Classification of Acute Lymphoblastic Leukemia using Persistent Homology

Waqar Hussain Shah

Collaborators: Abdullah, Rider Jaimes Reategui, Sohail Iqbal, and Syeda Rafia Fatima

June 26, 2024, Lyapunov Seminar

Outline

Introduction

- Shape of data
- Interpretable and intutive
- Persistent Homology

Simplicial Homology

Simplex

An affine independent subset let say, $\{u_0, u_1, \ldots, u_m\}$ of $\mathbb{R}^n$. Convex set spanning by this set, represented by $[u_0, u_1, \ldots, u_m]$, is called affine m-simplex, and with vertices $u_0, u_1, \ldots, u_m$.

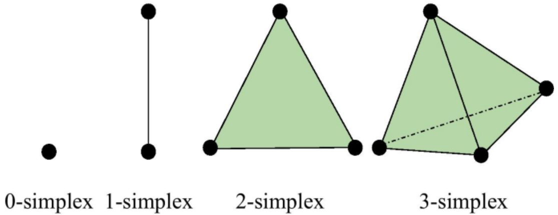
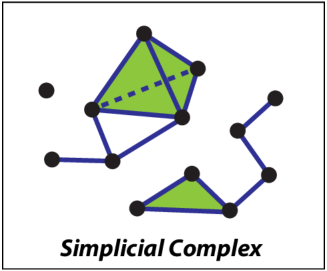


Figure: Simplexes

Simplicial Complex

Consider K is a simplicial complex in space of Euclidean and its a collection of finite simplexes such that;

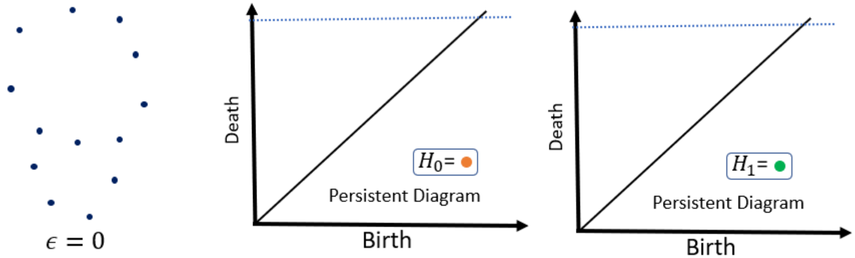
1. If $x \in K$ then each face of s belongs to K.
2. If $x, y \in K$, then $x \cap y$ is either empty or a common face of x and y.



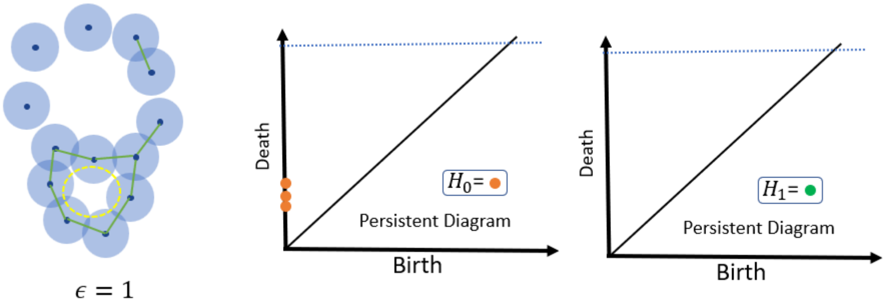
Persistent Homology

- point cloud
- Construct a filtration of simplicial complexes
  $\triangle_0 \subset \triangle_1 \subset \triangle_2 \ldots \subset \triangle_n$
- Apply the Homology functor H_n,
  $H_n(\triangle_0, F) \rightarrow H_n(\triangle_1, F) \rightarrow H_n(\triangle_2, F) \ldots \rightarrow H_n(\triangle_n, F)$
- Summarize birth and death times of above homological features in a persistent diagrams [8](or barcodes).

