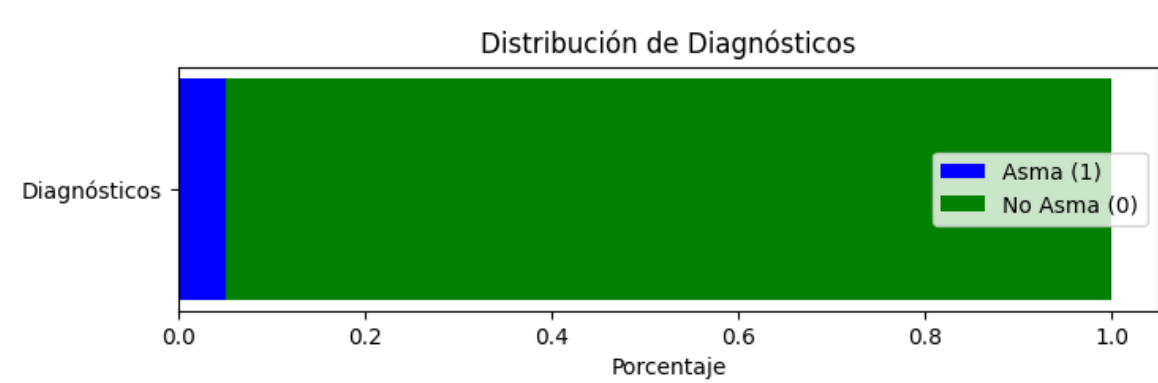
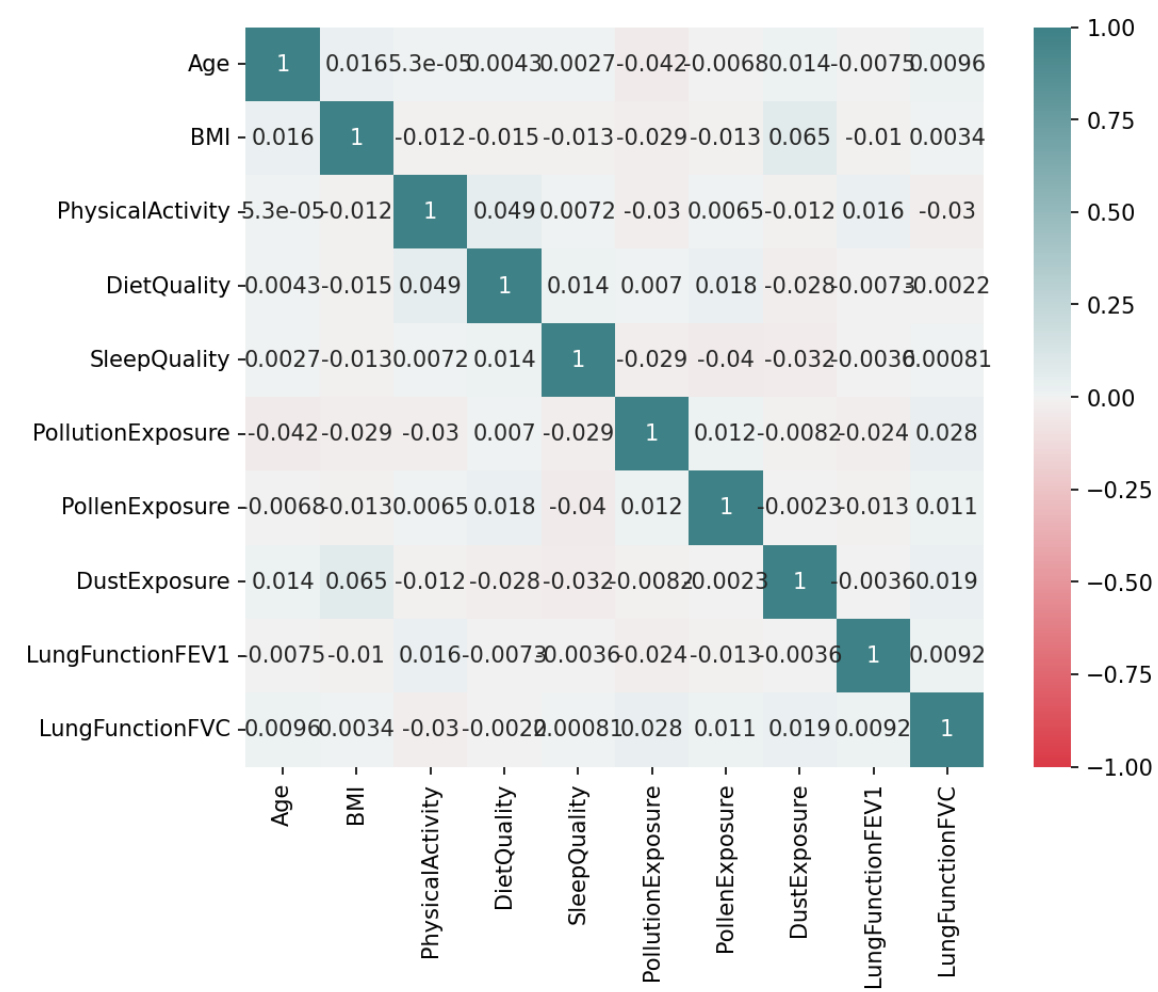
