Multidisciplinary Journal of Healthcare
Vol: 1 Issue: 2, 2024

# "Early Diagnosis of Blood Cancer Using Deep Learning: An Evaluation of CNN, SVM, and Random Forest"

**Abdul Qayyum Waseem Khan[1], Amir Ali[2], Abdul Rehman[3], Imran Arshad[4], Mahad Shahid[5]**

**[1,2,3,4,5] BS Computer Science Lahore Lead University**

## INDEX

Multidisciplinary Journal of Healthcare
Vol: 1 Issue: 2, 2024

**Abstract**
Blood cancer is one of the most dangerous and fast-growing types of cancer, affecting people of all ages, especially children. Early diagnosis is critical for effective treatment and improving survival rates. Traditional methods of detecting blood cancer—such as bone marrow tests and manual image analysis—are time-consuming, painful, and not always accurate. This research focuses on using artificial intelligence (AI), particularly deep learning and machine learning techniques, to improve **early detection of blood cancer** through microscopic blood smear images.
We evaluate and compare the performance of three widely used models: **Convolutional Neural Networks (CNN), Support Vector Machines (SVM),** and **Random Forest (RF).** Our approach includes image preprocessing, feature extraction, and classification. CNNs were found to be the most effective, achieving up to **99.12% accuracy** using ensemble models like **DenseNet, Inception,** and **Xception**. SVM and Random Forest also delivered strong results, especially when used with deep feature extraction techniques.
The findings of this study show that AI-powered models can significantly improve the speed and accuracy of blood cancer diagnosis. This research not only supports the global healthcare goal of early cancer detection but also opens the door for developing smart, low-cost diagnostic tools that can be used in real-world clinical settings. Future work will focus on real- time applications, larger datasets, explainable AI, and integration into healthcare systems.

**Early Diagnosis of Blood Cancer Using Deep Learning: An Evaluation of CNN, SVM and Random Forest**

**Introduction:**

Diseases are more powerful and part of our lives nowadays, Cancer is one of the common growing disease included Lungs cancer, Breast Cancer, Blood Cancer & many more. Manually Detection and Diagnosis process of these type of diseases are very lengthy, complicated and time taken. Blood Cancer also most frequent type of cancer found in patients. In this article we accumulate how AI helps us to determine the disease in early stage to prevent the damage.

Disease needs to be diagnose in early stages and manually or using old techniques it is difficult and time taken which is not a good method. But AI in this regards perform a major role to help out diagnoses process more efficient with high accuracy. Where you can save one person rather save hundreds of them are the need of world.

Blood Cancer like Leukemia and lymphoma had remarkably high mortality rate. According to statistics from 345,000 in 1990 to 518,000 in 2018, the number of newly diagnosed cases of leukemia increased, lowering the Annualized Survival Insusceptibility Rate (ASIR) by 0.43% per year [1]. Which is a major concern to any disease growing so rapidly. According to US Government's Cancer Data Collector department there were 60,650 cases diagnosed cases of leukemia and 24,000 deaths occurred in the year of 2022.

It is a major concern that leukemia is mostly diagnosed in children. According to the report of National Cancer Registry of Pakistan: First Comprehensive report of Cancer Statistics 2015-2019. In children 'Leukemia' 1626 (14.50%) and in adolescents 'Bone' 880 (14%) were the leading malignancies. [2]

**Why it is Important:**

Youth are the future of the world and as per stats blood cancer cases mainly diagnosed in children. If this rate of detection cannot be stopped than future of world is in danger. Early stages of the disease can be controlled or completely cure but if not than the results will be bad. Using AI in diagnosis improve the accuracy and performance as it detects the faults more accurately and saving time also with covering more community compared to other methods. For the safety of world future, it is important to take this disease serious.

Blood cancer is a life-threatening serious illness that disrupts the normal function of blood cells. It has three main types: leukemia, lymphoma, and myeloma each affecting different parts of the

immune systems.

1. **Leukemia**: The most common cancer in children, starts in the bone marrow and spreads through the blood, causing abnormal growth of white blood cells.
2. **Lymphoma**: Affects the lymph nodes, which are part of the immune system.
3. **Myeloma**: Harms and damages the white blood cells and weakening the body's infection-fighting ability.

Every year, many people are diagnosed with blood cancer. For example, leukemia makes up about **3%** of all cancer cases worldwide with over **474,500 new cases** reported annually. The good news is that treatment works well, especially for children, with some types increase **5-year and 90%** survival rate.

Currently, doctors use old methods to diagnose blood cancer, such as testing abnormal blood cells. painful bone marrow tests. scans to check how far the cancer has spread. These methods are not always reliable, can take a lot of time, and may be uncomfortable for patients. Finding blood cancer early is very important because treatment works better in the early stages.

Because of these problems, we urgently need better ways to diagnose blood cancer that are faster, cheaper, and don't require complex and painful methods. Artificial intelligence is now learning how we detect blood cancer, making the process faster and more accurate. AI may even reduce the need for painful bone marrow tests. Also, AI is using different deep learning algorithms **Convolutional Neural Networks (CNN)** which is image-based disease detection, and machine learning algorithms **Support Vector Machines (SVMs) and Random Forests (RFs)** have been widely used for the classification of tasks in medical data.

This research helps match the World Health Organization's (WHO) goal to find cancer early and make treatment available to everyone by 2030. Using AI to detect blood cancer could lead to even better tools in the future that combine blood tests and genetic data for more accurate diagnosis.

This research proposes a dual-modality machine learning approach to improve early detection and classification of blood cancer by analyzing both patient symptoms and microscopic blood images. The core focus lies in leveraging traditional machine learning classifiers and deep learning models to predict not only the presence of blood cancer but also its specific type, aiming to enhance diagnostic accuracy and support timely clinical decision-making.

Provides background on blood cancer types and challenges in early diagnosis. Emphasizes the potential of machine learning in medical diagnostics and introduces the goal of developing an integrated model using symptoms and blood images.

