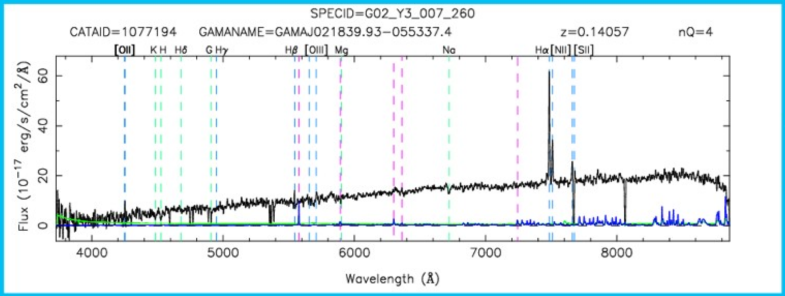
JULIA ZIEGLER

CLASSIFICATION OF GALAXY SPECTRA WITH ARTIFICIAL NEURAL NETWORKS

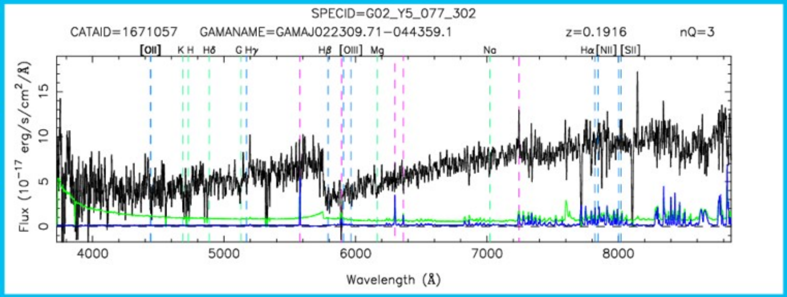
Bachelor Thesis Defense

CLASSIFICATION OF GALAXY SPECTRA

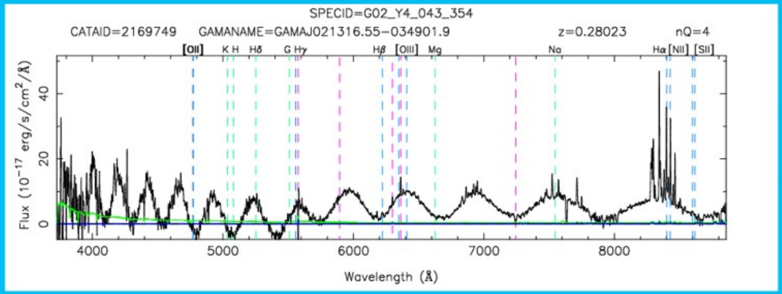
- GAMA Spectroscopic survey
- 300,000 spectra
- AAOmega spectrograph on 3.9m Anglo Australian Telescope
- From 2008 to 2014
- Down to magnitude r<19.8 mag



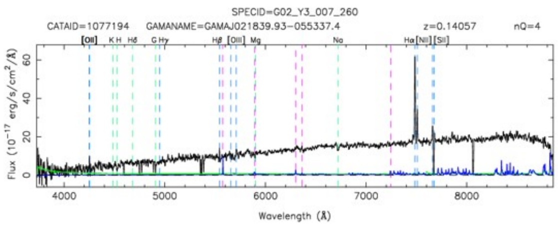
No Problem



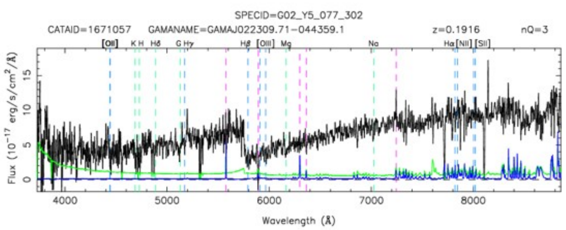
Bad Splice



Fringed

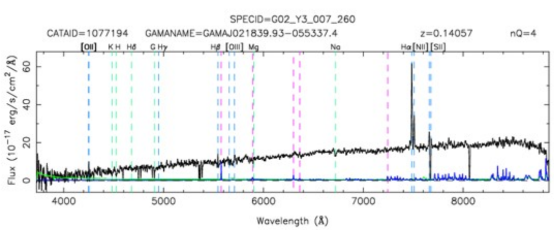


CLASS 0

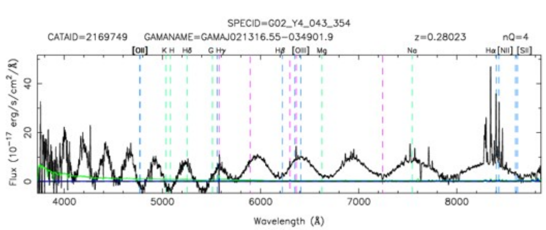


CLASS 1

CLASSIFIER 1



CLASS 0

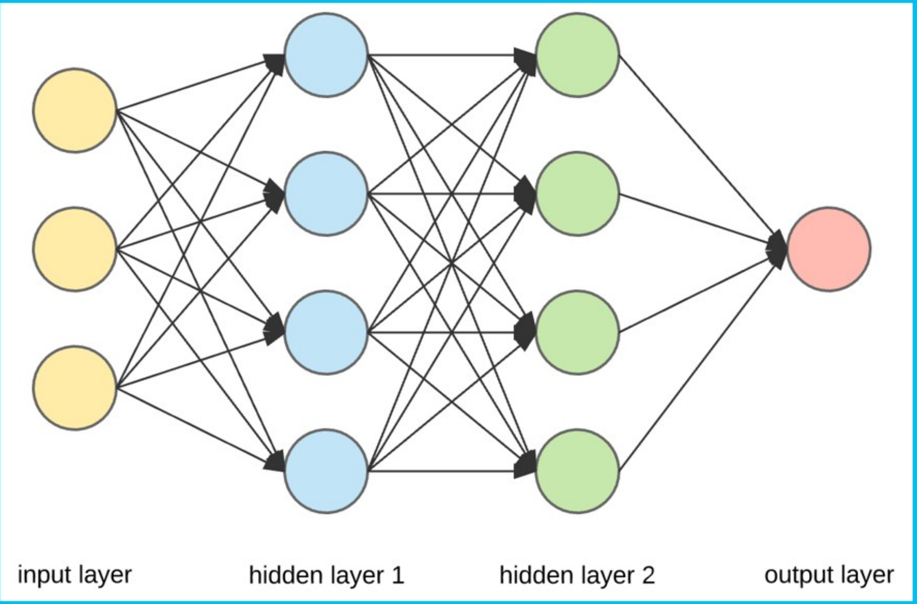


CLASS 1

CLASSIFIER 2

ARTIFICIAL NEURAL NETWORKS (ANN)

