SVM Model for Blood Cell Classification using Interpretable Features Outperforms CNN Based Approaches

William Franz Lamberti [1]

George Mason University

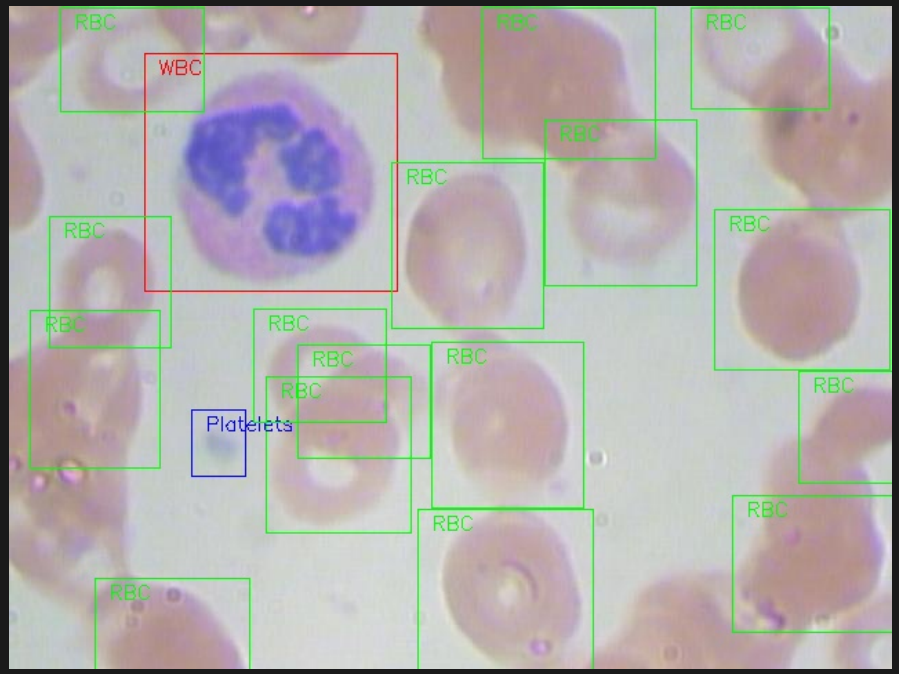
June 4, 2020

[1]MS Statistical Science
PhD Candidate Computational Sciences and Informatics

Outline

Introduction: Blood Cell Classification

- Important task in Health Sciences
- Counts are used to measure overall health of patient
- Often done by hand, which is tedious



Introduction: CNNs

- Convolution Neural Networks (CNNs) very popular method in Computer Vision
- Lots of different applications and very powerful
- Difficult to interpret and explain
- Require a large amount of data

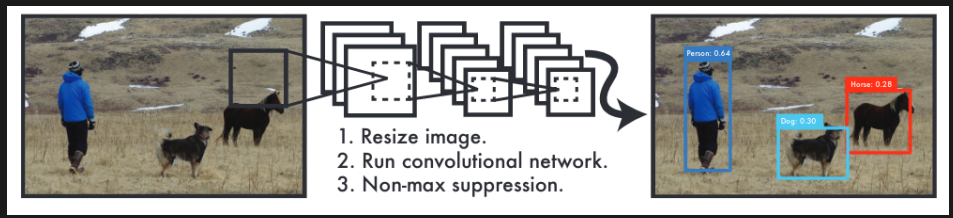


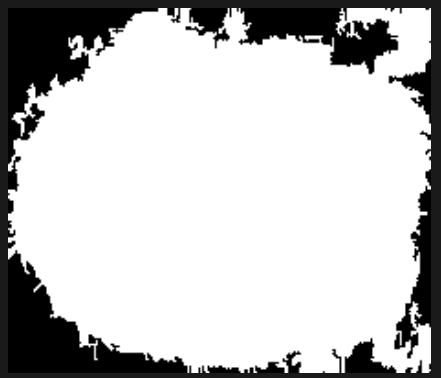
Image from: Redmon, Joseph et al. "You Only Look Once: Unified, Real-Time Object Detection." arXiv.org (2016): n. pag. Web.

