# Application of machine learning algorithm on binary classification model for stroke treatment eligibility

by

**Joon Ho Han**

Submitted in partial fulfilment of the requirements for the degree of

Master of Applied Science

At

Dalhousie University

Halifax, Nova Scotia

April 2023

# Table of Contents

## List of Tables

## List of Figures

## List of Abbreviation Used

| | |
|---|---|
| EVT | Endovascular Treatment/Thrombectomy |
| QEII | Queen Elizabeth II Heath Centre |
| tPA | Tissue Plasminogen Activator |
| DASH | Designated Acute Stroke Hospitals |
| mRS | Modified Rankin Scale |
| SVM | Support Vector Machine |
| AUC | Area Under Curve |
| ROC | Receiver Operating Characteristic |
| TPR | True Positive Rate |
| FPR | False Positive Rate |
| ASPECTS | Alberta Stroke Program Early CT Score |

# Abstract

In Canada, stroke is the leading cause of adult disability and the third leading cause of death. Ischemic stroke is the most common type, making up approximately 85% of all stroke patients. Endovascular treatment (EVT) is effective for severe ischemic stroke patients. Unfortunately, EVT requires specialized equipment and personnel, which limits its availability.

There are several clinical and imaging factors that are critical in determining eligibility for EVT. Furthermore, in stroke, minutes matter as the brain dies quickly after onset, making EVT treatment's effectiveness highly time dependent. For this reason, timely across to EVT is critical. This study is to create a binary classification model to predict the EVT eligibility of stroke patients and discover attributes of the patient information that help to make efficient decision on transfer EVT eligible patient. Following algorithms applied to dataset: Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine.

## 1.0 Introduction

Stroke is a catastrophic disease that affects 13.7 million people worldwide each year, and 5.5 million people die as a result (1). Stroke can be classified into two types: hemorrhagic stroke and ischemic stroke. Hemorrhagic stroke occurs when the blood vessel burst within the brain and ischemic stroke occurs when a blood clot blocks the blood flow in artery in the brain (2). Age is the most well-known key factor that affects incidence of stroke, the average age of stroke patients was 69 years old in 2005 (3). Furthermore, the Public Health Agency of Canada (PHAC) presented the increase in incidence of stroke with age, which is shown on Figure 1 from PHAC from 2012/13. They found that 10% of Canadians aged 65 years or older will experience a stroke and 20% for age group 85 years or older will experience a stroke (4).

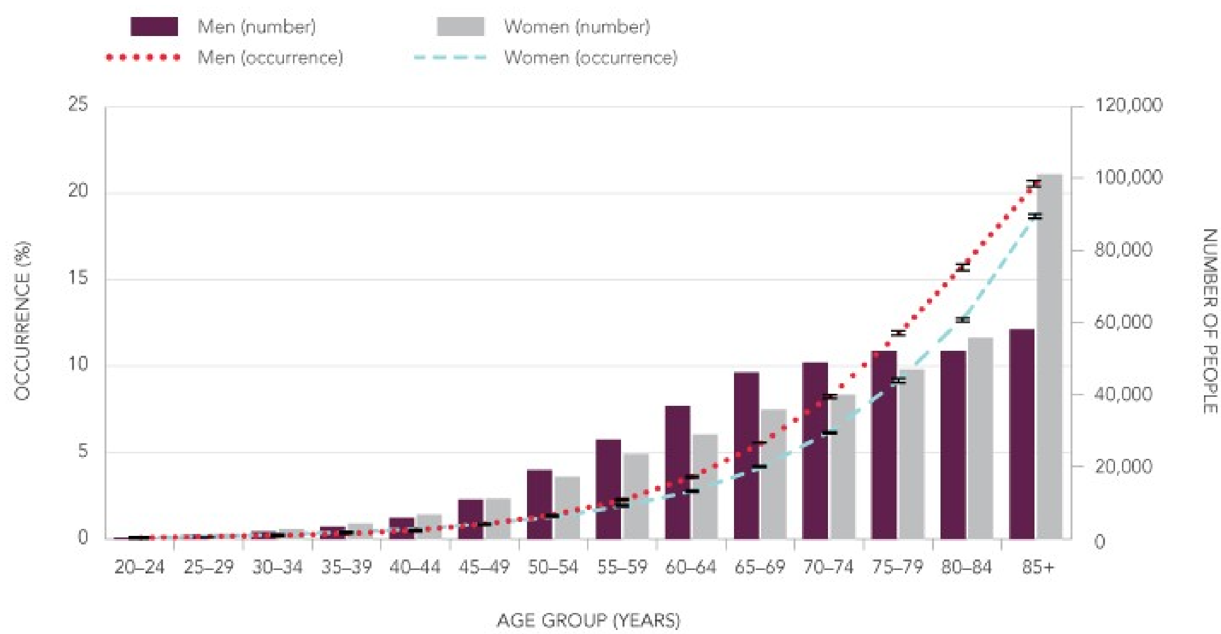


*Figure 1: Stroke occurrence (%) and number of people y five-year age group and sex in Canada 2012-2013* (4)