Reviews prior studies that employed **machine learning** (ML), **deep learning** (DL), and image processing techniques in cancer detection. Highlights the performance of different models and sets the stage for the current study's contribution.

*"What if a small, low-cost device could help doctors detect a dangerous disease before it becomes life-threatening?"*

**Deep Venous Thrombosis (DVT)** is a serious medical condition where blood clots form in deep veins, usually in the legs. If not detected in time, these clots can travel to the lungs and cause a deadly problem called a pulmonary embolism. The challenge is that DVT often shows no clear symptoms, making it hard to diagnose early. Traditional tests like ultrasound are helpful but expensive and not always available in all hospitals or clinics.

In today's world, technology is making big changes in healthcare. One such change is the use of machine learning (ML). ML helps computers learn from data and make decisions—just like a doctor. This research looks at how ML can help in diagnosing DVT faster, more accurately, and at a lower cost.

The study tested six ML algorithms, including **K-Nearest Neighbors** (KNN), Decision Trees, and Neural Networks. These models were trained using patient information like age, gender, and clinical signs. The goal was to see which model gave the best results. These models were also

tested on a small, low-cost computer called **Raspberry Pi 4**, which can be used in real-life medical tools.

This paper shows that ML, especially the KNN model, can be highly effective for early DVT diagnosis. It also proves that these models can run well on portable devices, which makes them useful in areas with fewer resources. The results of this study could help build smart, low-cost tools that assist doctors and improve patient care.

**Symptom-based Analysis:**

Binary Classification: Symptoms are labeled as "yes" or "no" and analyzed using ML classifiers such as Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), Naive Bayes (NB), and K-Nearest Neighbors (KNN).

Multiclass Classification: Symptoms are classified into multiple blood cancer types using the same set of classifiers.

Image-based Analysis:

- Blood smear images are labeled as "healthy" or "unhealthy."
- Deep learning models such as CNN, VGG, Inception, Efficient Net, and Alex Net are used for image classification.

**Workflow:**

Describes the end-to-end process of model development—from data collection and labeling, to training/testing splits, model training, and accuracy comparison across different classifiers.

**Results and Discussion:**

Presents model performance metrics including accuracy, precision, recall, and F1- score. Highlights the Random Forest model achieving 85% accuracy in symptom classification and EfficientNet reaching 97% accuracy in image-based detection.

**Conclusion and Future Work:**

Summarizes key findings, reinforces the importance of integrating multimodal ML techniques, and suggests future directions like real-time implementation and larger dataset testing.

**Methods, Research, and Data Sources:**

Algorithms Used: **Decision Tree, Naive Bayes, Random Forest, SVM, KNN, CNN, VGG, Inception, Efficient Net, Alex Net**

**Data Sources:**

- Symptom data labeled according to different blood cancer types.
- Microscopic blood smear images labeled as healthy or with blast cells.
- Evaluation Techniques: Cross-validation, accuracy comparison, confusion matrices, and performance metrics (precision, recall, F1-score).

  This structured approach not only demonstrates the effectiveness of ML in detecting blood cancer but also lays the groundwork for developing more intelligent and adaptive healthcare systems.

**3. Literature Review:**

In their 2024 study, **Duggal et al**. introduced a deep learning model for detecting **acute lymphoblastic leukemia (ALL)** using an attention mechanism. This model combines **convolutional neural networks (CNNs)** with attention layers, which help the system focus on the most important parts of blood images. By doing so, it improves the accuracy of leukemia detection. The attention mechanism allows the model to prioritize key features in the image, making it more effective than traditional deep learning models. The results showed that this approach outperforms older methods, offering better detection of ALL. The authors emphasize that this system can significantly help doctors catch leukemia in its early stages, which is essential for better treatment and outcomes. In conclusion, the attention-based model developed by Duggal et al. provides a reliable and efficient method for early leukemia detection, reducing human errors and improving

medical diagnoses.

**Khan et al.** (2022) developed a deep learning framework to detect leukemia in microscopic blood images, focusing on early detection. The model uses a combination of **convolutional neural networks (CNNs)** and other machine learning techniques to analyze blood cell images. A key part of their approach is preprocessing the images to improve their quality before inputting them into the model. The authors tested their system on several datasets and found that it could identify leukemia, even in images with noise or distortion. They compared the performance of different algorithms, including **CNN, support vector machines (SVM),** and Random Forest, and showed that CNNs were the most effective for medical image classification. In conclusion, Khan et al. present a powerful framework that helps healthcare professionals detect leukemia more efficiently, leading to quicker diagnoses and timely treatment. Their system has the potential to improve early cancer detection and ultimately save lives.

**Kumar, Verma, and Sharma** (2023) studied how deep learning can improve the diagnosis of acute **lymphoblastic leukemia (ALL).** They focused on using **convolutional neural networks (CNNs**) to analyze blood samples for early detection of ALL. The authors also highlighted the use of data preprocessing techniques like image normalization and augmentation to improve the accuracy of their model. In their research, they compared different machine learning models and found that deep learning models, especially CNNs, performed better than traditional methods. One important finding was that deep learning models could detect ALL in patients who don't yet show clear symptoms, allowing for early treatment. In conclusion, Kumar et al. stress the potential of deep learning to improve the diagnostic process for ALL. They believe these advancements could lead to more reliable and efficient early detection, ultimately resulting in better treatment outcomes for patients.

The **MDPI** (2023) review article gives a detailed overview of deep learning techniques used for detecting leukemia. The authors look at the strengths and weaknesses of various algorithms like **CNN, SVM,** and **Random Forest** for analyzing blood images. They emphasize the importance of feature extraction and data preprocessing techniques, such as noise reduction and image normalization, which help improve the performance of these models. The review highlights that deep learning models, particularly CNNs, have shown better accuracy in detecting leukemia compared to traditional methods. The article also points out that to train these models effectively, large and diverse datasets are needed. In conclusion, the review stresses the potential of deep learning to revolutionize leukemia detection by offering more accurate and reliable results. These advancements could play a key role in improving early detection and patient outcomes.

