CANCEROUS BLOOD CELL IMAGE RECOGNITION

Edric Ma

Whoa!

According to the American Cancer Society, an estimated **23,670** people in the United States are expected to die from leukemia in 2024. This includes: Acute lymphocytic leukemia (ALL): 1,400, Acute myeloid leukemia (AML): 11,090, Chronic lymphocytic leukemia (CLL): 4,460, and Chronic myeloid leukemia (CML): 1,290.

Without treatment, ALL median survival is around **1-2 years.**
Without treatment, AML median survival is around **3-6 months.**

Early detection and treatment significantly improve leukemia survival rates, with overall 5-year survival rates for all leukemia types around **67%**.

CRITICAL QUESTIONS:

01 Can we help patients feel more assured with their diagnosis?

02 Can early detection be achieved with more efficiency?

03 Can initial screenings be more cost efficient?

THE ROLE OF MACHINE LEARNING

Machine Learning Models

A model could identify cancerous cells based off a blood cell smear with relatively high accuracy.

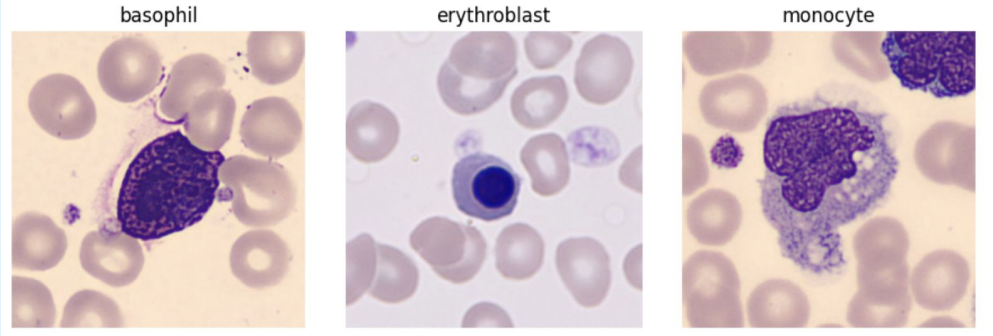
Health Portal Implementation

Potential for seamless screening if labs can upload images of blood cell smears to a health portal, where the model is already built into the system.

THE DATASET

Dataset consists of 5,000 images belonging to 5 classes of cancerous blood cell smears:

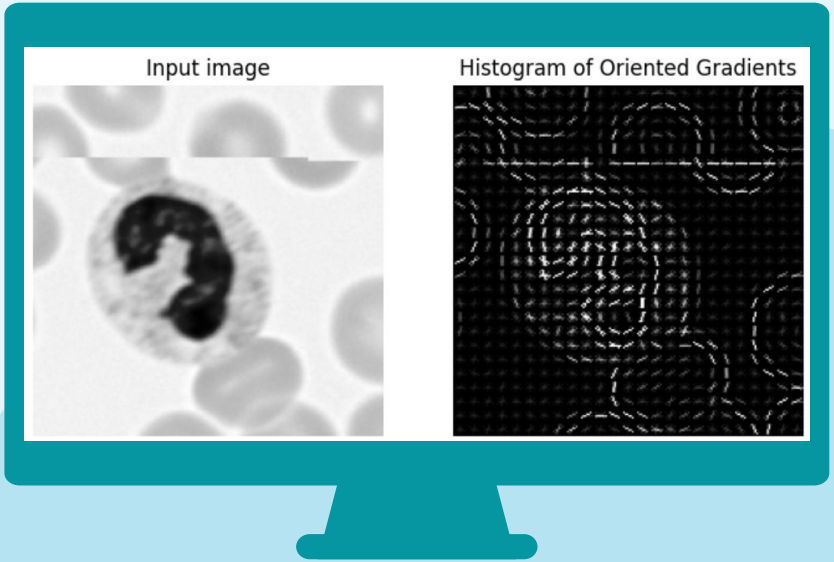
- Basophil
- Erythroblast
- Monocyte
- Myeloblast
- Segmented Neutrophil



CLASSICAL MACHINE LEARNING WITH HOG FEATURE EXTRACTION

- Extracted key information from complex images using **Histogram of Oriented Gradients (HOG)** features.

