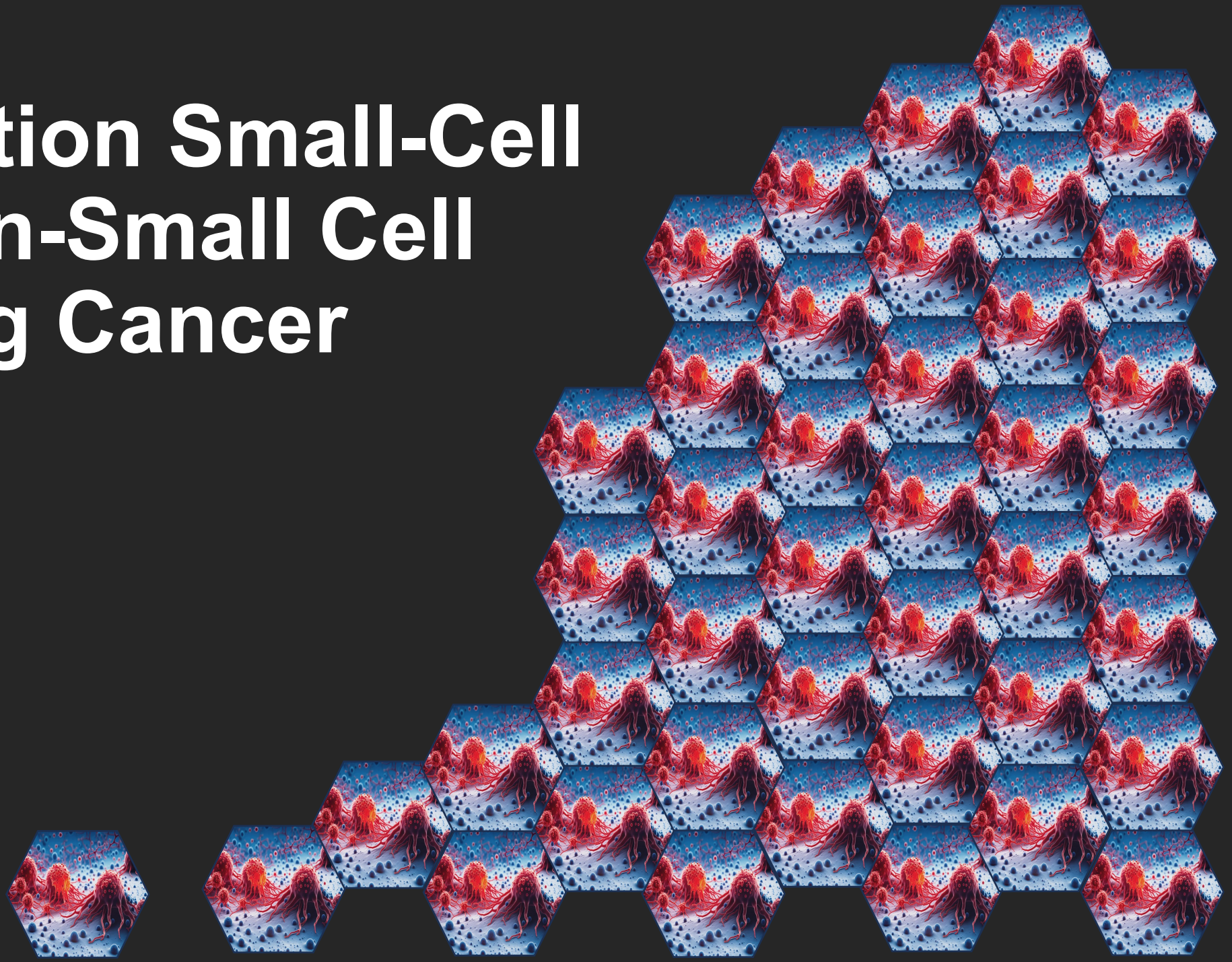
Classification Small-Cell and Non-Small Cell Lung Cancer