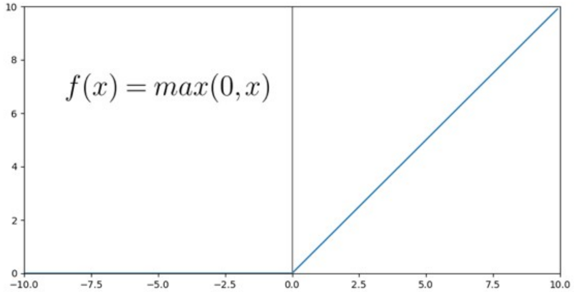
- Inspired by biological neural Networks
- Ability to learn
- Ability to handle probabilistic information
- Often used for: Image recognition, classification

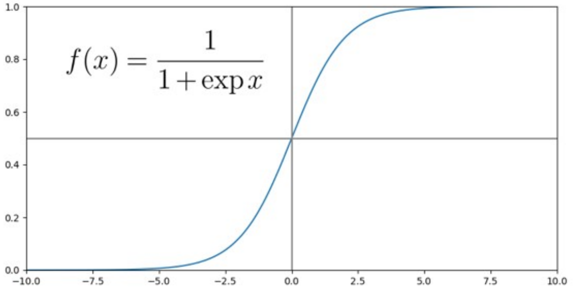


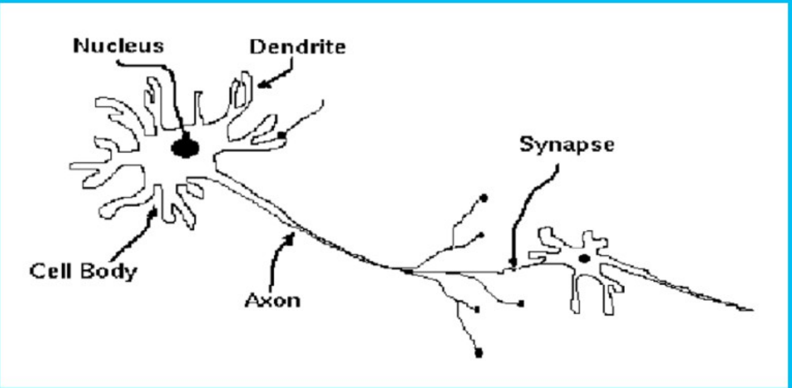
HTTPS://TOWARDSDATASCIENCE.COM/APPLIED-DEEP-LEARNING-PART-1-ARTIFICIAL-NEURAL-NETWORKS-D7834F67A4F6

NEURON

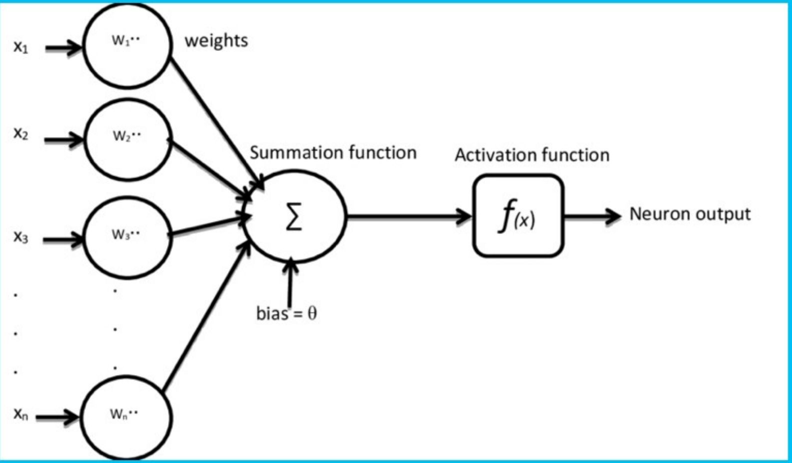
- Activation Functions:







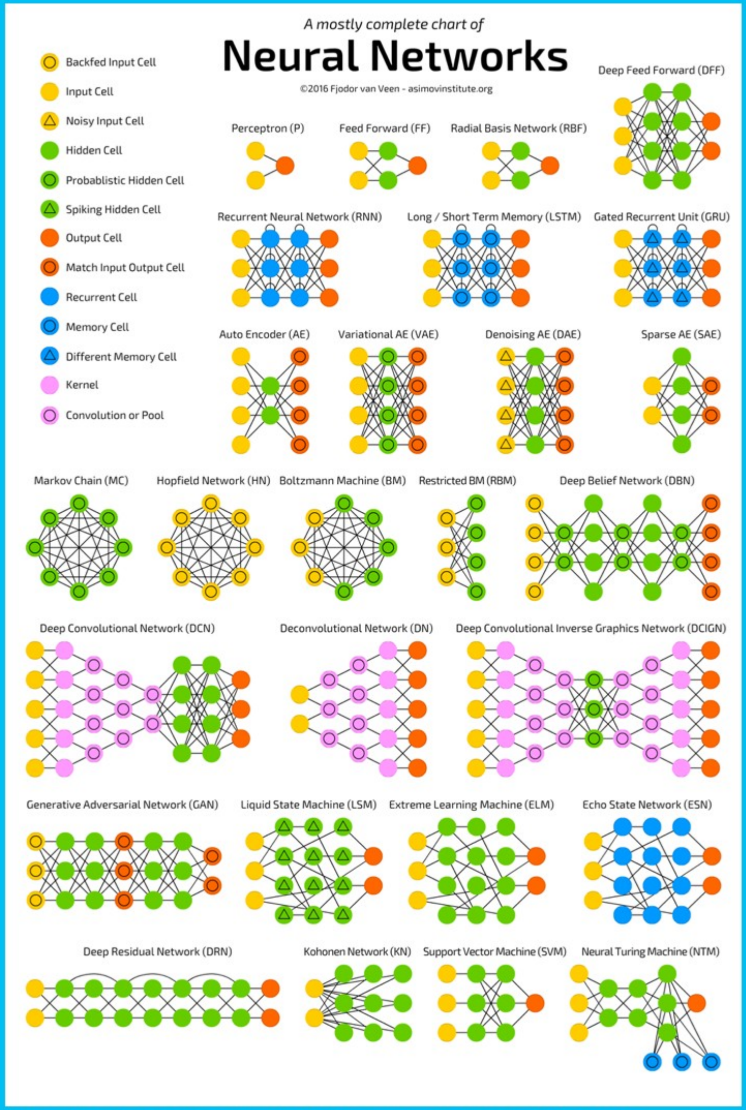
ANDRIES P. ENGELBRECHT. COMPUTATIONAL INTELLIGENCE: AN INTRODUCTION. JOHN WILEY SONS LTD, 2007.



