5.1 Conclusion

Cerebrovascular disease or Stroke or brain damage are very related with carotid artery disease. As a top three disease that cause disability and death, people should be pay a lot of attention to the symptoms of this disease. Because there are no symptoms of carotid artery, it is important to do physical examination and medical test to diagnostic the disease. Doctor can use Ultrasound and Magnetic Resonance Imaging (MRI) that already common as medical examinations that are used to detect carotid artery disease associated with cerebrovascular or stroke disease. But, such as the other surgical procedure, complications might always be occurring during or after the procedure done. Based on North Medical Center’s official website, the price of ultrasound and MRI test is very expensive. Therefore, this study will help doctors to analyze and predict the presence of carotid artery disease in patients. By the result of this research, we hope the doctor can predict the symptoms accuracy of this disease, so patient will know whether they have indication of Stroke or not.

Because the imbalance data is common in the medical case, we used under-sampling and over-sampling method. Under-sampling is used to reduce the majority data, and over-sampling to produce synthetic data so it will increase minority data. At the first model, SMOTE for over-sampling the minority data, and then this model is combined with Random Remove Sample to reduce the majority data. The number of dataset that produced with this method is 1,325. For the second model, K-Means Cluster for under-sampling majority data. This method only under-sampling the majority data, so the number of data that we get is similar with the minority, which is 325. Those two methods basically used for balancing the imbalance data. With balanced data in medical dataset, it will affect the performance of
classification, so high classification accuracy will be reached out in order to support an accurate medical decision.

Genetic Algorithm and Gain Ratio are used for selecting features. In order to improve the comprehensibility of the resulting classifier models significantly and improve the specificity in the classification task, features selection is needed. GA and GR are applied in the SMOTE_RRS and KMeansCluster balanced dataset, and the result that we get is Health Examine Age and Systolic Blood Pressure in each dataset is always selected. Also, Prothrombin Time and HBA1C are the most selected features among the others because they are chosen at least in three combination model dataset. But, Heart Disease and APTT has not been selected by any methods.

At the last section, we calculate the accuracy, sensitivity, and specificity based on method combinations. From the final result we can get that the highest correct rate, sensitivity, and specificity occur when we use BPN (NN trainscg) as the classifier. The average value is 83.59% with 90.85% specificity and 84.69% sensitivity. The high values indicate that our model performs well on the medical data that provided by Brain Health Checkup at North Medical Center Hospital. Thus, doctors can used this model to predict the presence of carotid artery disease in patients.

5.2 Future Research

In this study, the development of the model is to predict whether Carotid Artery Disease is occurring in patient or not. Related with that, as future research we can look for performing the same experiments on imbalance dataset with larger number of dataset in order to validate that our method is stable and consistent with the nowadays result.