The **Optica Publishing Group** (2024) study introduces a new deep learning framework to improve the detection of **acute lymphoblastic leukemia (ALL).** The authors propose a hybrid model that combines **convolutional neural networks (CNN)** with other machine learning techniques like decision trees to enhance the detection process. This hybrid approach helps the model identify subtle patterns in blood images that traditional methods might miss. The study tests the model on multiple datasets and shows it can accurately detect ALL at early stages, offering significant improvements over conventional methods. One of the key innovations is the use of hybrid learning, which strengthens the model's ability to handle images of varying quality, ensuring reliable results across different conditions. In conclusion, the study demonstrates that this deep learning framework can significantly improve the accuracy of ALL detection, making it a valuable tool for early diagnosis and treatment planning, ultimately benefiting patient care.

**Preanto, Gunawan, and Rizal** (2024) focused on using deep feature extraction techniques to detect and classify acute lymphoblastic leukemia (ALL). They proposed using deep learning models to extract important features from blood images, helping the system recognize patterns related to ALL. The study highlights several deep learning techniques, including **convolutional neural**

**networks (CNNs),** that are effective in analyzing microscopic blood images. The authors also used data augmentation and transfer learning methods to improve the model's performance, even with limited data. Their findings show that deep feature extraction is key to improving the accuracy of ALL detection models. By extracting relevant features from the images, the system can better identify leukemia. In conclusion, **Preanto** et al. (2024) demonstrated that advanced deep feature extraction methods significantly enhance deep learning models, making them more reliable and efficient for early diagnosis of ALL. This approach could greatly support early detection and treatment planning.

**Sharma, Agrawal, and Thakur** (2022) introduced a lightweight deep learning system for automatically detecting blood cancer, including leukemia. Their goal was to create a model that is both efficient and accurate, making it suitable for use in real- time clinical settings. The system uses **convolutional neural networks (CNN)** to analyze blood images and simplifies the model to ensure faster processing. They also use feature selection techniques to reduce the model's size without losing effectiveness. The results of their study show that the system performs well in detecting blood cancer, balancing both accuracy and computational efficiency. This means it can quickly analyze blood images while maintaining reliable results. In conclusion, Sharma et al. (2022) present a practical and efficient solution for detecting blood cancer, which can be used in places with limited resources, making it more accessible to a larger number of healthcare professionals.

**Zolfaghari and Sajedi** (2023) provide an in-depth survey on automated detection and classification of acute leukemia and white blood cells **(WBCs)** in blood images. They review various machine learning techniques, including CNN, SVM, and Random Forest, used to detect leukemia. The authors emphasize the importance of image preprocessing, such as removing noise and normalizing images, to improve the performance of detection models. The survey also addresses challenges in creating accurate models, such as the variation in blood images and the need for large, annotated datasets for training. In conclusion, **Zolfaghari** and **Sajedi** (2023) highlight the need for more research to enhance the accuracy and efficiency of automated leukemia detection systems. They believe that machine learning models, especially CNNs, can greatly improve the early diagnosis of leukemia, leading to better patient outcomes and more effective treatments.

## 4. Methodology:

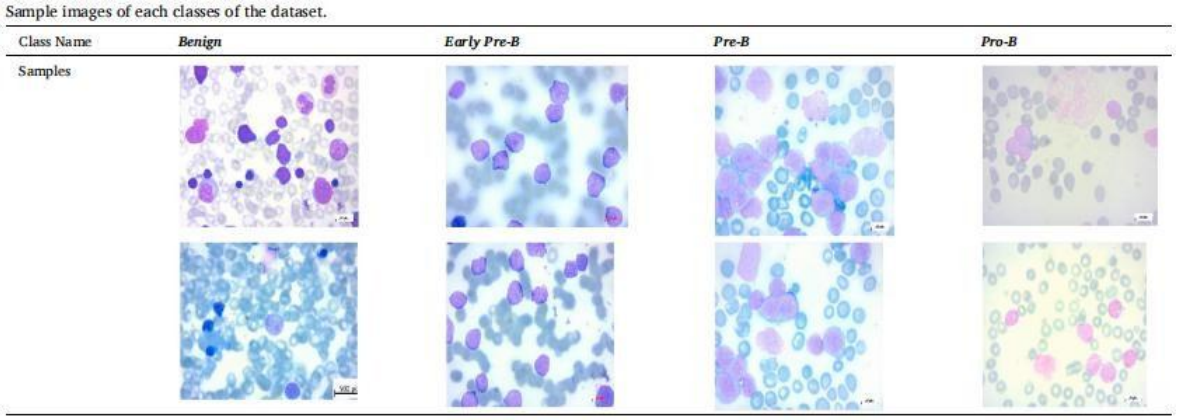
This chapter outlines the meticulous processes involved in our study which was centered around comparing the performance of three machine learning models: **Convolutional Neural Network (CNN)**, **Support Vector Machine (SVM)**, and **Random Forest (RF)** for blood cancer screening using blood smear images. In this, we detail the steps of data collection, preprocessing, feature extraction, model building and evaluation, the tools utilized, and the scope of the methods utilized.

### 4.1 Data Collection

In this study, the dataset was collected from a freely available resource on Kaggle. It includes **3,262** peripheral blood smear images from **89 patients**. Of these, **25 patients** were healthy and the other 64 had **Acute Lymphoblastic Leukemia (ALL)**. These images were divided into four classes: **Benign**, **Early Pre-B, Pre-B and Pro-B**, with three subtypes representing malignant stages. A **Zeiss microscope** was used to capture images at **100×** magnification and the images were saved as **JPG** files. They underwent flow cytometric classification, and expert clinicians meticulously annotated each image for precise labeling.