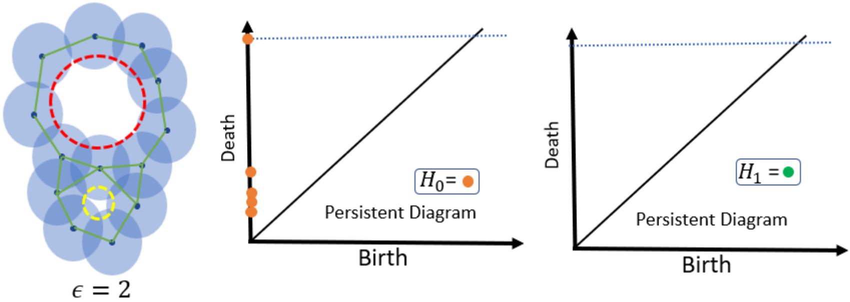
Persistent Diagram



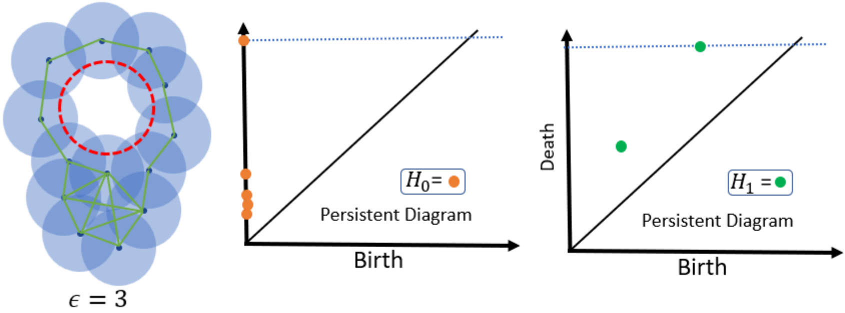
Persistent Diagram



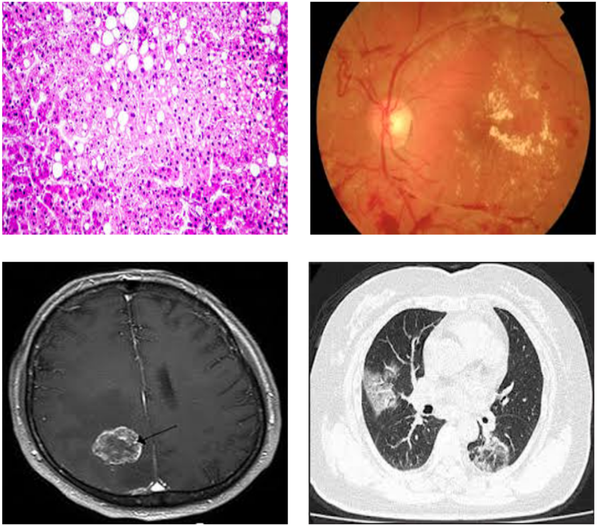
Persistent Diagram



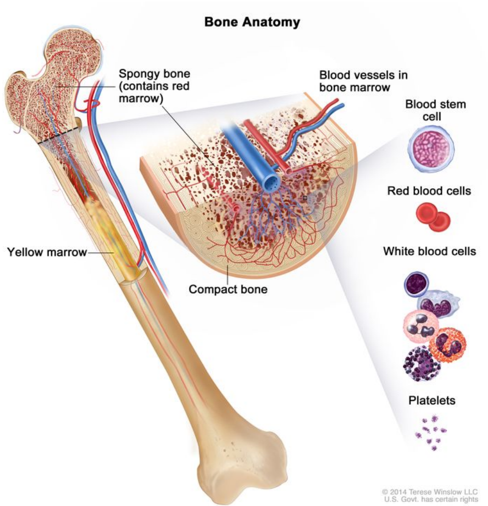
Persistent Diagram



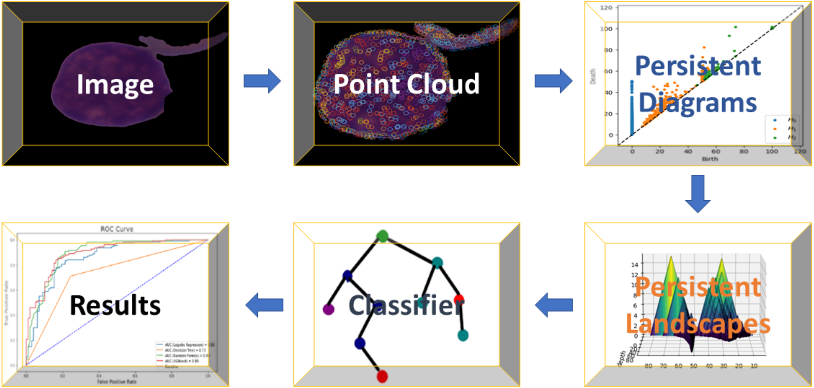
Biomedical Images



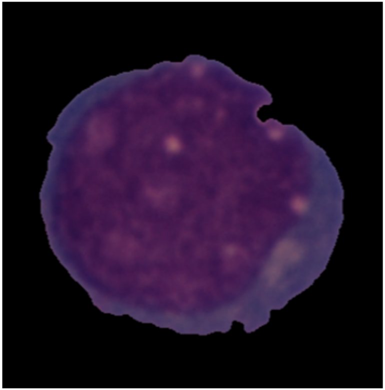
Acute Lymphoblastic Leukemia



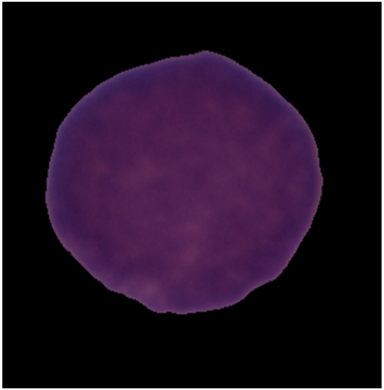
Methodology



Microscopic images of blood cells