Goal

- Build model that outperforms state of the art in classifying objects
- Use interpretable metrics

Data

- Publicly available BBCD dataset: `https://github.com/Shenggan/BCCD_Dataset`
- Classes
  - Red Blood Cells (RBCs): 4153
  - White Blood Cells (WBCs): 372
  - Platelets:361
- Objects are extracted with given annotation file and universal segmentation operators are applied



Model: Metrics

$\vec{m}_{q,i}$	Metric
1	White EI
2	Black EI
3	SP value
4	1^{st} Eigenvalue
5	2^{nd} Eigenvalue
6	Eccentricity
7	White Bounding Box Count
8	Black Bounding Box Count
9	Circularity
10	Number of Corners

Model: Algorithm

- Split data into training (70%) and validation data (30%)
  - Training data = data used to build the model
  - Validation data = data never used to build the model
- Perform 5-folds CV on training data to determine SVM Polynomial parameters
- Using learned parameters, build model on all of training data
- Evaluate model on validation data

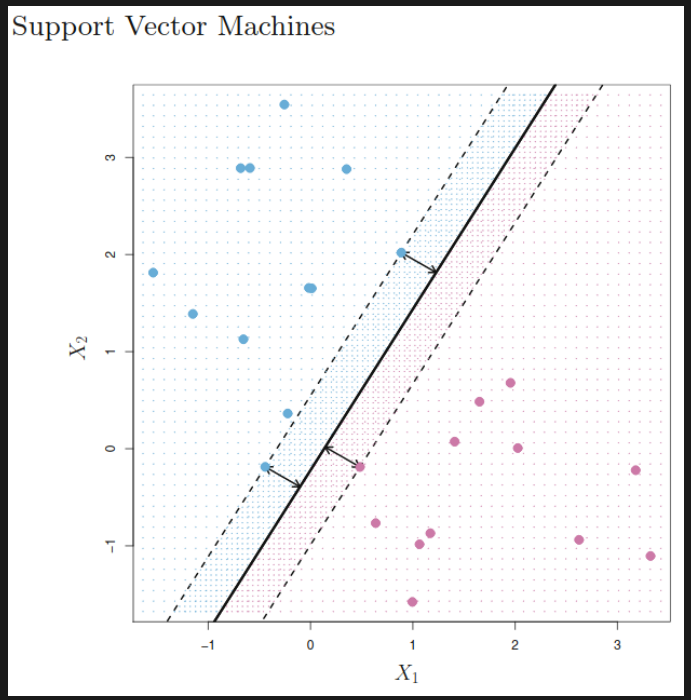


Image from: An Introduction to Statistical Learning with Applications in R

Results: Confusion Table

Predicted/Truth	Platelet	WBC	RBC
Platelet	249 **(106)**	1 **(1)**	1 **(6)**
WBC	0 **(0)**	93 **(93)**	6 **(3)**
RBC	4 **(2)**	17 **(17)**	2901 **(1236)**

Results: Classification Rates

Approach	Platelet	WBC	RBC
SVM (Lamberti)	**98.1%**	83.8%	**99.3%**
Tiny YOLO	96.1%	86.9%	96.4%
VGG-16	73.0%	**100%**	90.9%
ResNet50	79.8%	95.1%	87.3%
InceptionV3	87.8%	**100%**	96.4%
MobileNet	74.2%	93.4%	83.6%

CNN approaches from: Mohammad Mahmudul Alam, and Mohammad Tariqul Islam. "Machine Learning Approach of Automatic Identification and Counting of Blood Cells." Healthcare Technology Letters 6.4 (2019)

Conclusion

- SVM outperforms all other approaches
  - Overall Mean Outperformance: 5%
  - Underperforms for WBC
- Future Work
  - Develop segmentation technique without need for annotations
  - Improve classification
  - Develop for other applications such as COVID-19
- Code and Manuscript available
  - Code: `https://github.com/billyl320/bccd_svm`
  - Manuscript: `https://github.com/billyl320/bccd_svm/blob/master/Lamberti_SDSS_short_paper.pdf`

