AI-Powered Leukemia Detection

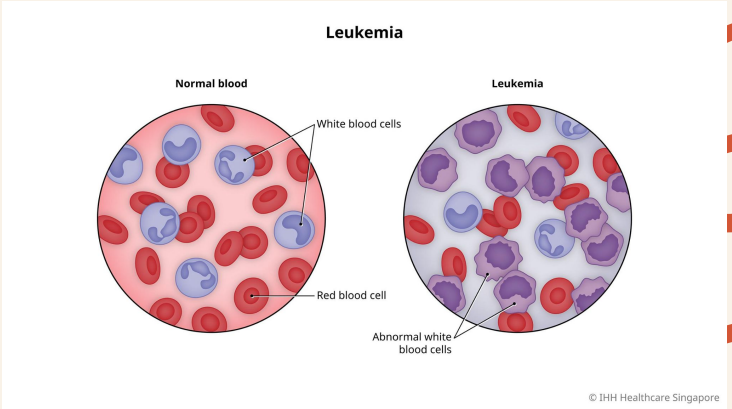
Mia Beltran, Alexa Boksiner, Sophia Florea, Juliana Galvez, Alice Stryker

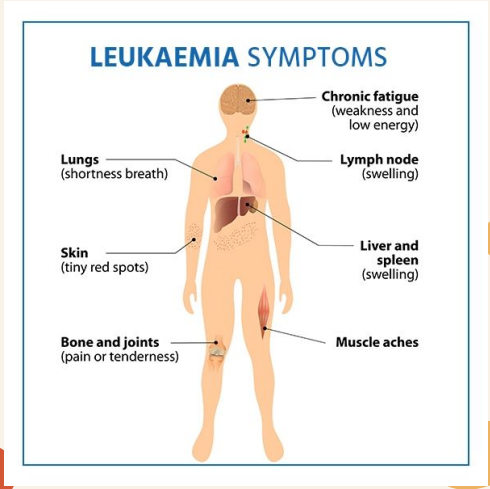
Project Overview

- Leukemia is a type of blood cancer that affects white blood cells, and early detection is critical for effective treatment.
- Our project focuses on using machine learning to improve the early diagnosis of leukemia by analyzing blood test results.

Problem Statement

- Leukemia symptoms including fatigue, bruising, and bone pain are vague and often misdiagnosed or detected too late, especially in chronic cases.
- Delayed diagnosis leads to more invasive treatments and lower survival rates.
- There is a critical need for faster, more accurate diagnostic tools so that patients can begin receiving treatment sooner.



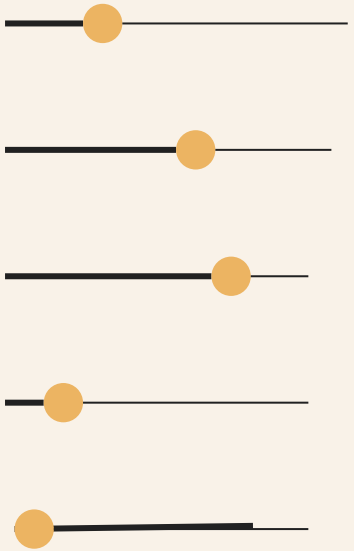


Significance

- ~24,000 deaths from leukemia occur annually in the U.S.
- Blood plays a critical role in homeostasis and immune function.
- Cancer in the bloodstream can rapidly affect multiple systems.
- Earlier diagnosis allows for earlier treatment which leads to less invasive, more effective outcomes.
- Biomedical engineers aim to revolutionize diagnosis using machine learning & data analysis for faster, more accurate detection.

Data

- Dataset included various biomarkers: individual's lactate dehydrogenase levels, white blood cell count, gene expression, and the CRP protein levels.
- Data was labeled with either a 0 (false) or 1 (true) to denote whether or not the patient had leukemia.
- Dataset was synthetic (artificially generated), but based on real, accurate values of biomarker levels in leukemia and non-leukemia patients.



Methodology

Our code:

- Loads the synthetic dataset
- Explores the dataset

- Preprocesses the data, keeping all relevant features (all features were relevant in this case)

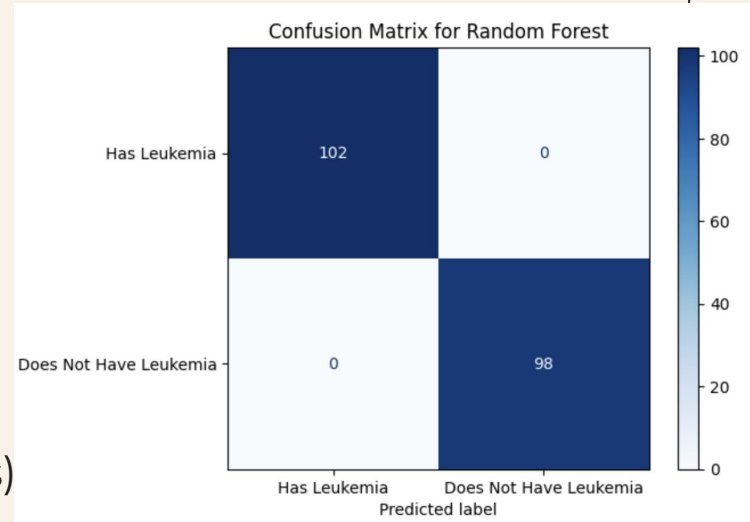
- Trains the models and tunes settings for best precision

- Evaluates the model performance (confusion matrix, accuracy, precision, recall, and F1 score)

- Takes new user input and predicts leukemia status using trained model

Results

- **Accuracy**
  - **Definition:** percentage of correct predictions
  - **Our model:** 100% accurate
- **Precision**
  - **Definition:** accuracy of model's positive predictions
  - **Our model:** 100% precise (no missed diagnosis)
- **Recall**
  - **Definition:** percent of true positives identified by the model
  - **Our model:** 100% recall (no false negatives)
- **F1 Score**
  - **Definition:** harmonic mean of precision and recall
  - **Our model:** 100%



Link to Code

[Github Repository Link]