SVM-Based Models for Pill Shape Classification

William Franz Lamberti [1]

University of Virginia

June 3, 2021

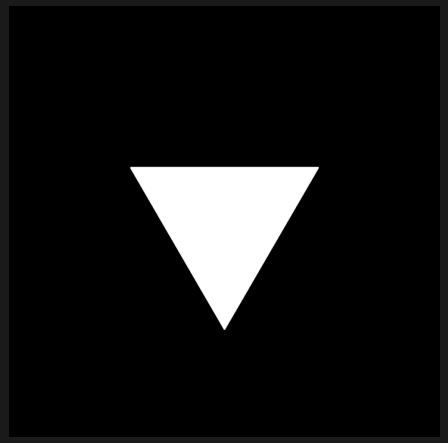
[1]MS Statistical Science
PhD Computational Sciences and Informatics

Outline

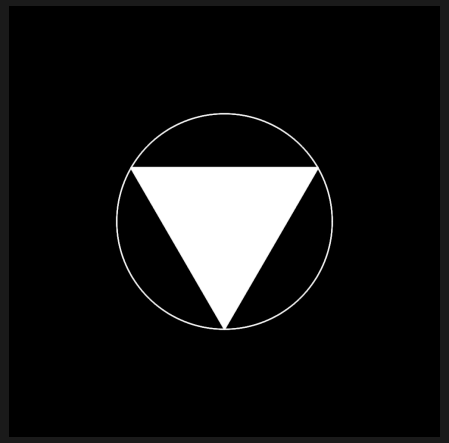
Introduction

- Pill identification not where it needs to be
- Pill shape classification best via Maddala et. al.
- Goal: Develop a model for pill shape classification

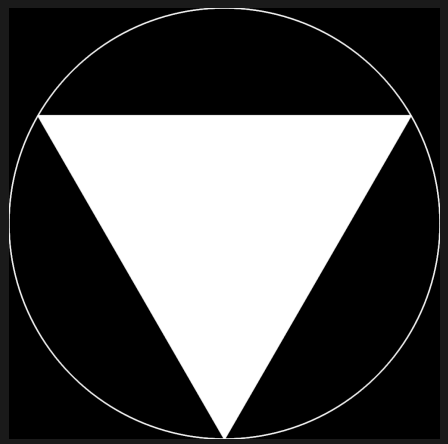
SPEI: Overview



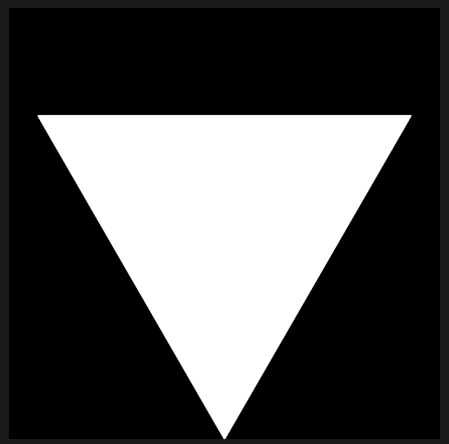
(a) Original input image.



(b) Minimum encompassing circle.



(c) Minimum encompassing square.



(d) Final image.

Figure 1: SPEI overview. Starts with Figure 1a and ends with 1d going from left to right.

SPEI: Data

Triangle

Square

Pentagon

Hexagon

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

Data from National Library of Medicine

SPEI: Experimental Setup

- 100 runs per Training Size
- k different Training Sizes per class
- Definitions
  - Training: Data used to build the model
  - Validation: Data not used to build the model
- Build SPEI-based SVM model
- Compare to CNN model

SPEI: Results

Model	3	4	5	6	Data Set
CNN	0.98	0.92	0.90	0.84	Train
SVM	1.00	1.00	1.00	1.00	Train
CNN	**0.73**	0.72	0.74	0.74	Valid
SVM	0.70	**0.75**	**0.80**	**0.85**	Valid

- SPEI-based SVM with radial kernel outperformed CNN by about 5.76%
- CNNs struggle in small data scenarios
- Github link: `https://github.com/billyl320/SPEI-Paper/tree/nml_nih`