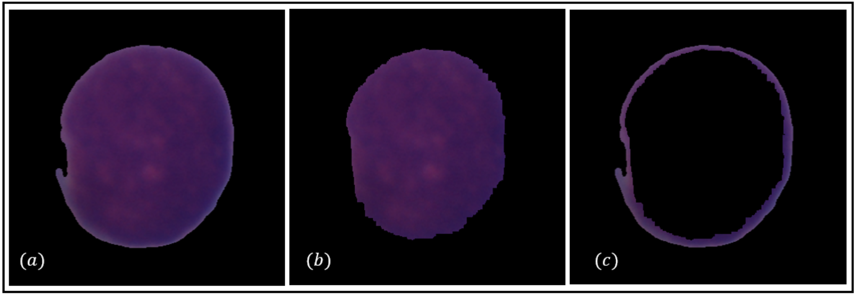


Cancer Cell

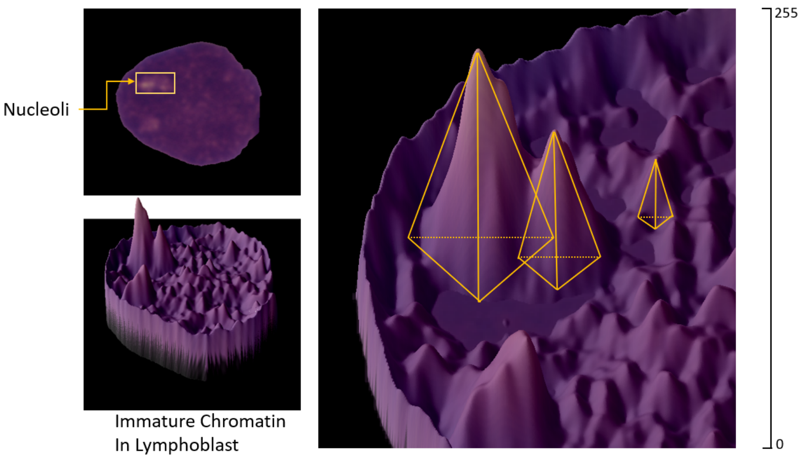


Normal Cell

Quantification of the cell



Quantification of the cell



Point cloud of a cell

Feature Point Cloud: Let (X, d) be a point cloud, and (F, d_F) be a compact feature space. We define projection map $\alpha : X \to F$ such that $\alpha(p)$ is a feature of the point p. Let $U = \{U_i\}_{i \in I}$ be a covering of F, where I is finite.

