SINGLE CELL RNA SEQUENCING OF STEM CELL-DERIVED RETINAL GANGLION CELLS

FILIPPIDOU MARINA-THALASSINI, PAPADOPOULOU MARIANNA, VOSSOS CHARALAMPOS

PAPER SUMMARY

Use of scRNA seq to characterize the transcriptomes of 1,174 human embryonic stem cell-derived retinal ganglion cells (RGCs)

Use of Hierarchical Clustering and discovery of three distinct subpopulations of cells, with various degrees of maturity

Potential to enhance our comprehension of (RGCs) and their implication in pathologies such as glaucoma optic neuropathies

AUTHORS' RESULTS

Subpopulation 1

- Genes associated with **neural cell adhesion molecule signaling** and **Hedgehog pathway**

Subpopulation 2

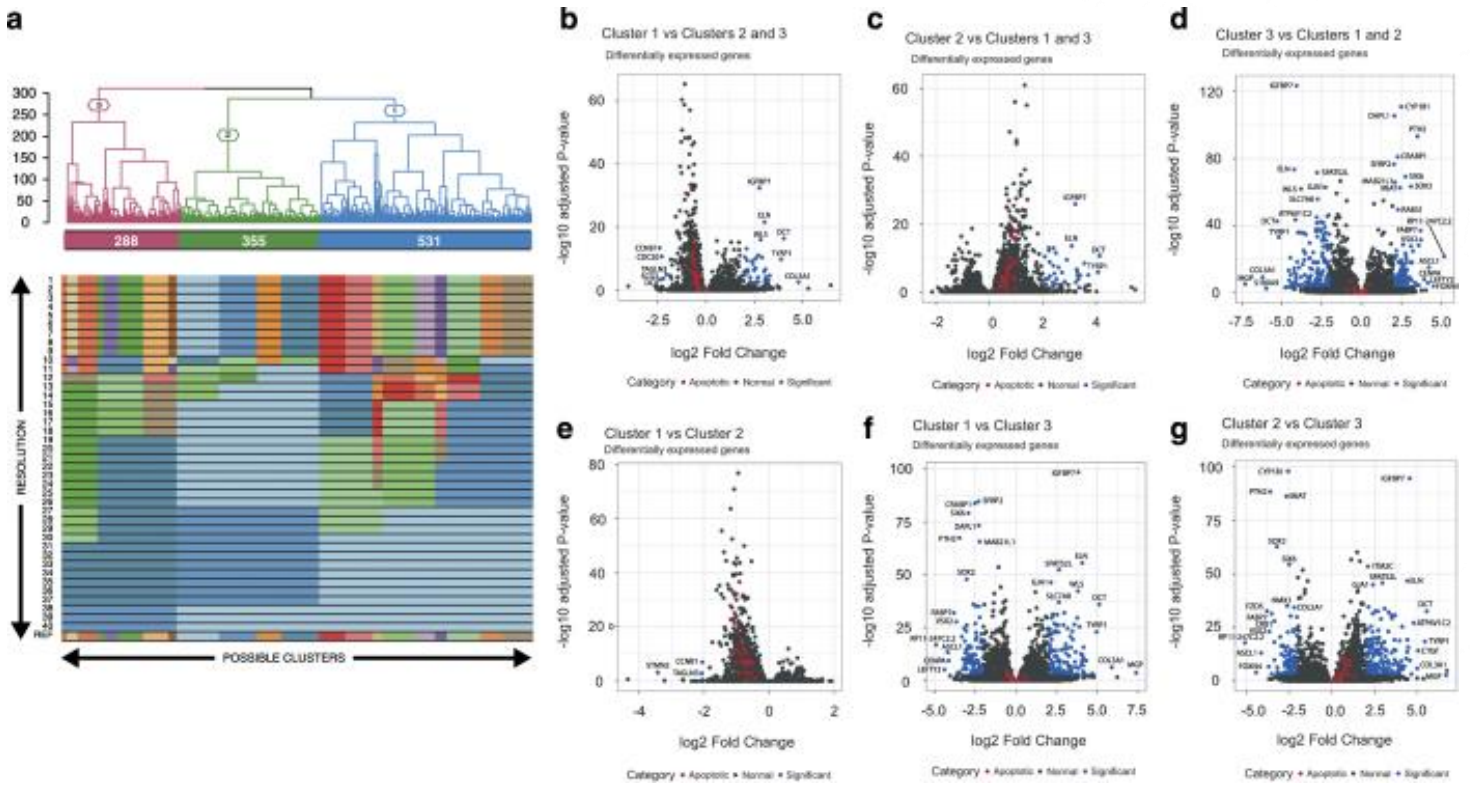
- Regulation of **Notch protein** implicated in **neuronal function** and **development**, and **DNA repair**

suggests a more specialized phenotype of retinal ganglion cells

Subpopulation 3

- **Upregulated** genes involved in **axon guidance** and **extracellular matrix proteoglycans**
- **Downregulation** of genes associated with the **cell cycle**

indicates a more mature neuronal phenotype



RESEARCH MOTIVATION, OBJECTIVES, AND ANTICIPATED BENEFITS

Motivation for more research :

- Gain a deeper understanding of stem cell-derived retinal ganglion cells and their gene expression patterns.
- Contribute to the scientific community by replicating and expanding upon previous research.

"Gap" we try to address:

•Bridge the gap between research findings and practical applications by providing a robust classification tool for stem cell-derived retinal ganglion cells.