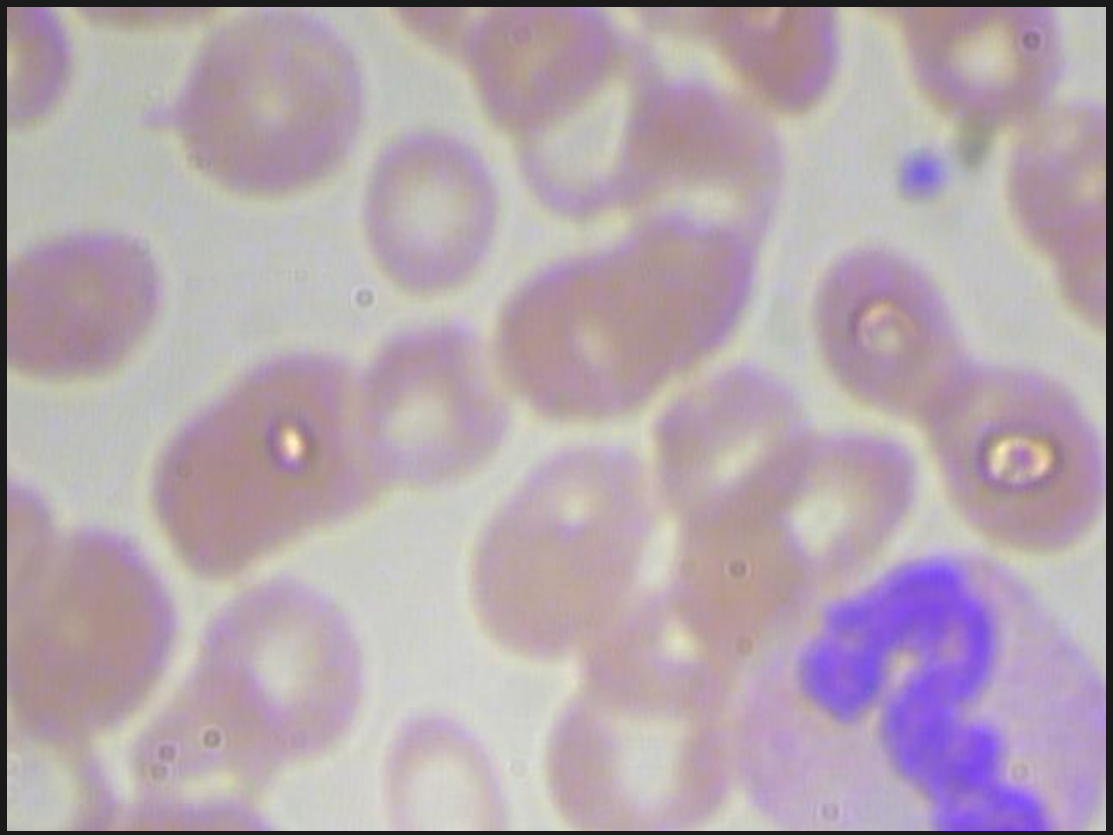
Acknowledgements

- Committee
  - Jason Kinser
  - William Kennedy
  - Michael Eagle
  - David Holmes
- Sounding Board
  - John Schuler

Acknowledgements

- GMU, Office of the Provost
- 2020 SDSS Student and Early-Career Award

Any Questions?