$$X \text{ has a covering } C(X) = \{\alpha^{-1}(U_i) : i \in I\}$$

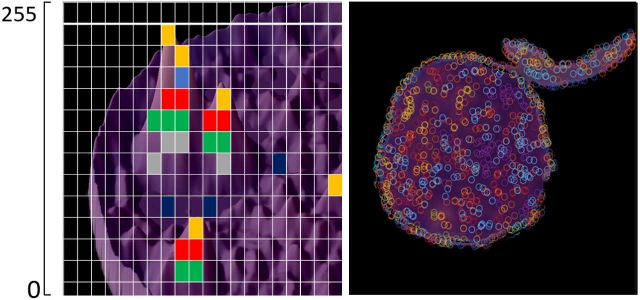
Let $X(U)$ be the connected components of $C(X)$. The vertices of FPC are naturally consider as the filtration parameter of the simplexes.

Point cloud of a cell

For any image M, we take point cloud in $\mathbb{R}^3$ defined as,

$$X_M = \{(i, j, p) : p \text{ is the intensity value at the point } (i, j)\}$$

and $F_M = [0, 255]$. For some finite cover $W = \{W_i\}_{i \in I}$ we get the FPC, denoted as $X_M(W)$.



Point cloud of a cell

For any image M, we take point cloud in $\mathbb{R}^3$ defined as,

$$X_M = \{(i, j, p) : p \textit{ is the intensity value at the point } (i, j)\}$$

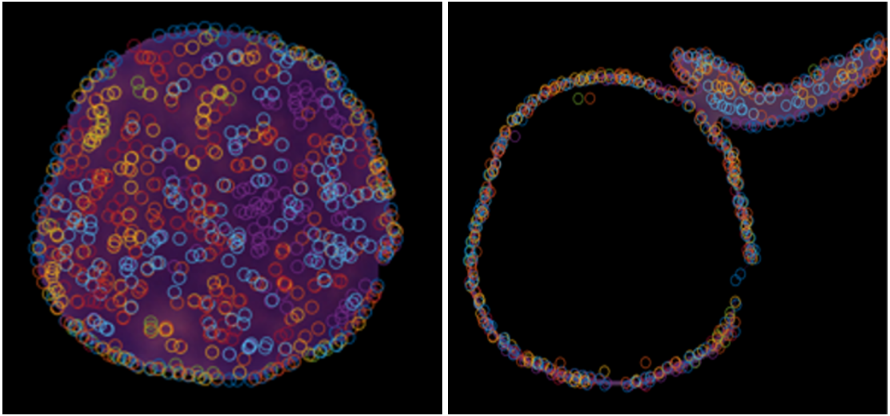
and $F_M = [0, 255]$. For some finite cover $W = \{W_i\}_{i \in I}$ we get the FPC, denoted as $X_M(W)$.
To calculate the PH we use the following Vietoris-Rips filtration,

$$VR_0(X_M(W)) \subset VR_{0.5}(X_M(W)) \subset \ldots \subset VR_{500}(X_M(W)),$$

where

$$VR_\epsilon(X_M(W)) = \{A \subset X_M(W) : d(a_1, a_2) < 2\epsilon, \forall a_1, a_2 \in A\}.$$

Point cloud of a cell

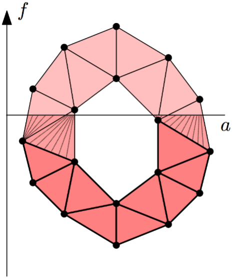


Lower Star Filtration

1. Simplicial complex witha real value function on its vertices $f : Ver(\triangle) \to \mathbb{R}$
2. Extend f to $\triangle$
3. Simplicail complex for ϵ