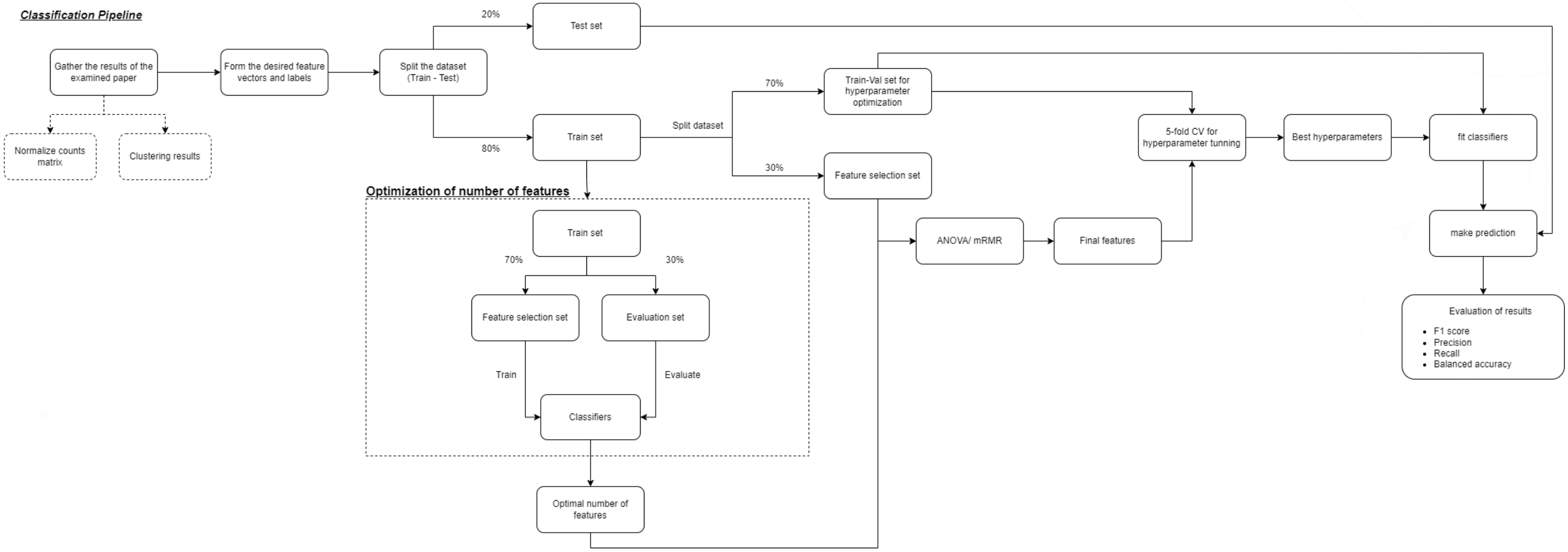
Anticipated Benefits?

Enhanced understanding of stem cell-derived retinal ganglion cells: Validating subpopulations and gene expression patterns contributes to knowledge and retinal ganglion cell development understanding.

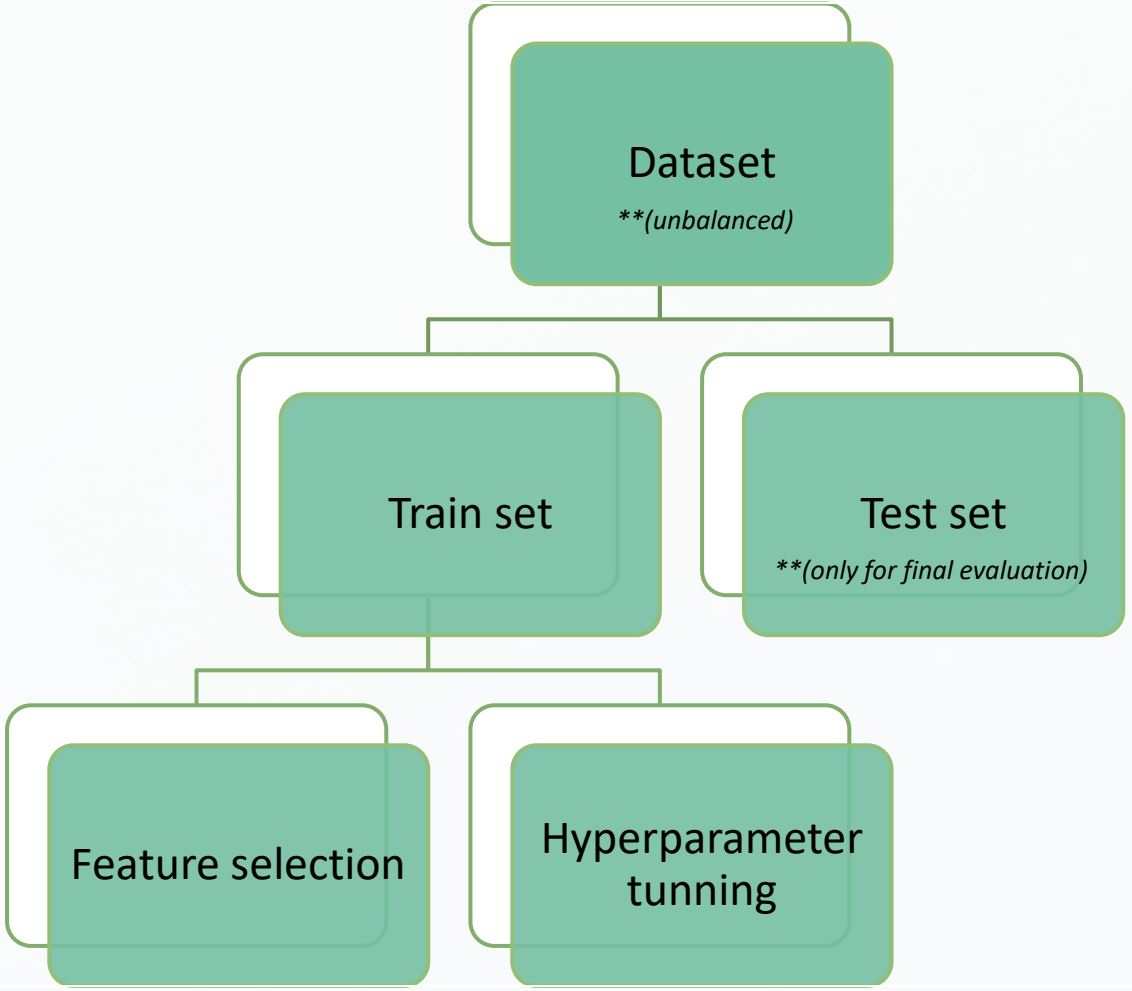
Development of a practical classifier: Enables accurate classification of cells, aiding future research, drug development, and regenerative medicine applications.