According to the annual statistic report of American Heart Association, 87% of all strokes are ischemic stroke. Fortunately, ischemic strokes are treatable, the medical treatment with thrombolysis using either tPA (tissue plasminogen activator) also known as alteplase. Alteplase and TNK (Tenecteplase). Thrombolysis treatment is given through an intravenous (IV) in the arm, and it dissolve the blood clot that blocks the blood flow in the brain. Thrombolysis treatment should be administered to stroke patients as soon as possible, and it can only be given with in 4.5 hours of symptom onset. The faster administration of a thrombolysis drug to restore blood flow in the brain helps to limit the risk of brain damage and functional impairment (5). National Institute of Neurological Disorders and Stroke conducted one of the key randomized clinical trials to prove the effectiveness of alteplase, the result showed that 39% of alteplase-treated patients will not have disability as compared to 26% of patients who are not treated. Thus approximately 25% of ischemic stroke patients can be effectively treated with alteplase (6). Among the cases of the ischemic stroke approximately 30-40% is due to a large vessel occlusion (LVO) which is the most severe type of ischemic stroke (7). A new highly effective treatment for ischemic stroke due to LVO is called endovascular treatment (EVT), which mechanically removes the clot using stent retriever the procedure called a thrombectomy. The blood clot may be removed by trapping it in a stent then pull it out with the clot. The Figure 2 shows how does the EVT removes the blood clot (8).

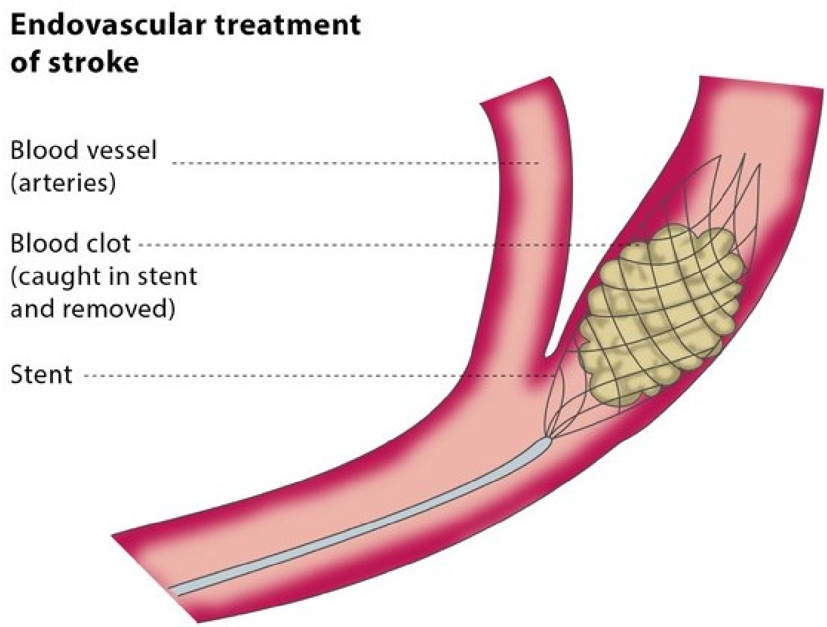


*Figure 2: Procedure of Stent Remove the Blood Clot* (8)

EVT is a highly effective treatment method when compares to alteplase, about 26.9% of EVT treated stroke patients will recover with no disability compared to 12.9% of patients who did not receive EVT and 46.0% of EVT treated patients will only have minor disability compared to 26.5% who did not receive EVT (9). It should be noted that both treatments are synergistic, and ischemic stroke patients can receive both treatments if they are eligible, as each treatment has different contraindications. Eligibility for EVT is determined using a CT Angiogram (CTA) which shows the location of the blood clot, and it allows for the evaluation of the collateral circulation of the brain. The clot needs to be accessible through the EVT procedure, and it needs to be in a large vessel in the brain. However, EVT requires specialized equipment and personnel to conduct the procedure. Thus, EVT has limited availability to ischemic stroke patients that live outside of the catchment of a hospital that provides EVT (10).

Not all the hospitals have the ability to treat ischemic stroke, hospitals with a CT (computed tomography) scanner and expertise to treat stroke patient are called as Designated Acute Stroke

Hospitals (DASH) in Nova Scotia. For example, in Nova Scotia in Canada, there are 11 DASH across the province, and among them only one hospital is capable of treating stroke patients with EVT, the QEII. Figure 3 from Nova Scotia Health Authority that shows the location of all DASH around Nova Scotia.


*Figure 3: DASH around the Nova Scotia*

The yellow boxed hospitals are only able to administer thrombolysis treatment to ischemic stroke patients red boxed hospital QEII is only one that is capable of EVT as well as thrombolysis treatment. Stroke patients outside the catchment of the QEII, will be transported to a DASH where yellow boxes, if patients are eligible for EVT, the DASH urgently transfers the patient to QEII. Similar to Nova Scotia, region across Canada and internationally must transfer ischemic stroke patients to EVT-capable hospitals. A study from Ontario showed that 34% of patient who were transferred for EVT ended up to receiving EVT, and the remaining 66% of patients were ineligible for treatment upon arrival (11). These cases are called futile transfer: a small but costly portion of overall transfer volume. Futile transfer incurs significant burden cost to healthcare system and could cause delay for other patients. Therefore, optimizing the decision on the ischemic stroke patients will hopefully decrease the futile transfer rate, and lessen on the cost and delay burdens.