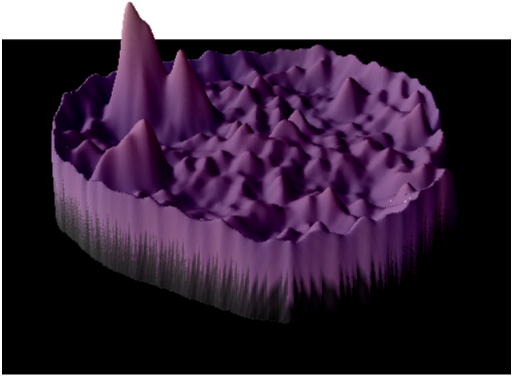
$$\triangle_\epsilon = \{\sigma \in \triangle : max_{v \in \sigma} f(v) \leq \in\}$$

Each pixel is taken as vertex with its intensity value $f : V \to \mathbb{R}$. Each pixel is connected to its 8 or less neighbours.

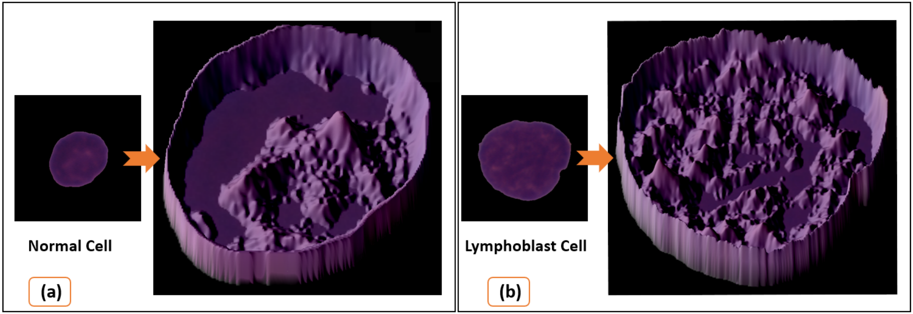


Lower Star Filtration

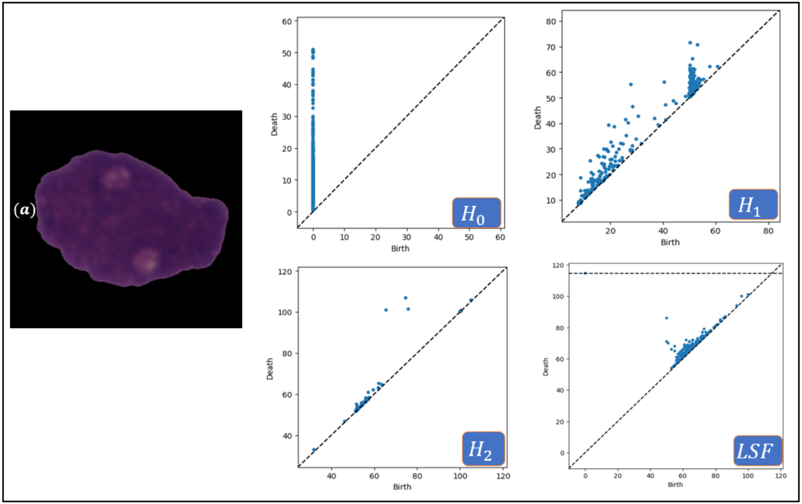
Birth times in persistent diagram are the local minimums, and deaths times are saddle points in a 0-dimensional PD.