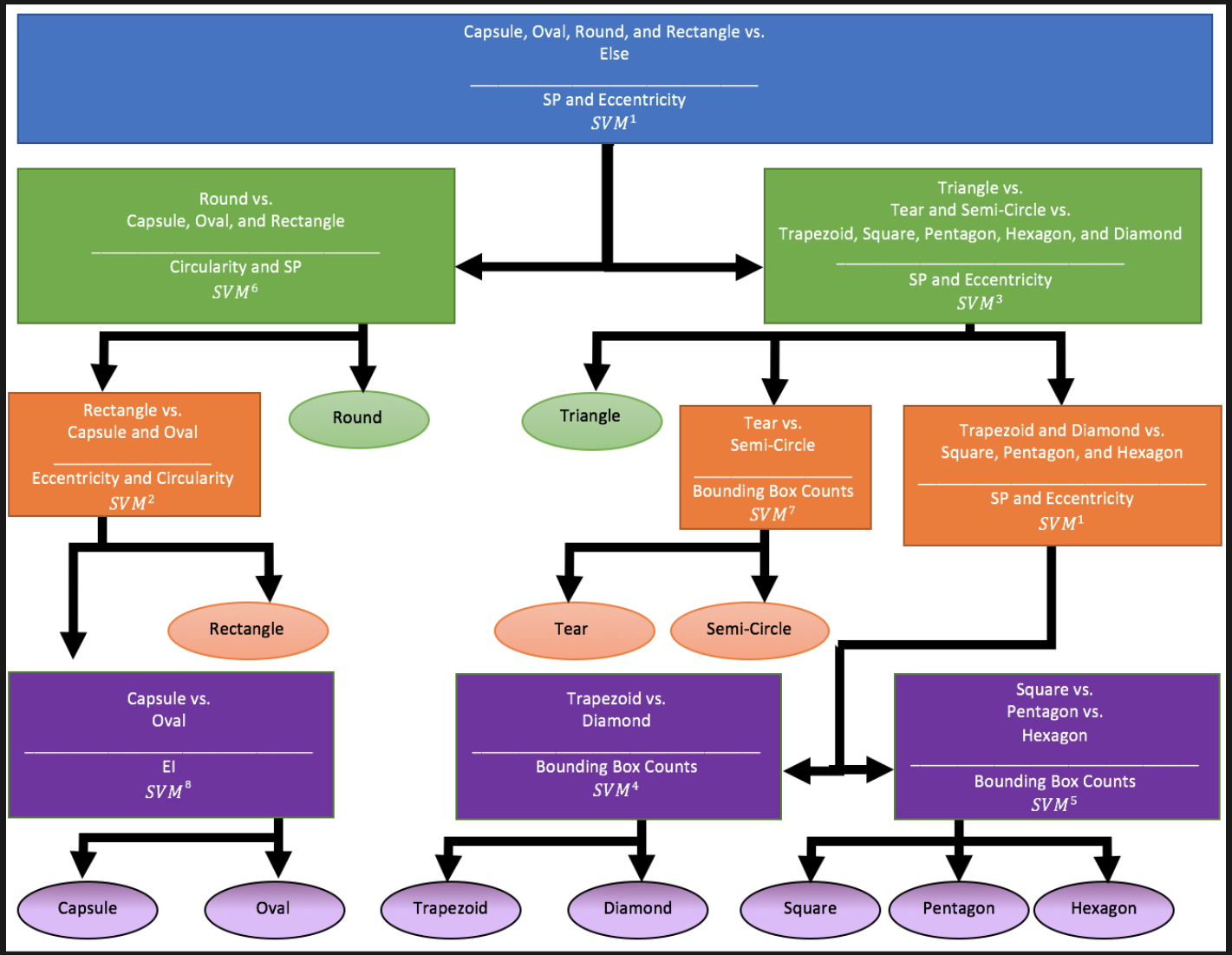
HILT Model: Data

Class	Training Count	Validation Count
Capsule	25	307
Diamond	6	6
Hexagon	4	4
Oval	25	661
Pentagon	6	6
Rectangle	3	3
Round	25	881
Semi-circle	2	2
Square	4	4
Tear	5	5
Trapezoid	2	2
Triangle	6	6
Total	113	1887

HILT Model: Metrics

$\vec{m}_{q,i}$	Metric
1	White EI
2	Black EI
3	SP value
4	Eccentricity
5	White Bounding Box Count
6	Black Bounding Box Count
7	Circularity