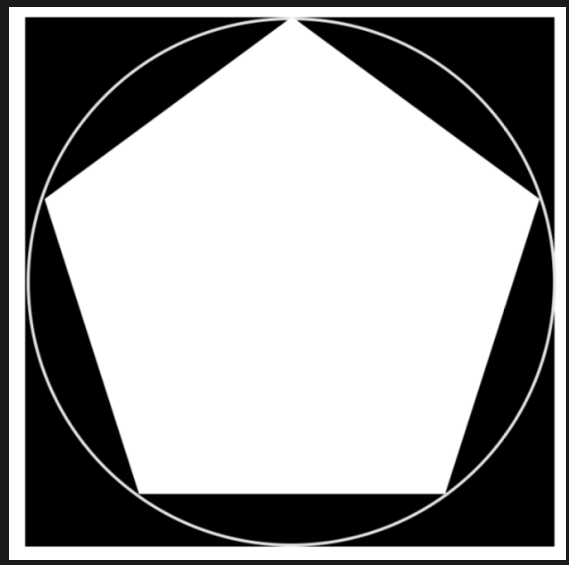


Email: wlamber2@gmu.edu

Extra

Metrics: SPEI

- Shape Proportion (SP)
  - Describe shape as a proportion
- Encircled Image-Histogram (EI)
  - Black and White pixel counts
  - Realization of SP



Metrics: Eccentricity and Eigenvalues

- First Eigenvalue = Relative measure of major axis
- Second Eigenvalue = Relative measure of minor axis
- Eccentricity = ratio of first over second
- Ch. 18 of "Image Operators: Image Processing in Python", Kinser for more details

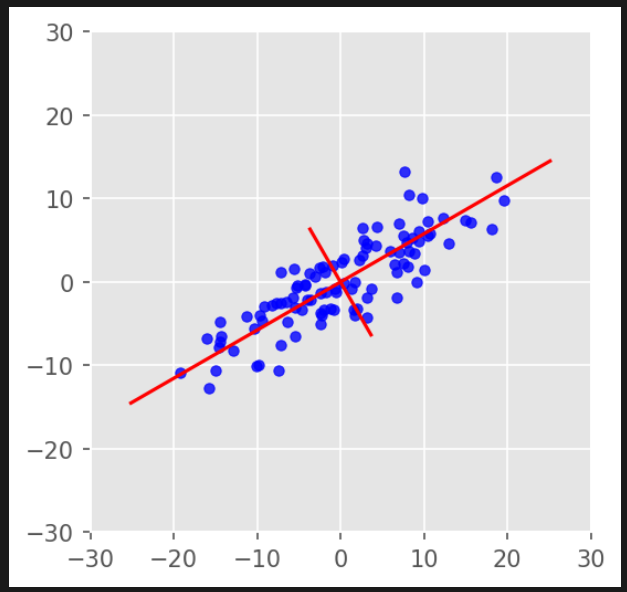


Image from: `https://bit.ly/35FH8h5`

Metrics: Rectangularity

- Minimum bounding box (MBB) counts
  - Black and White picel counts
  - Jointly measure how similar object is to a rectangle

Image from: "Minimum Bounding Rectangle." Encyclopedia of Geographic Information Science, edited by Karen K. Kemp, SAGE Publications, 2008, pp. 286-287. Gale eBooks

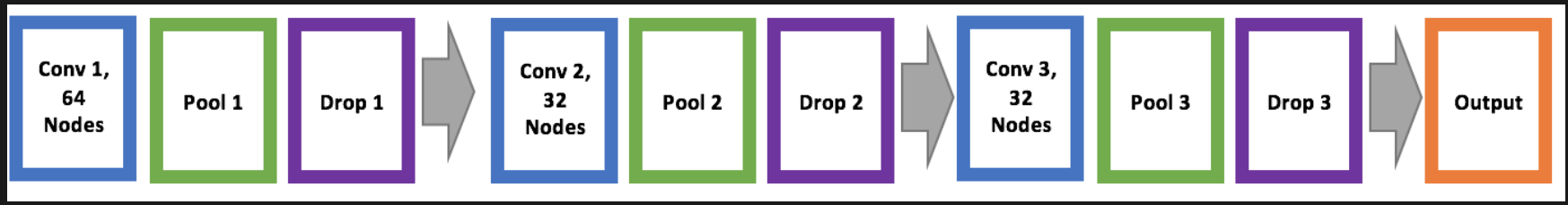
Metrics: Circularity

- Circularity = measure of how circular a shape is
- Has unique values for regular polygons and circles
- Based on area and perimeter

Metrics: Number of Corners

- Finds the edges in the x and y direction
- Takes Gaussian smoothing kernel of both
- Create tensor image: A
- Obtain final image by: $H = det(A) - k * trace(A)^2$