The objective of this research project is to figure out the adaptability of the machine learning algorithms on transfer patient for EVT prediction through the creation of classification machine learning models with following algorithms logistic regression, support vector machine, decision tree, and random forest. Furthermore, discovering the valuable attributes that are correlated with the prediction for EVT eligibility. A historical dataset of ischemic stroke patient data who were transferred to get EVT is used to build the machine learning model. Therefore, this machine learning model could support decision making on EVT eligibility of ischemic stroke patients to minimize the futile transfer rate which will result in better transfer cost efficiency and less delay on patient transfers in the healthcare system.

## 2.0 Literature Review

In the following section, the literature related to the application of machine learning to the clinical area of stroke is described. The primary purpose of this section is to review studies where machine learning has been applied to ischemic stroke patients in order to determine the state-of-the-art in this area and to identify research gaps. The outline of this section is as follows—Section 2.1 discussed research methodology used to extract the related studies and papers; and the rest of these sections are categorized by each study's: diagnostic, prognostic, patient outcome prediction, and stroke prediction.

### 2.1 Search Methodology

Google Scholar, PubMed, and Novanet were used to identify the relevant papers and studies. The keyword included 'machine learning, 'ischemic Stroke,' 'logistic regression,' 'random forest', 'decision tree', 'support vector machine', and 'XG Boost.' The search with these keywords on three databases produced 517 records. Thirty-five duplicated records were removed during the initial screening process. During the screening session, the titles and abstracts of the documents were reviewed, then 407 articles were excluded as they were focused on an unrelated topic. Unfortunately, five papers could not be retrieved due to limited access, and 45 articles were excluded due to insufficient information, poor model performance, and a lack of relationship between the research objective. Therefore, the search ended up with 35 articles, which are discussed in the rest of this chapter. The search methodology is overviewed in a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. The detailed process and results are shown in the Figure 4.

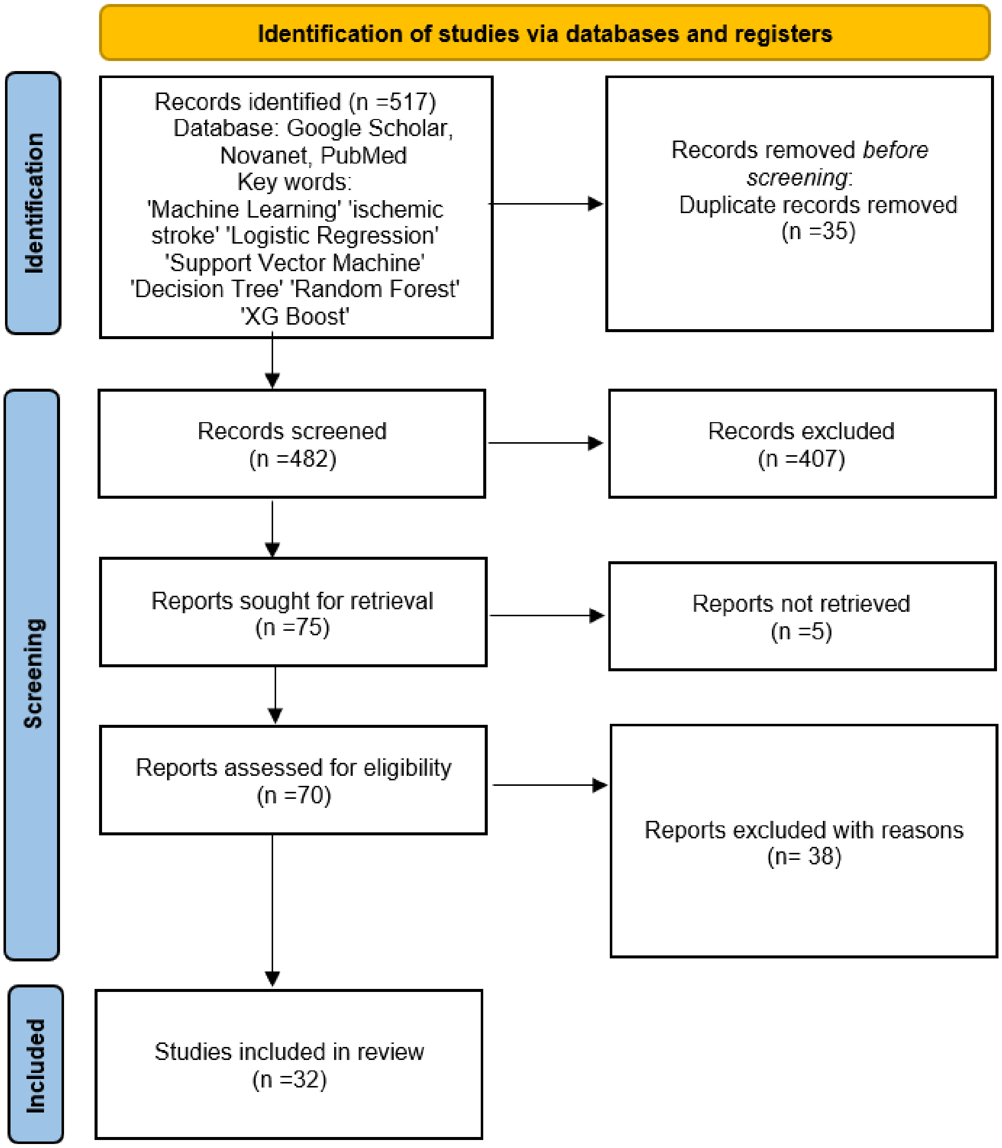


*Figure 4: PRISMA Flow Diagram* (12)