Accelerated study of retinal ganglion cells: Potential advancements in ocular health and vision-related disorder research

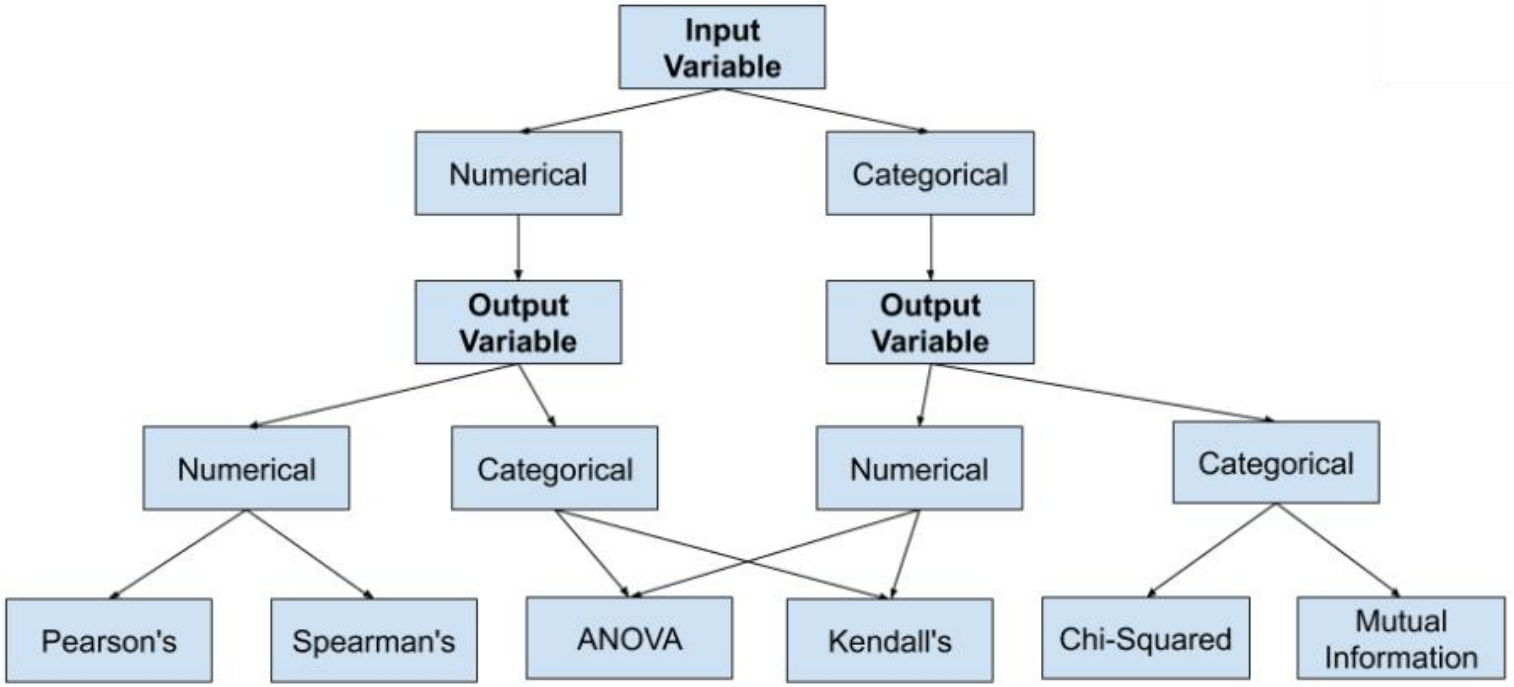
CLASSIFICATION PIPELINE



DATASET SPLIT



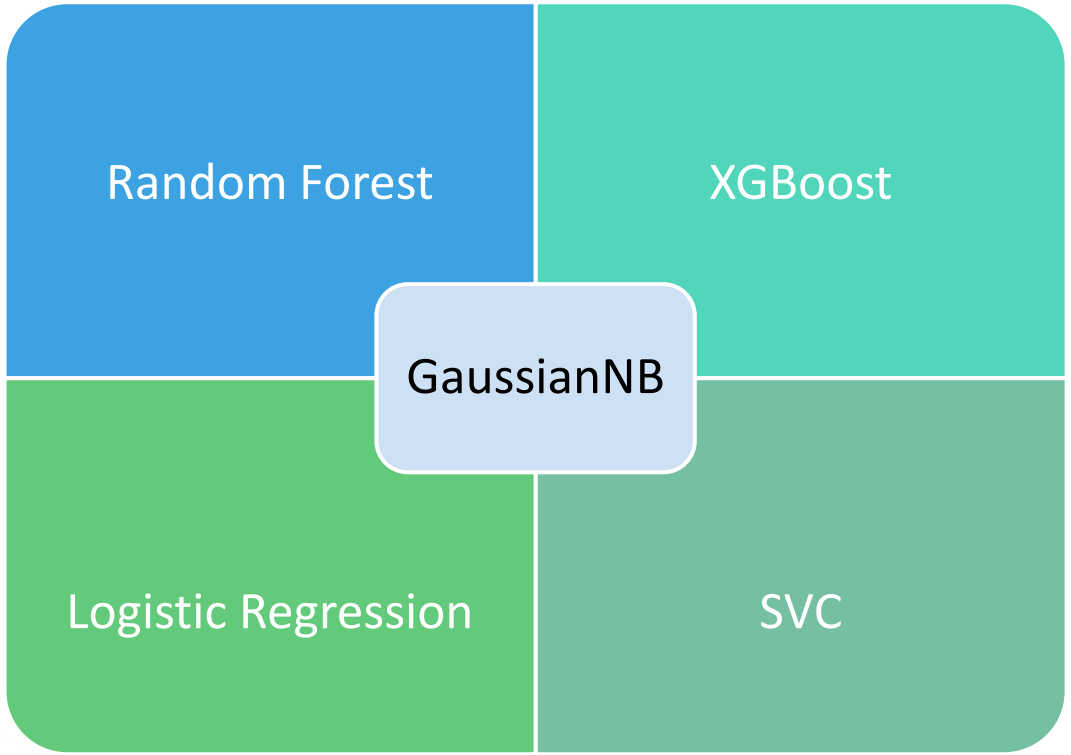
FEATURE SELECTION



We used Optuna to find the best number of features to keep