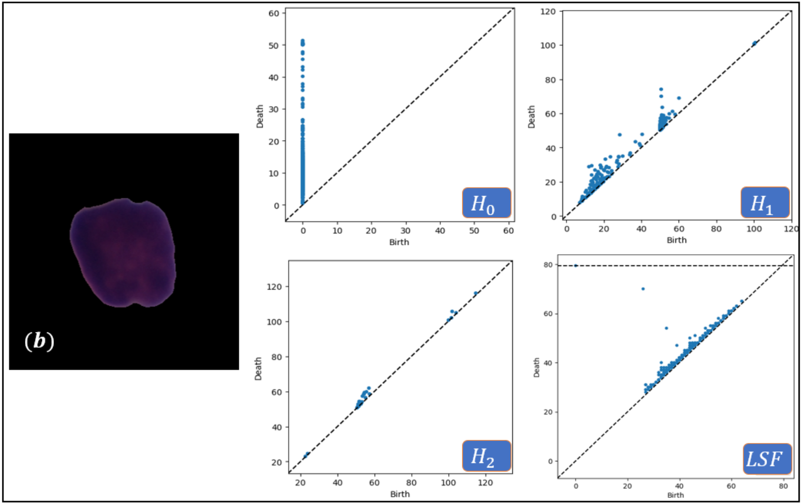
Lower Star Filtration



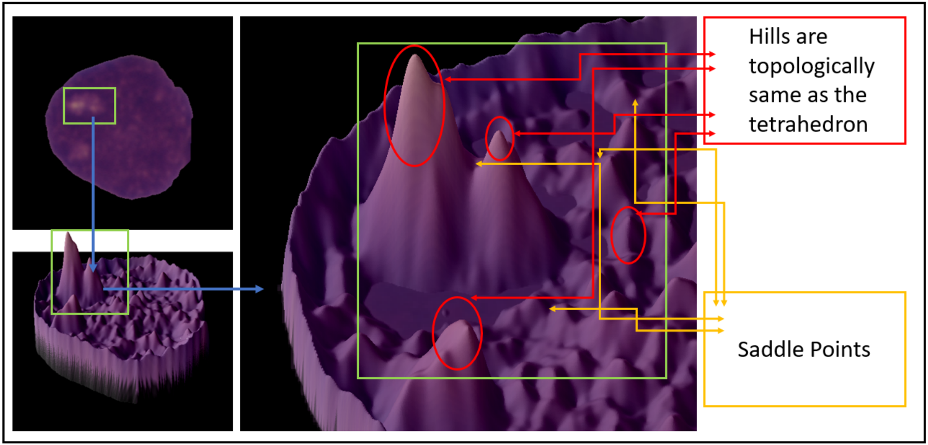
Persistent diagams



Persistent diagams



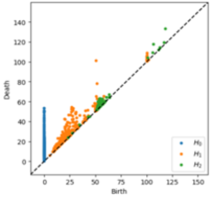
Persistent diagams



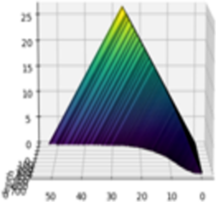
Persistent Landscape

- To delvelop a classification model, we need a kernel on the space of persistent diagrams.
- A PD is a plotting of birth rate against death rate for key features of a specific type of homology groups.
- An alternative function in which the points of a PD are rotated from birth-death pairs (b, d) to $(x, y) = ((d + b)/2, (d - b)/2$.
- For each point, a linear function is developed and the kth landscape function [1] is the largest value of the tents at any point in the t horizontal axis.