## 2.2 Focus

The final 32 articles that are included in this review were categorized into four categories: diagnostic model, prognostic model, patient outcome prediction model, and stroke prediction model. Diagnostic models' objective is to diagnose the current state of ischemic stroke using patient data, predictive models aim to use the given data of patients to predict the recurrence of ischemic stroke (prognostic), patient outcome prediction models predict the results of treatments and post-stroke patients' disabilities, and the stroke prediction models are simply making a binary decision for whether patients will have a stroke or not based on the primary patient dataset. There are nine articles in the diagnostic model, five in the prognostic model, 14 in the patient outcome prediction model, and four on stroke prediction model. Table 2.1 categorized the articles based on the focus of the articles.

| Topic | Articles | Authors for each paper |
|---|---|---|
| Diagnostic Model | 9 | Maier et al., Grosser et al., Subudhi et al., Lee et al., You et al., Allen et al., Bandi et al., Adam et al., Hayashi et al.(13–21) |
| Prognostic Model | 5 | Abedi et al., Lin et al., Ntaios et al., Kim et al. Mitra et al., (22–26) |
| Patient Outcome Prediction Model | 14 | Monteiro et al., Meier et al., Jiang et al., Li et al., Su et al., Chiu et al., Kappelhof et al., Ramos et al., OS et al., Brugnara et al., Hung et al. Lozano et al., Alqudah et al., Rui et al.(23,27–39) |
| Stroke Prediction Model | 4 | Emon et al., RVS technical Campus, Islam et al., Yu et al.(40–43) |

*Table 1: Summary and distribution of articles in final review*

## 2.3 Diagnostic Model

There are two main types of strokes: ischemic and hemorrhagic stroke. However, ischemic, and hemorrhagic stroke also classifies into different type of stroke depending on the condition. For example, ischemic stroke is a blockage of an artery, but depends on the inital location of the blood clot, they classify into two different group: thrombotic stroke (where blood clot developed in the blood vessels insider the brain) and embolic stroke (where blood clot develops elsewhere in the body and travels to one of blood vessels in the brain). Bandi et al. and Adam et al. created the stroke type classifiers that support a faster and more accurate diagnosis of stroke (19,20). Bandi et al. applied machine learning techniques to identify, classify, and predict the type of stroke from various clinical information on the stroke patient. They created two types of models; one is for attribute extraction from medical records based on NIHSS (National Institute of Health Stroke

Scale). Then the second model, the random forest model based on multiple algorithms, which classifies the type of stroke by using three hierarchical modules: 1. Ischemic Stroke 2. Intra Cerebral Hemorrhage 3. Subarachnoid Hemorrhage. The best performance model classifies stroke type with an accuracy of 96.67% (19). Adam et al. created the machine learning classification model that classifies the ischemic stroke type based on the patient information with CT and MRI (magnetic resonance imaging) results. Their model also classifies the stroke type into ischemic and hemorrhage based on the patient information. The model showed an excellent performance that classifies types of strokes with higher than 0.95(95%) accuracy and ROC area (20). From Japan, Hayashi et al. created a diagnostic model that identifies the large vessel occlusion in acute ischemic stroke patients. The model has based on a patient dataset with 51 attributes that categorize into past medical history, vital signs, and presence of symptoms. The LVO prediction model has a decent result as AUROC with 0.898 and an accuracy of 89.7% (21).

Ischemic stroke is usually diagnosed with assistance from CT imaging, using non-contract CT(NCCT) and MR (magnetic resonance) imaging. NCCT or MRI is used to distinguish between an ischemic stroke and hemorrhagic stroke, and to obtain information about the progression of the stroke. Additionally, CT Angiogram and CT perfusion provide additional information about potential treatment options for ischemic stroke. MRI using DWI (Diffusion Weighted Image), and FLAIR (fluid-attenuated inversion recovery) can also be used to determine treatment eligibility for ischemic stroke when MRI is being used. NCCT and MRI is used to diagnose ischemic stroke patients by identifying the lesion segmentation and the infect core size. CTA is used to detect the vessel that is occluded, CTP or DWI/FLAIR on MRI is used to determine the size of the penumbra (salvageable brain tissue) compared to the size of the core (dead brain tissue). Many studies that have machine learning applications support interpretation of imaging of ischemic stroke patients from CT, CTA, CTP, DWI, and FLAIR. Maier et al. compared the nine classification methods to determine the best fit model for stroke lesion segmentation. They used Thirty-seven multiparametric MRI datasets of ischemic stroke patients for evaluation. The random decision forest and convolutional neural networks classification approach outperformed all previously published results(13). Grosser et al. investigated the addition of spatial attributes on image-based parameters for the lesion outcome prediction model to determine whether the spatial attributes improve the outcome. The combination of spatial attributes and image-based parameter showed better performance on multi-parametric tissue outcome prediction, which outperformed the model with an image-based attributes only (14). Subudhi et al. created the automated segmentation method based on machine learning model to detect the ischemic stroke using the DWI sequence of MRI. They preprocessed the MRI dataset with morphological and statistical datasets through the expectation-maximization process, and then they trained the model to identify the cerebral infarction and to categorize them into three groups: total anterior circulation stroke syndrome, partial anterior circulation stroke, and lacunar stroke syndrome. The random forest model achieved 93.4% accuracy, which could be used in the decision-making process in treating the ischemic process (15). Another study from Asia made a similar approach, which used machine learning to interpret the brain imaging, Lee et al. created the binary classification model to identify stroke within 4.5 hours, which is the recommended time window for thrombolysis(16). You et al. created the hierarchy evaluation system of the LVO detection model for ischemic stroke patients, which used a deep learning model to transform unstructured data from the non-contrast CT image into a structured data format. (17). Allen et al. created a machine learning model to support the decisions on eligibility for thrombolysis for ischemic stroke patients. The binary classification model used the emergency stroke patients' admission data, and six binary classification models resulted in 81% to 86% accuracy (18).