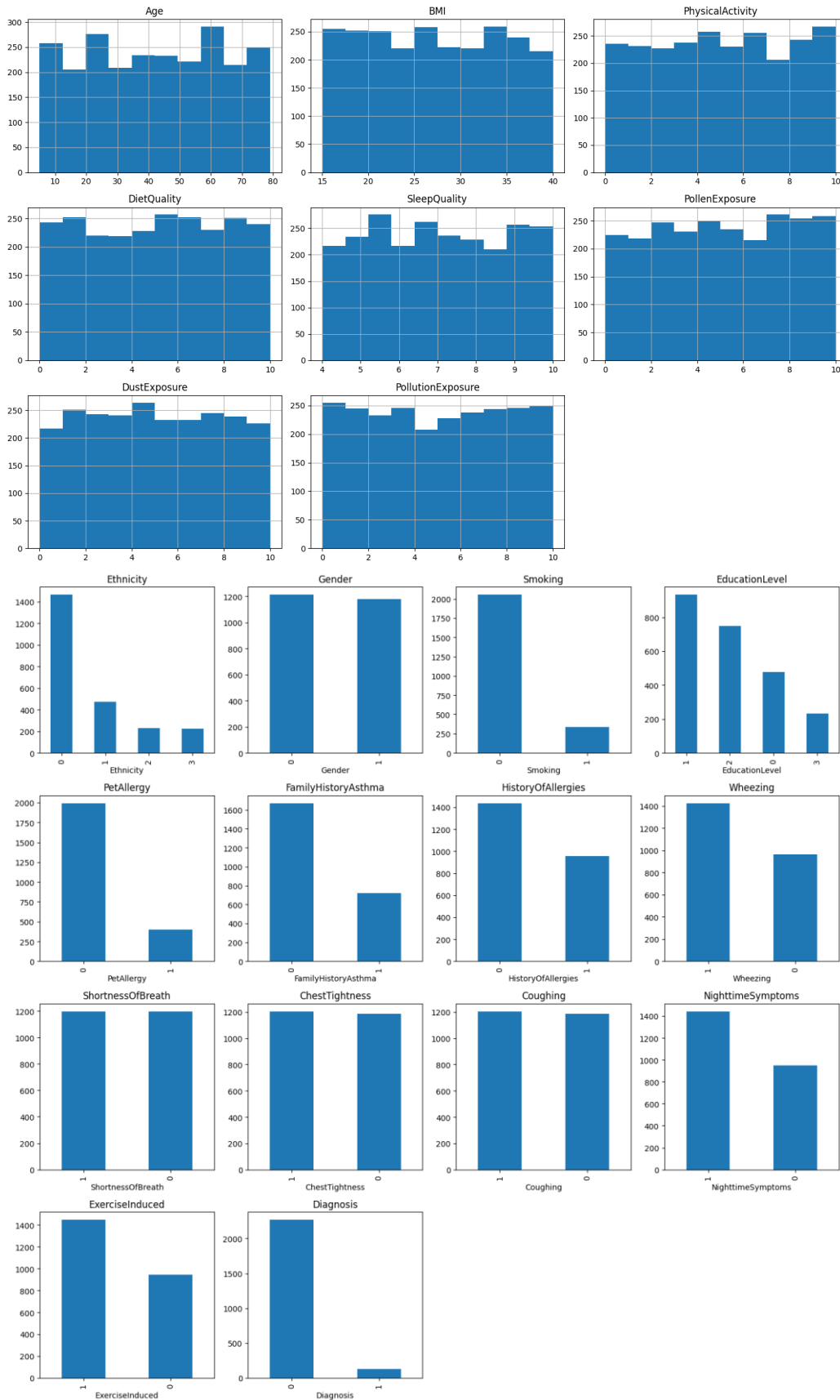