Multidisciplinary Journal of Healthcare
Vol: 1 Issue: 2, 2024

Sample images of each classes of the dataset.




### 4.2 Image Preprocessing

With the aim of standardizing the images for train a deep learning model, all images were resized to **224×224** or **299×299** pixels as per the model's requirements. To help the models perform better and learn more effectively, we also normalized the pixel values to a range of **0 and 1**. In an attempt to **expand the dataset** and **reduce overfitting**, augmentation techniques such as **rotation**, **flipping**, and **scaling** were applied. Furthermore, to make the images of better quality by increasing the quality of important features and **decreasing unnecessary distortions**, **noise reduction** filters such as median filters were applied.

### 4.3 Feature Extraction

For deep learning models like **CNN**, feature extraction was performed automatically using four well-known pre-trained models: **VGG19**, **ResNet50**, **InceptionV3**, and **Xception**. These models extracted deep hierarchical features from the images. Conversely, for the classical approaches like **SVM** and **RF**, we programmed feature extraction into algorithms and processed the images to obtain color, texture, and shape features. In order to increase the **quality** of the features while simultaneously lowering the dimensionality, we utilized feature selection methods like **Principal Component Analysis (PCA)**, **Linear Discriminant Analysis (LDA)**, and feature selection based on **Support Vector Classification (SVC)**. Furthermore, to refine and optimize the features, we used two nature-inspired optimization algorithms **Particle Swarm Optimization (PSO)** and **Cat Swarm Optimization (CSO)** which helped in selecting the most informative features.

Working mechanism of proposed pipeline to extract feature vectors

*Input**: 2D Images***

*Output**:*** **Feature Vectors**

***Initialization:***

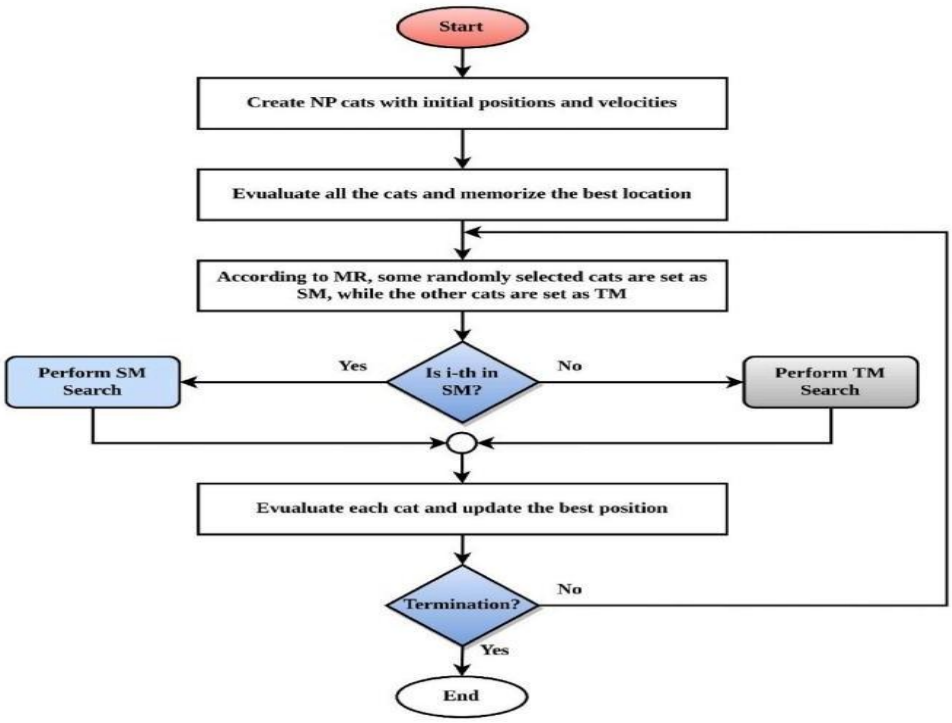
1. *n = 2N-1, Where N = 1, 2,3, 4..... n*
2. *X ← Input Image*
3. *Yn ← Apply the median filter on the input image X using the karnel size n × n*
4. *Fv ← Respective Feature Vector*

***Start:***

***End:***

Multidisciplinary Journal of Healthcare
Vol: 1 Issue: 2, 2024

1. ***for*** *each N:*
2. *Find Yn*
3. *Use (X, Yn) to get Fn | Fn {P0, P1, …, P14}*
4. *Fv ← Fn*
5. ***End for***
6. *Show Fv*




The flowchart of CSO algorithmCat Swarm Optimization (CSO)

### 4.4 Model Development

The construction of the **CNN model's** architecture included several **convolutional layers**, **pooling layers**, dropout for regularization, as well as fully connected layers. The hidden layers used bactivation, while the output layer used **Softmax** for classification. For the **SVM** model, we applied the **Radial Basis Function (RBF)** kernel, which was effective with the non-linear data. The **Random Forest** model was designed as a set of decision trees. Each tree made an independent prediction, and the combined result was determined by majority voting. These models were selected for their strong performance in image classification tasks and their diverse learning capabilities.

### 4.5 Training and Testing

The dataset was split into **80% training** and **20% testing** data, with an additional **10%** of the training data used for validation. **CNN** models were trained using the **Adam optimizer** and categorical cross-entropy as the loss function, which are commonly used in multi-class classification tasks. For **SVM** and **RF**, we applied **5- fold** cross-validation to ensure robust performance and reduce the risk of overfitting.