## 2.4 Prognostic Model

The prognostic models for ischemic stroke patients focused on predicting the probability of a recurrent acute ischemic stroke and identifying the directly related attributes that affects risk of recurrence of an ischemic stroke. Lin et al. and Abedi et al. studied recurrent stroke prediction from a machine learning classification model (22,44). The model from Lin's project used patient data who underwent a documented carotid ultrasound within 30 days of experiencing an acute first stroke. The model predicts recurrent stroke inpatients based on 43 attributes of carotid sonography(44). Abedi's research team created six classification models based on the patient-level data that indicate the stroke patient's long-term recurrence. The models were categorized into 1,2,3,4, and 5 years of window time for stroke recurrence from the first stroke. Among the five window times, the 1-year recurrence model performed better than other time windows (22). Ntaios et al. developed a machine learning-derived prognostic model for predicting the cardiovascular risk of ischemic stroke patients. The 12 variables for the model were selected by the LASSO (Least Absolute Shrinkage and Selection Operator). The outcome of the model assessed was a major adverse cardiovascular event: non-fatal stroke, non-fatal myocardial infarction, and cardiovascular death during a 2-year follow-up (24). Mitra et al. created the automated diagnostic model to identify chronic ischemic infarcts that may be used to aid the development of post-stroke management strategies. Their approach is based on Bayesian-Markov Random field classification to segment probable lesion volumes present on FLAIR. Furthermore, a random forest classification of the information from multimodal (T1 and T2 weighted, FLAIR, and apparent diffusion coefficient) MRI images and context-aware attributes extract the areas with a high likelihood of being classified as lesions(26).

## 2.5 Patient Outcome Prediction Model

The patient outcome prediction model is the most active area of study in machine learning application in ischemic stroke. This section includes 14 papers relating to outcome prediction for ischemic stroke patients, which takes up the largest portion of the extracted literature. This section is divided into three sections: Patient functional outcome prediction, treatment outcome prediction, and other prediction models.

### 2.5.1 Patient Functional Outcome Prediction

The mRS(modified Rankin Scale) is a severity score representing the degree of disability in patients who had a stroke; the scale is in the range of 0 to 6, with a higher score indicating more severely handicapped status and 6 indicating death (27). When predicting the stroke patient's functional outcome, studies usually look for a 3-month and 6-month time window from the patient's admission. Meier et al., Monteiro et al., and Jiang et al. applied the classification machine learning algorithms to patient data who had an acute ischemic stroke about 3-month ago to predict the functional outcome of stroke patients based on mRS(27–29). Meier's model yields a binary decision from mRS, the mRS $< 2$ for a good outcome and mRS $\geq 5$ for a poor outcome(28). Monteiro's classification model's decision standard is as follows: mRS $\leq 2$ is a good outcome, and mRS $> 2$ is a poor outcome(27). Jiang et al. test out the combination of the data type with clinical and imaging attributes for the prediction model to figure out the combination that performs best. Their mRS score standard for classification as 0-2 mRS are favorable outcomes, and 3-6 mRS is unfavorable (27,29). Furthermore, these three studies compared the existing functional outcome prediction model for stroke patients, such as SPAN-100, ASTRAL score, THRIVE score, and DRAGON score. Through the comparison between existing model and machine learning model, all of the machine learning models outperformed on classification(27–29). Research from South Korea showed that Vitamin D is a well-known predictor of poor outcomes for cardiovascular disease. Thus, Kim et al. created the classification machine learning model to figure out the reciprocal relationship between 25-hydroxyvitamin D level and the prognosis of acute ischemic stroke. As the outcome value, the mRS (modified Rankin Scale) score was used to indicate a score between 3 to 6 as a poor outcome. The research outcome, a low 25-hydroxyvitamin D level was associated with poor outcomes in patients with acute ischemic stroke (25). Li et al. developed the machine learning model to predict the 6-month unfavorable outcome of acute ischemic stroke. The model predicts the mRS score of patients and then classifies them into two groups: mRS 3 to 6 as a poor outcome and otherwise good outcome at 6-month. Afterward, the machine learning model was compared with other scores that predict the patient outcome at 6-months, such as the HIAT score, THRIVE score, and the NADE nomogram. The machine learning model had significantly better results(30). Su et al. created the machine learning classification model that predicts the mRS at hospital discharge and then discretized the mRS into two groups a good outcome with mRS 0-2 and a poor outcome with mRS ≥3. Various attributes were used to build the model: demographic characteristics, medical comorbidities, NIHSS total scores, initial physiological parameters at admission, initial laboratory parameters of blood tests, and data of urine tests (31).

