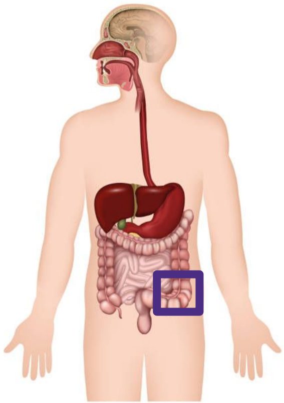
Cancer Progression Modeling

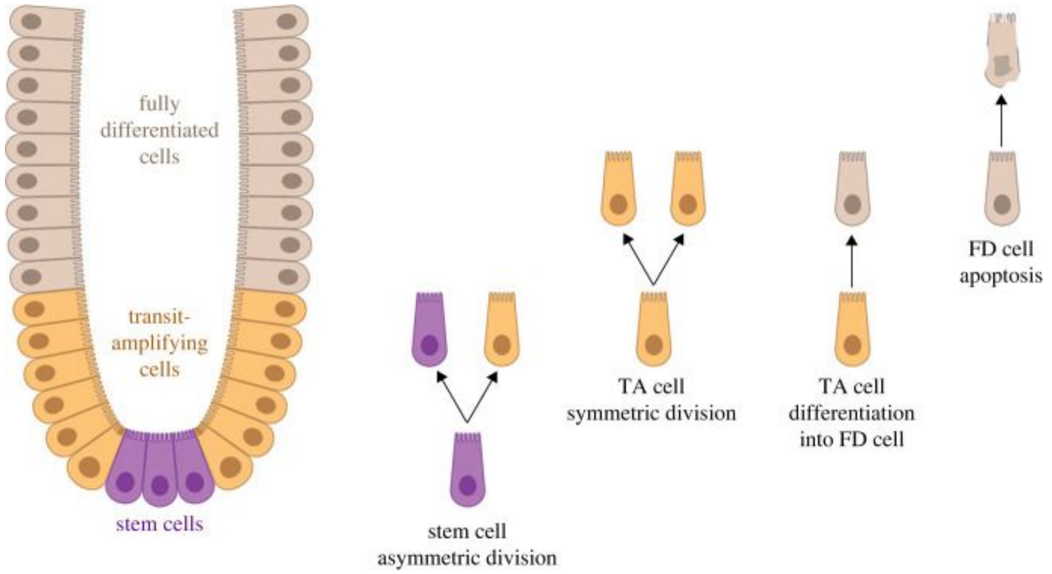
Rebecca Wang, Alexis Georgiades

Mentor: Anja Vogt

UNIVERSITY *of* WASHINGTON

Intro & Biological Background





Deterministic Modeling (ODEs)

UNIVERSITY *of* WASHINGTON

System of Equations

- N_0 = Initial Stem Cell count (18)
- r = Stem Cell asymmetric division rate (1/2.5 d^{-1})
- λ = TA symmetric division rate (1/30 h^{-1})
- d = TA to FD differentiation rate
- γ = FD cell apoptosis rate (1/3.5 d^{-1})

$$\left.\begin{array}{c} \mathrm{S} \xrightarrow{r} \mathrm{S}\ \mathrm{TA} \\ \mathrm{TA} \xrightarrow{\lambda} \mathrm{TA}\ \mathrm{TA} \\ \mathrm{TA} \xrightarrow{d} \mathrm{FD} \\ \mathrm{FD} \xrightarrow{\gamma} \emptyset. \end{array}\right\}$$