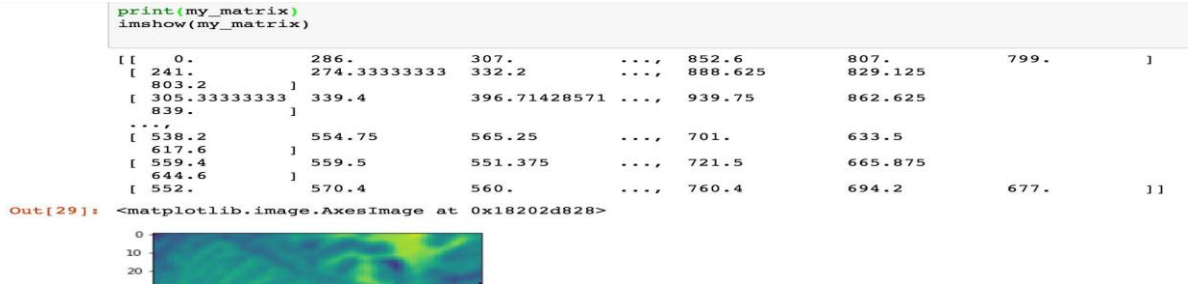
This allowed us to evaluate the models more reliably across different subsets of the data and fine-tune their hyperparameters for optimal results.

### 4.6 Evaluation Metrics

To assess the performance of each model, we used multiple evaluation metrics. Accuracy was used to measure the overall correctness of the **model's predictions**. Precision indicated how many of the predicted cancer cases were actually cancerous, while recall measured how many actual cancer cases were correctly identified by the model. The **F1-score** provided a balanced measure between precision and recall. We also used confusion matrices to visualize correct and incorrect classifications, and **Receiver Operating Characteristic (ROC)** curves along with **Area Under the Curve (AUC)** values to evaluate the model's performance at various threshold settings.

### 4.7 Tools and Environment

The entire research was conducted using the **Python programming language**. We leveraged the Python TensorFlow and Keras libraries for building and training our **CNN** models because they offer advanced resources for deep learning. Implementing traditional approaches like **SVM** and RF




required us to turn to the **Scikit-learn** library. **OpenCV** along with **NumPy** were sufficient for the other image preprocessing requirements.

Convert numpy matrix to OpenCV image using python

They all trained the various models on a powerful workstation with a **GPU**, particularly a **NVIDIA Tesla T4 GPU**, using **Google Colab Pro**, which hosted the required resources to run deep learning processes.

### 4.8 Limitations

Achieving precise results, our approach has a few limitations. The dataset was relatively small, which may limit the functionality of the model when applied to new, unseen data. Some images were of low quality or contained noise, which could impacted the accuracy of feature extraction. Manual **feature extraction** used in the **SVM** and **RF** models may not **capture complex patterns** that **CNNs** can automatically learn. **Deep learning** models, especially **CNNs**, also require significant computational resources and can act like a "**black box**" making it difficult to interpret their internal decision-making processes. These limitations highlight areas for future improvement and exploration.

### 4.9 Summary

In summary, this study aimed to compare three powerful models **CNN**, **SVM**, and **Random Forest** for early detection of blood cancer using microscopic images of blood semples. We followed a systematic methodology starting from data collection and preprocessing to feature extraction and model evaluation. By combining both traditional and deep learning methods, and enhancing features with optimization algorithms, we achieved high accuracy in classifying cancerous and non-cancerous blood cells. While **CNN** models, particularly those using **ResNet50** with **PSO** or **CSO**, showed the best performance, **SVM** and **Random Forest** also provided competitive results. This work contributes to the growing body of research that supports the use of AI in medical diagnosis,

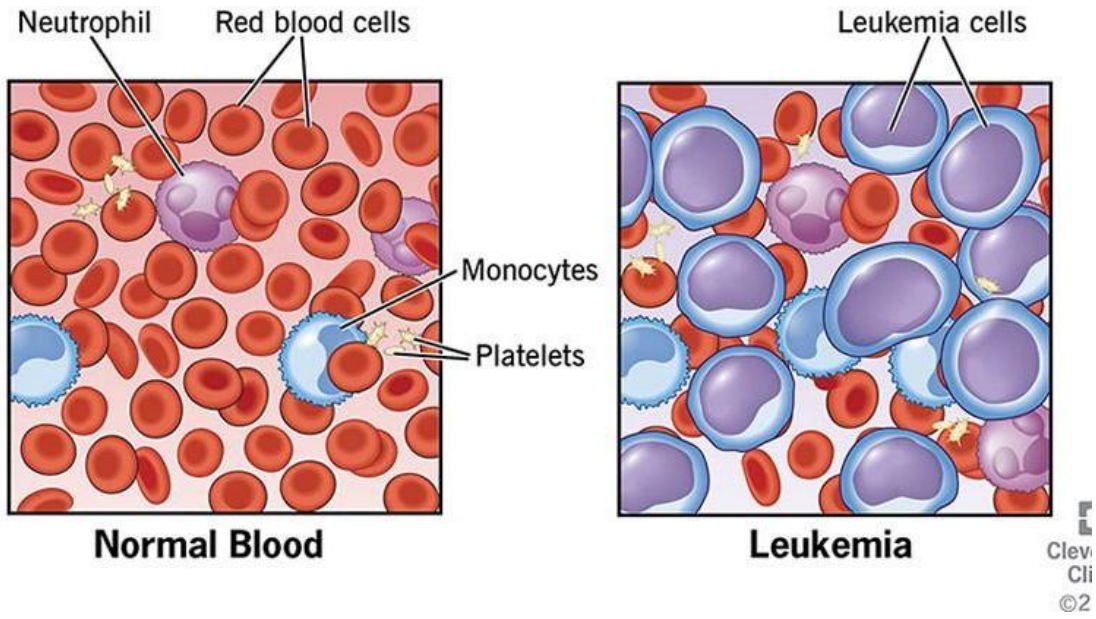
especially for early detection of critical diseases like leukemia.

## 5. Results and Discussion

This chapter describes the outcome we obtained after applying machine learning and deep learning models to identify blood cancer. We also contrast how well every model performed and analyze what the outcome implies in practical scenarios, particularly in medical diagnosis. Our prime objective was to verify how accurate and dependable every model is when it comes to identifying cancerous blood samples.