### 2.5.2 Treatment Outcome Prediction

The mRS was also widely used to predict the outcomes after reperfusion therapies for acute ischemic stroke; the reperfusion therapies are thrombolysis and endovascular thrombectomy (EVT). Chiu et al. developed the classification machine learning model to predict reperfusion therapy for acute stroke patients. First, they used the data of patients who were required for reperfusion therapy. The model classifies the outcome into three different groups based on mRS: favorable outcome (mRS 0-2), intermediate outcome (mRS 3-4), and miserable outcome (mRS 5-6). Then they compared the existing prediction model DRAGON score with the machine learning model (32). Kappelhof et al. and Ramos et al. created the classification machine learning model that predicts the poor outcome of endovascular treatment for acute ischemic stroke patients. Both studies' standards for poor outcome was the mRS ≥5 to make it to the binary classification problem. Kappelhof's model performance was significantly better than Ramos' model, with about a 30-40% difference in accuracy(33,34). Os et al. also developed the classification model for endovascular treatment outcomes. Still, they created two models: one that predicts functional independence and one that indicates the reperfusion status. The outcome variable for the reperfusion model was modified TICI-score ≥ 2b as a good outcome. And the model for functional independence in 3-month used the mRS ≤2 as a good outcome (35). Brugnara et al. developed the endovascular treatment outcome prediction model by predicting the mRS score 90 days after treatment. Their model classifies outcomes into two groups: favorable outcomes (mRS ≤2) and unfavorable outcomes(mRS>2), which is like other studies. However, they approached with an integrative assessment of clinical, multimodal imaging, and angiographic characteristics for model build-up. Their research gradually adds data attributes to the base model, which builds with baseline clinical and conventional imaging characteristics. Thus, the model shows higher AUC and accuracy each time they add more attributes (36).

### 2.5.3 Other Prediction Models

There were also a number of studies that developed other predictive models in the application area of stroke. Hung et al. and Lozano et al. created the predictive machine learning model for mortality and morbidity of acute ischemic stroke patients 3-months after admission (23,37). Hung's group made two models predicting M&M (Mobility and Mortality) and readmission rates. In addition, they implemented various resampling methods to balance the class distribution. This study also identifies the essential predictive factors for M&M and readmission rates for acute ischemic stroke patients (23). Lozano created the random forest model for the M&M prediction according to three groups of the stroke patient: 1) ischemic stroke, 2) ischemic stroke + intracerebral hemorrhage 3) intracerebral hemorrhage. Group 2 showed the most stable mortality prediction, and group 1 as well. Furthermore, they tested the top 7 critical variables to predict mortality and morbidity, and NIHSS scores at 48 hours and 24 hours showed the best performance (37). BI (Barthel Index scale) is an effective method for measuring the performance of active daily living. Alqudah et al. created the classification machine learning model that predicts BI of post-stroke patient through machine learning algorithms. They applied the chi-squared test to reduce the number of parameters while developing the model, ending with a better prediction performance on each model (38). In China, the research group from Hebei university made a creative application of machine learning for a patient outcome prediction model. Rui et al. created the XG boost model to predict the length of hospital stay for acute ischemic stroke patients. They had plenty of data, around 18,000 patients, to build the model. They applied the 10-fold cross-validation method to train the model, and the model showed excellent performance of 96% accuracy, 0.82 recall rate, and 0.79 F1 scores (39).

## 2.6 Stroke Prediction Model

As mentioned in the previous section, early awareness of the sign of strokes will significantly reduce the mortality of stroke patients. Thus, the prediction of stroke would be helpful to react faster when a stroke occurs or to be prepared so they can save their lives and reduce the risk of functional disabilities. Machine learning algorithms are actively applied to create the stroke prediction model using patient data. Emon et al. created the supervised machine learning model that predicts stroke based on various patient data variables. The patient dataset was collected from the medical clinic of Bangladesh, which has 12 different attributes related to patient history by whether patients had a stroke or not. They applied ten machine learning algorithms to find the best model for stroke prediction. The weighted voting algorithm made the best performance on stroke prediction with 97% of accuracy (40). Stanford University's research team also created the stroke classification model with an automatic attribute selection algorithm that selects robust attribute based on the proposed heuristic method: conservative mean. Through their algorithm, they tried out various scenarios with different numbers of attribute for the machine learning model in order to figure out the best attribute selection scenario. Then they created two different types of a prediction model with SVM (Support Vector Machines) and MCR (Margin-based Censored Regression) algorithms. The combination of attribute selection and MCR showed the best performance with accuracy and the highest AUC (area-under-curve) (41). Islam et al. also created the integrated stroke prediction model. They applied the SMOTE (Synthetic Minority Oversampling Technique) to handle the imbalanced data to overcome the over-fitting issue. Then they compared the model performance before the SMOTE application and after. Therefore, the accuracies were increased by 1.5~3% depending on the model type, and there was an improvement in other values such as precisions and recall rates (42). The research team from South Korea created a supervised machine learning classification model to predict the stroke severity of patients over 65 years use the baseline characteristics of patients and some attribute from NIHSS. Most stroke prediction models were binary classification models, but they created a model that predicts the five different classes and shows the severity of the stroke(43).