ANOVA and mRMR were tried out as feature selection techniques

HYPERPARAMETER TUNNING –CLASSIFIERS – PREDICTIONS



We used a Stratified 5-fold cross validation with Optuna to hyper tune the parameters of each classifier

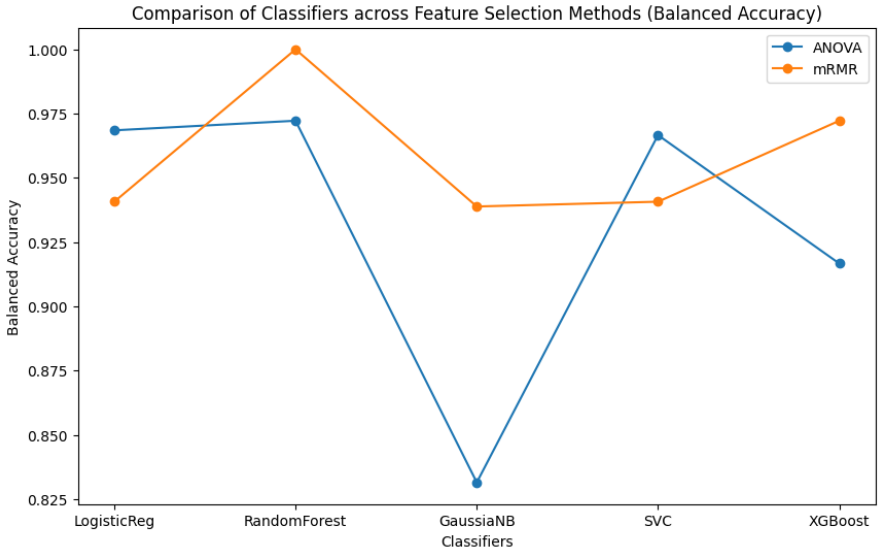
MCC was used as the optimization metric

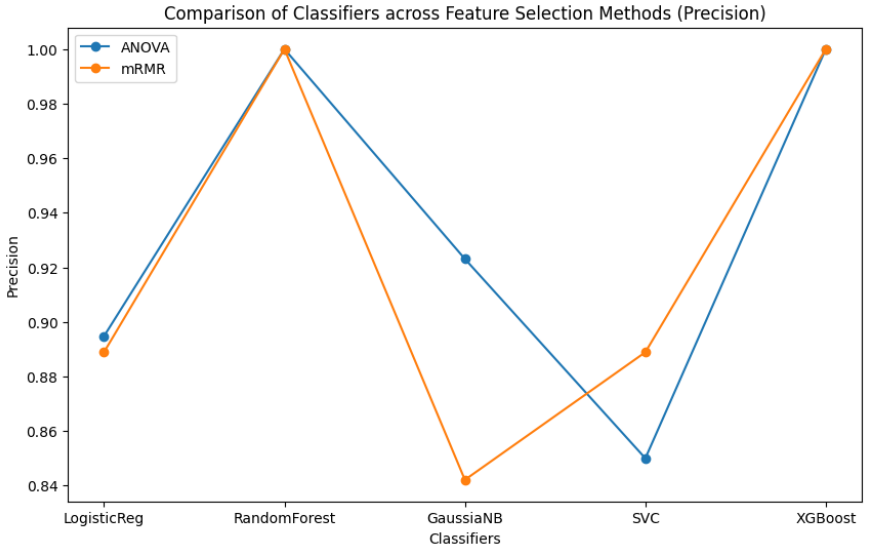
Evaluation metrics:

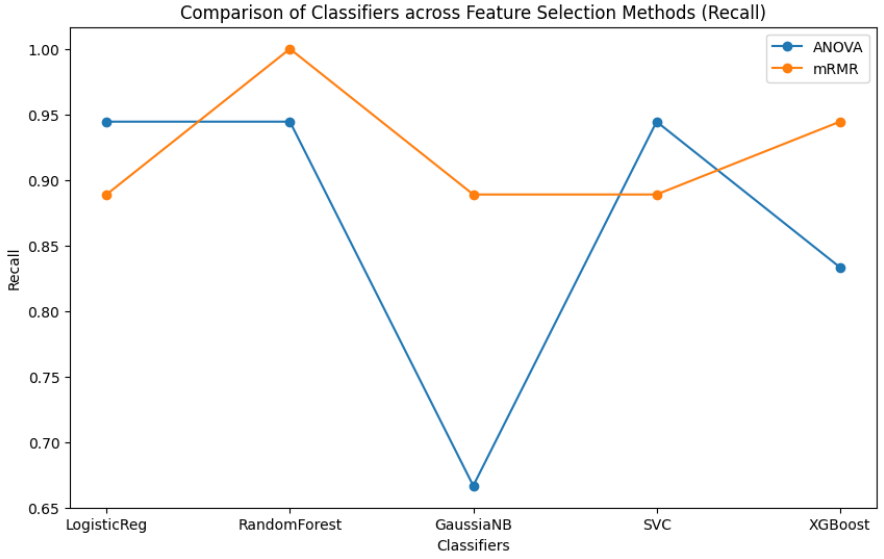
- Balanced accuracy
- Precision
- Recall
- F1-score

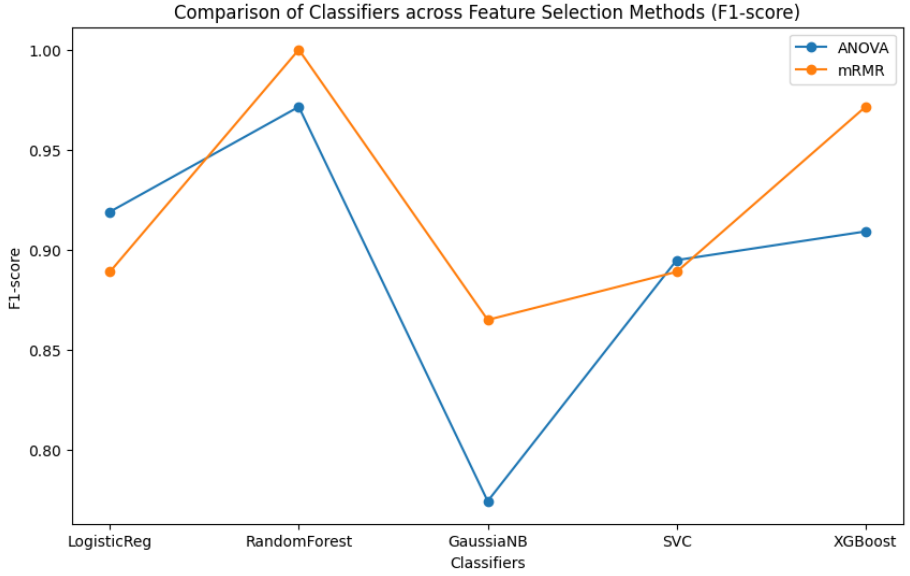
RESULTS

2 class problem









RESULTS

2 class problem

Feature importance threshold: 0.01

Marker genes (for the 2 classes): 118

Feature selection **ANOVA:**

RF: returned 22 genes, **all of them were marker genes from *class 1***

XGBoost: returned 30 genes

17/30: marker genes from *class 1*

1/30 : marker gene for *class 2*

12/30 : no marker genes

Feature selection **mRMR:**

RF: returned 20 genes

17/20: marker genes for *class 1*

2/20: marker genes for *class 2*

1/20: no marker gene

XGBoost: returned 17 genes

11/17: marker genes for *class 1*

2/17: marker genes for *class 2*

4/17: no marker genes

RESULTS

2 class problem

20 most important features

Marker genes (for the 2 classes): 118

Feature selection **ANOVA:**

LR:
7/20: marker genes from *class 1*
4/20: marker genes from *class 2*
9/20: no marker genes

SVC:
7/20: marker genes from *class 1*
4/20: marker genes from *class 2*
9/20: no marker genes