JOSEPH YACIM AND DOUW BOSHOFF. IMPACT OF ARTIFICIAL NEURAL NETWORKS TRAINING ALGORITHMS ON ACCURATE PREDICTION OF PROPERTY VALUES. JOURNAL OF REAL ESTATE RESEARCH, 40:375–418, 11 2018.

NEURAL NETWORKS

- Supervised Training
- Minimize Loss
- By adjusting weights



FJODOR VAN VEEN. NEURAL NETWORK ZOO PREQUEL: CELLS AND LAYERS, 2017.

EVALUATING ANN

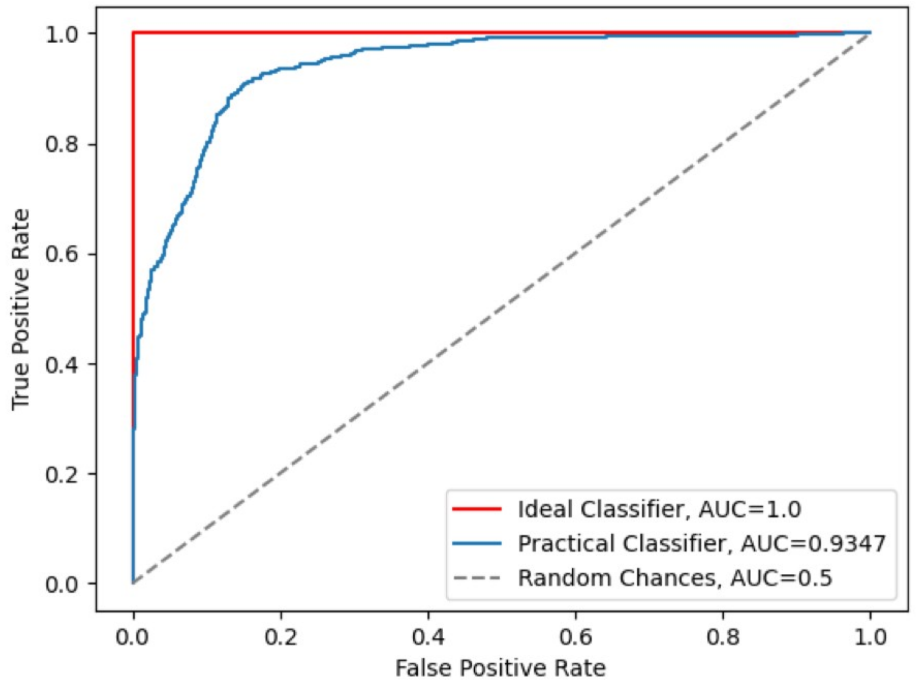
$$ACC = \frac{TP+TN}{TP+FP+TN+FN}$$

$$PRE = \frac{TP}{TP+FP}$$

$$REC = \frac{TP}{TP+FN}$$

$$F1 = \frac{2 \cdot PRE \cdot REC}{PRE+REC}$$

ROC curve



UNBALANCED DATA PROBLEM

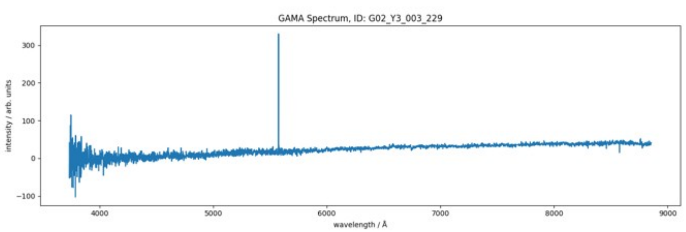
- ANNs perform best on balanced data

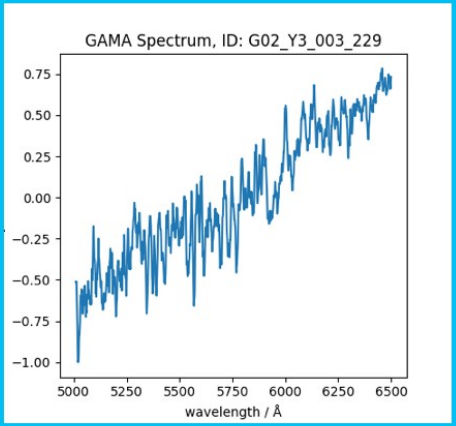
SOLUTION

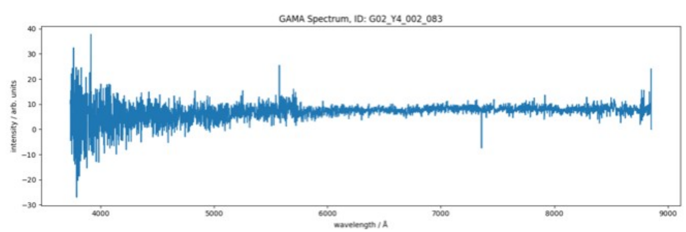
- Random under sampling
- Training Set:
- - Classifier 1: 800/800 (other/ bad splice)
- - Classifier 2: 400/400 (other/ fringed)
- Test Set:
- - 3,213 (364 bad splice (11%), 92 fringed (3%), 2,757 other (86%))