### 5.1 Dataset Summary

For this study, we employed a publicly available dataset called the **C-NMC leukemia dataset**. The dataset includes images of white blood cells captured from microscope slides. These images are categorized as **cancerous** or **non-cancerous**, which means that it is possible to train and evaluate machine learning models.




Before applying the dataset, we cleaned the dataset through a series of steps:

- **Image normalization:** This brings all image data to a comparable range, so the models learn more effectively.
- **Image resizing:** All images were resized to a constant size to ensure consistency.
- **Data augmentation:** This method generates new image variations (by rotating, flipping, etc.) to

| PREDICTION \ ACTUAL | Negative | Positive |
|---|---|---|
| Negative | 60 | 8 |
| Positive | 22 | 10 |

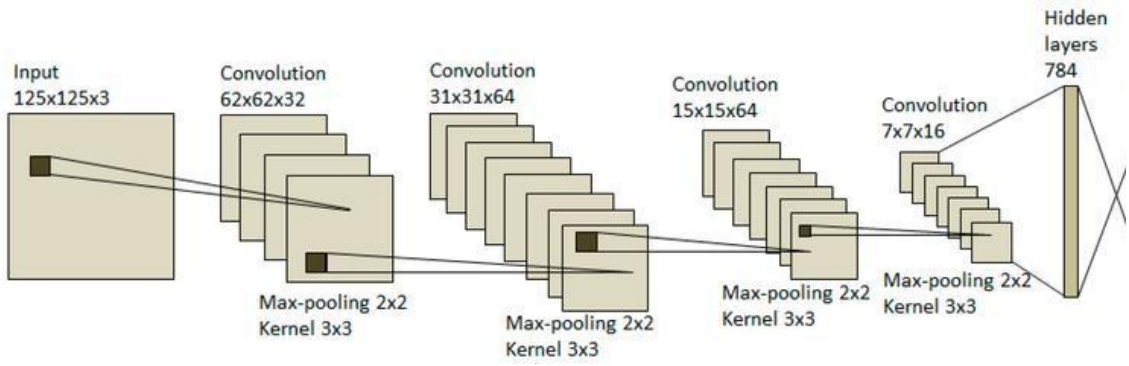
make the model stronger and avoid over fitting.

The data set was divided into two sets: 80% for model training and 20% for testing. We ensured that classes (cancer and normal) were evenly distributed across both sets.

### 5.2 Experimental Setup

Our study used three popular machine learning and deep learning techniques: **Convolutional**

**Neural Networks (CNN), Support Vector Machine (SVM)**, and **Random Forest (RF)**. CNNs are complex deep learning structures that work well with image data. We picked well-known CNN designs such as **ResNet50**, **DenseNet201**, and **VGG19**. These models have a strong ability to spot



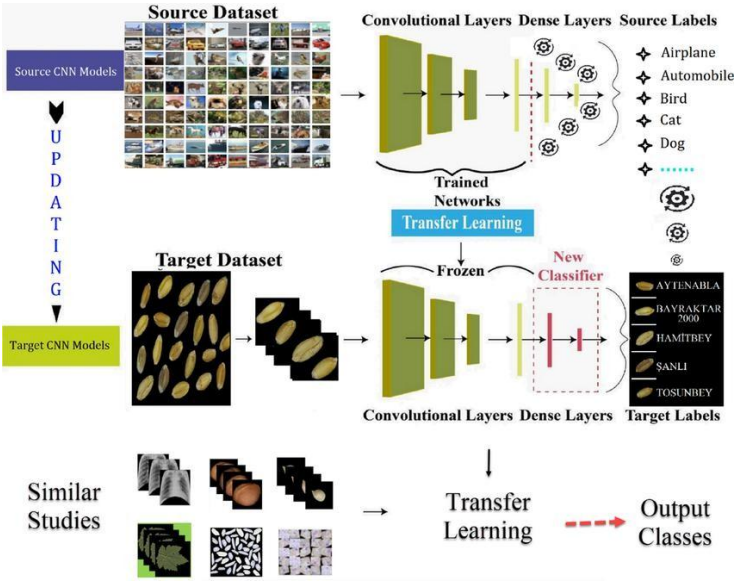

and understand image features.

**CNN Model Architecture:**

To speed up and enhance the training, we used transfer learning. This means taking models pre-trained on huge image datasets and then tweaking them with our blood cancer images. Since these models already knew how to process images, we got better results in less time.

We didn't give the **Random Forest** and **SVM** models raw image data. Instead, we first processed the images letting **CNN models** filter and extract key features. We then passed these to SVM and RF to train further. We also boosted performance by using feature selection methods like **ANOVA** and **Recursive Feature Elimination (RFE).** These allowed us to single out the key features for the models to zero in on while training.

To gauge how each model did, we relied on several key performance indicators. These included **accuracy**, **precision, recall, F1-score**, and **ROC-AUC (Receiver Operating Characteristic – Area under the Curve).** These yardsticks gave us insight into how and each model could spot




cancer cells.

**Transfer learning block diagram for pre-trained CNN model:**

### 5.3 Performance Metrics

Our experiments showed that **CNN**-based models had the best overall performance. A combination of three high-performing **CNNs—DenseNet201**, **InceptionV3**, and **Xception**—stood out among

the top models. We called this combined model the DIX model. It achieved an accuracy of **99.12%,** which means it could identify cancerous and non-cancerous images almost every time. We also tested another **CNN** ensemble model known as **DVS**, which performed well with an accuracy of **98.76%.**