HILT Model: Setup

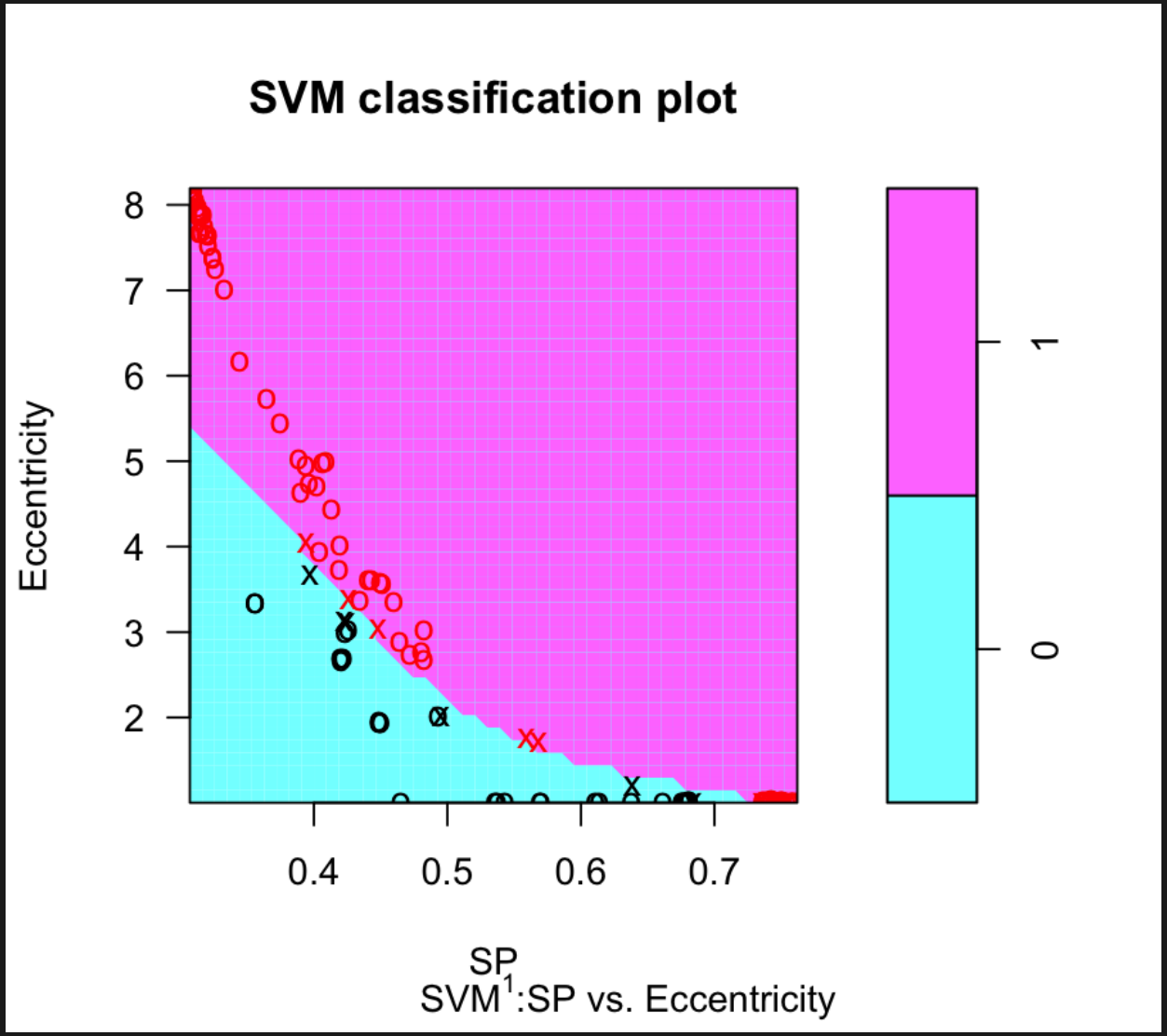


HILT Model: Results

Predicted/Truth Train	Capsule	Oval
Capsule	23	0
Oval	2	25

Predicted/Truth Valid	Capsule	Oval
Capsule	307	41
Oval	0	622

HILT Model: Plot



HILT Model: Results

Model	SVM - P	SVM - R	SVM - S	NB
MP	0.355	0.757	0.269	0.623

Model	LDA	Maddala	Lamberti
MP	0.801	0.897	**0.990**

- HILT SVM with polynomial kernel outperformed other models by about 95%
- MP = mean precision
- Breaking down multinomial into series of binary classification problems
- Github link: `https://github.com/billyl320/human_decision_tree_pills`

Conclusion

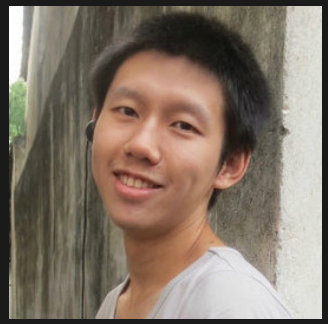
- HILT outperforms all other approaches
  - Overall Mean Outperformance: 95%
- HILT model important variables
  - SP
  - Eccentricity
- Future work would provide a completely machine driven approach
- Lamberti, "Algorithms to Improve Analysis and Classification for Small Data", 2020, `http://search.proquest.com/docview/2476825035/abstract/90CC4207B46B4068PQ/1`
- US Patent pursued by GMU
- My chapter in "AI Assurance: Towards Valid, Explainable, Fair, and Ethical AI" (Spring 2022)

Acknowledgements - George Mason University

- Dr. Jason Kinser
- Dr. William G. Kennedy
- Dr. Michael Eagle
- Dr. David I. Holmes
- George Mason University's Office of the Provost for funding

Acknowledgements - University of Virginia: Zang Lab

Chongzhi Zang, PhD



Shengen Hu, PhD

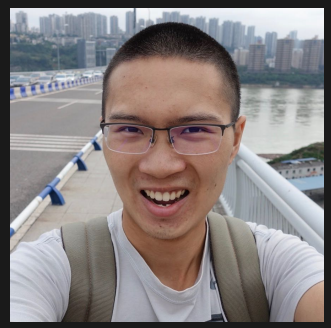
Qingying Chen, PhD

Mengxue Tian

Zhenjia Wang, PhD

Chris Dampier, MD

William F. Lamberti, PhD



Hongpan Zhang