Gordon Chan
Sai Shamanth Chedde
Kadi-Ann Hinds
Ankit Srivastava



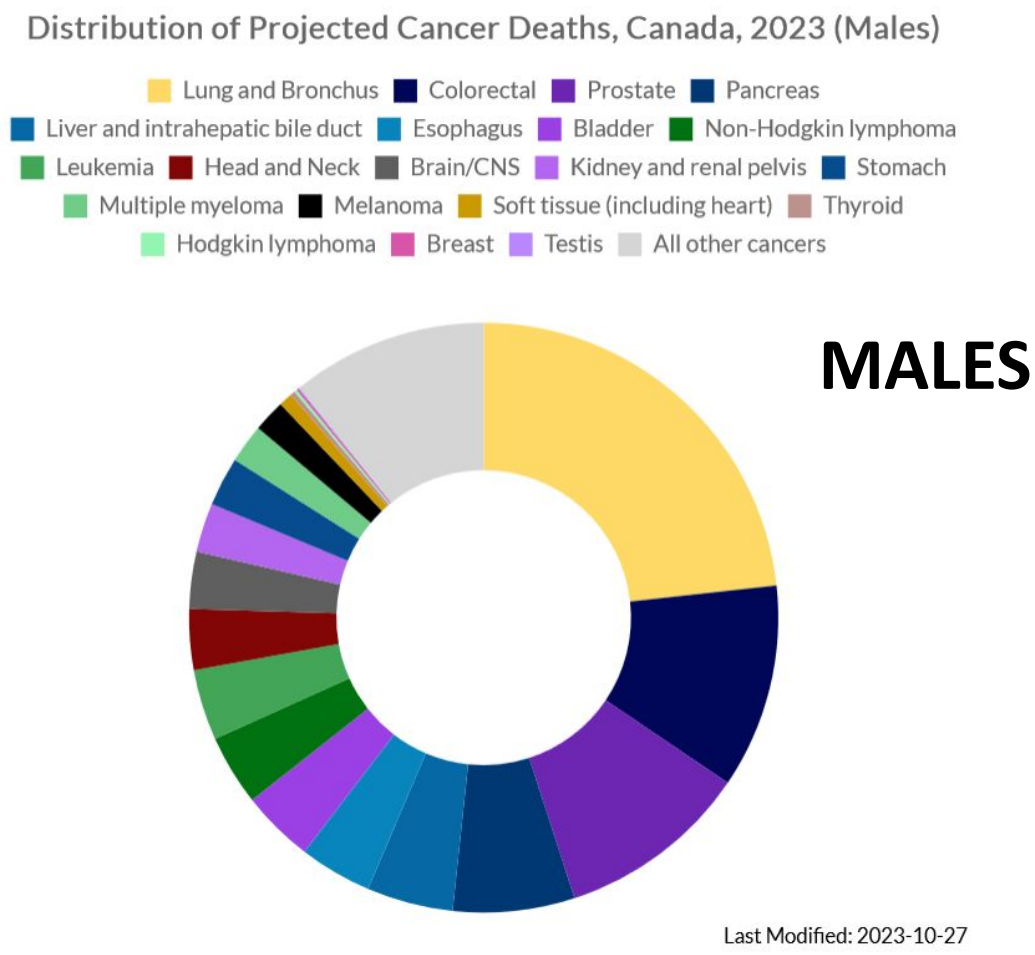
Lung Cancer

1 in 14 Canadian will develop lung cancer during their lifetime

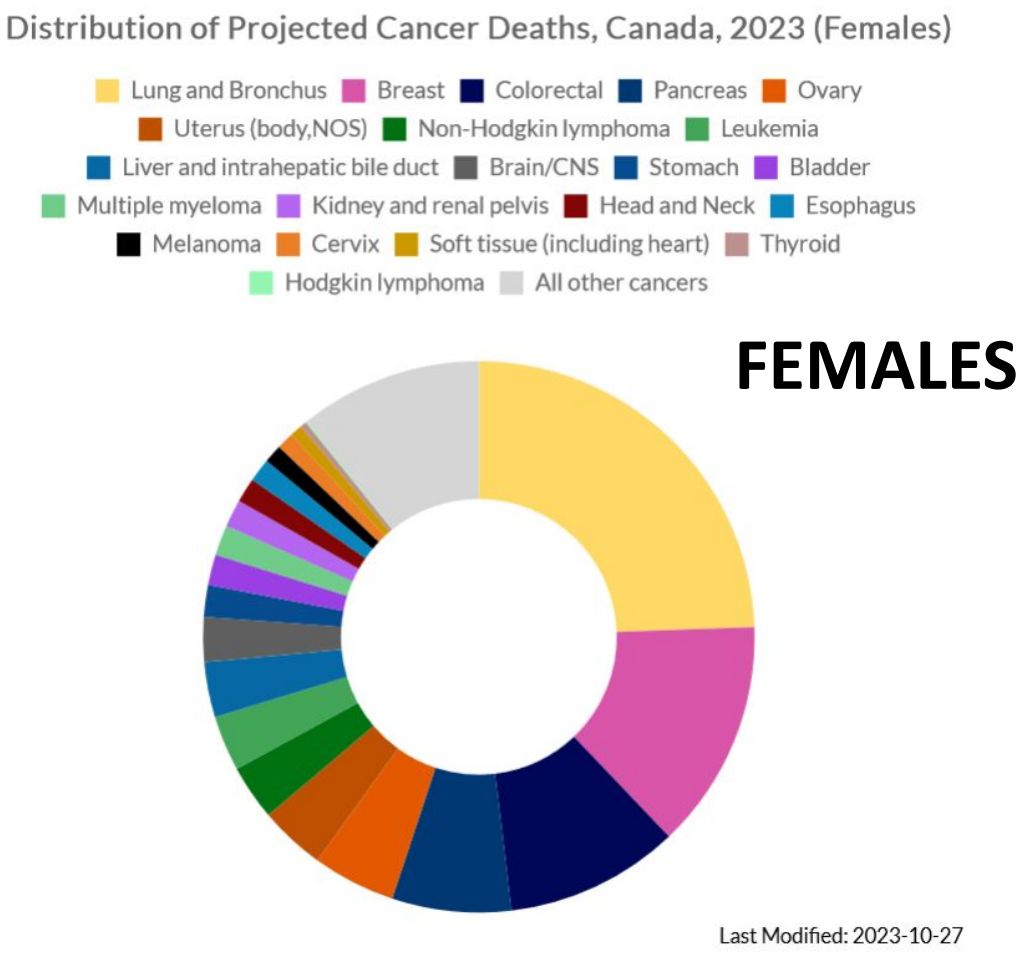
1 in 20 will die from it.

Canadian Cancer Society

Distribution of Projected Cancer Deaths in Canada, 2023



MALES



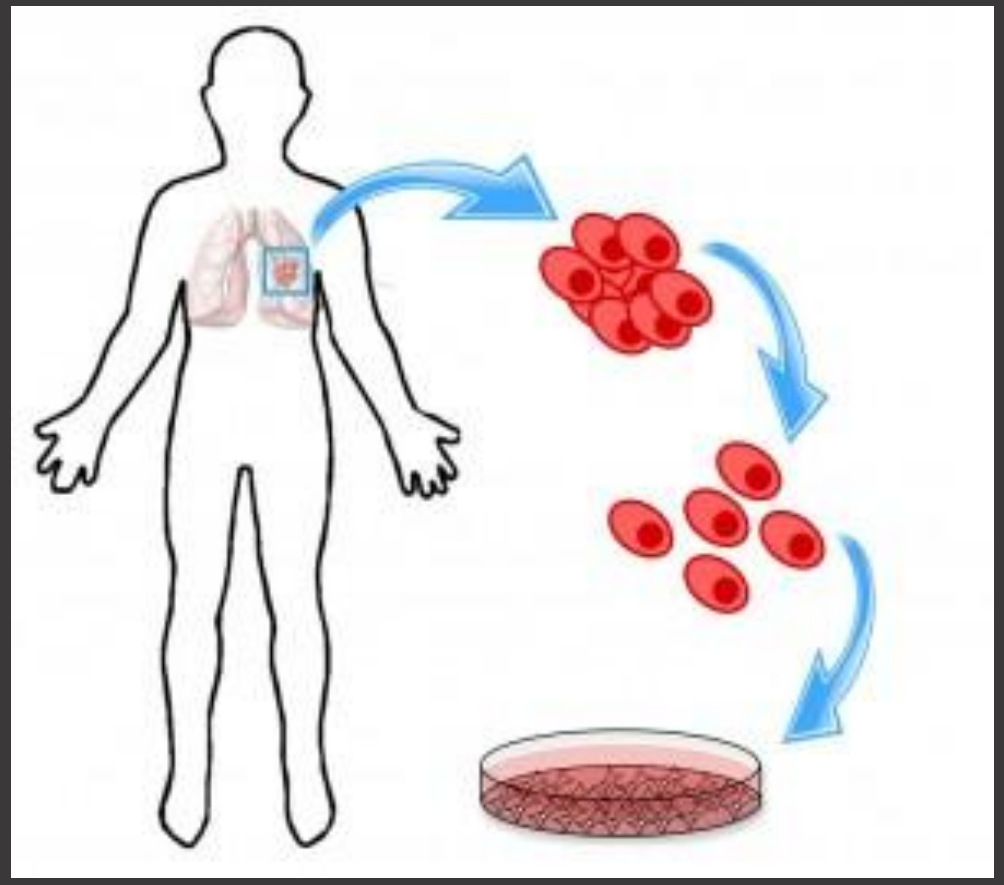
FEMALES

Lung Cancer

Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC) are the two main types of lung cancer

- **Most lung cancers are NSCLC – usually about 80 – 90%.**
- **Only about 15% of lung cancers are SCLC.**

Making Tumor Samples as models of the diseases



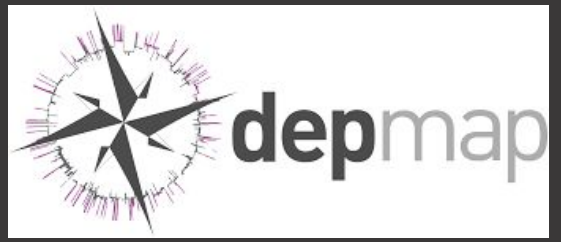
Lung Cancer Tumor Samples

Small Cell **49**

Non-Small Cell **119**

959 features

(concentration of different molecules in the cells)



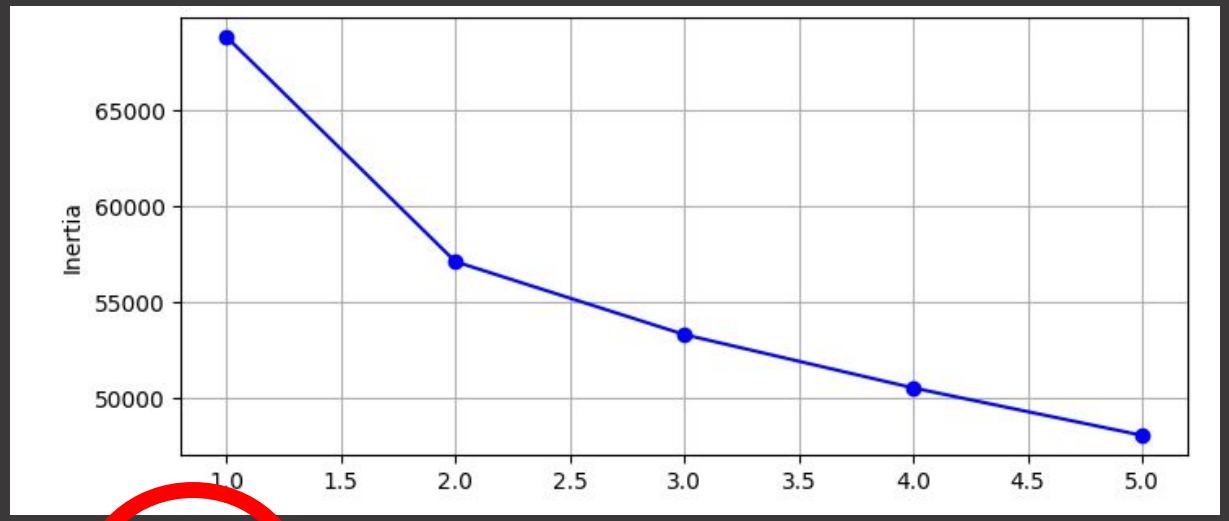
Objective:

To identify the molecular differences between Small Cell and Non-Small Cell Lung Cancer

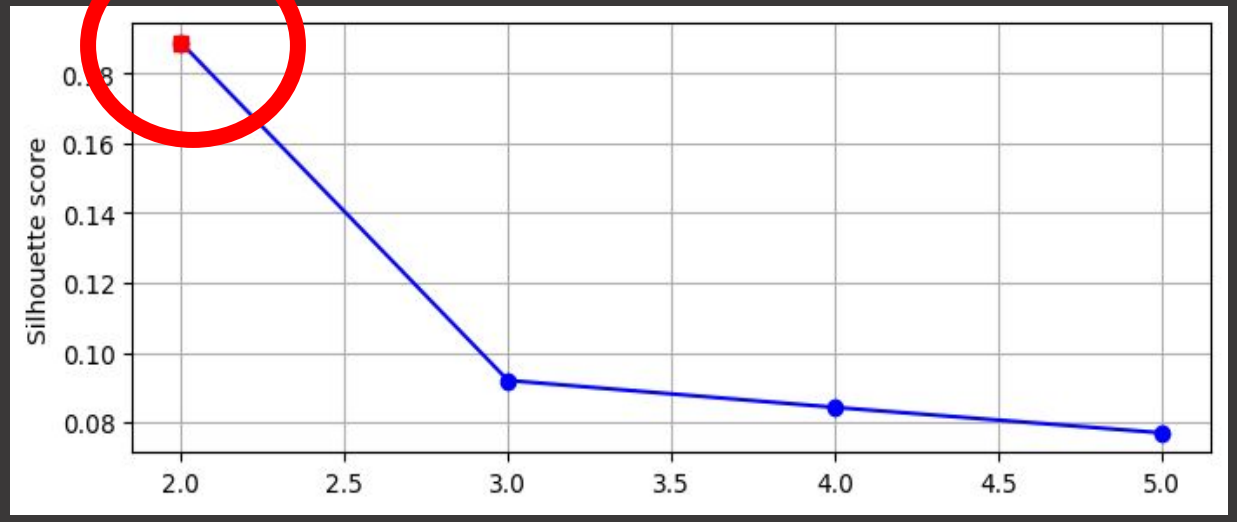
If we use these features, how many clusters do we get?

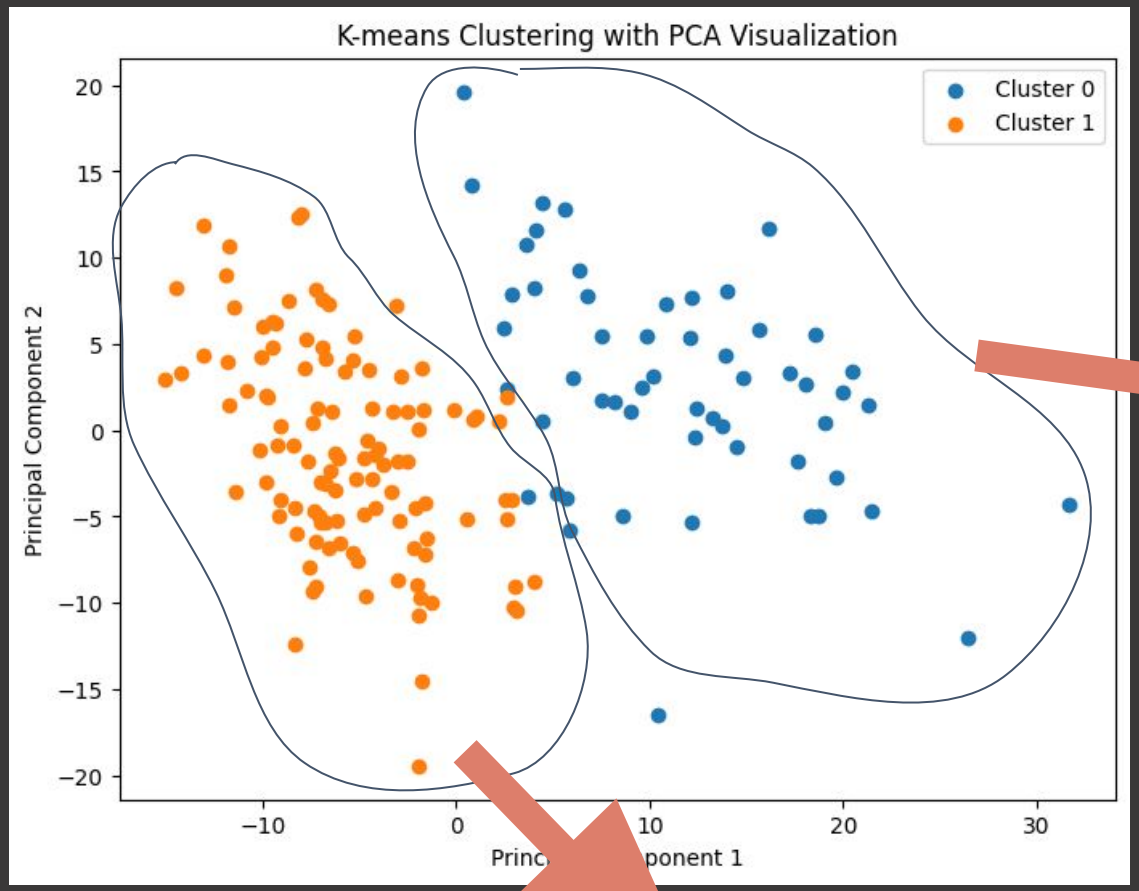
Clustering by KMeans

Inertia:

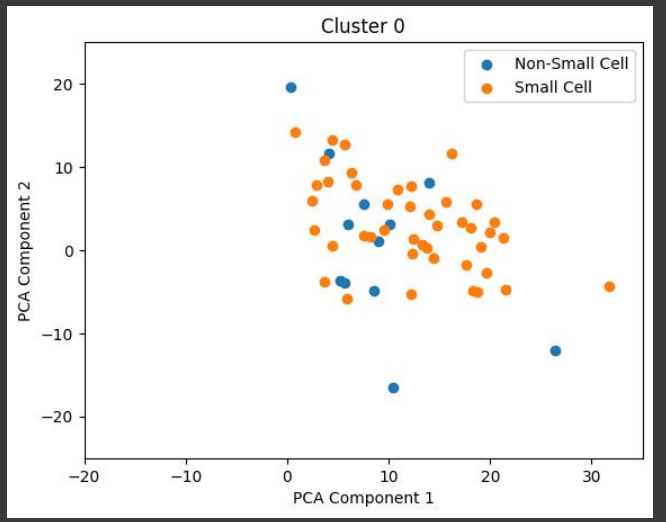


Silhouette Score:

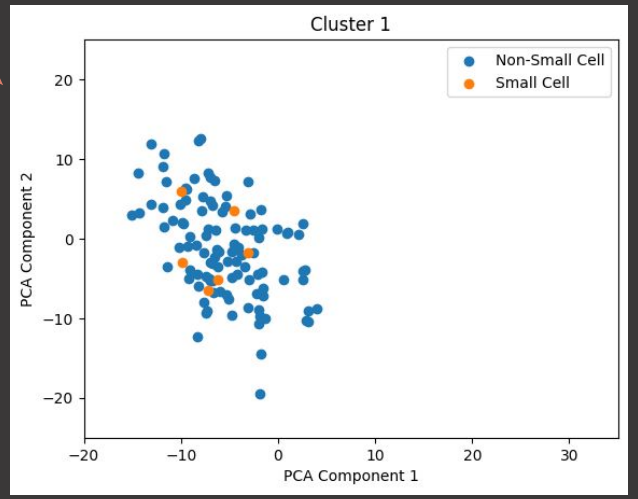




Cluster 0:
Small Cell 43,
Non-Small Cell 12

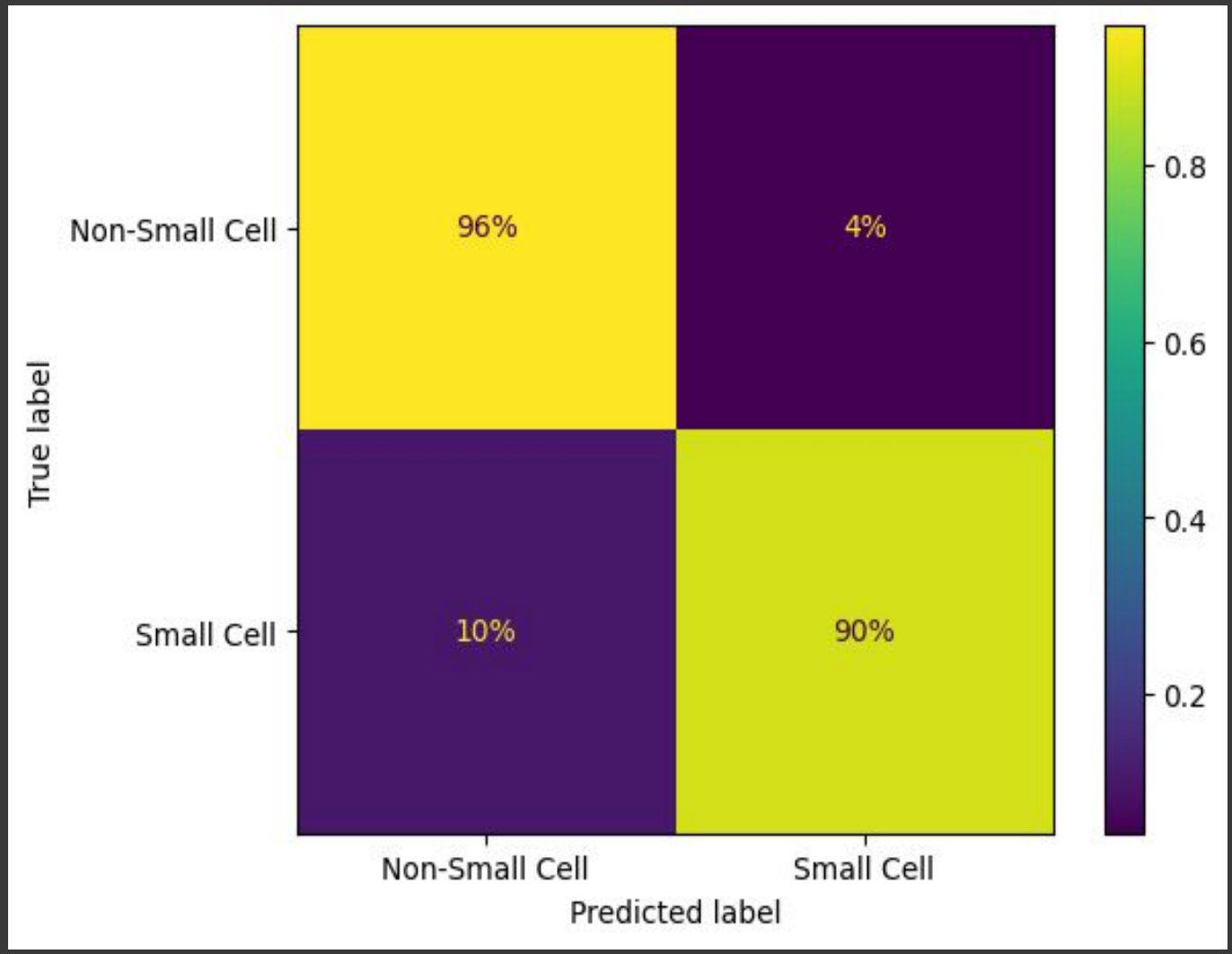


Cluster 1:
Small Cell 6,
Non-Small Cell 107



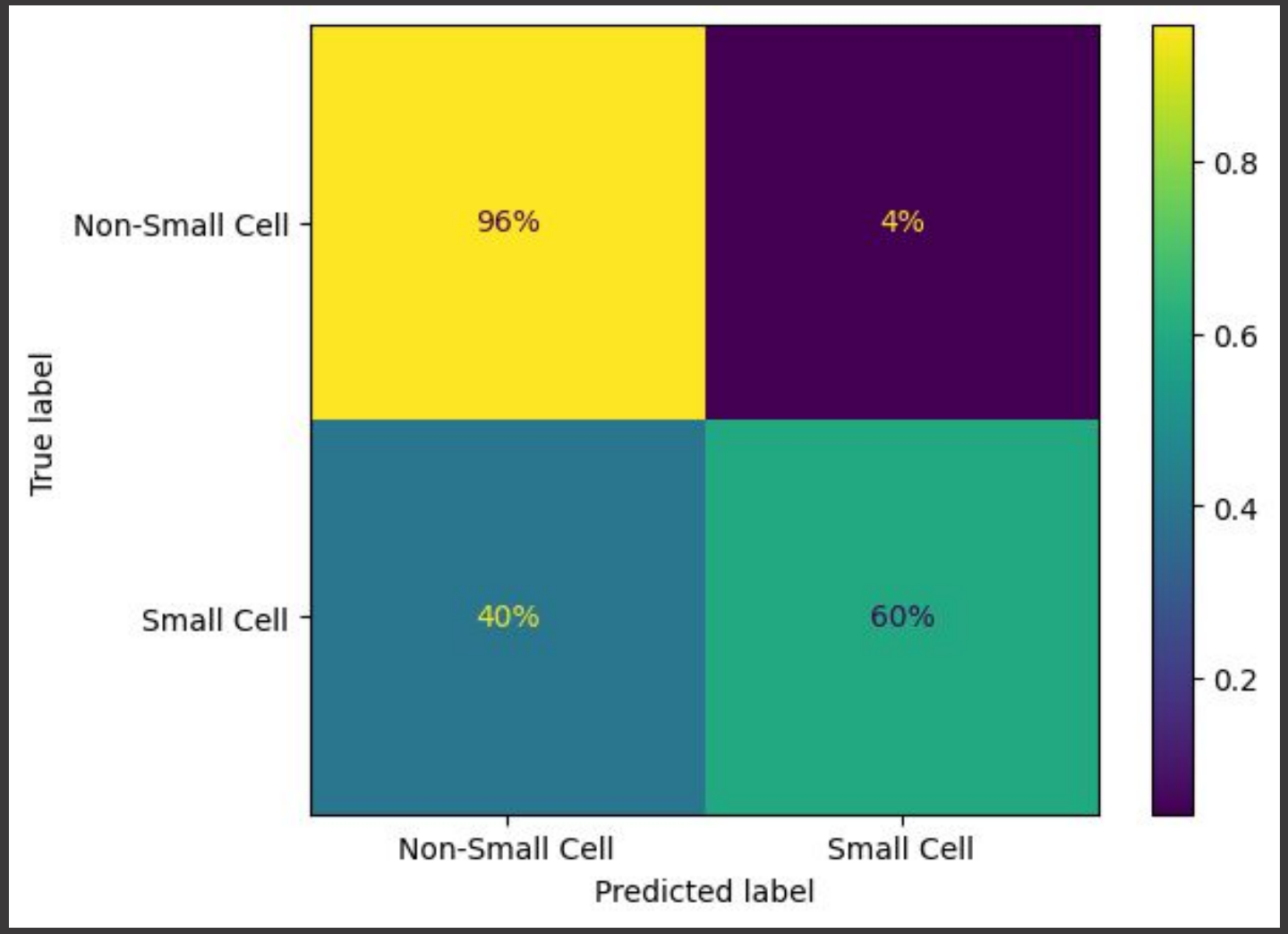