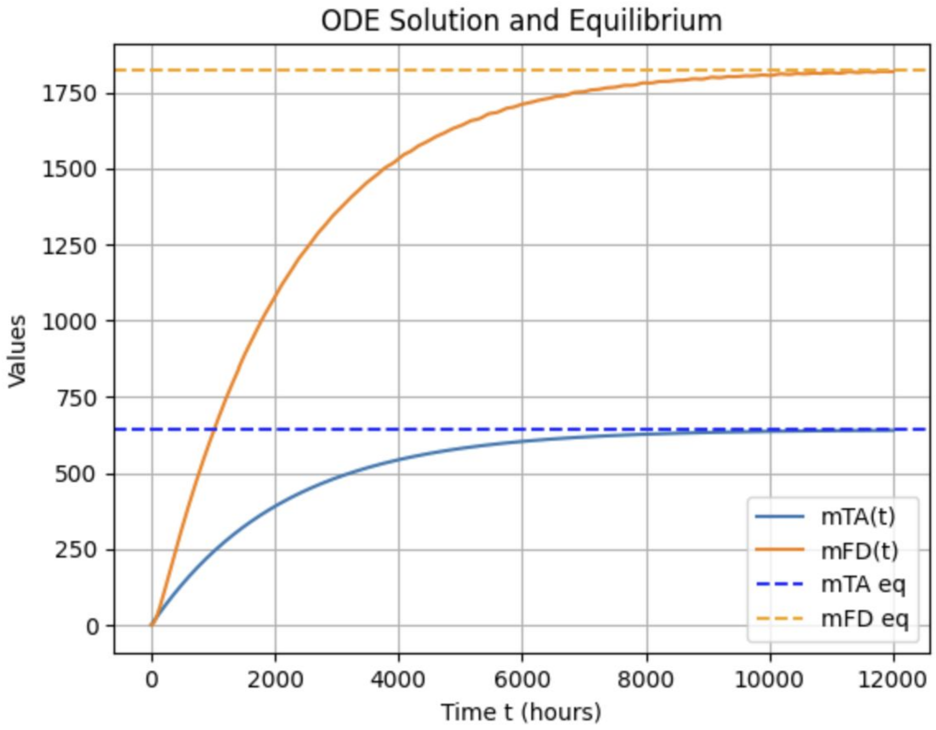
d is adjusted so that the total crypt size stays around 2300

$$\left.\begin{array}{l} m'_{\mathrm{TA}}(t) = rN_0 + (\lambda - d)m_{\mathrm{TA}}(t) \\ m'_{\mathrm{FD}}(t) = dm_{\mathrm{TA}}(t) - \gamma m_{\mathrm{FD}}(t). \end{array}\right\}$$

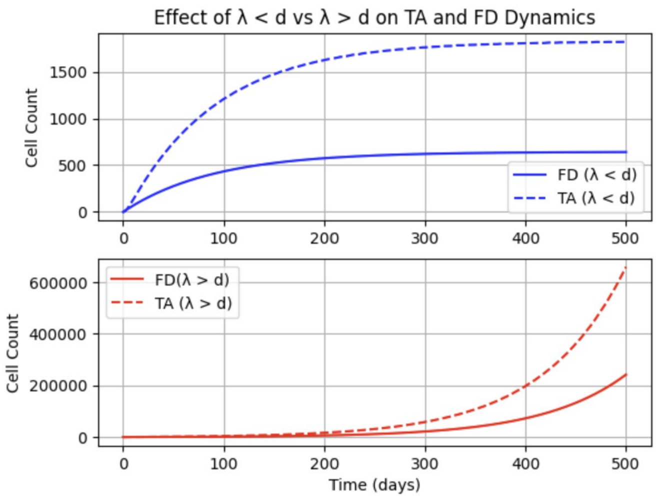
UNIVERSITY *of* WASHINGTON

Solution to ODE

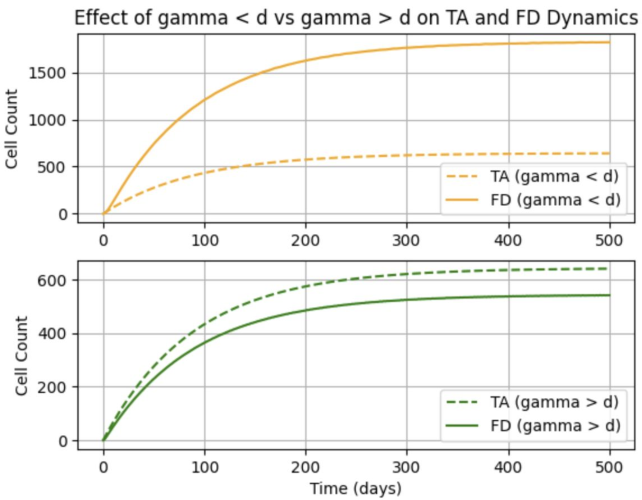
This system starts with only Stem Cells (N_0) and grows to a steady state of ~1800 FD cells and ~625 TA cells. This result is verified when the equilibrium points, calculated via the jacobian, are plotted as well.