Feature selection **mRMR:**

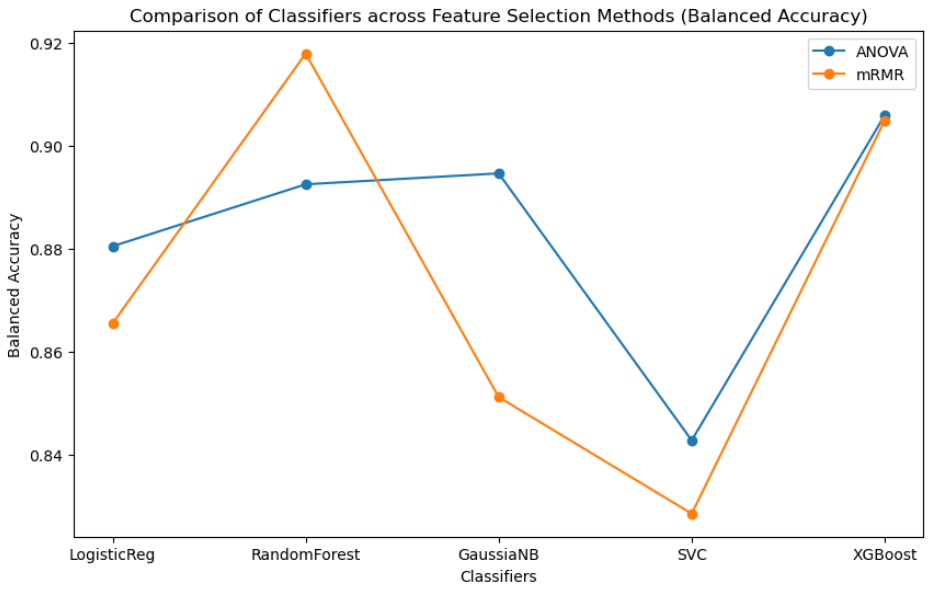
LR:
2/20: marker genes from *class 1*
7/20: marker genes from *class 2*
11/20: no marker genes

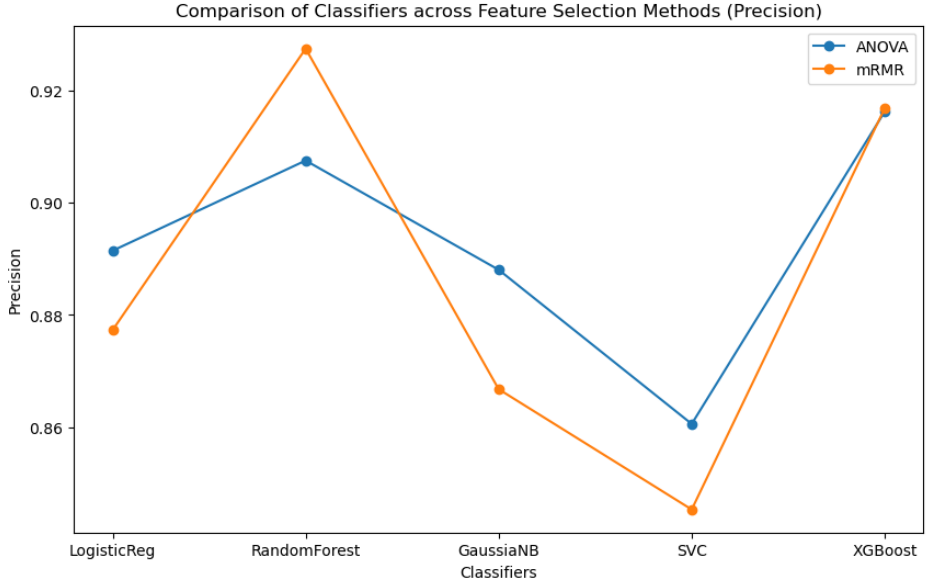
SVC:
1/20: marker genes for *class 1*
8/20: marker genes for *class 2*
11/20: no marker genes

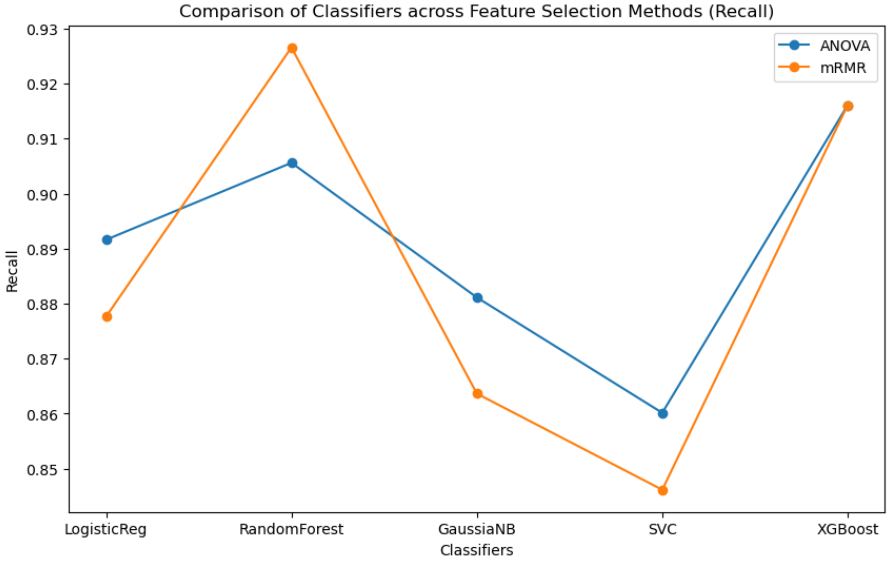
! Five out of eight times was given importance to the gene '**ENSG00000111640**' which is **not** identified as a **marker gene**