- “Shotgun” approach five different classifiers:
  - Logistic Regression
  - Decision tree
  - K-Nearest Neighbor
  - Random Forest
  - Ada Boost

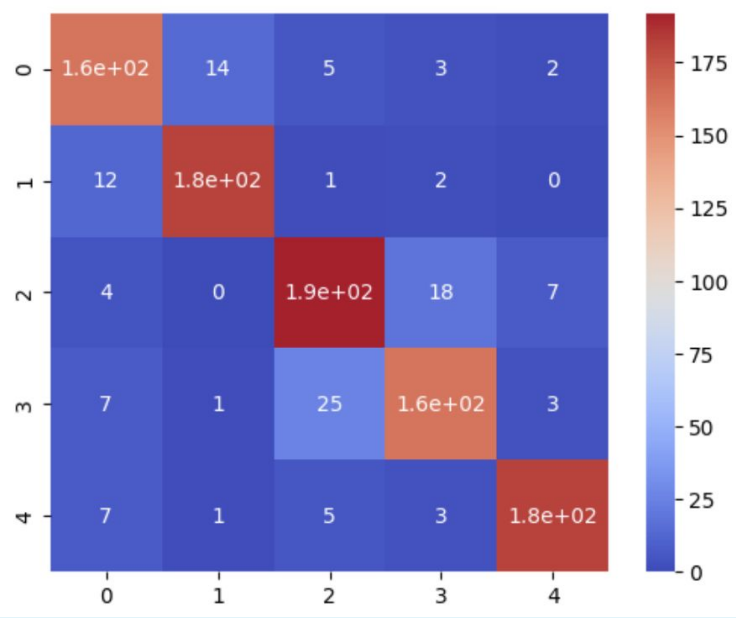


BEST MODEL: LOGISTIC REGRESSION

Logistic regression achieved 88% accuracy!

Potential reasons:

- Nature of HOG features
- Linearity of feature space
- Dimensionality reduction

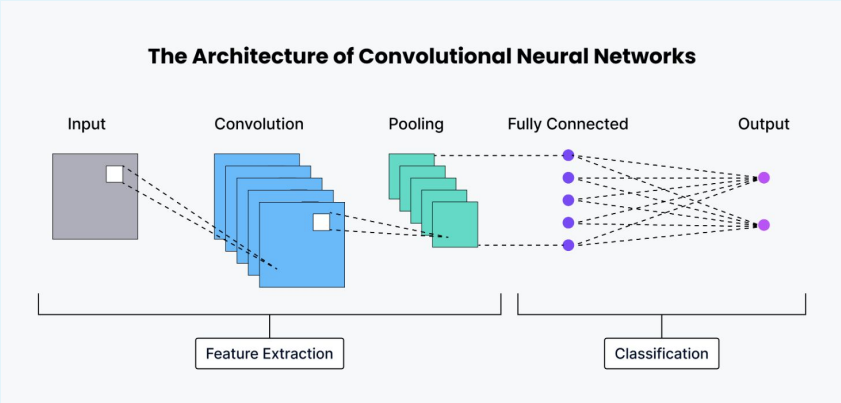


A bit of trouble deciphering basophil and erythroblasts as well as monocyte and myeloblast.