CNN Architecture



- Convolution Layers: ReLU Activation and l_2-norm regularization
- Pooling Layers: Batch size of 2×2 and max pooling
- Output Layer: Softmax operation
- Dropout Layer: 0.20
- Early stopping: Max of 100 epochs

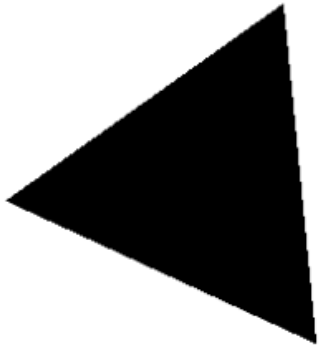
J. Gu et al., "Recent advances in convolutional neural networks," Pattern Recognition, vol. 77, pp. 354–377, May 2018.

CNN Results

(a) Figure is the input triangle with a side length of 25.

(b) Figure is an example of the input heptagon with a side length of 25.

(c) Figure is an example of the input octagon with a side length of 25.

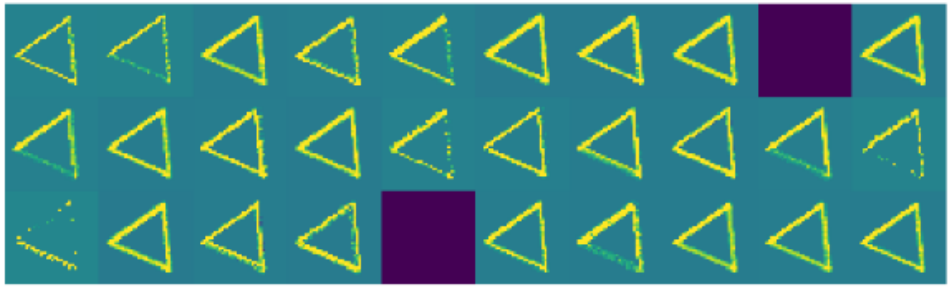


(d) Figure is an example of the input triangle with a side length of 120.

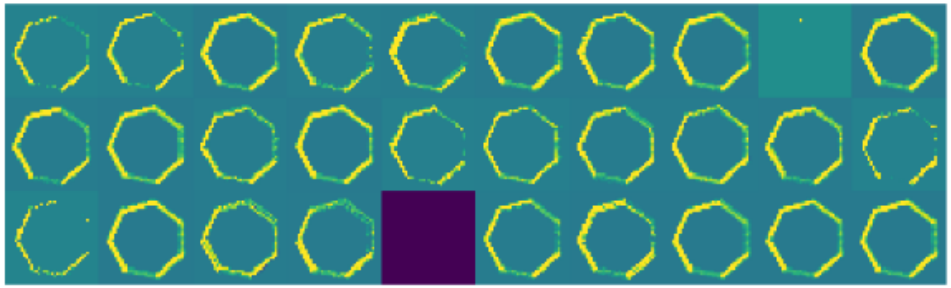
(e) Figure is an example of the input heptagon with a side length of 120.

(f) Figure is an example of the input octagon with a side length of 120.

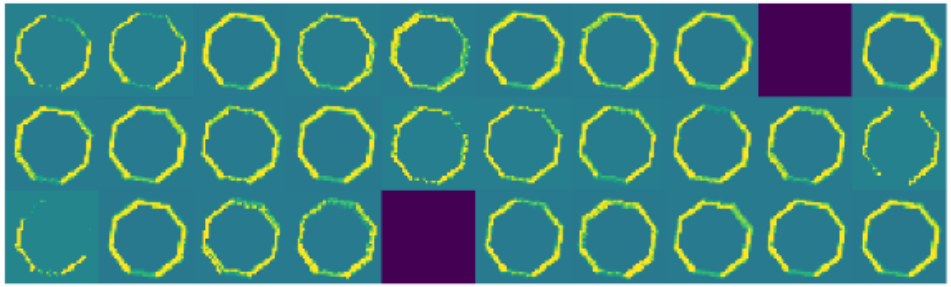
CNN Results



(a) Figure shows the third dropout layer of the triangle with a side length of 120.