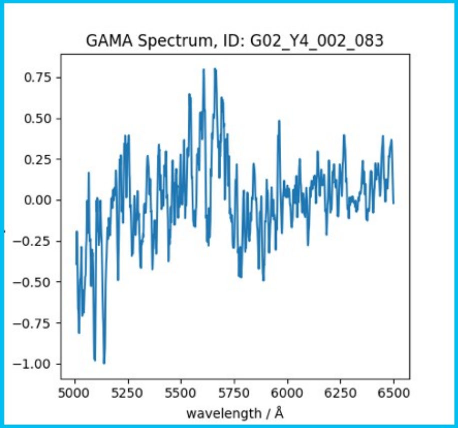
DATA PREPARATION

CLASSIFIER 1



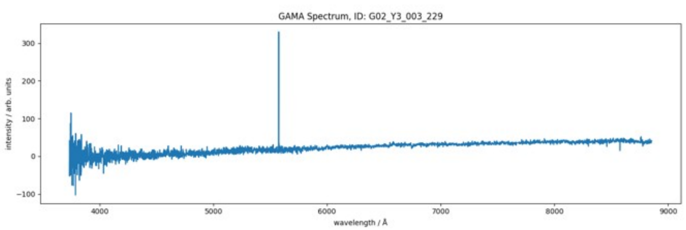


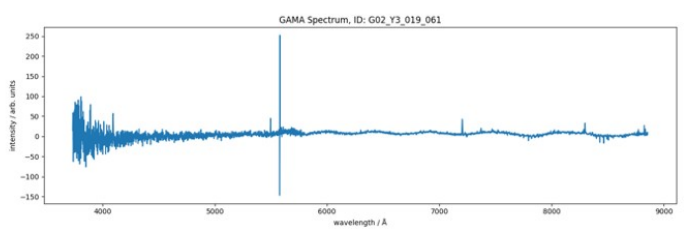


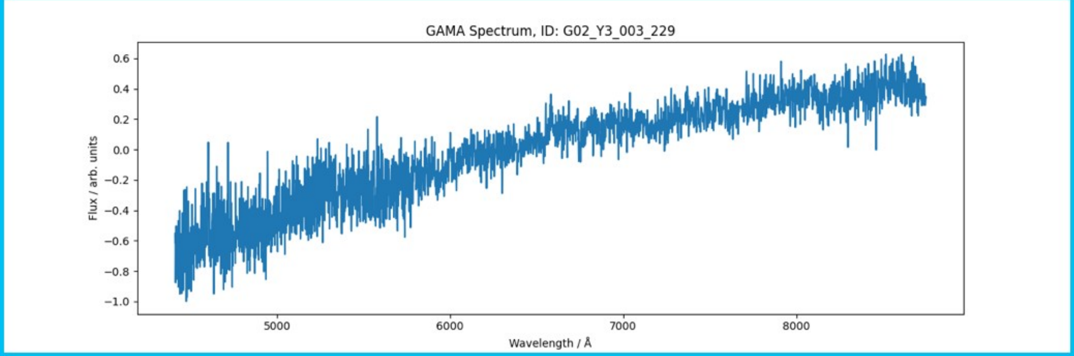


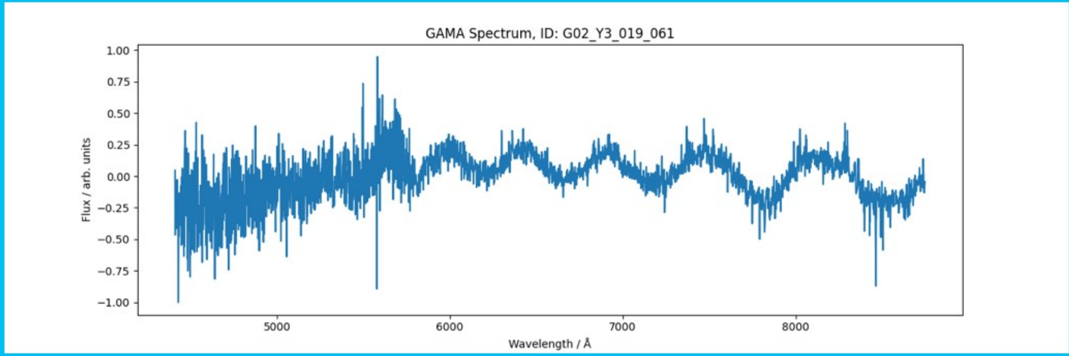
DATA PREPARATION

CLASSIFIER 2



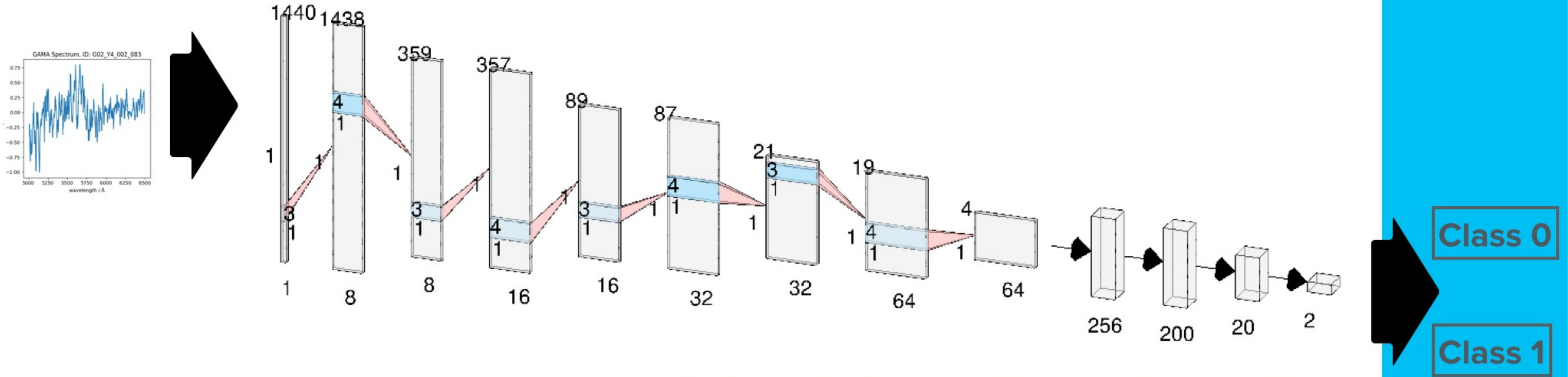




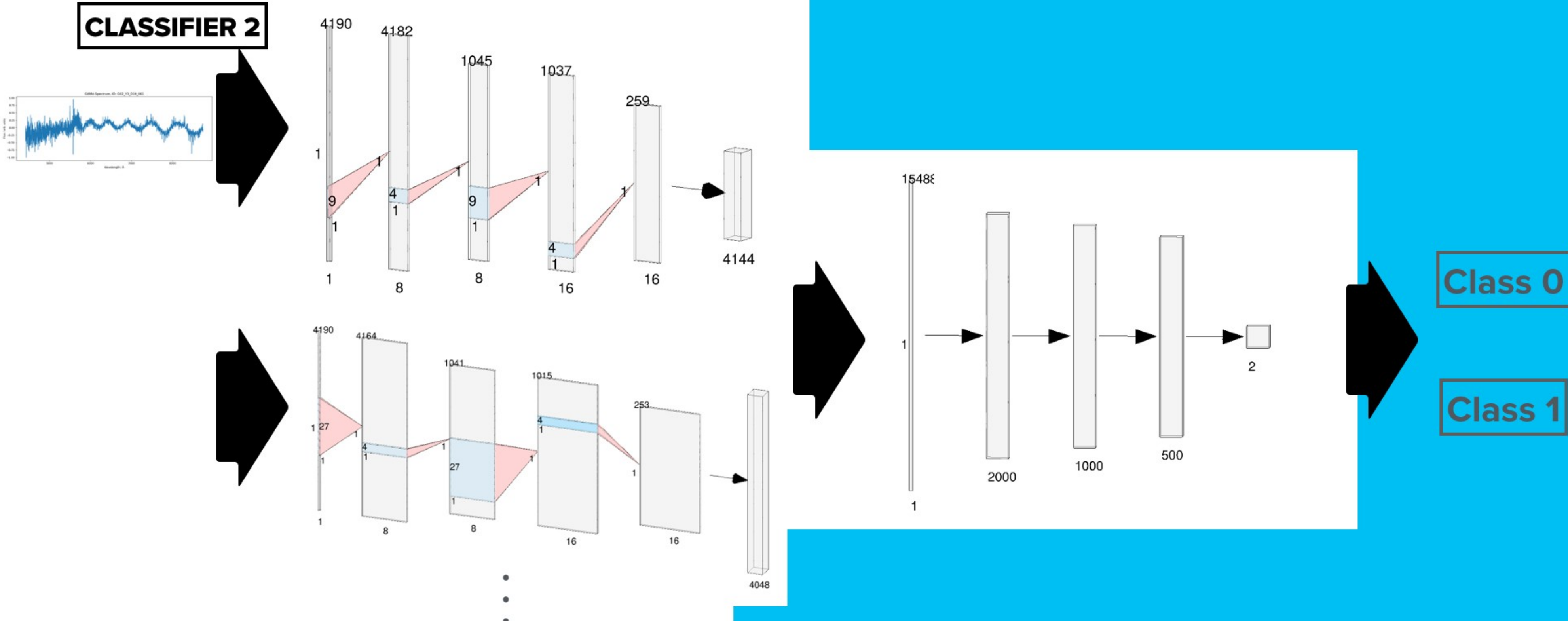


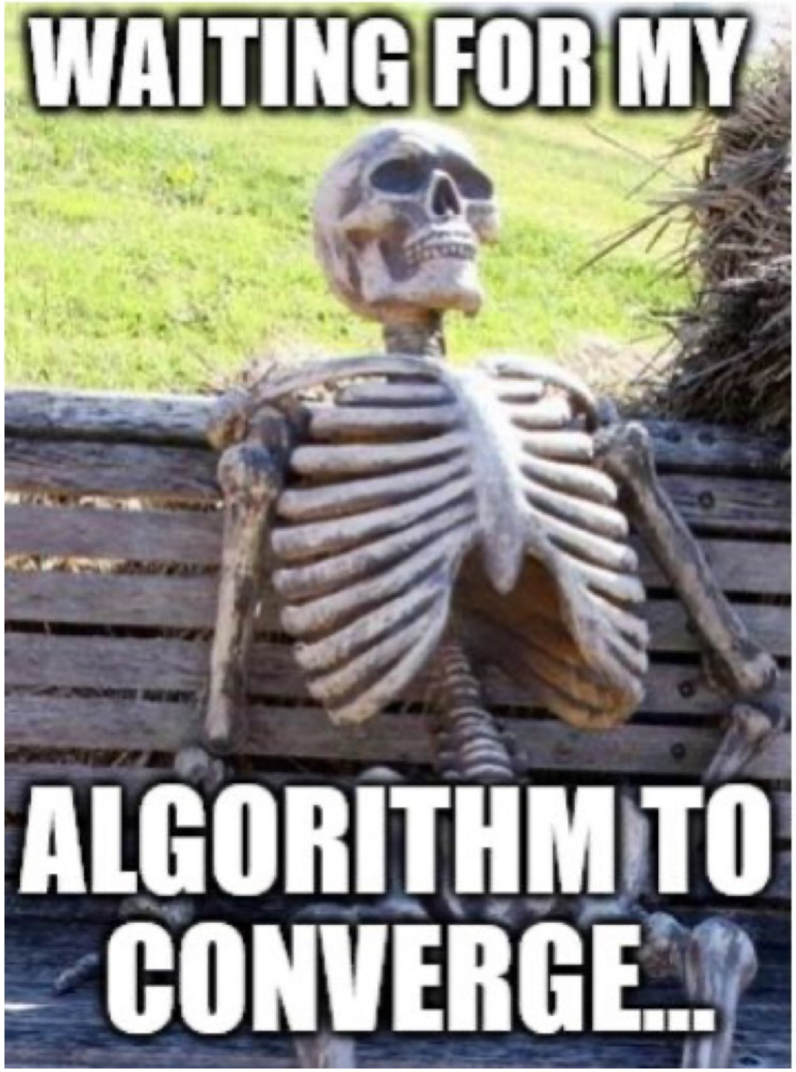
ANN ARCHITECTURE

