tbody>
</table>
Table 2. Cross Matrix

<table>
<thead>
<tr>
<th>Case Number</th>
<th>Patient Data Collected</th>
<th>Image uploaded</th>
<th>Cancer Predictions</th>
<th>Percentage Affected</th>
<th>Cancer Stage</th>
<th>Original Image</th>
<th>Recent Patient Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>Yes</td>
<td>Malignant</td>
<td>85%</td>
<td>Stage 3</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
<td>Benign</td>
<td>20%</td>
<td>Stage 1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>Yes</td>
<td>Yes</td>
<td>Malignant</td>
<td>95%</td>
<td>Stage 4</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>Yes</td>
<td>Yes</td>
<td>Benign</td>
<td>15%</td>
<td>Stage 1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>Yes</td>
<td>Yes</td>
<td>Malignant</td>
<td>70%</td>
<td>Stage 2</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

USER INTERFACE

User ID and Password Authentication: Users authenticate their identity through a unique user ID and password combination, providing an initial layer of security.

Patient Data Input: Once authenticated, healthcare professionals can input patient data seamlessly. The intuitive interface guides users through entering essential demographic information, medical history, and contact details.
Image Upload: Users can effortlessly upload medical images for analysis. The system supports various image formats and ensures smooth uploading for diagnostic purposes.

Result Display: After data input and image upload, users gain access to the complete report, which includes detailed patient data and prediction results. The report offers insights into cancer likelihood, affected percentage, and cancer stage, aiding clinicians in making informed decisions.

Additional functionalities include:

- **Print Report:** Users have the option to print the complete report directly from the interface. Three dedicated buttons offer flexibility in printing options to suit user preferences.
- **Test Once Again:** Users can initiate a new diagnostic process by returning to the patient data input stage, enabling seamless testing for multiple patients or scenarios.
- **Logout:** Upon completion of tasks, users can log out securely. The logout function redirects users to the authentication page, ensuring privacy and security.

This comprehensive workflow streamlines the diagnostic process, prioritizing security, usability, and efficiency for healthcare professionals using our software.

**TESTING**

A cross-matrix is a tabular representation of data that shows the frequency or proportion of occurrences for different combinations of categorical variables, aiding in analyzing patterns and relationships between them. It allows for the systematic comparison of variables, facilitating insights into dependencies and correlations within datasets.

The cross-matrix table offers a comprehensive overview of a predictive cancer diagnosis system, detailing ten distinct scenarios encountered during its application. In the first case, both patient data and an image were uploaded, resulting in a diagnosis of cancer with an 85% certainty level and identification of Stage 3 cancer. Conversely, the second scenario demonstrates the inability to generate a cancer prediction when patient data was collected but no image was uploaded, highlighting the crucial role of image analysis alongside patient data. Similarly, in the third case, where only an image was uploaded without patient data, a cancer prediction could not be provided, emphasizing the necessity of holistic data collection. In contrast, case four showcases a successful diagnosis when both patient data and an image were available, leading to a benign cancer diagnosis with a 20% certainty level and identification of Stage 1 cancer. Subsequent cases demonstrate variations in data availability and diagnosis outcomes, reinforcing the importance of comprehensive data inputs for accurate predictions. Overall, the scenarios underscore the system's reliance on comprehensive data inputs, including both patient data and image analysis, to ensure effective diagnosis and treatment planning for cancer patients.

**RESULTS**

![Figure 9. Microscopic Image Inputs](image-url)
The findings of our research effort show encouraging progress toward the development of an intelligent cancer biomarker kit. Using cutting-edge transfer learning methods, our system achieves an astonishing 96.66% cancer prediction accuracy rate. Figure 9 depicts the microscopic input images used for testing, demonstrating the diversity of the input data.

Figures 10 and 11 show the predicted outcomes, which distinguish between cancer-detected and non-detected patients. Using modern imaging techniques and predictive analytics, our technology identifies malignant tissues and anomalies with high sensitivity and specificity.

Figure 12 depicts the user interface for prediction results in the form of a detailed report that includes patient data. This user-friendly interface provides healthcare practitioners with deep insights into forecast outcomes, supporting informed decision-making and personalized treatment strategies.

Our system's robustness and reliability in real-world circumstances are clearly demonstrated by rigorous examination against benchmark datasets. Finally, our findings represent a substantial advancement in oncology diagnostics, providing a promising answer for improving patient care and treatment regimens. Our methodology's rigorous validation and evaluation highlight its efficacy and potential for wider clinical adoption, with the ultimate goal of positively impacting cancer care and patient outcomes.

ADVANTAGES

1. Enhanced Diagnostic Accuracy: Utilizing AI-driven algorithms, the project aims to significantly improve the accuracy of blood cancer detection by analyzing microscopic images. This could potentially lead to early detection and more precise diagnoses.

2. Speed and Efficiency: The integration of machine learning models allows for quicker analysis of blood smear images compared to traditional manual methods. Rapid processing could expedite diagnosis, aiding in timely medical interventions.

3. User-Friendly Interface: The development of an intuitive and engaging frontend interface makes the application more accessible and user-friendly. This can facilitate ease of use for healthcare professionals and patients alike.

4. Compact Deployment: Choosing the MobileNetV2 architecture for the CNN model allows for
its deployment on compact devices like the Raspberry Pi. This portability enhances the system's feasibility for implementation in various medical settings.

**DISADVANTAGES**

1. **Data Variability:** Challenges related to variability in staining procedures, resolutions, and morphologies among datasets might affect the model's generalization and accuracy. Addressing these variations could be complex and time-consuming.

2. **Ethical Concerns:** Utilizing patient data for training models raises ethical and privacy concerns. Ensuring compliance with strict regulations and maintaining patient confidentiality is crucial but challenging.

**CONCLUSION**

The completion of this project marks a significant milestone in the field of cancer detection and diagnosis. From its inception, the primary objective was to develop an innovative solution capable of revolutionizing the way cancer is identified and treated. Through meticulous planning, rigorous research, and relentless dedication, the team has successfully achieved this goal.

Beginning with the conception of the project, extensive market research was conducted to identify gaps and opportunities in existing cancer detection methods. This initial phase laid the foundation for the development of the Epigenetically Reprogrammed Intelligent Cancer Biomarker Kit. The kit's design was meticulously crafted by a team of expert engineers, drawing upon cutting-edge advancements in mechatronics and biomedical engineering.

Central to the project's success is the novel algorithm developed for analyzing biomarker data obtained from patient samples. Leveraging machine learning and data analytics techniques, the algorithm can accurately identify cancer biomarkers with unprecedented precision. This breakthrough not only enhances diagnostic accuracy but also enables personalized treatment strategies tailored to individual patients' needs.

Furthermore, the project's innovative approach extends beyond diagnostic capabilities to encompass predictive analytics for cancer prognosis. By analyzing historical patient data and correlating it with treatment outcomes, the algorithm can forecast disease progression and recommend optimal treatment regimens. This proactive approach empowers healthcare professionals to intervene early, potentially improving patient outcomes and survival rates.

**REFERENCES**