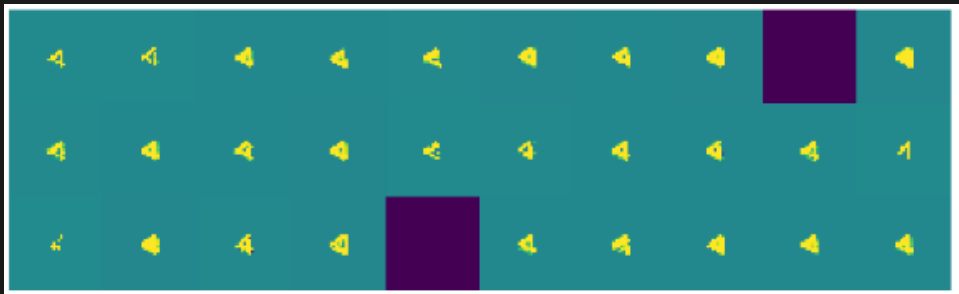


(b) Figure shows the third dropout layer of the heptagon with a side length of 120.

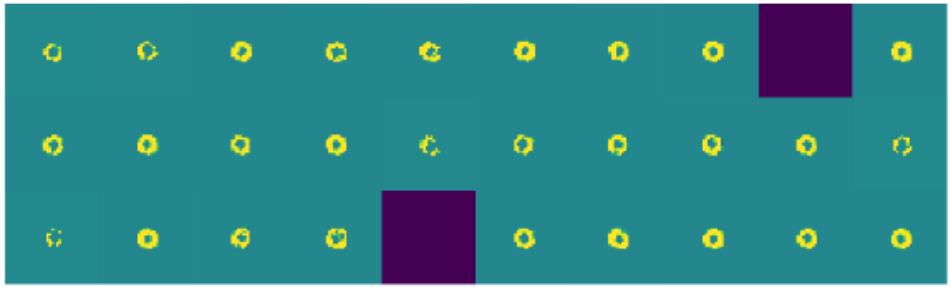


(c) Figure shows the third dropout layer of the octagon with a side length of 120.

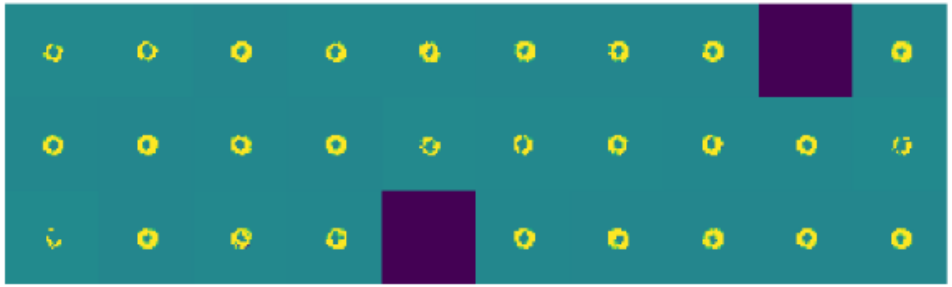
CNN Results



(a) Figure shows the third dropout layer of the triangle with a side length of 25.



(b) Figure shows the third dropout layer of the heptagon with a side length of 25.