Stability analysis



If λ is greater than d, system is unstable and grows exponentially



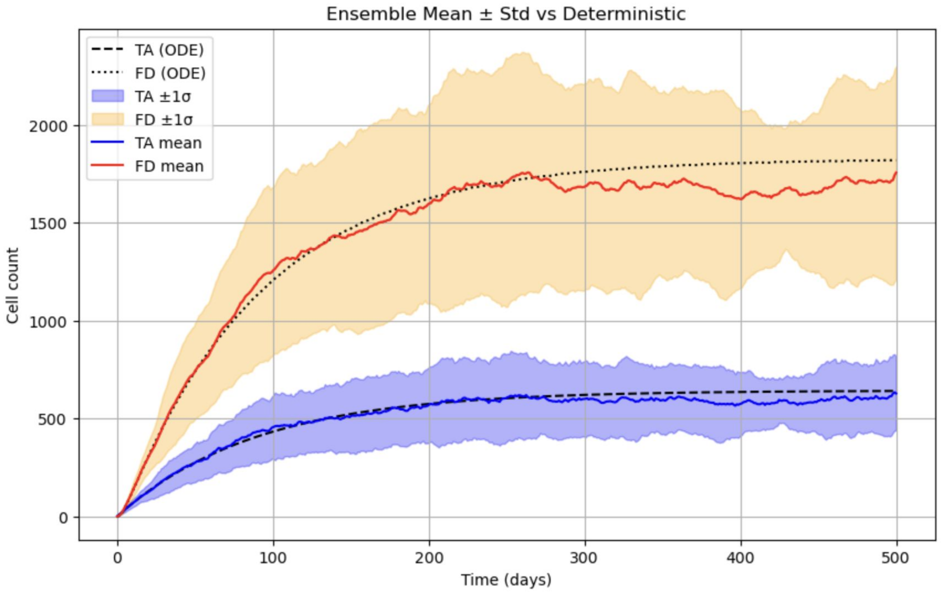
Various γ and d values always result in steady system

Stochastic Modeling (Gillespie Algorithm)

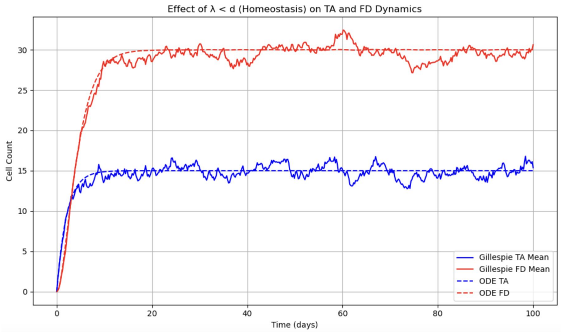
UNIVERSITY of WASHINGTON

ODE v.s. Stochastic Simulation

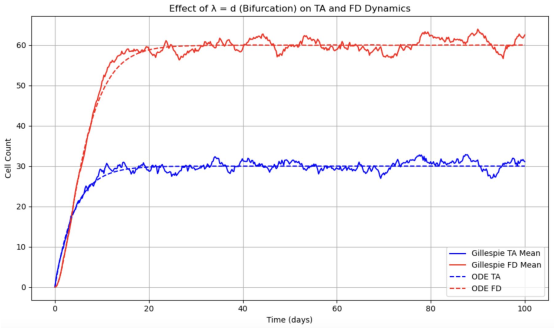
- **Solid Lines**: Show the average number of TA and FD cells from many Gillespie simulations.
- **Shaded Bands**: Represent the variability around the average for TA and FD.
- **Dashed Lines**: Show predictions from the deterministic ODE model.
- **Comparison**:
  - The means from the simulations closely match the ODE predictions.
  - The stochastic simulations have more spread as time goes on, especially for FD cells.
- **Insight**: Deterministic models capture the average trend, but Gillespie simulations show how much real systems can vary due to randomness.



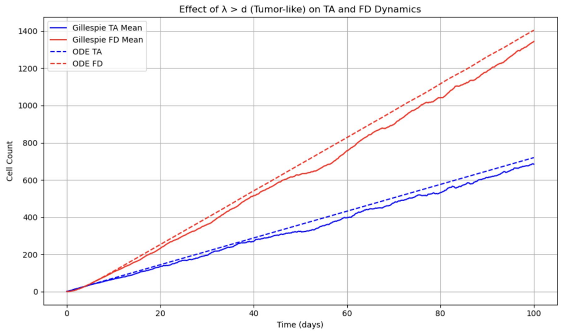
TA and FD Dynamics with Stochasticity



Homeostasis ($\lambda < d$)