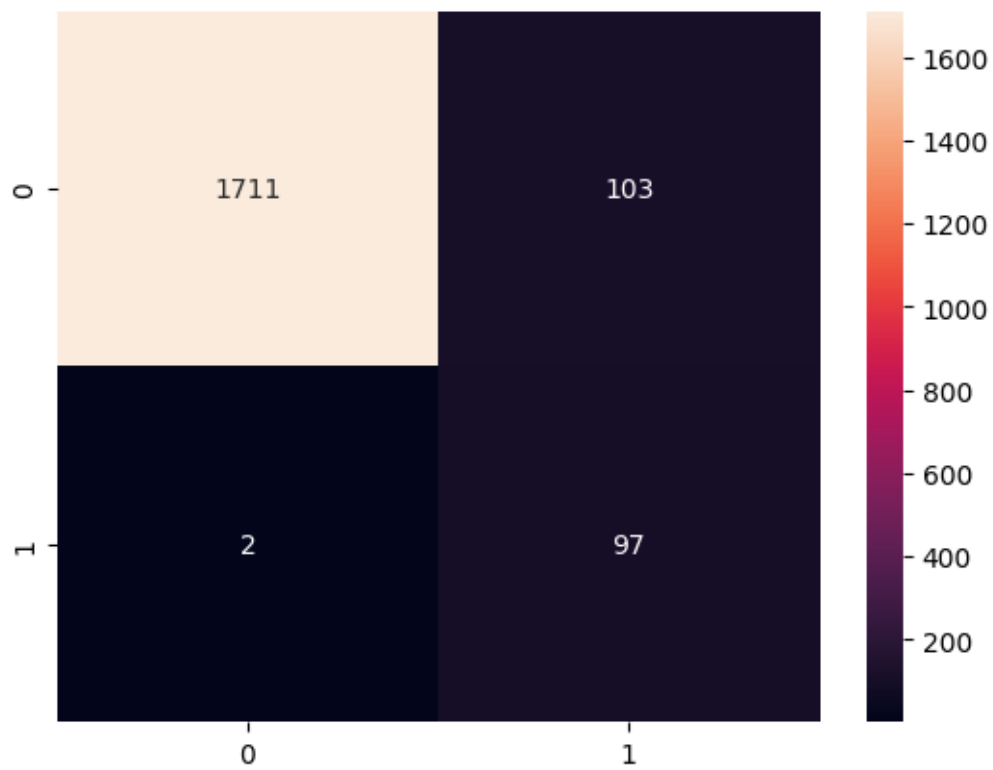
	Age	BMI	PhysicalActivity	DietQuality	SleepQuality	PollutionExposure	PollenExposure	DustExposure	LungFunctionFEV1	LungFunctionFVC
0	63	15.848744	0.894448	5.488696	8.701003	7.388481	2.855578	0.974339	1.369051	4.941206
1	26	22.757042	5.897329	6.341014	5.153966	1.969838	7.457665	6.584631	2.197767	1.702393
2	57	18.395396	6.739367	9.196237	6.840647	1.460593	1.448189	5.445799	1.698011	5.022553
3	40	38.515278	1.404503	5.826532	4.253036	0.581905	7.571845	3.965316	3.032037	2.300159
4	61	19.283802	4.604493	3.127048	9.625799	0.980875	3.049807	8.260605	3.470589	3.067944