## 2.7 Discussion/Research Gap

This literature review provided a brief overview of the current state of machine learning applied to the clinical area of ischemic stroke. There were two purposes of this literature review. First, identify the research trends, such as important factors or attributes of the dataset related to ischemic stroke patients, and find out what algorithms are applied while developing the machine learning models. The following are the key points of the current state of the machine learning application for ischemic stroke patients:

1. Most machine learning models outperform the existing models or scores related to ischemic strokes while both look for the same objectives.
2. Data preprocessing or data cleaning methodologies directly affect the model qualities.
3. Reducing the number of attributes for the outcome value takes less time for model development and increases accuracy. The higher number of attributes causes the model to be more complex and reduce performance.
4. The machine learning model is mostly focused on patient outcome prediction, and the mRS is the most well-known measure of outcome for stroke patients.

Secondly, the literature review allows us to determine the research gaps in the literature. Studies about outcome prediction for ischemic stroke patients was on research trend, then diagnostic, prognostic, and stroke prediction models follows. However, there was a lack of studies that applied machine learning to decision making through the treatment process for acute ischemic stroke; specifically, transfer eligibility of ischemic stroke patients to access EVT. Among the literature in this review, only one paper was related to the model that supports the decision on thrombolysis eligibility in ischemic stroke patients(18). There are two treatments for ischemic stroke patients: thrombolysis (with Alteplase and Tenecteplase) and EVT. Depending on the circumstances, these treatments could be limited to the patients based on various clinical factors; thus, a better decision-making process could be made with the support of the classification models to prevent unnecessary expense and time.

## 3.0 Method

This section describes the given dataset's brief description and model development procedure for this project. The main processes included data preprocessing, description of the machine learning algorithms, and evaluation method. Figure 5 represents the procedure of machine learning model development.

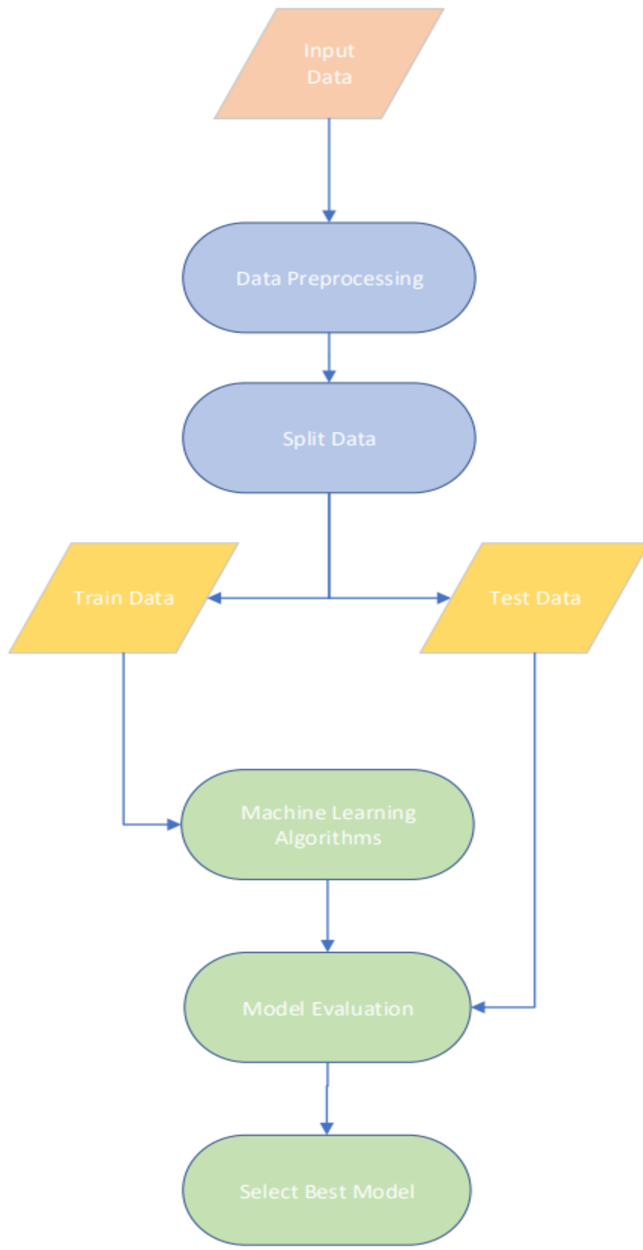


*Figure 5: Process of EVT Eligibility Prediction Model Development*

## 3.1 Data Description and Ethics

Data from patients in Nova Scotia that were transferred for EVT is acquired from NS Health. The data from each patient that is obtained from them to build the model is described below. This study was conducted under an existing approved ethic application (File Number: 1028274). Additionally, the data access request was also previously approved by NS Health.

| Attribute | Description |
|---|---|
| MRN | Unique number represents the patient |
| Sex | Sex of patient |
| Age | Age of patient |
| PMHX | Other medical history |
| Pre-Stroke (mRS) | Pre-stroke disability |
| Multiphase CT | Whether patient took a multiphase CT |
| Date/time of first CT | Time of first CT at thrombolysis centre |
| Date/time of first CTA | Time of first CTA at thrombolysis centre |
| ASPECTS (Alberta Stroke Program Early CT Score) | Brain condition measured from first imaging |
| Covers Arch | Whether covers arch |
| Covers Whole Brain | Whether covers whole brain |
| Venous Contamination | Whether are contaminated |
| Occlusion Location | The large vessel location of the clot |
| Collateral Status | The extent of collaterals from first imaging |
| Referring Center | Origin medical center of patient |
| Patient Transferred | Whether patient transferred to EVT center |
| Re-Image | Whether patient got another CT after transferring |
| Collateral Status 2 | The extent of collateral from second imaging |
| Repeat ASPECTS | Core size measured after second imaging |
| Thrombectomy Performed | Whether thrombectomy (EVT) performed on patient or not |
| Date | Year of patient information recorded |
| Miscellaneous | Any special note from the doctor |

*Table 2 : Description of Data*

## 3.2 Data Preprocessing

Data preprocessing is the essential preparation for the machine learning model. The primary purpose of data preprocessing is to check the dataset's quality. Six standards in data preprocessing:

1. Accuracy: To check whether the data entered is correct or not.
2. Completeness: To check whether the data is available or not recorded.