Classification

Logistic Regression



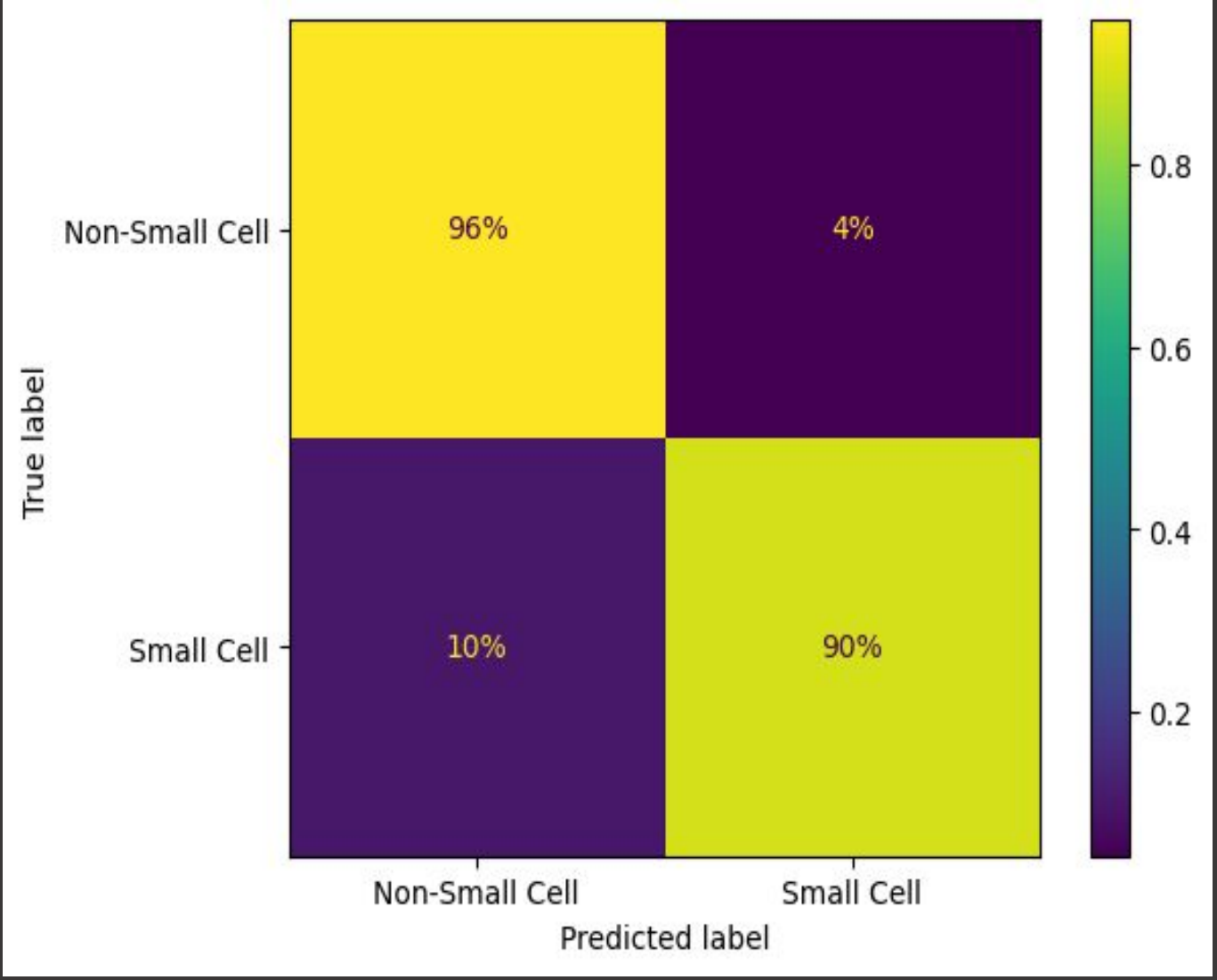
F1 = 0.94

SVC


Non-Small Cell
Small Cell
True label
96%
4%
40%
60%
Non-Small Cell
Small Cell
Predicted label
0.8
0.6
0.4
0.2