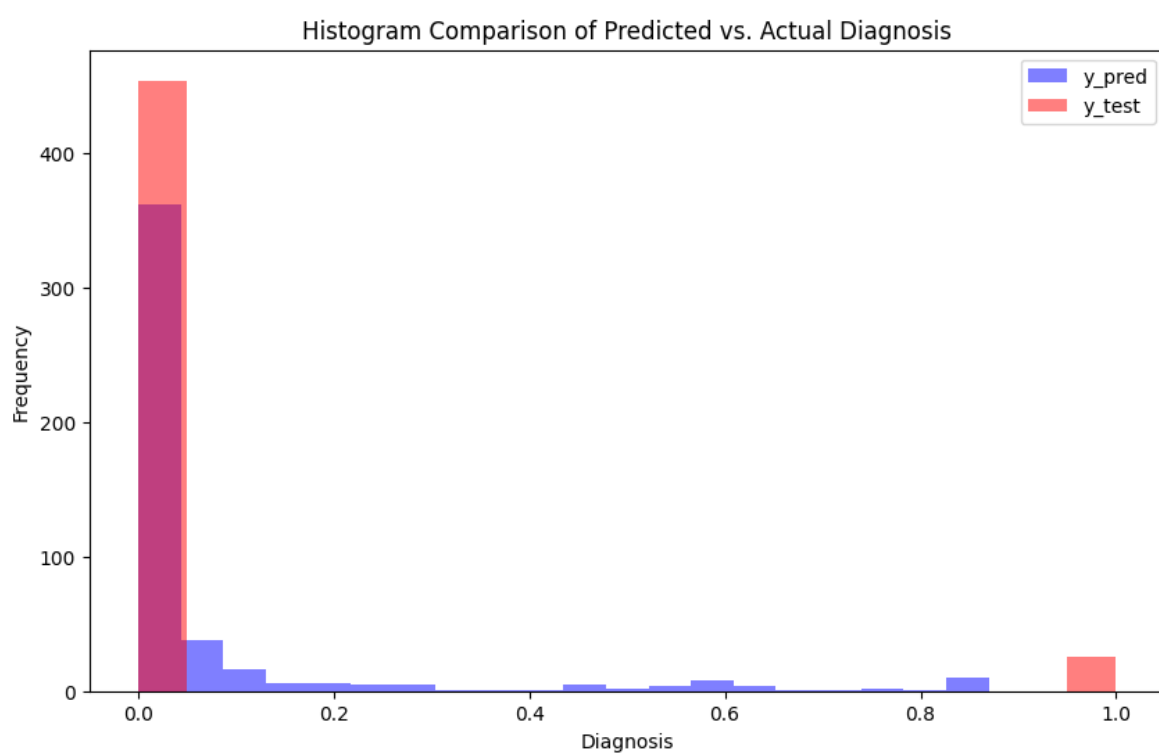
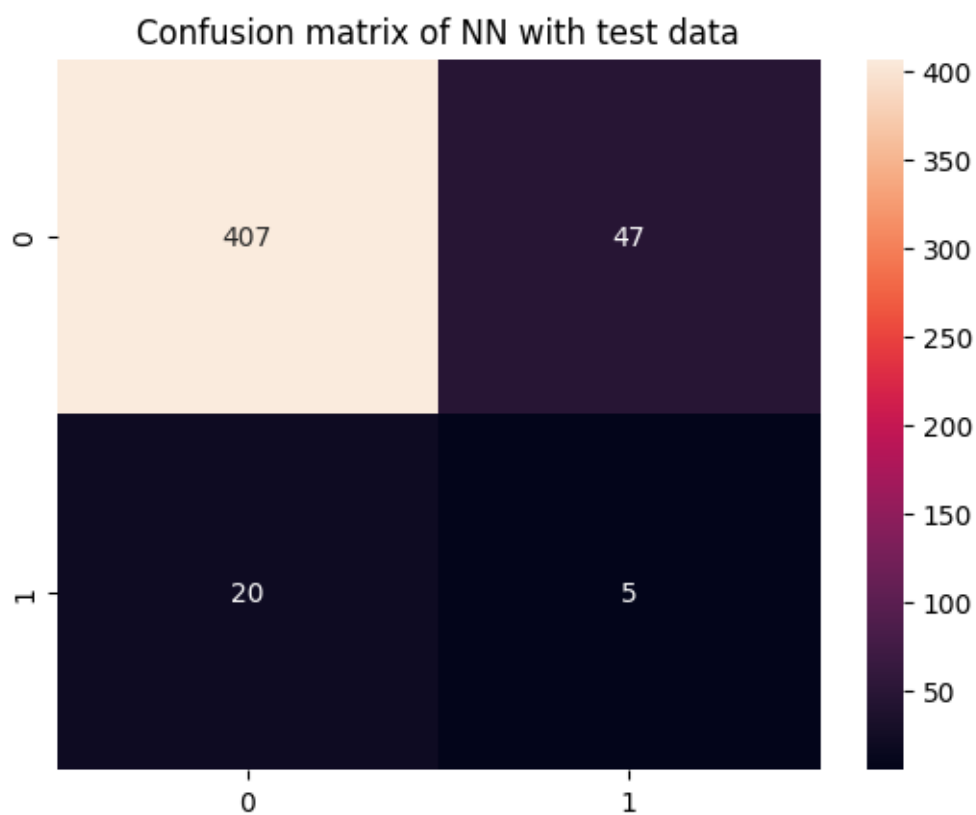


```
NA counts for each column:
Age                                0
Gender                            0
Ethnicity                         0
EducationLevel                    0
BMI                               0
Smoking                          0
PhysicalActivity                  0
DietQuality                       0
SleepQuality                      0
PollutionExposure               0
PollenExposure                   0
DustExposure                     0
PetAllergy                       0
FamilyHistoryAsthma              0
HistoryOfAllergies                0
Wheezing                         0
ShortnessOfBreath                0
ChestTightness                   0
Coughing                         0
NighttimeSymptoms                0
ExerciseInduced                   0
Diagnosis                        0
dtype: int64
```