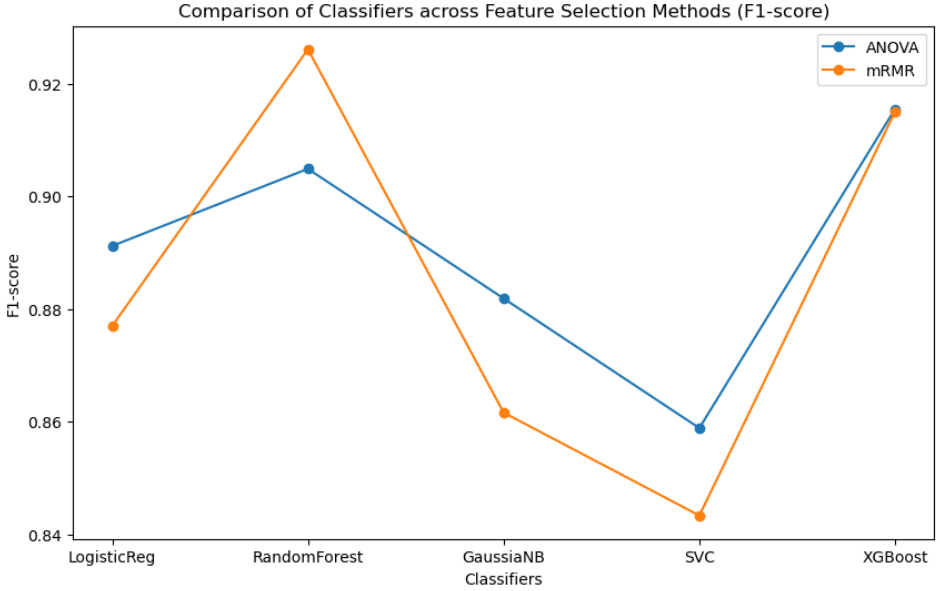
RESULTS

4 class problem









RESULTS

4 class problem

Feature importance threshold: 0.01

Marker genes (for the 2 classes): 302

Feature selection **ANOVA:**

RF: returned 25 genes

14/25: marker genes for *class 1*

9/25: marker genes for *class 3*

2/25: marker genes for *class 4*

XGBoost: returned 12 genes

9/12: marker genes for *class 1*

2/12: marker genes for *class 3*

1/12: marker genes for *class 4*

Feature selection **mRMR:**

RF: returned 22 genes

3/22: marker genes for *class 3*

1/22: marker genes for *class 4*

18/22: no marker genes

XGBoost: returned 11 genes

11/11: no marker genes

RESULTS

4 class problem

20 most important features

Marker genes (for the 2 classes): 302

Feature selection **ANOVA:**

LR: **Class 1:** 3/20: marker genes from *class 1*

17/20 no marker genes

Class 2: 2/20: marker genes from *class 1*

6/20: marker genes from *class 2*

2/20: marker genes from *class 3*

2/20: marker genes from *class 4*

8/20: no marker genes

Class 3: 2/20: marker genes from *class 1*

11/20: marker genes from *class 3*

7/20: no marker genes

Class 4: 2/20: marker genes from *class 3*

11/20: marker genes from *class 4*

7/20: no marker genes

SVC: 1/20: marker gene from *class 1*

1/20: marker gene from *class 4*

18/20: no marker genes

RESULTS

4 class problem

20 most important features

Marker genes (for the 2 classes): 302

Feature selection **mRMR:**

LR: **Class 1:** **4**/20: marker genes from *class 1*
16/20 no marker genes

Class 2: 2/20: marker genes from *class 1*
5**/20: marker genes from *class 2***
2/20: marker genes from *class 3*
3/20: marker genes from *class 4*
8/20: no marker genes

Class 3: 2/20: marker genes from *class 1*
11/20: marker genes from *class 3*
7/20: no marker genes

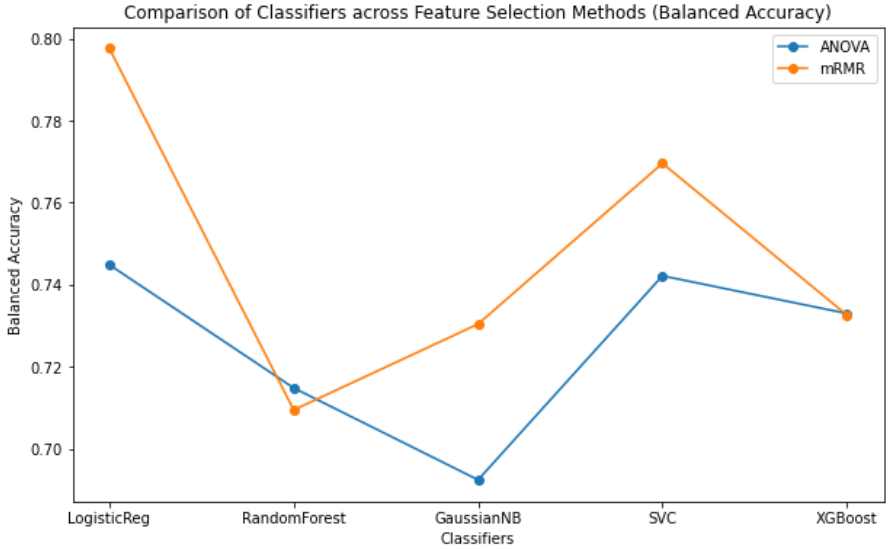
Class 4: 1/20: marker genes from *class 1*
1/20: marker genes from *class 3*
11/20: marker genes from *class 4*
7/20: no marker genes