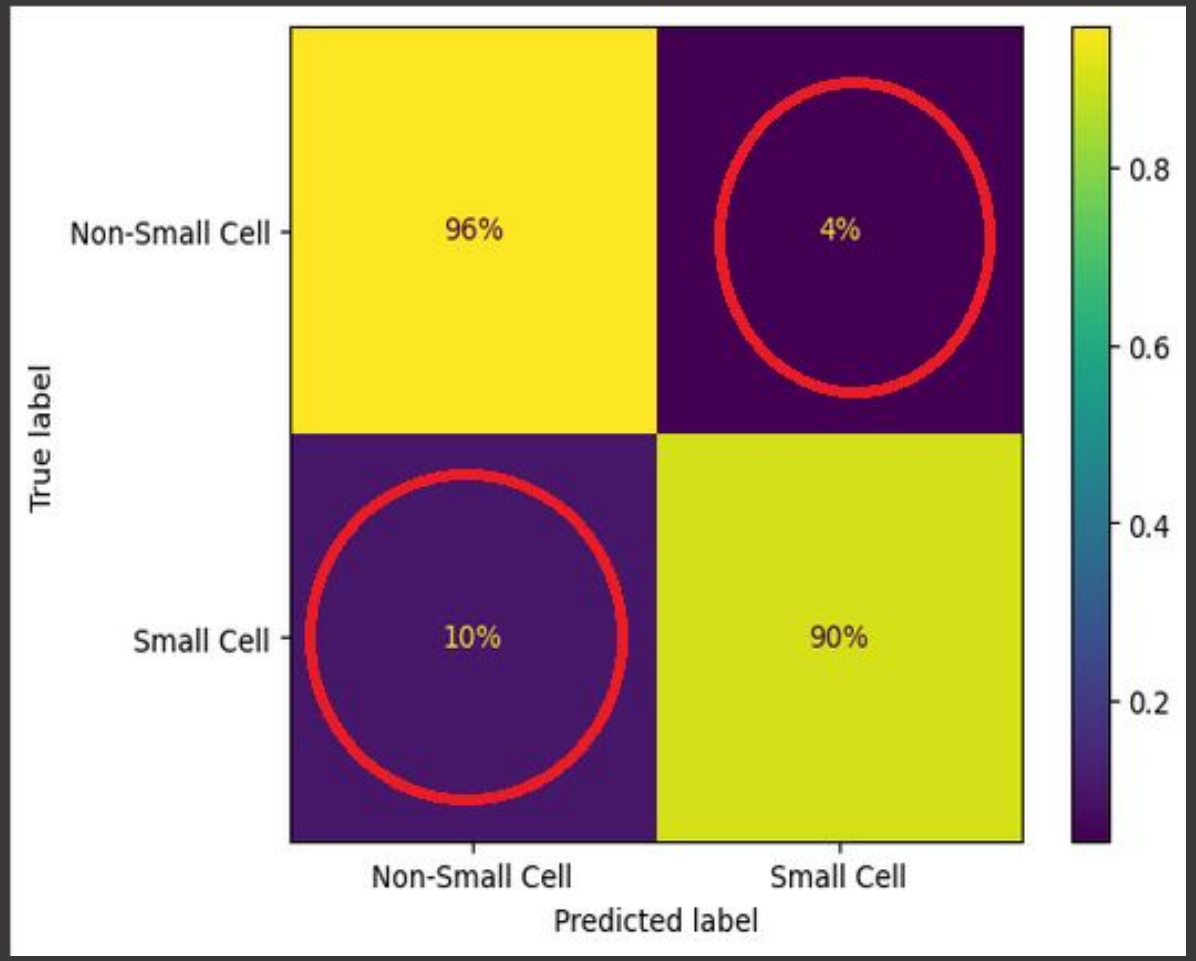
F1 = 0.84

RandomForest

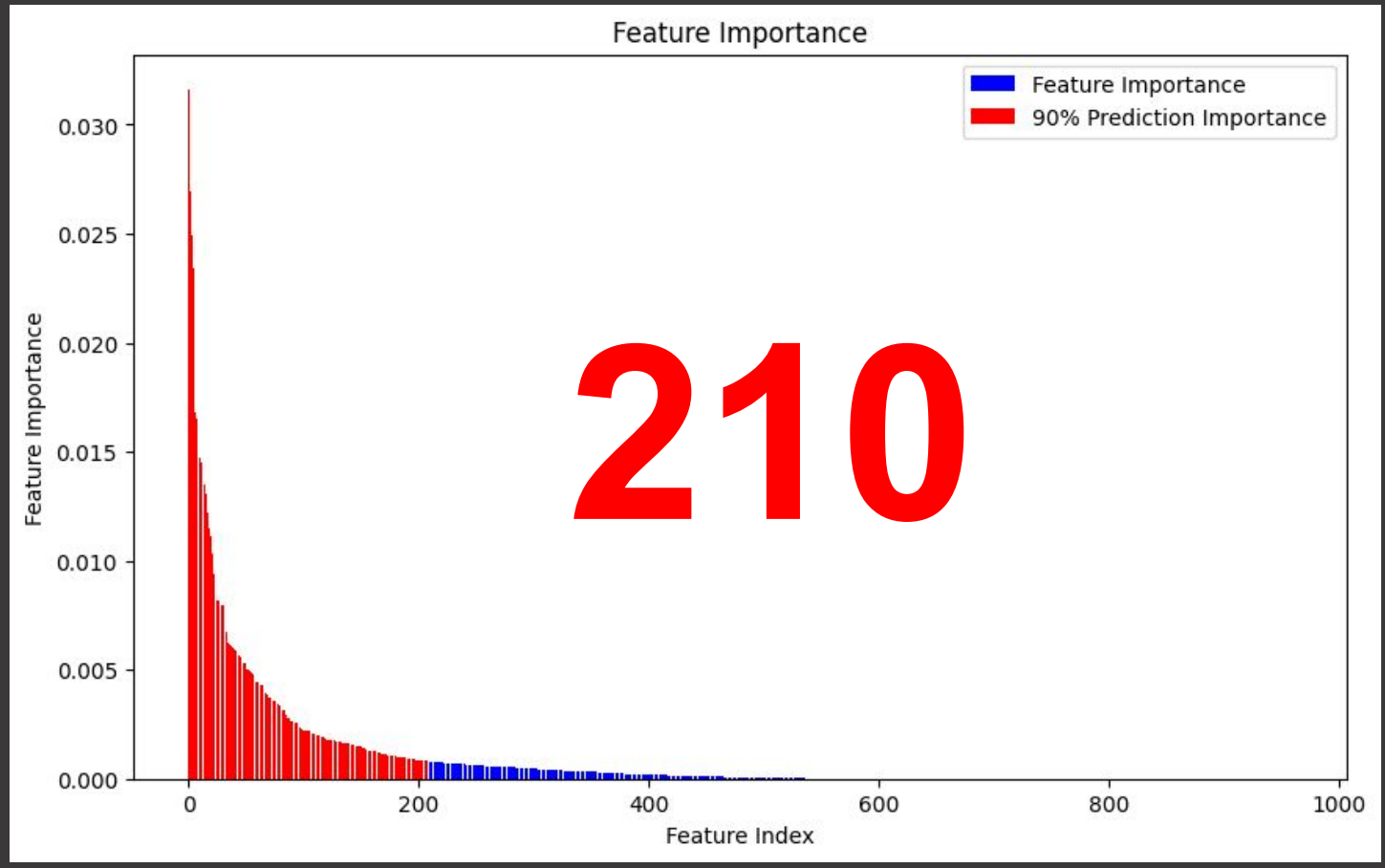


F1 = 0.94

Exploring Misclassified Instances

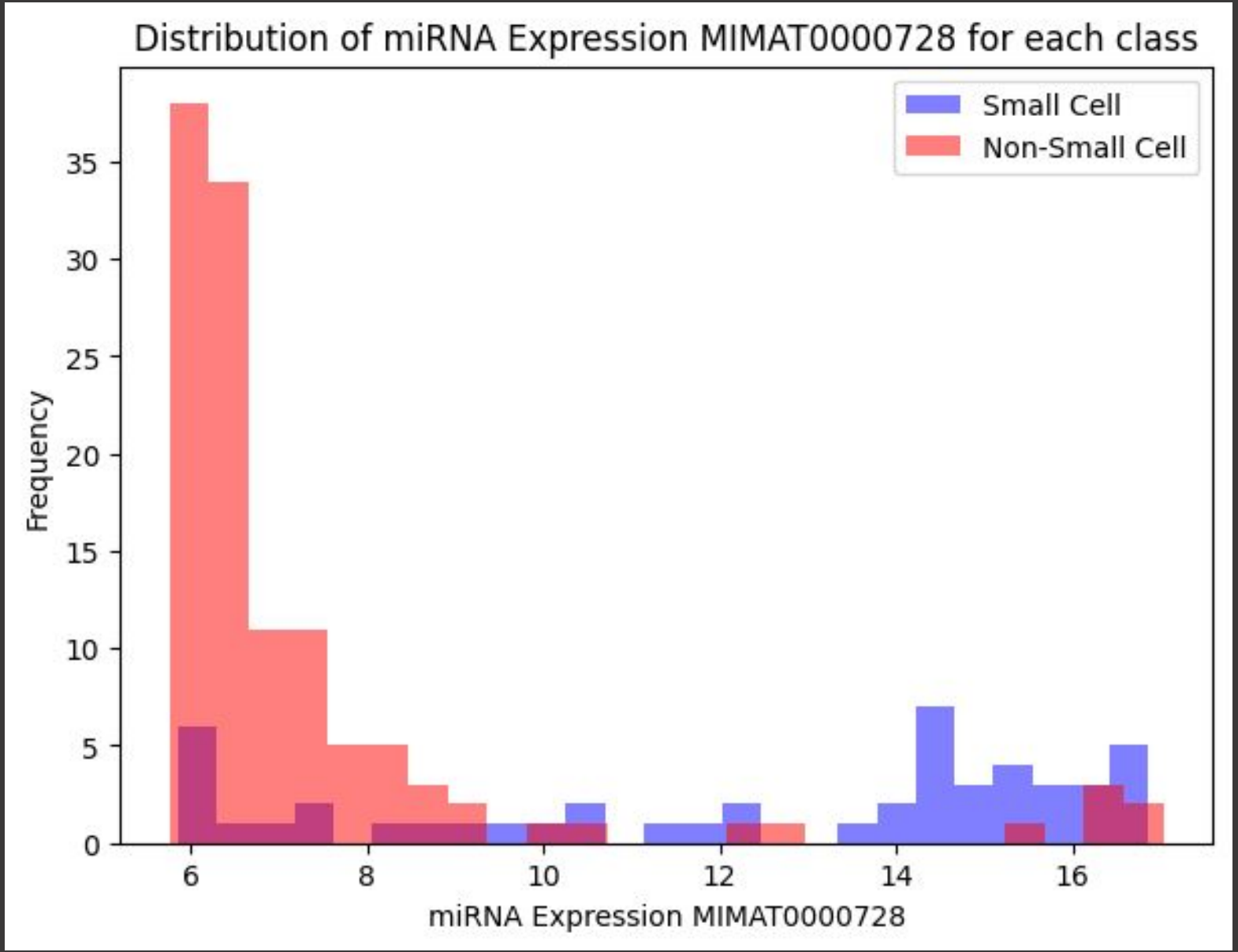


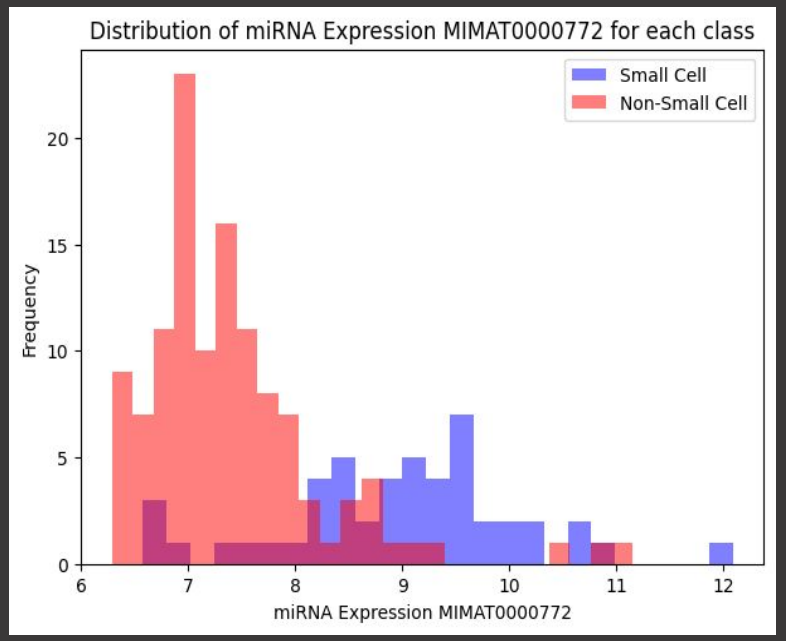
Using RandomForest, how many of the 959 features accounts for 90% of the prediction?

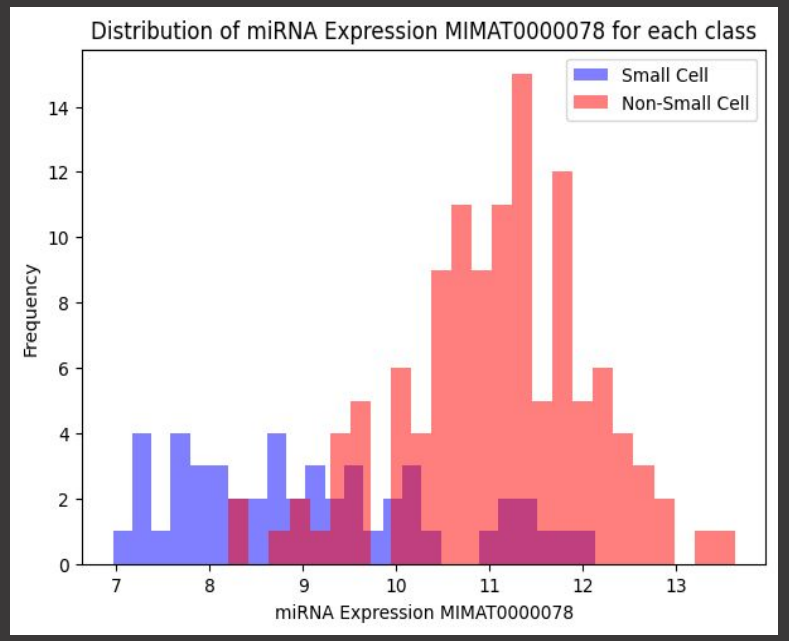


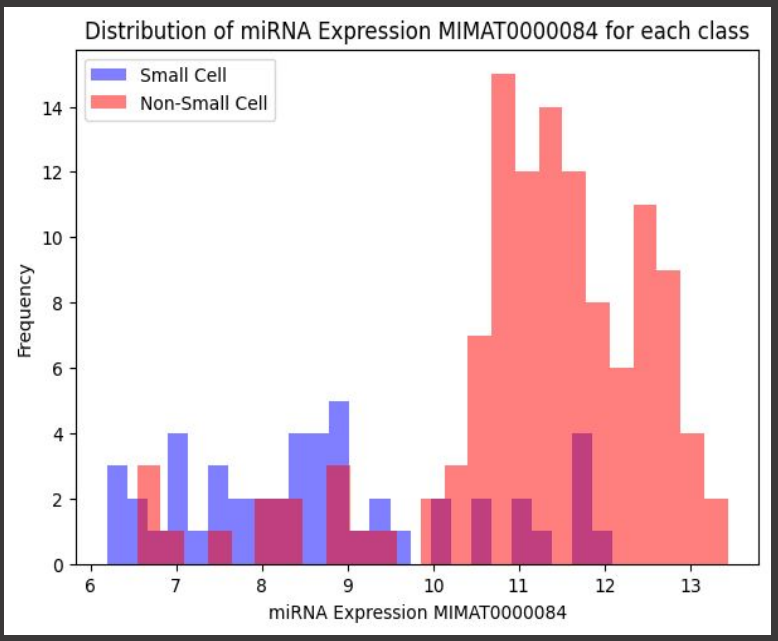