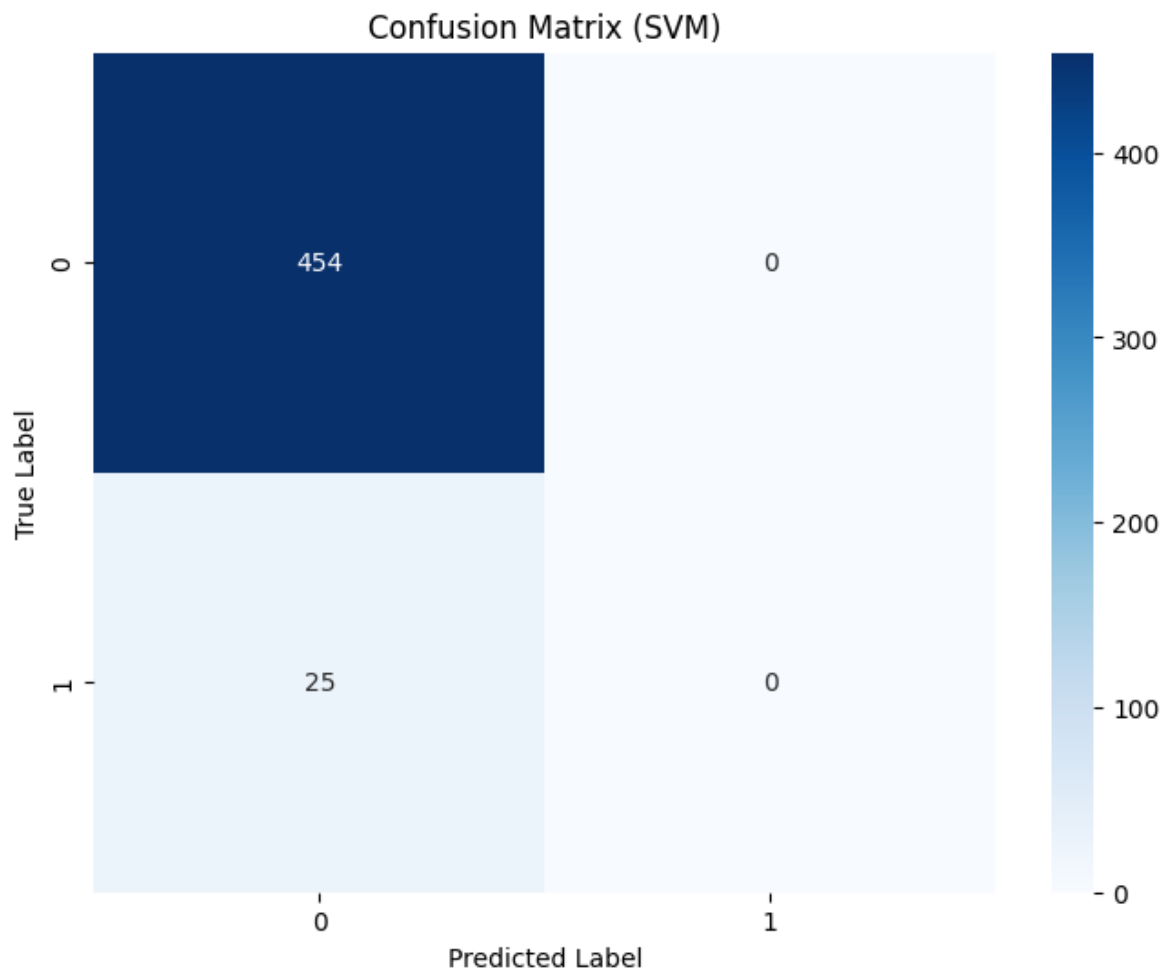


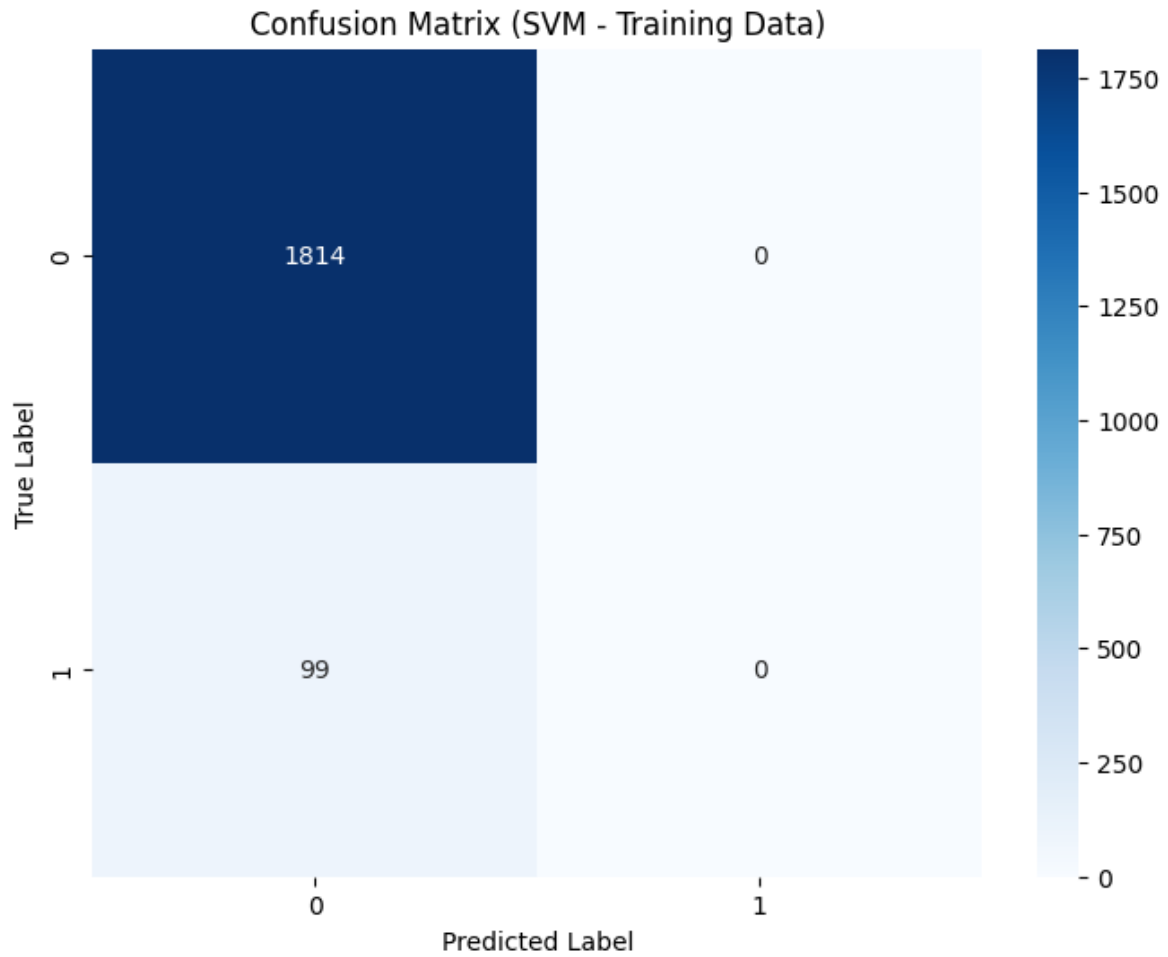
Confusion matrix of NN with train data





SVM Accuracy: 0.9478079331941545					
		precision	recall	f1-score	support
	0	0.95	1.00	0.97	454
	1	0.00	0.00	0.00	25
	accuracy			0.95	479
	macro avg	0.47	0.50	0.49	479
	weighted avg	0.90	0.95	0.92	479