CLASSIFIER 1



ANN ARCHITECTURE





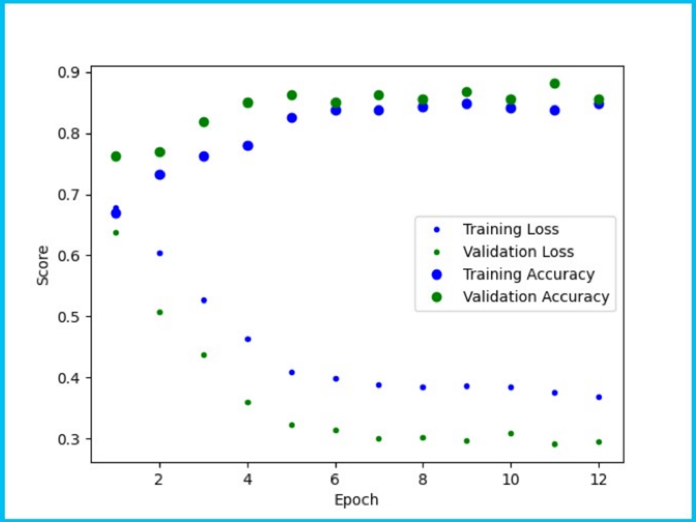
HTTPS://I.PINIMG.COM/736X/F6/D6/19/F6D6191DF7B4239DE8C5DC76CA121D33.JPG

RESULTS

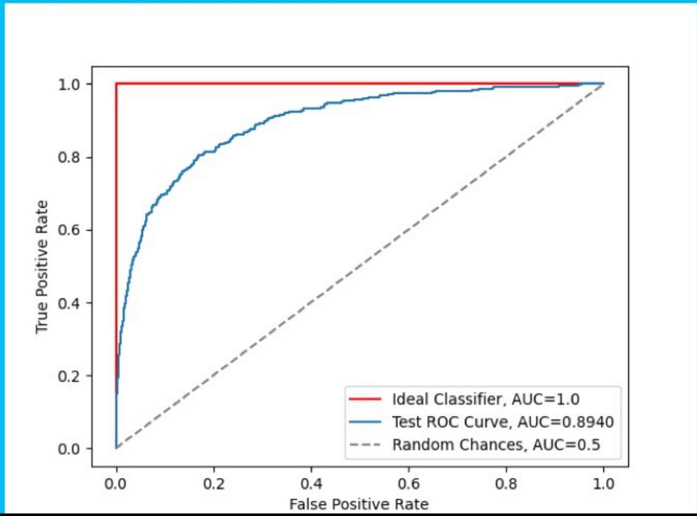
CLASSIFIER 1

- Test Results:
- ACC: 0.8880
- F1_class_0: 0.9353
- F1_class_1: 0.5765

- Losscurve



- Test ROC curve

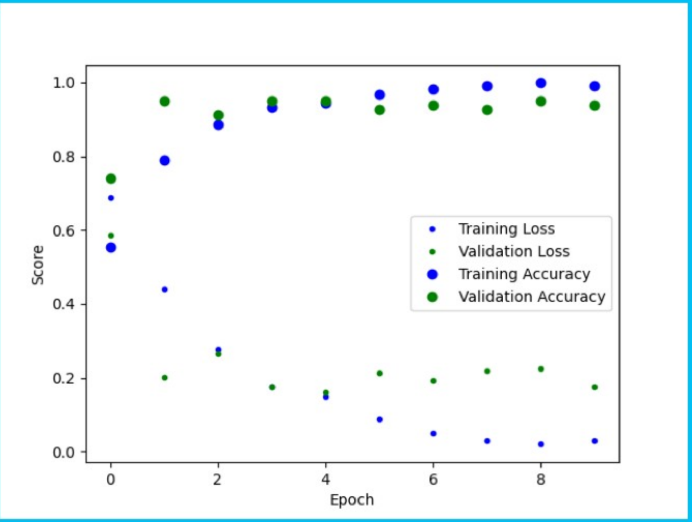


RESULTS

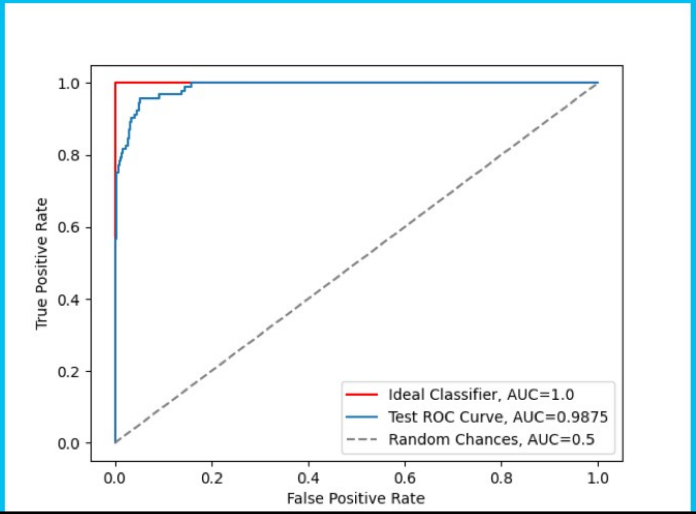