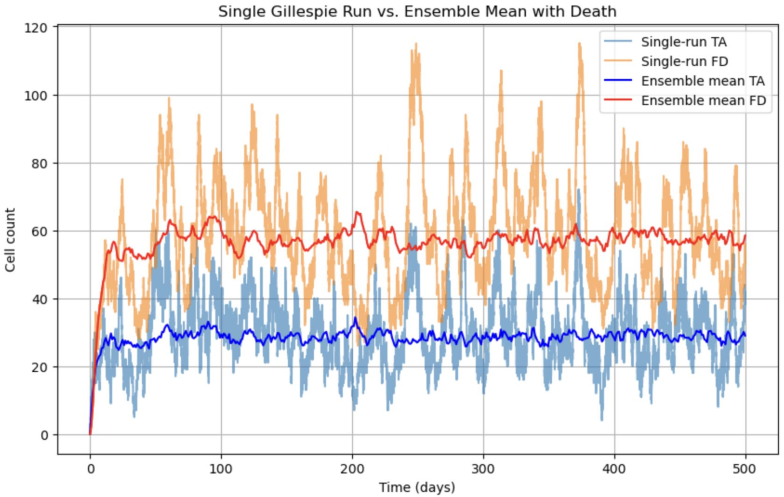
Bifurcation ($\lambda = d$)

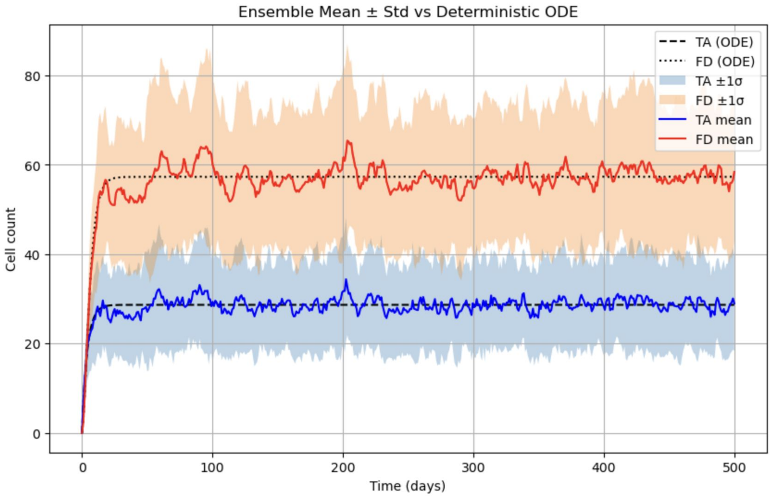


Tumor-like Growth ($\lambda > d$)

The stochastic models show the same dynamics that we would expect based on the ODE stability analysis.

Gillespie Simulation With TA Cell Death





- **More fluctuation:** TA and FD curves show bumpier paths with more ups and downs.
- **Stabilized growth:** Cell counts settle around a steady level instead of growing forever.
- **Wider spread:** Variability (±1σ) increases, especially for FD, due to added randomness from cell death.
- **Biologically realistic:** Models natural cell divide and die to maintain tissue balance.

UNIVERSITY of WASHINGTON

Future Directions with Stochasticity!

Introducing stem cell dynamics

- Asymmetric Division (default behavior): $S \rightarrow S + TA$
- Symmetric Renewal: $S \rightarrow S + S$
- Symmetric Differentiation (Loss): $S \rightarrow \varnothing$

Alternative additions of stochasticity

- Uses alternative methods for determining probability of an event happening such as a binomial distribution.

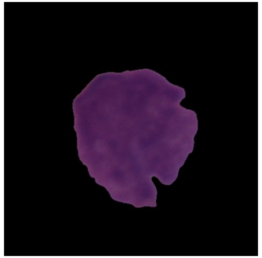
Machine Learning Modeling

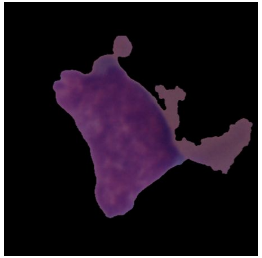
UNIVERSITY *of* WASHINGTON

Background

Dataset Overview: Acute Lymphoblastic Leukemia (ALL)

- **Source:** ALL Challenge Dataset – ISBI 2019
- **Total Images:** 15,135 segmented cell images from 118 patients
- **ALL**: One of the most prevalent forms of childhood cancer.
- **Classes:**
  - Normal cells
  - Leukemia blast cells
- **Annotations:** Verified by expert oncologists
- **Image Format:** .bmp files from microscopic scans
- **Image Quality:** Preserves real-world imperfections (e.g., staining, lighting)