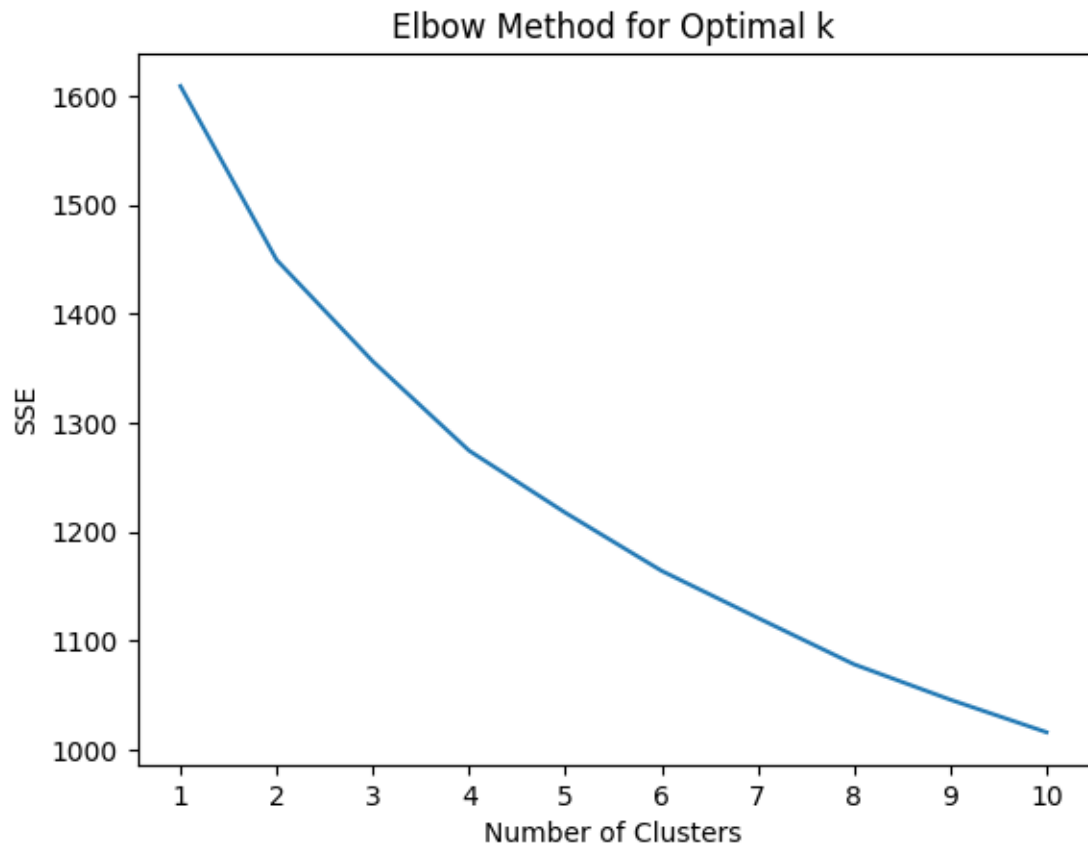




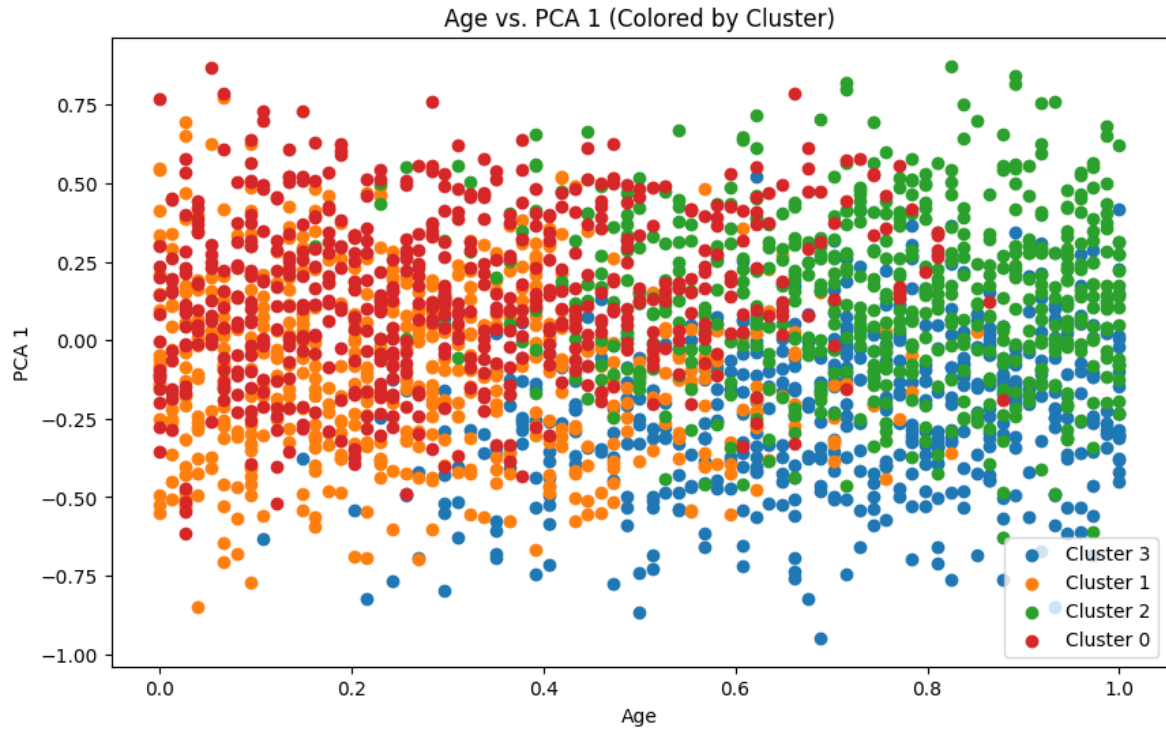
SVM Cross-validation scores: [0.9519833 0.94572025 0.95606695 0.95606695 0.93096234]
SVM Mean accuracy: 0.9481599566740332

Neural Network Cross-validation scores: [0.9519833 0.94572025 0.93096234 0.95606695 0.93096234]
Neural Network Mean accuracy: 0.9431390361719412

	Age	BMI	PhysicalActivity	DietQuality	SleepQuality	PollenExposure	DustExposure	PollutionExposure
0	0.783784	0.032738	0.089324	0.548738	0.783941	0.285523	0.097217	0.738898
1	0.283784	0.309582	0.589909	0.633997	0.192255	0.745783	0.658399	0.196922
2	0.702703	0.134793	0.674163	0.919608	0.473612	0.144769	0.544485	0.145987
3	0.472973	0.941078	0.140359	0.582532	0.041970	0.757202	0.396396	0.058100
4	0.756757	0.170395	0.460548	0.312499	0.938207	0.304948	0.826042	0.098005



	Age	BMI	PhysicalActivity	DietQuality	SleepQuality	PollenExposure	DustExposure	PollutionExposure	cluster
0	0.783784	0.032738	0.089324	0.548738	0.783941	0.285523	0.097217	0.738898	3
1	0.283784	0.309582	0.589909	0.633997	0.192255	0.745783	0.658399	0.196922	1
2	0.702703	0.134793	0.674163	0.919608	0.473612	0.144769	0.544485	0.145987	2
3	0.472973	0.941078	0.140359	0.582532	0.041970	0.757202	0.396396	0.058100	1
4	0.756757	0.170395	0.460548	0.312499	0.938207	0.304948	0.826042	0.098005	2