CLASSIFIER 2

- Test Results:
- ACC: 0.9788
- F1_class_0: 0.9890
- F1_class_1: 0.6937

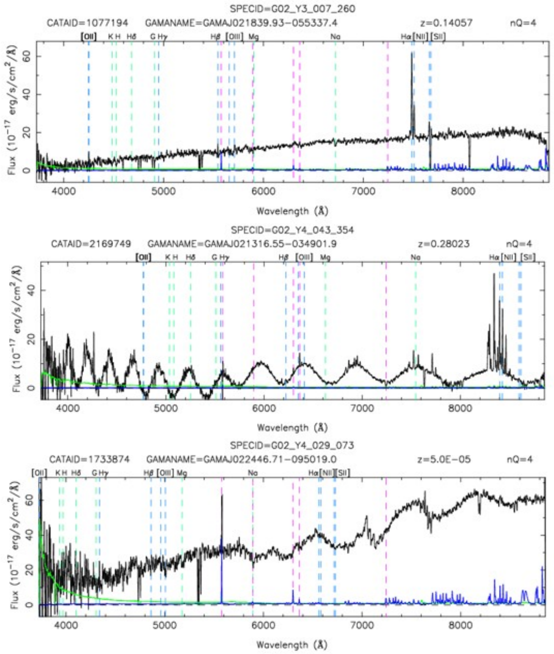
- Losscurve



- Test ROC curve



IMPROVING CLASSIFIER 2



CLASS 0

CLASS 1

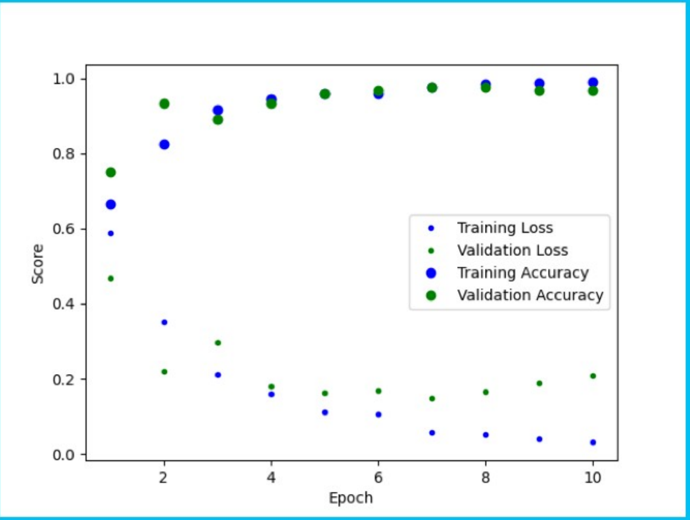
CLASS 2

RESULTS

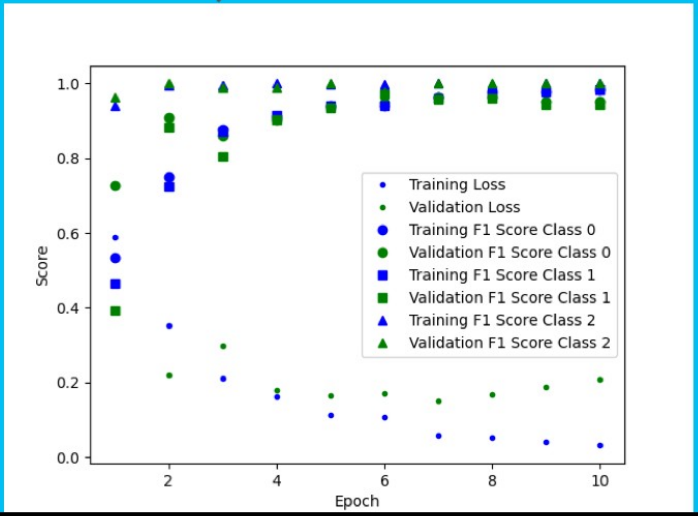
CLASSIFIER 2

- Test Results:
- ACC: 0.99667
- F1_class_0: 0.9813
- F1_class_1: 0.6312
- F1_class_2: 0.8966