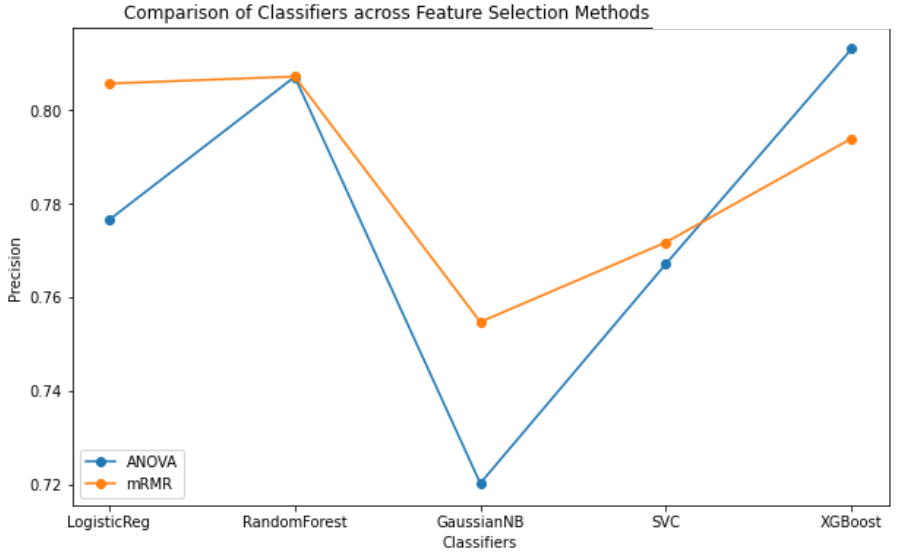
SVC: 1/20: marker gene from *class 1*
19/20: no marker genes

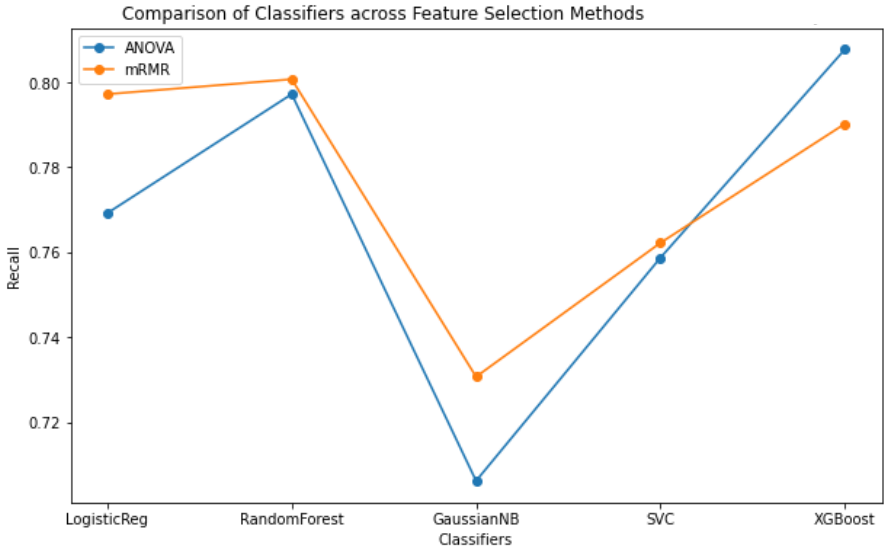
No marker genes identified as important from 6 classifiers: ENSG00000138326 (class2), ENSG00000205542 (class1)

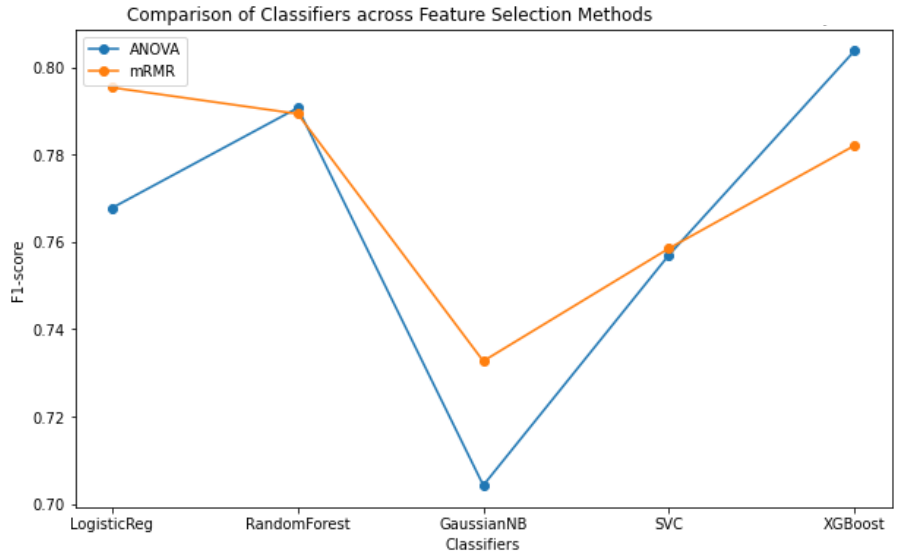
RESULTS

11 class problem









FURTHER RESEARCH?

Use the Random Forest or XGBoost classifier as feature selection methods

Explore why the no identified as marker genes are important to our classifiers

THANK YOU
FOR YOU ATTENTION!