	Age	BMI	PhysicalActivity	DietQuality	SleepQuality	PollenExposure	DustExposure	PollutionExposure	cluster
0	0.783784	0.032738	0.089324	0.548738	0.783941	0.285523	0.097217	0.738898	1
1	0.283784	0.309582	0.589909	0.633997	0.192255	0.745783	0.658399	0.196922	0
2	0.702703	0.134793	0.674163	0.919608	0.473612	0.144769	0.544485	0.145987	1
3	0.472973	0.941078	0.140359	0.582532	0.041970	0.757202	0.396396	0.058100	0
4	0.756757	0.170395	0.460548	0.312499	0.938207	0.304948	0.826042	0.098005	1

	Diagnosis	Cluster
0	0	1
1	0	0
2	0	1
3	0	0
4	0	1
...
2387	1	1
2388	1	0
2389	1	1
2390	0	0
2391	0	0

[2392 rows x 2 columns]

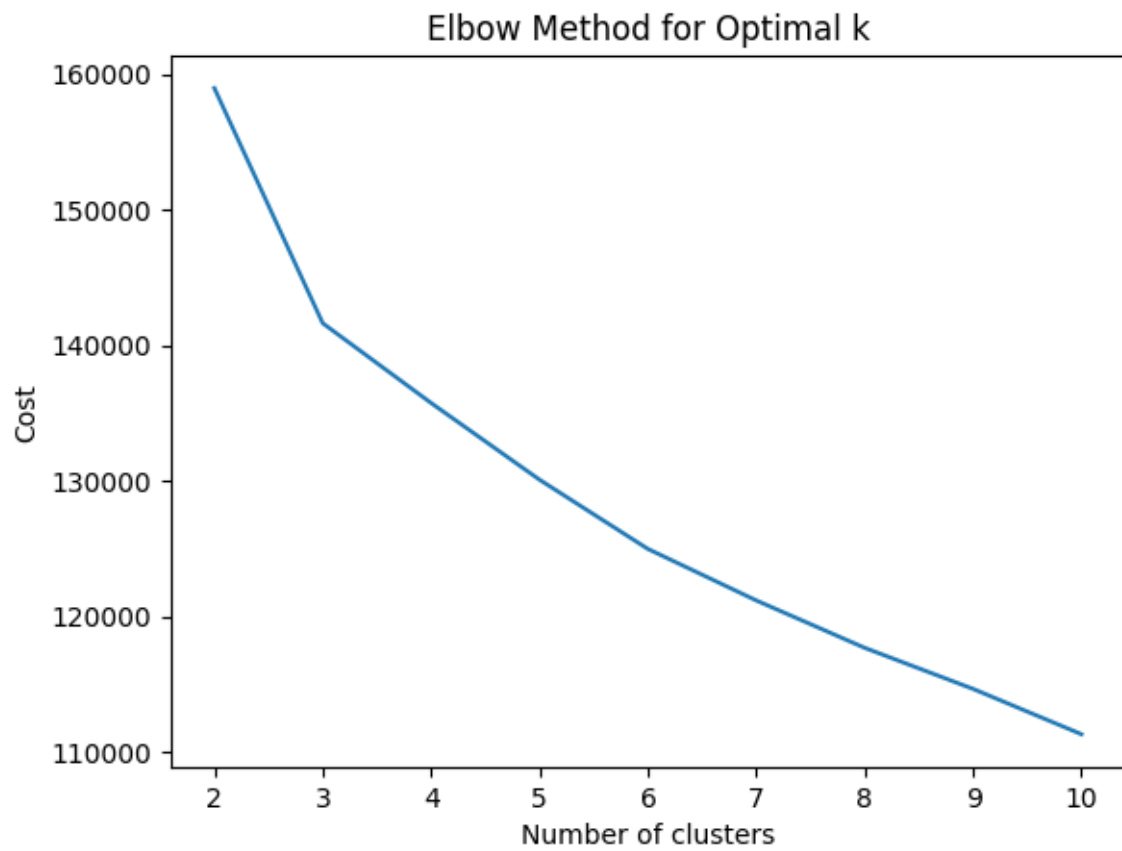
Contingency Table:

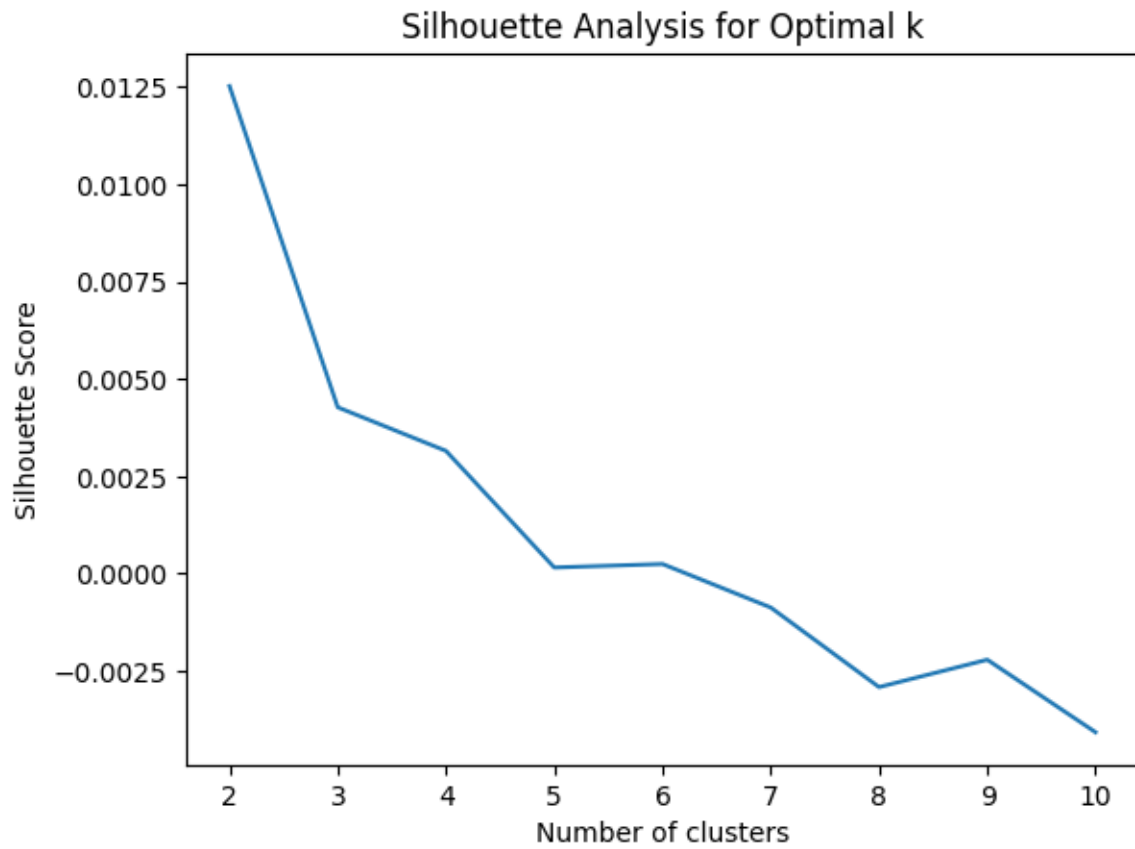
Cluster	0	1
Diagnosis		
0	1137	1131
1	66	58

Accuracy: 0.4995819397993311

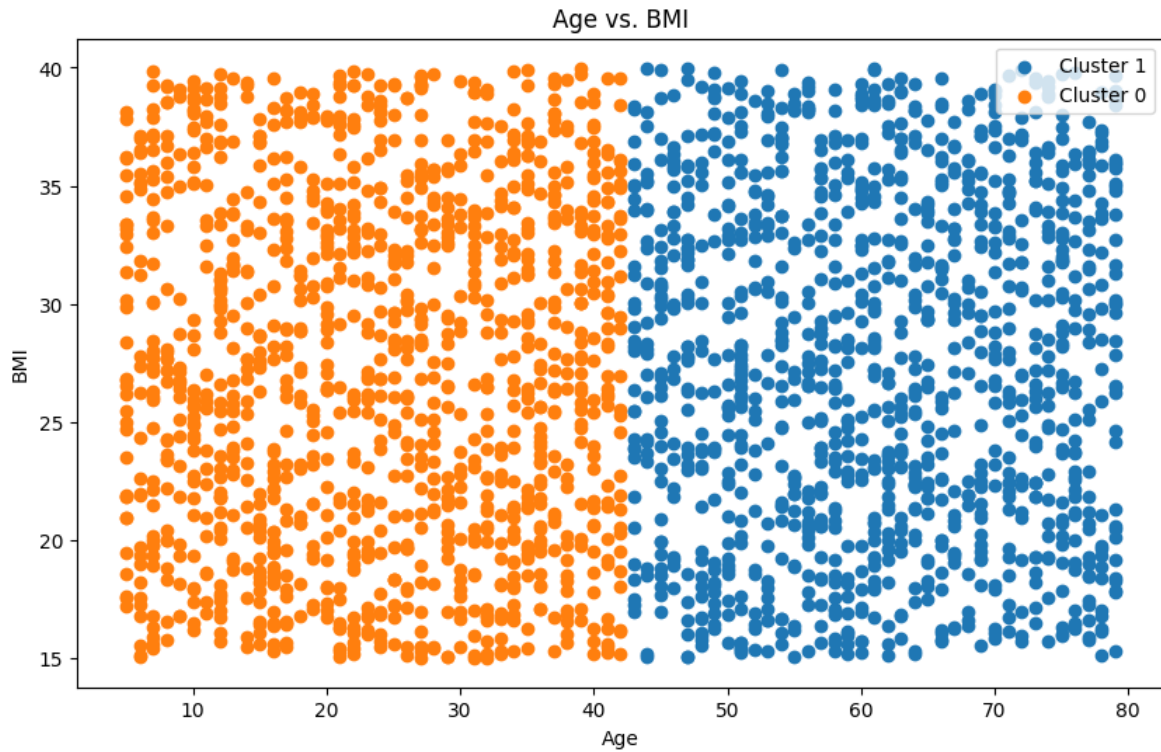
Recall: 0.46774193548387094

F1-score: 0.08834729626808835





Optimal k using silhouette: 2



Highly Correlated Variable Pairs:

BMI and DustExposure: Covariance = 1.3351224762671838

BMI and Age: Covariance = 2.5182047518516395

PollutionExposure and Age: Covariance = -2.683713558771351

DustExposure and Age: Covariance = 0.8554872985901283

	Age	PCA_Index
0	63	-0.119891
1	26	-1.572429
2	57	1.285219
3	40	1.226504
4	61	-0.200483
...
2387	43	-1.650841
2388	18	1.478054
2389	54	-1.458928
2390	46	-1.609384
2391	26	-0.198772
2392 rows × 2 columns		