- Losscurve



- Losscurve, F1 scores

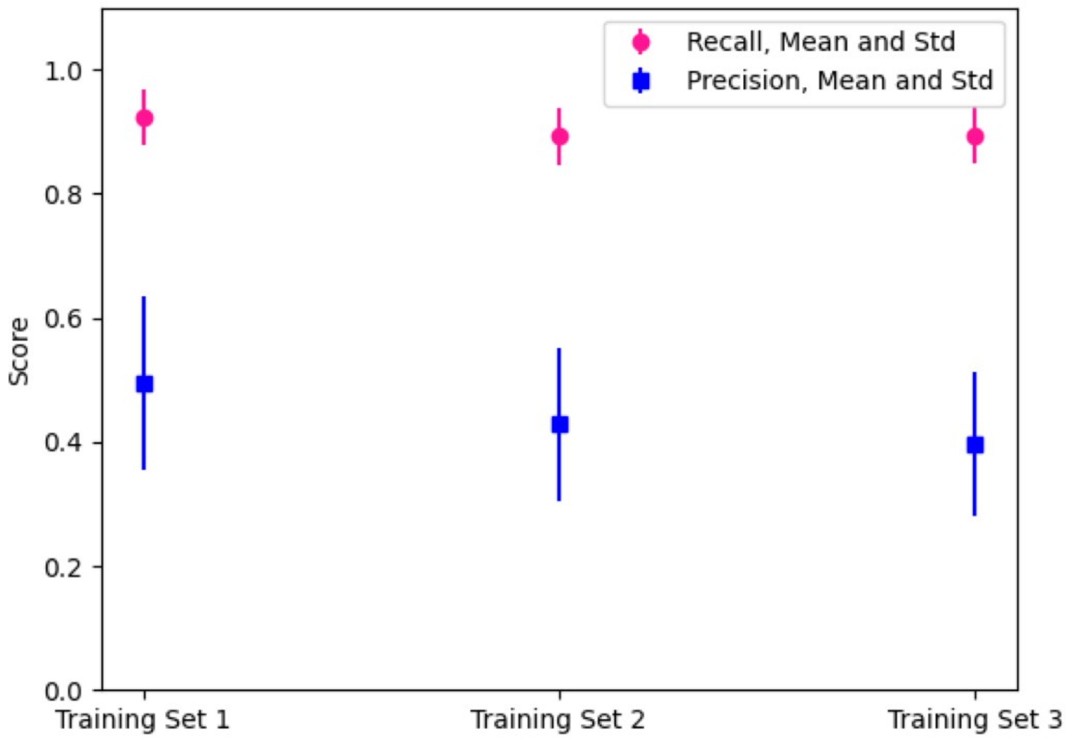


RESULTS

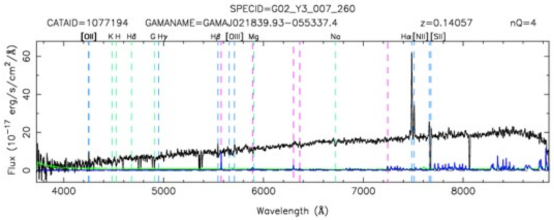
CLASSIFIER 2

- Training Set 1: 400/400/400 (other/fringed/M-star)
- Training Set 2: 400/400 (other/fringed)
- Training Set 3: 400/400 (other(containing 2% M-stars)/fringed)

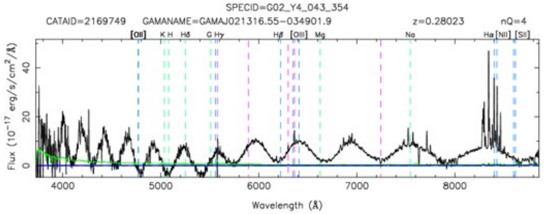
Test Recall and Precision, Class 1



IMPROVING CLASSIFIER 2



CLASS 0

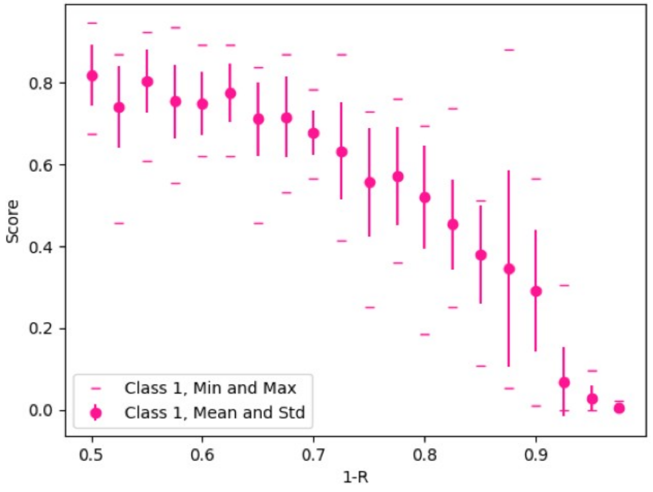


CLASS 1

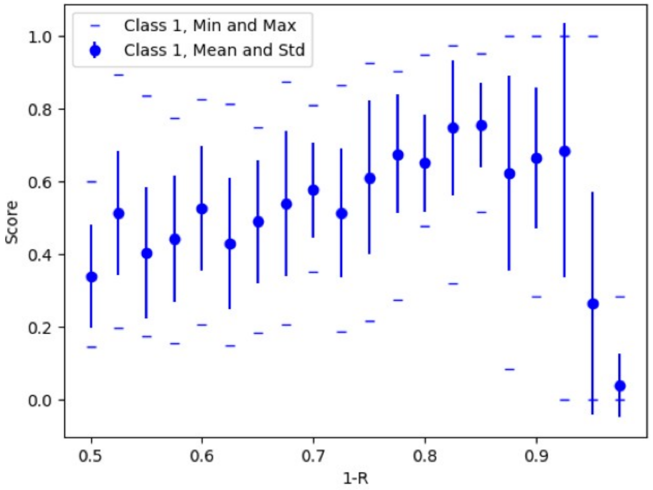
RESULTS

- Change Ratio of class 1 in training set
- Total set size: 400 (other and fringed)