Top 10 Features

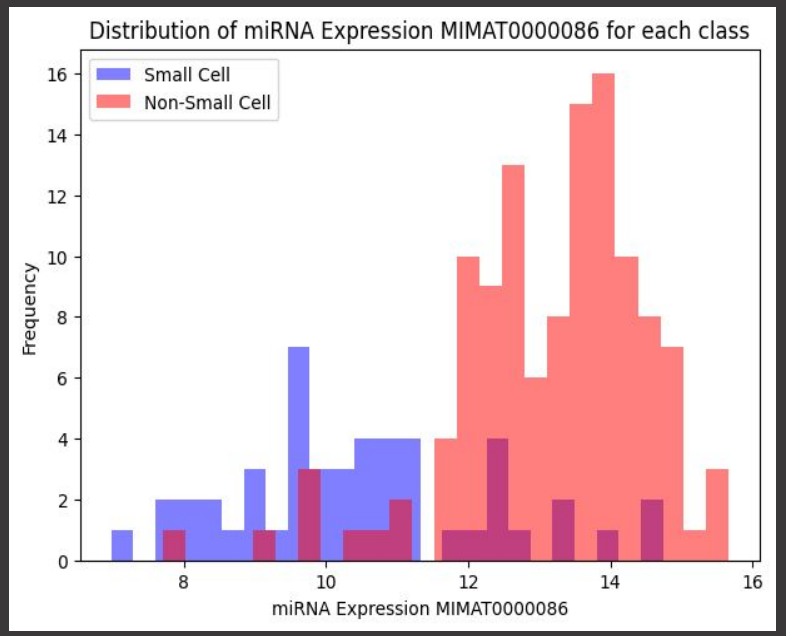
1. miRNA Expression MIMAT0000728
2. miRNA Expression MIMAT0000772
3. miRNA Expression MIMAT0000078
4. miRNA Expression MIMAT0000084
5. miRNA Expression MIMAT0000086
6. miRNA Expression MIMAT0000428
7. Metabolomics 2-deoxycytidine
8. miRNA Expression MIMAT0003266
9. miRNA Expression MIMAT0000077
10. miRNA Expression MIMAT0000094