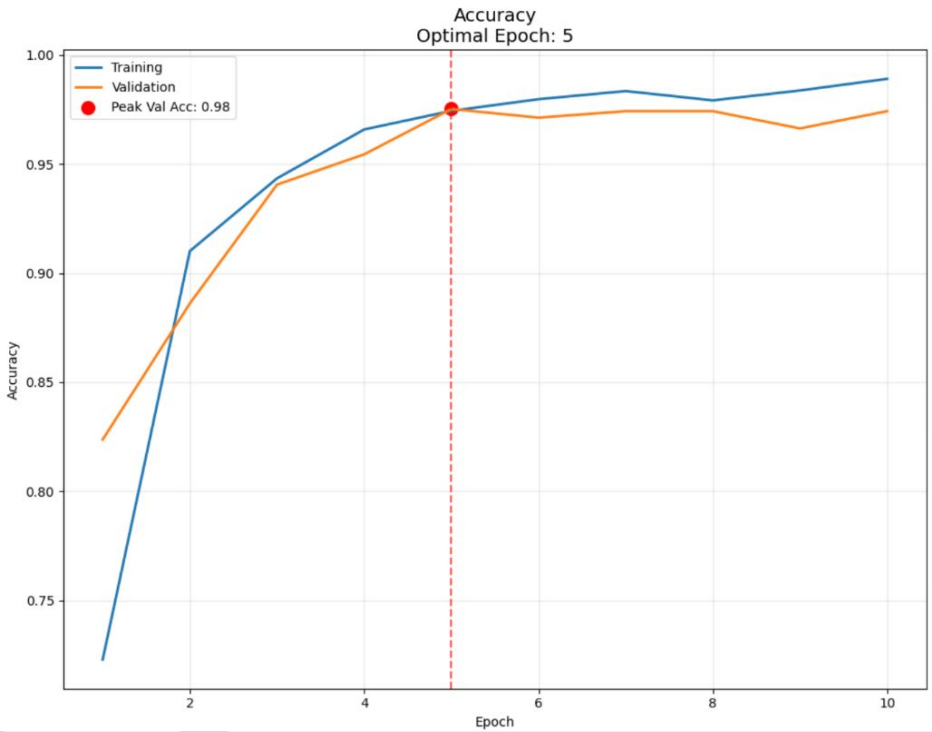
DEEP LEARNING: CONVOLUTIONAL NEURAL NETWORK

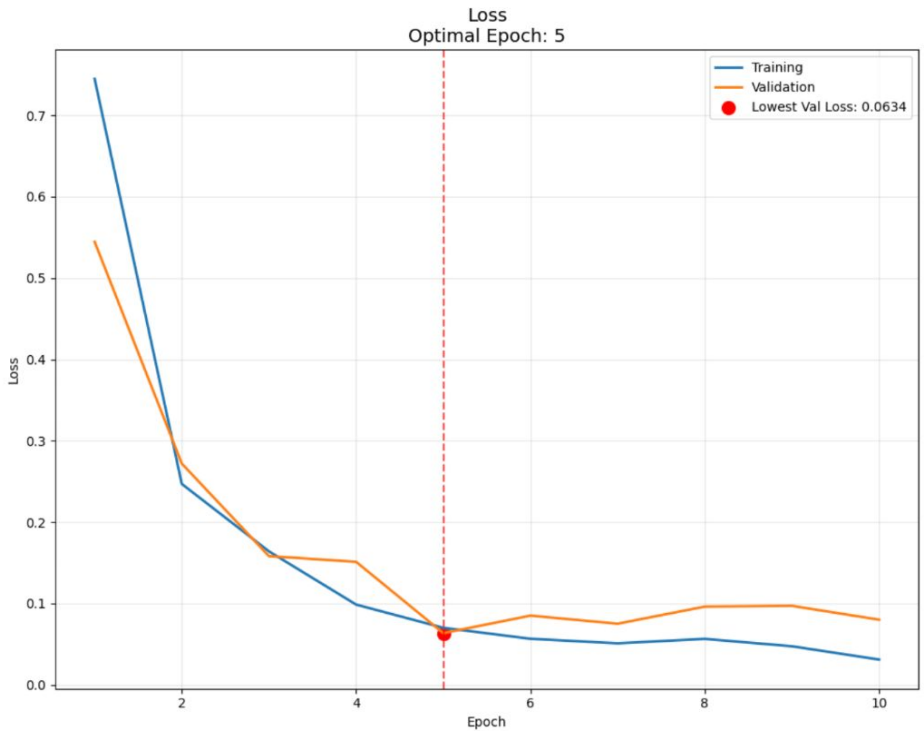
A convolutional neural network was created with the following layering and parameter tuning:

- Rescaling and resizing to 256x256 image
- 3 Convolutional 2D layers (ReLU activation, 2x2 kernel size, filter size 16, 32, 64)
- 3 Max Pooling 2D layers
- 2 Dropout layers
- 1 Flattening layer
- 2 Connective layers (ReLu activation)
- 1 output layer (softmax activation)

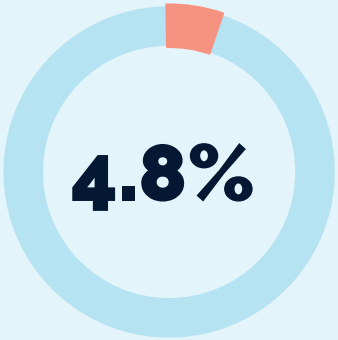


TRAINING/VALIDATION ACCURACY AND LOSS



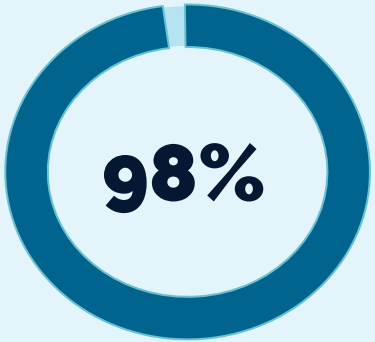


FINAL TEST DATA



Test Data

240 out of the 5,000 images were saved as test data to see how our model would do after training and validation



Testing Accuracy

Our carefully tuned and trained CNN was able to reach 98% testing accuracy!

Resources

Slides:

- https://slidesgo.com/theme/medical-collaboration-research-creative#search-medical+research&position-1&results-910&rs=search

Cancer Facts:

- https://www.cancer.org/cancer/types/acute-myeloid-leukemia/about/key-statistics.html#:~:text=About%2022%2C010%20people%20will%20be,is%20about%20%C2%BD%20of%201%25.
- https://pocketdentistry.com/disorders-of-white-blood-cells/#:~:text=Medical%20Management%20of%20Acute%20Leukemia%20Factor%20Type,Poor%20Survival%20time%2C%20mean%20%E2%80%94%202%20years

Images:

- https://zilliz.com/glossary/convolutional-neural-network