Test Recall, Class 1



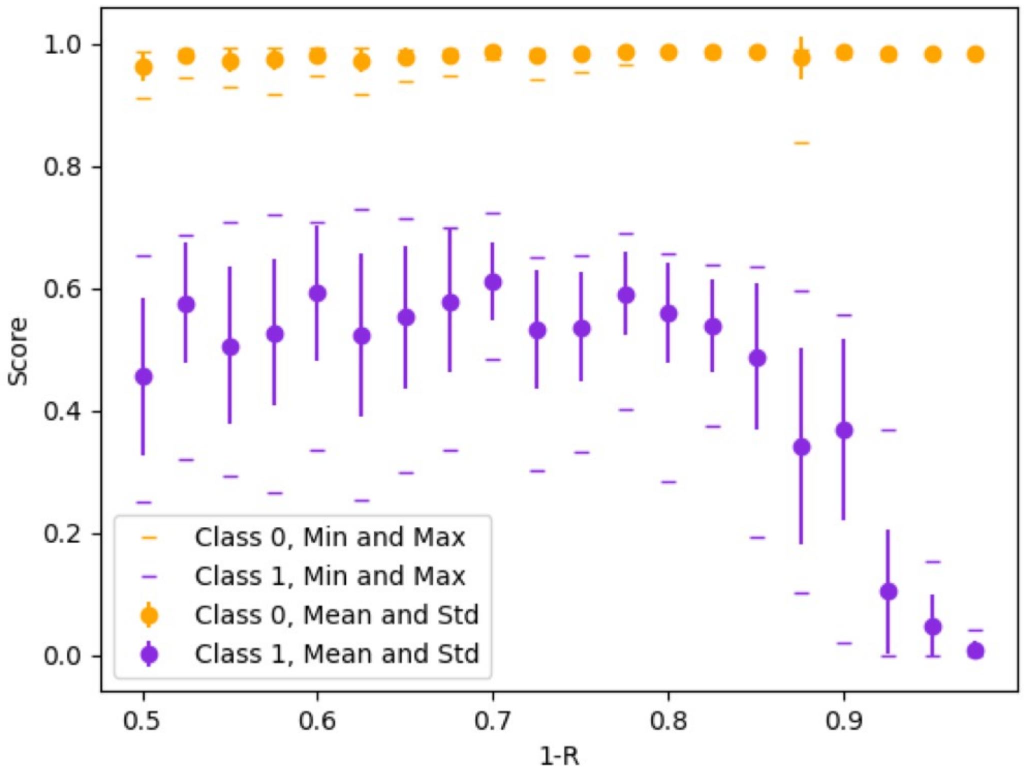
Test Precision, Class 1



RESULTS

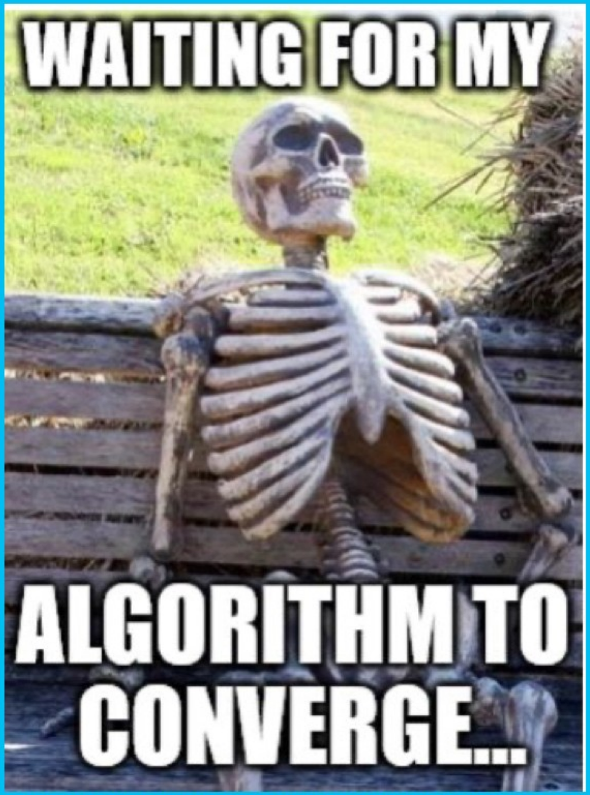
- Change Ratio of class 1 in training set
- Total set size: 400 (other and fringed)

Test F1 scores, Class 0 and 1



FINAL RESULTS

- Unbalance in training set could improve precision
- Highest scores for balanced training:
- - splicing: AUC=0.8880
- - fringing: AUC=0.9875



HTTPS://I.PINIMG.COM/736X/F6/D6/19/F6D6191DF7B4239DE8C5DC76CA121D33.JPG

ROC CURVE

True class

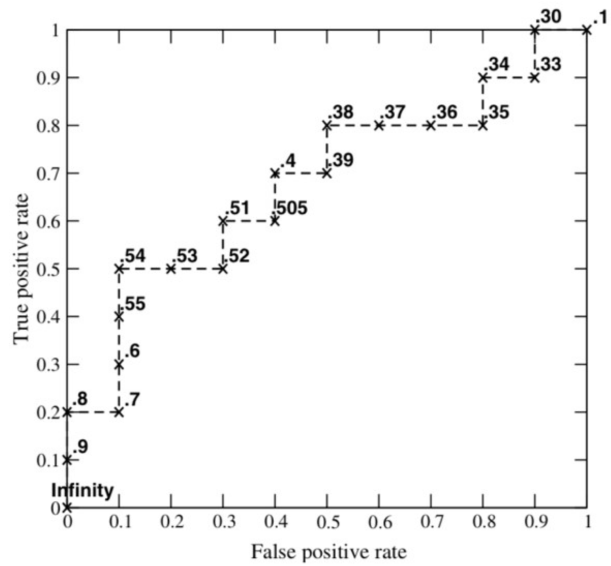
	p	n
Y	True Positives	False Positives
N	False Negatives	True Negatives

$$\text{fp rate} = \frac{FP}{N} \qquad \text{tp rate} = \frac{TP}{P}$$

$$\text{precision} = \frac{TP}{TP+FP} \qquad \text{recall} = \frac{TP}{P}$$

$$\text{accuracy} = \frac{TP+TN}{P+N}$$

Inst#	Class	Score	Inst#	Class	Score
1	p	.9	11	p	.4
2	p	.8	12	n	.39
3	n	.7	13	p	.38
4	p	.6	14	n	.37
5	p	.55	15	n	.36
6	p	.54	16	n	.35
7	n	.53	17	p	.34
8	n	.52	18	n	.33
9	p	.51	19	p	.30
10	n	.505	20	n	.1



TOM FAWCETT. AN INTRODUCTION TO ROC ANALYSIS. PATTERN RECOGNITION LETTERS, 27(8):861 – 874, 2006. ROC ANALYSIS IN PATTERN RECOGNITION.