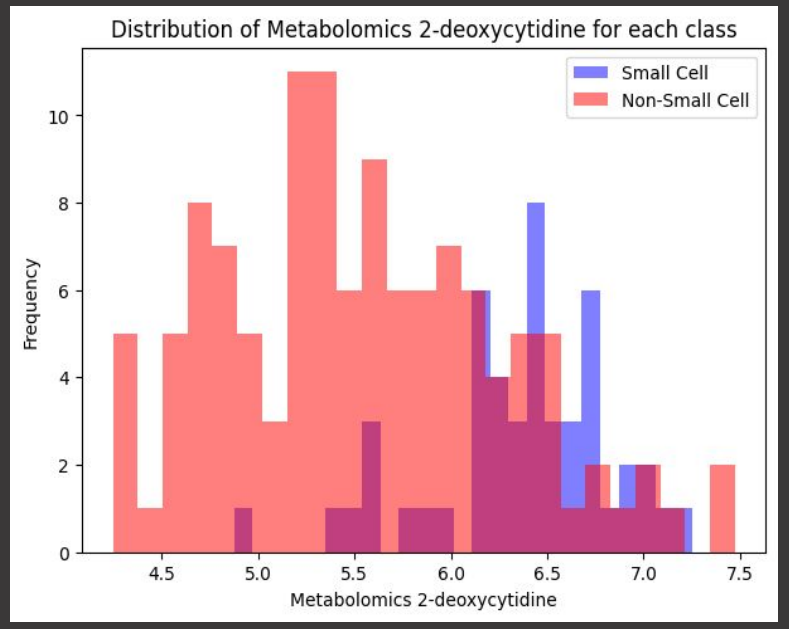

Distribution of miRNA Expression MIMAT0000728 for each class
Small Cell
Non-Small Cell
Frequency
0
5
10
15
20
25
30
35
6
8
10
12
14
16
miRNA Expression MIMAT0000728

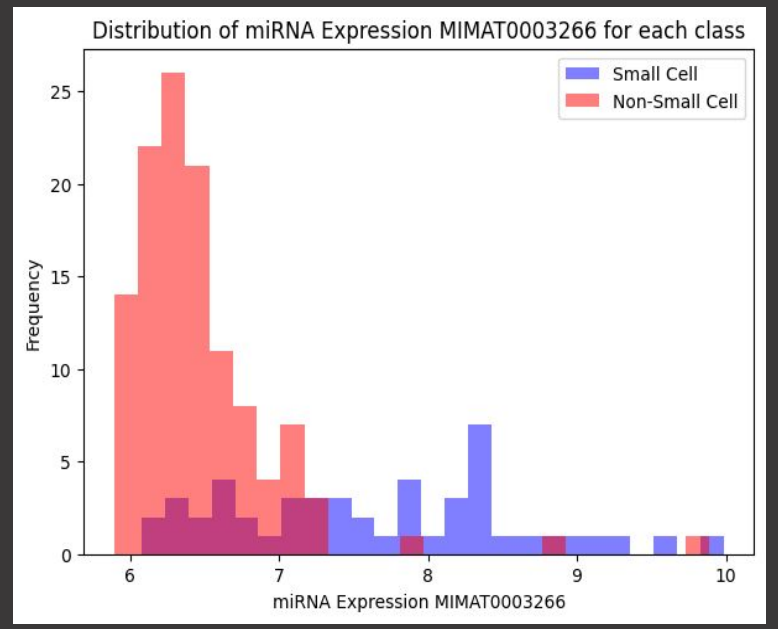

Distribution of miRNA Expression MIMAT0000772 for each class
Small Cell
Non-Small Cell
Frequency
miRNA Expression MIMAT0000772

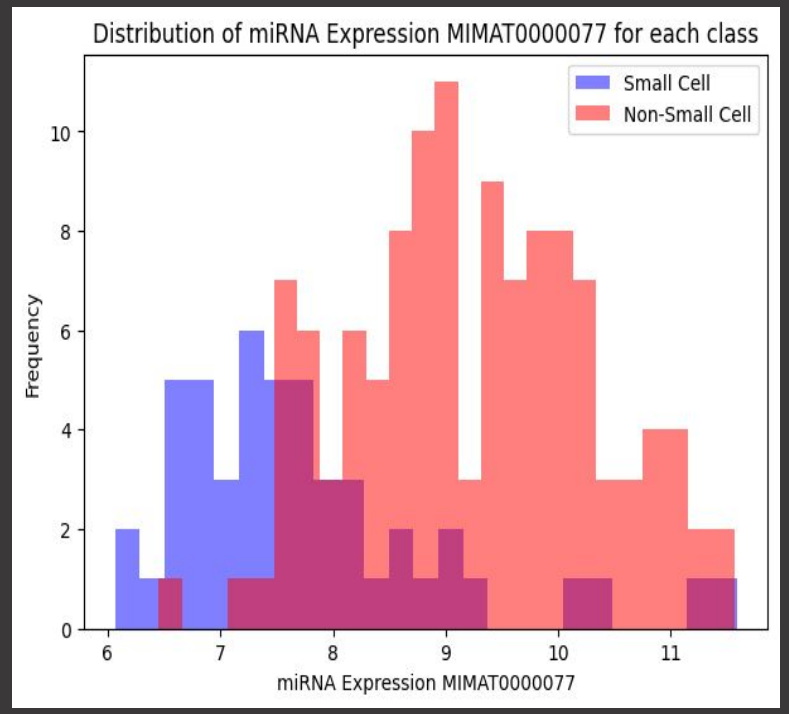

Distribution of miRNA Expression MIMAT0000078 for each class
Small Cell
Non-Small Cell
Frequency
miRNA Expression MIMAT0000078


Distribution of miRNA Expression MIMAT0000084 for each class
Small Cell
Non-Small Cell
Frequency
miRNA Expression MIMAT0000084


Distribution of miRNA Expression MIMAT0000086 for each class
Small Cell
Non-Small Cell
Frequency
miRNA Expression MIMAT0000086


Distribution of Metabolomics 2-deoxycytidine for each class
Small Cell
Non-Small Cell
Frequency
Metabolomics 2-deoxycytidine


Distribution of miRNA Expression MIMAT0003266 for each class
Small Cell
Non-Small Cell
Frequency
miRNA Expression MIMAT0003266



ID for MiR-22-3p

The clinical significance and mechanism of microRNA-22-3p targeting TP53 in lung adenocarcinoma

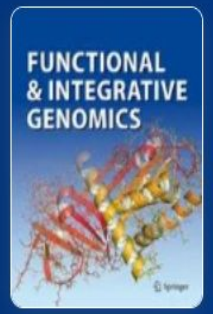
Cite

Article type: Research Article

Authors: Lin, Rui[a] | Li, Guo-Sheng[b] | Gan, Xiang-Yu[a] | Peng, Jun-Xi[a] | Feng, Yue[a] | Wang, Li-Ting[a] | Zhang, Chu-Yue[a] | Huang, Kun-Ying[a] | Huang, Shi-Hai[a] | Yang, Lin[a] | Kong, Jin-Liang[c] | Zhou, Hua-Fu[b] | Chen, Gang[a] | Huang, Wan-Ying[a;*]

ical University,
ed Hospital of
ical Care
i Medical

MiR-22-3p suppresses NSCLC cell migration and EMT via targeting RAC1 expression



Original Article | Open access | Published: 25 August 2023

Volume 23, article number 281, (2023) Cite this article

Lessons learnt

- Small datasets can have value, especially with lots of features
- Real data is messy, can still be useful

Challenges faced

- Small dataset, ~ 170 rows
- Finding a “good” dataset
- Lots of weak features