**Results**

After preprocessing our dataset, we trained each of our machine learning and neural network models using the dataset. We had split the dataset into training and testing parts and after developing the models using the training split we have carefully tested each model using the test split. We have accumulated the following parameters to evaluate the performance of the different algorithms against our dataset.

* Accuracy

Table I.

|  |  |
| --- | --- |
| **Name of Model** | **Accuracy** |
| KNN Classifier | 96% |
| SVM Classifier | 88% |
| Decision Tree Classifier | 87% |
| Random Forest Classifier | 91% |
| Naive Bayes Classifier | 89% |
| Logistic Regression Classifier | 88% |
| XGBoost Classifier | 89% |
| Gradient Boost Classifier | 87% |
| ANN Classifier | 98% |
| MLP Classifier | 96% |
| Residual Neural Network | 99% |

Fig. \*\*#\*\*. Accuracy of different models

We have plotted the accuracies of the different models for the training split in the table above. From the above table it is

evident that the Residual Neural Network and the ANN Classifier has the highest accuracy in classifying whether a

patient has Parkinson’s disease or not based on their attributes.

* Error

Error is the difference between the true result and the result predicted by the model. We have plotted the different error vs. iteration graphs for the different models.

From the above graphs it is evident that with increase in iteration the error steadily reduces to acceptable levels.

* Recall

Table II.

|  |  |  |
| --- | --- | --- |
| **Name of Model** | **Recall** | |
| **Parkinson’s Absent** | **Parkinson’s Present** |
| KNN Classifier | 0.86 | 1 |
| SVM Classifier | 0.57 | 1 |
| Decision Tree Classifier | 0.71 | 0.93 |
| Random Forest Classifier | 0.67 | 1 |
| Naive Bayes Classifier | 0.62 | 1 |
| Logistic Regression Classifier | 0.57 | 1 |
| XGBoost Classifier | 0.71 | 0.96 |
| Gradient Boost Classifier | 0.67 | 0.95 |
| ANN Classifier | 0.95 | 0.99 |
| MLP Classifier | 1 | 0.94 |
| Residual Neural Network | 0.99 | 0.99 |

Fig. \*\*#\*\*. Recall of different models

We have plotted the recall values of the different models for the training split in the table above. Recall measures the

proportion of actual positive cases that are correctly identified by the model as positive. It is a very important

metric for our use case as a high recall signifies that the model does not miss many positive cases as negative.

from the above table it is evident that Residual Neural Network and the ANN Classifier are the best performing

model in terms of recall.

* Precision

Table III.

|  |  |  |
| --- | --- | --- |
| **Name Of Model** | **Precision** | |
| **Parkinson’s Absent** | **Parkinson’s Present** |
| KNN Classifier | 1 | 0.95 |
| SVM Classifier | 1 | 0.86 |
| Decision Tree Classifier | 0.79 | 0.89 |
| Random Forest Classifier | 1 | 0.89 |
| Naive Bayes Classifier | 1 | 0.87 |
| Logistic Regression Classifier | 1 | 0.86 |
| XGBoost Classifier | 0.88 | 0.9 |
| Gradient Boost Classifier | 0.82 | 0.88 |
| ANN Classifier | 0.98 | 0.98 |
| MLP Classifier | 0.88 | 1 |
| Residual Neural Network | 0.97 | 1 |

Fig. \*\*#\*\*. Precision of different models

We have plotted the precision values of the different models for the training split in the table above. Precision is the

measure the fraction of correctly predicted positive instances out of all the instances the model has predicted as positive.

It is also a very important metric for our use case as a high precision value implies that the model does not misclassify

too many positive patients as negative and vice versa. From the above table it is also evident that the Residual Neural

Network is the optimal model in terms of precision.

* F1-Score

Table IV.

|  |  |  |
| --- | --- | --- |
| **Name of Model** | **F1 - Score** | |
| **Parkinson’s Absent** | **Parkinson’s Present** |
| KNN Classifier | 0.92 | 0.97 |
| SVM Classifier | 0.73 | 0.92 |
| Decision Tree Classifier | 0.75 | 0.91 |
| Random Forest Classifier | 0.8 | 0.94 |
| Naive Bayes Classifier | 0.76 | 0.93 |
| Logistic Regression Classifier | 0.73 | 0.92 |
| XGBoost Classifier | 0.79 | 0.93 |
| Gradient Boost Classifier | 0.74 | 0.91 |
| ANN Classifier | 0.96 | 0.99 |
| MLP Classifier | 0.94 | 0.97 |
| Residual Neural Network | 0.98 | 0.99 |

Fig. \*\*#\*\*. F1 - Score of different models

We have plotted the F1 scores of the different models for the training split in the table above. F1 - Score is calculated by combining both the Precision and recall values of the model and hence is a important metric to consider for our use case as a high F1 - Score implies that the model is making accurate predictions and also minimizes false negatives reducing the risk of miss-diagnosis. From the above table it is evident that Residual Neural Network is the optimal model as it has the highest F1 - Score for both the classes.

* AUC-ROC Curve

The AUC-ROC curve is made by plotting the model’s true positive rate against its false positive rate. A high AUC value in the ROC curve (Close to 1) implies that the model can precisely distinguish between Parkinson’s Positive and Parkinson’s Negative patients.

* AUC-PR Curve

The AUC-PR curve is made by plotting the model’s precision against recall of the model. In medical use cases it is a particularly important metric to consider as it can provide a more accurate evaluation of the performance model when the class distribution is imbalanced. The same is true for our dataset as well.