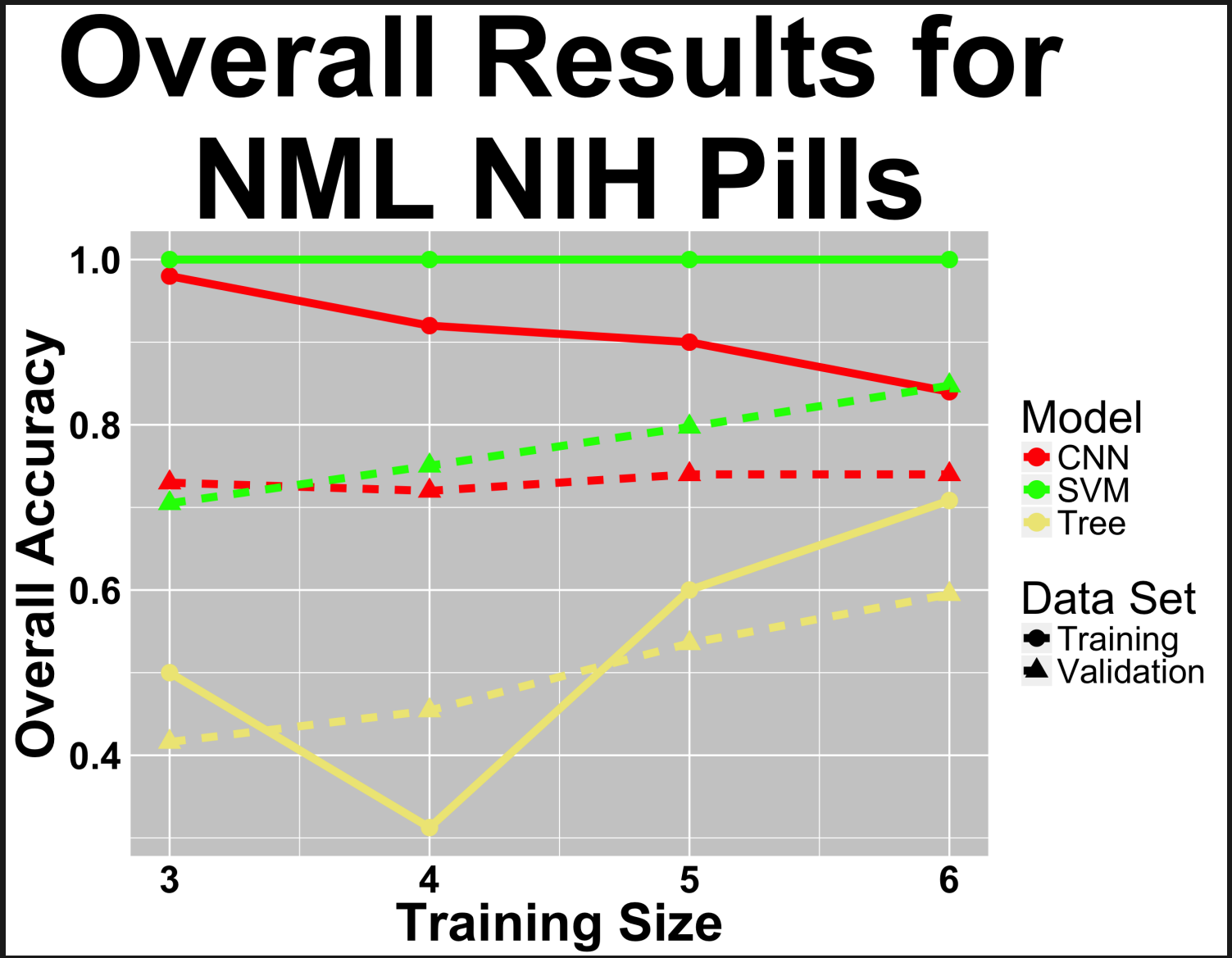


(c) Figure shows the third dropout layer of the octagon with a side length of 25.