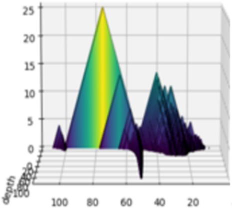
Persistent Landscape



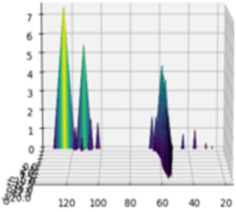
Persistent Diagram



PL of 0-dimensional



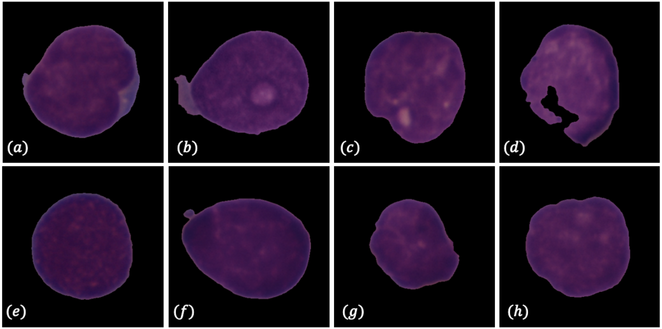
PL of 1-dimensional



PL of 2-dimensional

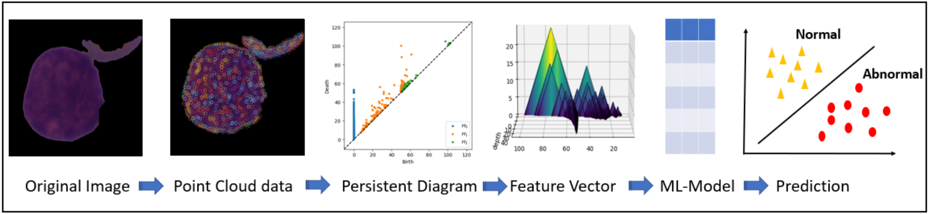
Dataset

The Cancer Imaging Archive (TCIA) provides the C-NMC-2019 dataset, used in the ISBI 2019 challenge for classifying normal vs. malignant cells in B-ALL white blood cancer images. The dataset includes 118 subjects, with training and testing sets divided into 73 and 45 subjects, respectively, totaling 10,661 training images and 1,867 preliminary images, plus 2,586 final images.



Summary of the model

As a result, each PD is converted into a vector length of 2500. This complete procedure is applied to each collection of persistent diagrams; at the end, a single image of each of the four PDs has a vector length of $[100]^2$.



ALL models with Topological Approach

Reference	Method	F1-score	Accuracy
	Our Model	94%	92%
[9]	ResNet	92%	–
[7]	Augmentation and Normalization	87%	–
[6]	ALNet	96%	91%
[4]	SDCT-AuxNet$^{\theta}$	94%	93%
[2]	ResNet101-9	88%	–
[10]	LeuFeatx	96%	–
[3]	ShuffleNet	96%	96%
[11]	ResNet50	87%	84%
[5]	ANN	–	89%

Individual Topological Features

Topological feature	Recall	Precision	Accuracy	F1-score
H_0	87%	72%	70%	79%
H_1	93%	81%	80%	87%
H_2	88%	80%	77%	84%
LSF	91%	85%	83%	88%

Table 2 Performance Metrics for Different Classes of topological features

CU LA GOS

Conclusion

- Introduced a TDA-based method to classify leukemia cells, distinguishing lymphoblasts from normal cells.
- Showed TDA techniques outperform many neural networks in classifying microscopic images.
- Proposed an interpretable model improving leukemia cell detection and diagnosis.
- Suggested integrating TDA with deep learning for future advancements.

[1] Peter Bubenik and Paweł Dłotko. "A persistence landscapes toolbox for topological statistics". In: *Journal of Symbolic Computation* 78 (2017), pp. 91–114.

[2] Yao-Mei Chen et al. "Classifying microscopic images as acute lymphoblastic leukemia by Resnet ensemble model and Taguchi method". In: *BMC bioinformatics* 22.Suppl 5 (2021), p. 615.

[3] Pradeep Kumar Das and Sukadev Meher. "Transfer learning-based automatic detection of acute lymphocytic leukemia". In: *2021 National Conference on Communications (NCC)*. IEEE. 2021, pp. 1–6.

[4] Shiv Gehlot, Anubha Gupta, and Ritu Gupta. "SDCT-AuxNetθ: DCT augmented stain deconvolutional CNN with auxiliary classifier for cancer diagnosis". In: *Medical image analysis* 61 (2020), p. 101661.

[5] Rana Zeeshan Haider et al. "Beyond the in-practice CBC: the research CBC parameters-driven machine learning predictive modeling for early differentiation among leukemias". In: *Diagnostics* 12.1 (2022), p. 138.

[6] Malathy

Thanks