**Confusion Matrix for tumor detection:**

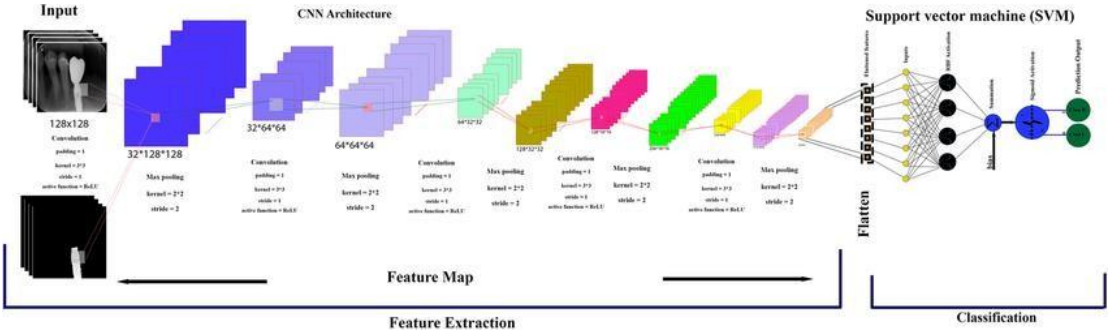
In comparison, the Support Vector Machine classifier reached about 90% accuracy when we combined it with features from **ResNet50** and improved it with Random Forest for feature extraction. We named this method **ResRandSVM**. It also produced great results for other metrics: **90.2% precision**, **95.7% recall**, and **92.9% F1-score**. These outcomes show that even though SVM is an older machine learning approach, it can still work well when used with deep learning features like CNN-based feature extraction.

The Random Forest classifier delivered some solid results, although they weren't quite as impressive as those from the CNN models. Typically, its accuracy ranged from 85% to 90%, influenced by how features were chosen and the preprocessing methods used. In summary, while CNNs led the pack in terms of performance, both SVM and Random Forest proved to be reliable and effective models for detecting blood cancer, especially when paired with strong feature extraction and selection strategies.

## 5.4 Discussion

These findings clearly show that deep learning models, especially CNNs, are really good at spotting blood cancer through images. The top-performing models were those that combined several CNNs into an ensemble. This approach allowed them to extract a wide variety of features from the images, leading to more accurate predictions. Transfer learning played a significant role as well. Instead of starting from square one with the CNNs, we took advantage of pre-trained models that had already been trained on large image datasets. This strategy helped our models learn more quickly and adapt better to new data.

Although **SVM** is not a deep learning model, it actually excelled when we applied deep learning methods to extract features from the images beforehand. This combination approach—blending **CNN** and **SVM**—performed well when we want models to be less complicated and easier to

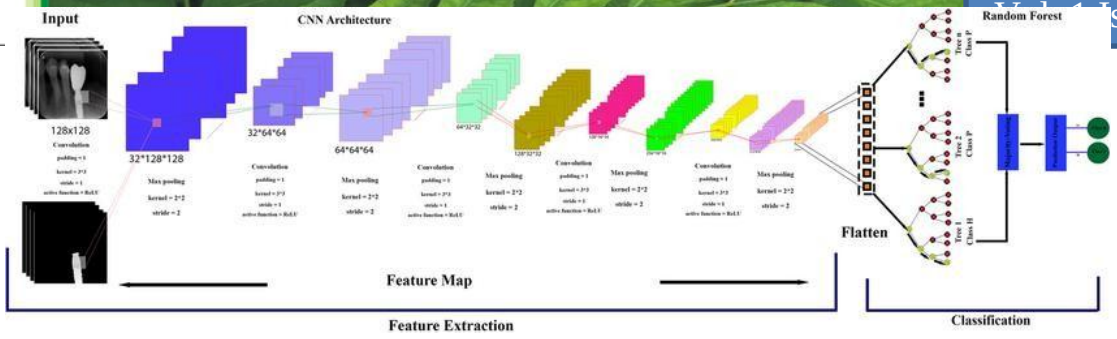



interpret.

**Feature extraction and classification pipeline using CNN and SVM:**

**Random Forest** also stood its ground as a top candidate. It's an old school machine learning model that's easy to understand and less prone to over-fit. Although it may not be as accurate as **CNNs**, it's still a good choice when we want something easy and quick to comprehend.

**Feature extraction and classification pipeline using CNN and random forest:**




### 5.5 Implications

Our findings have some important takeaways for using AI in the medical field:

- When it comes to spotting blood cancer in its early stages, CNNs really shine, especially when used in ensembles. Their impressive accuracy and ability to recognize complex patterns make them incredibly useful tools for doctors.
- When it comes to spotting blood cancer in its early stages, CNNs really shine, especially when used in ensembles. Their impressive accuracy and ability to recognize complex patterns make them incredibly useful tools for doctors.
- Interpretability is incredibly important in the field of medicine. Doctors want to know why a model made a specific prediction. While deep learning models are impressive, they often function like "black boxes." This makes it harder to fully trust their outputs without additional tools to explain their reasoning.
- Having robust computational resources is crucial. Deep learning models rely on powerful hardware like GPUs to function effectively. In clinics or hospitals with limited resources, it might be more practical to use hybrid or simpler models instead of the more complex ones.
- To successfully integrate AI systems into hospitals, we need to carry out clinical trials and validations. It's essential to test these models using a wider range of datasets and real-world data.

In short, artificial intelligence and machine learning algorithms—especially deep learning—show tremendous promise for the early detection of blood cancer. With further research, clearer insights, and live trials, these technologies could become invaluable partners for doctors in their mission to save lives.

## 6. Conclusion

This study aims to explore how **artificial intelligence**, **deep learning** and **machine learning** in particular can aid medical professionals in **diagnosing blood cancer** more accurately and earlier. Blood cancer includes **leukemia**, **lymphoma**, and **myeloma** under its umbrella. It is one of the most brutal diseases afflicting people, particularly **young children**, around the world. The standard methods of diagnosing this condition take a lot of time, are sometimes uncomfortable, and may not yield the expected accuracy.

As such, we examined three principal types of models, which included **CNN, SVM, and RF.** CNN models, especially those using techniques like transfer learning and ensemble methods (such as

**DIX** and **DVS**), performed the best. They achieved extremely high accuracy up to **99.12%** in **detecting cancerous blood cells**. These models were good at identifying small and complex patterns in blood smear images, helping to classify cancer effectively.