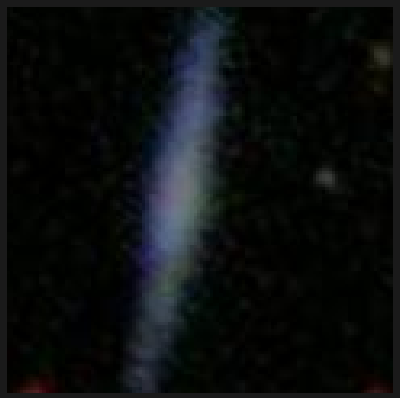
Pill Shapes - Examples

Class	Counts
Triangle	12
Quadrilateral	8
Pentagon	12
Hexagon	8
Total	40

Experiments: Pill Shapes - Results

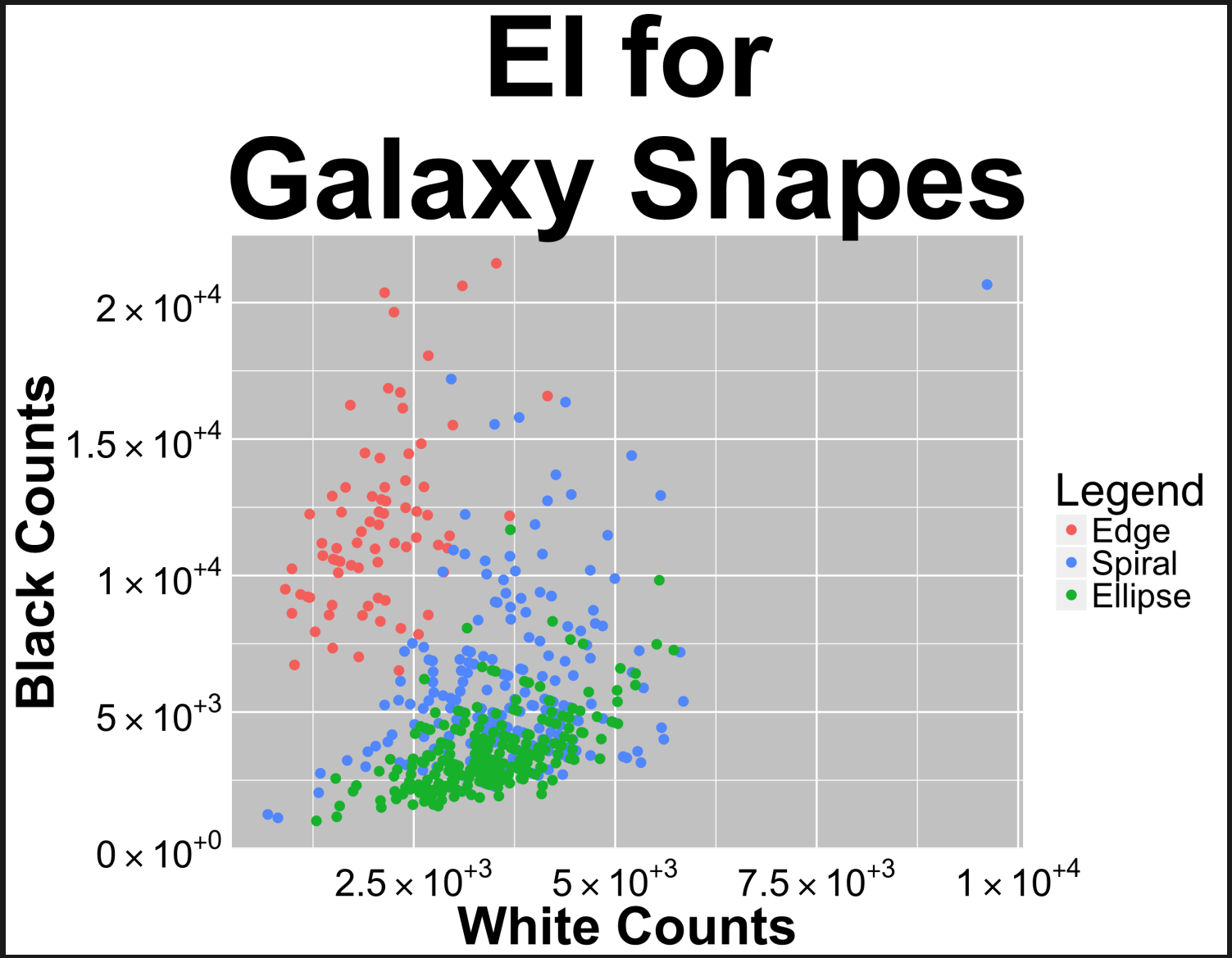


Experiments: Galaxy Shape - Examples



- Counts
  - Edge: 75
  - Spiral: 223
  - Ellipse: 225

Experiments: Galaxy Shapes - Results



Experiments: Galaxy Shapes - Results