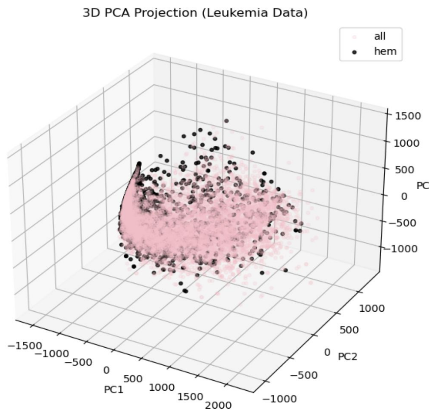


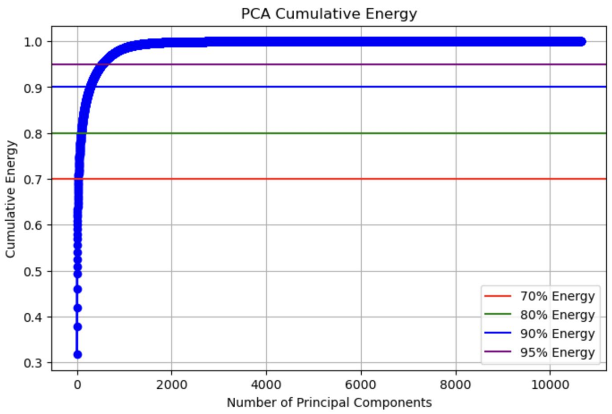


PCA

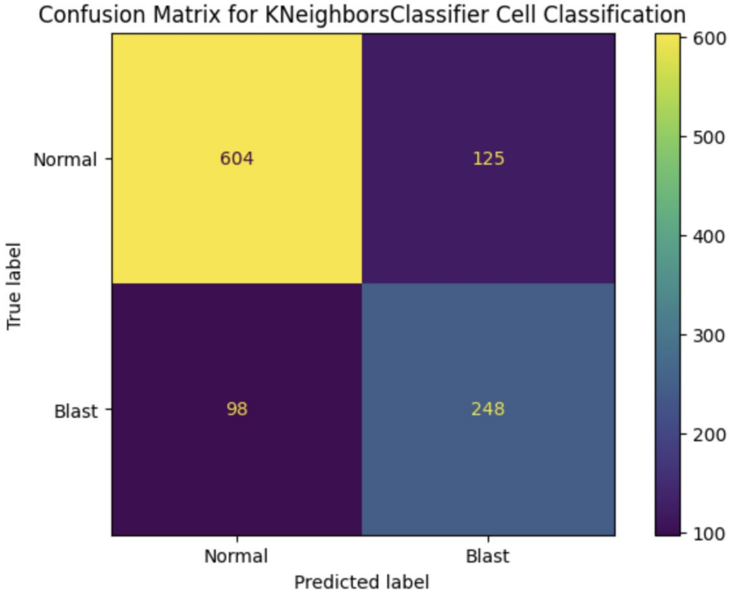
PCA Explained Variance:

- Indicates how many principal components are needed to retain a certain amount of total variance in the dataset.
- ~300 components retain 90% of the variance, and ~500 reach 95%, suggesting strong feature redundancy.
- Helps determine a suitable dimensionality reduction threshold before classification.
- The leukemia and normal cells are not linearly separable