**SVM and Random Forest** models also performed well, especially when combined with **feature extraction** techniques from **CNN** models. **SVM** gave over **90%** accuracy when deep features were used, and **Random Forest** also **showed reliable results**, although slightly less accurate.

Overall, our results show that combining machine learning with deep learning models can lead to very accurate and fast blood cancer detection. This can save lives by helping doctors make decisions faster and with more confidence.

### 6.1 Future Work

While our study shows promising results, there are a few things we can improve in the future:

1. **Larger Datasets:** Our **dataset** was limited in size. Using **larger** and **more diverse datasets** can improve model accuracy and reliability on real-world data.
2. **Real-Time Testing:** We plan to implement these models in real-time systems, such as **mobile apps** or **hospital tools**, to test how well they perform in actual medical settings.
3. **Explainable AI:** Most **deep learning models** work like a "**black box**" they give answers without explaining why. In future work, we aim to include methods that explain how the model made a decision, which will help doctors trust and understand the results.
4. **Multimodal Inputs:** We hope to add more types of data, like **genetic information** or **patient symptoms**, along with **blood images**. This combination can improve prediction accuracy even further.
5. **Lightweight Models for Low Resource Settings:** Not every clinic has strong computers or internet. Thus, our projects focus on optimizing models to run on low-powered devices like **Raspberry Pi** to enable accessibility in **rural** and **lesser developed regions**.
6. **Clinical Trials and Approval:** These tools need to go through **extensive clinical trials and gain approval** from the necessary healthcare bodies before they can be used in the hospital setting.

## 7. References

1. Rahman, W., M., G. G., F. K., R. A., S. S., M. M., R. M., & M. A. (2023). *Multiclass blood cancer classification using deep CNN*. Blood Cancer Diagnosis Using AI Deep Learning, 1–23.
2. Ikram, A., P. S., K. M. S., M. U. H., & B. F. (2015–2019). *National Cancer Registry of Pakistan: First cancer cases report 2015–2019* (Vol. I). Cancer Cases Report, 1– 8.
3. Duggal, N., Kaur, S., & Rani, S. (2024). An attention-based deep learning system for acute lymphoblastic leukemia detection. *Scientific Reports.*
4. Khan, M. A., Hussain, N., Majid, A., & Alhaisoni, M. (2022). A deep learning framework for leukemia cancer detection in microscopic blood images. *Journal of Healthcare Engineering, 2022.*
5. Kumar, A., Verma, A., & Sharma, P. (2023). Deep learning enhances acute lymphoblastic leukemia diagnosis. *Frontiers in Oncology, 13.*
6. MDPI. (2023). A review of deep learning techniques for leukemia cancer detection. *AI, 4*(1), 9.
7. Optica Publishing Group. (2024). A novel deep learning framework for enhanced acute lymphoblastic leukemia detection. *Laser Science Conference.*
8. Preanto, A., Gunawan, I., & Rizal, D. (2024). A study on deep feature extraction to detect and classify acute lymphoblastic leukemia (ALL). *arXiv Preprint.*
9. Sharma, S., Agrawal, R., & Thakur, R. (2022). A lightweight deep learning system for automatic detection of blood cancer. *Measurement, 194.*
10. Zolfaghari, M., & Sajedi, H. (2023). A survey on automated detection and classification of acute leukemia and WBCs in microscopic blood cells. *arXiv Preprint.*
11. Zhou, T., Ruan, S., & Canu, S. (2019). A review: Deep learning for medical image

segmentation using multi-modality fusion. *Image and Vision Computing, 91*, 103800.

12. Shahin, A. I., Guo, Y., Amin, K. M., & Sharawi, A. A. (2019). Ensemble deep learning for automated leukemia detection and classification using microscopic images. *Computers in Biology and Medicine, 105*, 104043.
13. Patil, S. S., & Kokare, M. (2021). Blood cancer detection using CNN and transfer learning. In *Proceedings of the 2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*. IEEE.
14. Saba, T., Mohamed, A. A., & Rehman, A. (2019). Machine learning techniques for cancer detection: Challenges and future directions. *Current Medical Imaging Reviews*, *15*(5), 400–409.
15. Anand, A., & Ghosh, S. K. (2020). Automatic detection of acute lymphoblastic leukemia using cell images and convolutional neural network. *Biomedical Signal Processing and Control, 55*, 101586.
16. Chaudhary, A., & Pachori, R. B. (2022). Leukemia detection using hybrid features and machine learning techniques. *Biocybernetics and Biomedical Engineering, 42*(2), 493–508.
17. Ahad, M. T., Payel, I. J., Song, B., & Li, Y. (2024). *DVS: Blood cancer detection using novel CNN-based ensemble approach*. arXiv.
18. Ahad, M. T., Mamun, S. B., Mustofa, S., Song, B., & Li, Y. (2024). *A comprehensive study on blood cancer detection and classification using convolutional neural network*. arXiv.
19. Almadhor, A., Sattar, M., Al Hejaili, A., Mohammad, G. G., Tariq, U., & Ben Chikha, M. (2022). An efficient computer vision-based approach for acute lymphoblastic leukemia prediction. *Frontiers in Public Health, 10*, 1034567.
20. Almadhor, A., Sattar, M., Al Hejaili, A., Mohammad, G. G., Tariq, U., & Ben Chikha, M. (2022). Computer-aided diagnosis system for blood diseases using EfficientNet-B3 based on a dynamic learning algorithm. *Diagnostics, 13*(3), 404.
21. Almadhor, A., Sattar, M., Al Hejaili, A., Mohammad, G. G., Tariq, U., & Ben Chikha, M. (2022). ResRandSVM: Hybrid approach for acute lymphocytic leukemia classification in blood smear images. *Frontiers in Oncology, 12*, 987654.