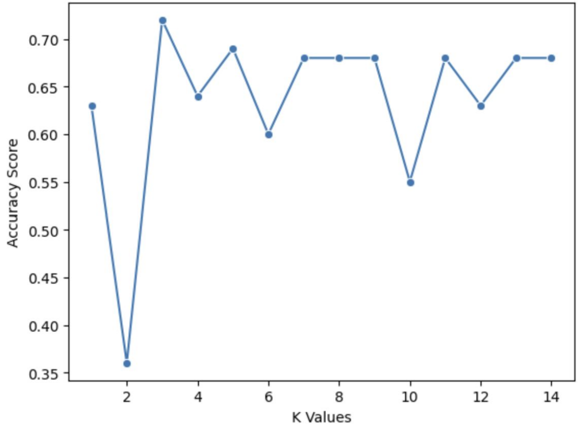




KNN

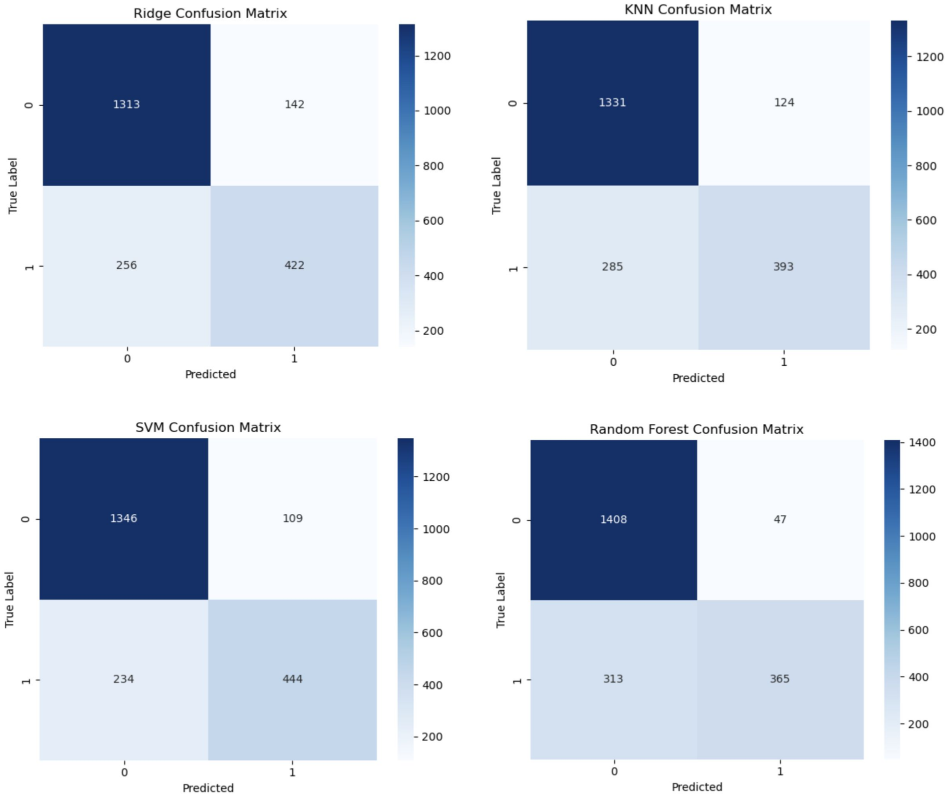


- Not linearly separable needed to use a higher order classifier
- Training Score: 0.8994413407821229
- Testing Score: 0.7925581395348837
- Alternative number of neighbors (k) as hyperparameter tuning
  - Optimal number of neighbors: 3



Classifiers Comparison

- **Top-left (TN):** Predicted normal, actually normal
- **Top-right (FP):** Predicted leukemia, actually normal
- **Bottom-left (FN):** Predicted normal, actually leukemia
- **Bottom-right (TP):** Predicted leukemia, actually leukemia



Takeaways:

- **SVM** is the most balanced and effective overall.
- **Random Forest** is very accurate on the normal class but struggles with leukemia class.
- **KNN** and **Ridge** are simpler and faster but less reliable in distinguishing complex patterns.

UNIVERSITY of WASHINGTON

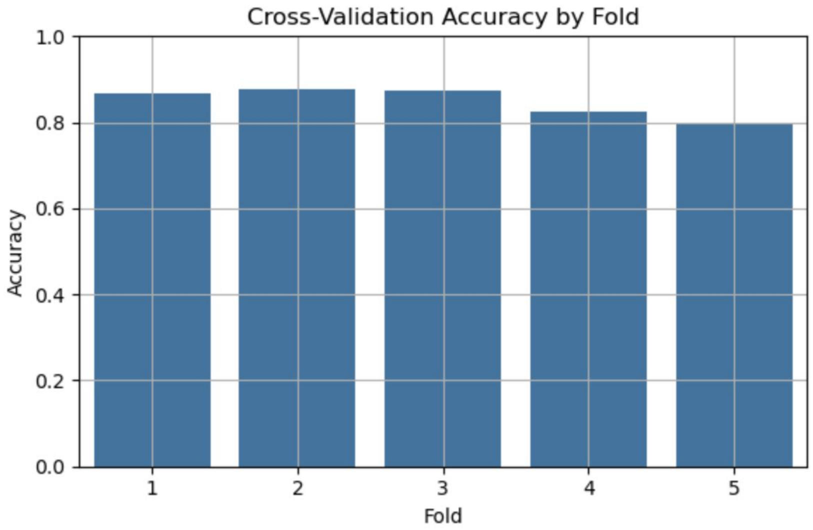
Classifiers Comparison

Model	Accuracy (%)	Precision (Leukemia)	Recall (Leukemia)	Strengths	Weaknesses
SVM	**83.92**	80%	**65%**	Highest overall accuracy, balanced performance	Still moderate recall on leukemia
Random Forest	83.12	**89%**	54%	High recall for normal class (97%)	Low recall for leukemia class
Ridge	81.34	75%	62%	Simpler and interpretable model	Lower performance on minority class
KNN	80.83	76%	58%	Performs comparably to Ridge	Sensitive to local variations

Model Evaluation – Random Forest

- Performed 5-fold Stratified Cross-Validation
- Mean Accuracy: 84.8%
- Standard Deviation: ±3.1%
- Accuracy remains strong across folds with mild variance, showing model stability and generalization.

Metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Accuracy	86.6	**87.9**	87.2	82.5	79.8

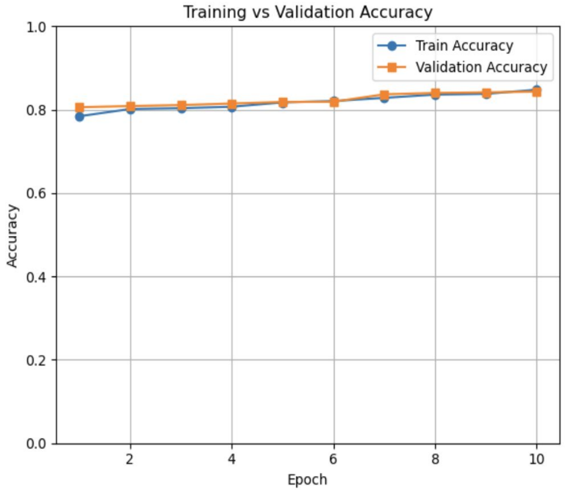


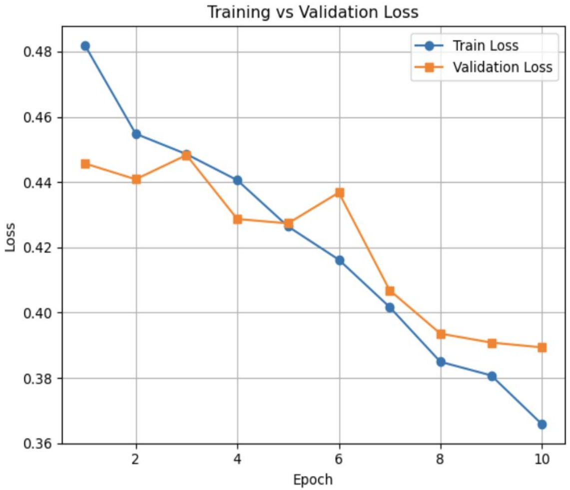
Convolutional Neural Network Training Performance

CNN

- **Model:** Simple CNN with 2 Conv layers and 2 FC layers
- **Input Size:** 64×64 RGB cell images
- **Epochs Trained:** 10
- **Train Accuracy:** Improved from **78.4% → 84.8%**
- **Validation Accuracy:** Rose from **80.6% → 84.3%**
- **Loss:** Steady decline in both training and validation loss
- **No signs of overfitting:** Validation closely follows training

Epoch 1 | Train Acc: 78.42% | Val Acc: 80.59%
Epoch 2 | Train Acc: 80.16% | Val Acc: 80.87%
Epoch 3 | Train Acc: 80.36% | Val Acc: 81.11%
Epoch 4 | Train Acc: 80.71% | Val Acc: 81.48%
Epoch 5 | Train Acc: 81.74% | Val Acc: 81.86%
Epoch 6 | Train Acc: 82.07% | Val Acc: 81.90%
Epoch 7 | Train Acc: 82.83% | Val Acc: 83.68%
Epoch 8 | Train Acc: 83.62% | Val Acc: 84.01%
Epoch 9 | Train Acc: 83.79% | Val Acc: 84.15%
Epoch 10 | Train Acc: 84.79% | Val Acc: 84.34%





Thank you for listening!

UNIVERSITY *of* WASHINGTON