3. Consistency: To check whether the same data is kept in all the places that do or do not match.
4. Timeliness: The data should be updated correctly.
5. Believability: The data should be trustable.
6. Interpretability: The understandability of the data.

### 3.2.1 Data Cleaning

Data cleaning is the first step of data preprocessing, which removes incorrect, incomplete, and inaccurate data from the dataset. As the first step of data cleaning is dealing with the missing values. The attributes with more than 20% missing values are deleted. Data sample that has the missing values on attributes with categorical values, such as collateral status and occlusion location are deleted since they can't be filled up with other values. Attributes with less than 20% of missing value such as ASPECTS were filled up with the average value.

### 3.2.2 Attribute Selection

Selecting the dataset's attribute affects the quality of the machine-learning model. First, the attribute "Patient transferred" was dropped since the machine learning model predicts the patient should transfer for EVT to the EVT center. Thus, the attribute "Patient transferred" and any samples with the label of "No" In "Patient transferred" attributes were dropped from the dataset. In the given dataset, some of the dropped attributes are unrelated to predicting whether the patient got the EVT. Also, the model predicts the eligibility of the EVT from the first imaging, any attributes related to the second imaging also dropped because this was done when they were already transferred after the decision to transfer was made. Therefore, following attributes were dropped: MRN, PMHx, Multiphase CT, CTA, Patient Transferred, Arch Coverage, Brain

Coverage, Venous Contamination, Re-imaged, collateral status2, Repeat ASPECT, Onset to Recanalization Time and mRS.

### 3.2.3 Transformation of values/data split

All data should be numerical to create a model. There are two types of categorical values which are binary or multiple categorical values:

1. Binary: sex, patient transferred, thrombectomy received.
2. Multiple: collateral status, occlusion location, referring center.

The values in binary forms substitute with 0 and 1. However, the categorical value with more than two types of value is transformed in the following tables.

| Collateral Status | |
|---|---|
| **Categorical Value** | **Numerical Value** |
| Good | 0 |
| Intermediate | 1 |
| No | 2 |
| Poor | 3 |

*Table 3: Transformation of Categorical Value for Collateral Status*

| Occlusion Locations | |
|---|---|
| **Categorical Value** | **Numerical Value** |
| CAROTID T | 0 |
| ICA | 1 |
| M1 | 2 |
| M2 | 3 |
| No Occlusion | 4 |
| Tandem | 5 |

*Table 4: Transformation of Categorical Value for Occlusion Location*

The referring center attribute is dropped, and two new attributes were added: driving distance and Euclidean distance to the EVT center from the referring medical center. This was calculated using Google Maps for each referring centre to the QEII. There were 14 referring center locations; the transformation proceeded as shown Table 5.

| Referring Centre | Driving Distance (Km) | Euclidean Distance (Km) |
|---|---|---|
| DGH | 8 | 3.28 |
| Cobequid | 17.8 | 13.71 |
| CEHHC Colchester | 92.7 | 81.5 |
| SSRH | 101 | 79.1 |
| VRH | 105 | 87.4 |
| Aberdeen | 157 | 127.57 |
| Cumberland Regional | 195 | 137 |
| St. Martha's Reginal | 217 | 166.9 |
| Prince County Hospital | 297 | 197.02 |
| YRH | 302 | 221.7 |
| QE PEI | 329 | 182.88 |
| CBRH | 400 | 312.58 |
| Newfound Land Regional Health Centre | 1054 | 768.85 |
| Eastern Health St. John | 1480 | 896.72 |

*Table 5: Distance Information from the Referring Hospital to EVT Centre*

After transforming the categorical values, the dataset has nine attributes. The dataset is split into training and test sets with a ratio of 3:7 and 4:6. This result in two sets of a machine learning model for each algorithm.

## 3.3 Additional Study

The additional studies proceeded with additional attributes in the additional dataset provided from NS stroke registry to determine new attributes affect the quality of the model. Through merge the original with new dataset to test out the algorithms as well.

### 3.3.1 Data Merge

Three attributes were chosen from the new dataset to add to the original dataset. First, the healthcare number of the patient linked the two datasets. The new attributes are in following:

1. Onset to First CT Time: The time patient took to get a first CT onset time in minutes (Numeric)

2. Onset to First Hospital Arriver: Patient transfer time to get to the first hospital in minutes (Numeric)
3. tPA: Whether patient got tPA or not (Binary Numeric)

The health card number were used for the data linkage of two dataset through match the health card number of patient. The comparison between the model from the original dataset and merged dataset's result will be discussed in chapter 4.

## 3.4 Machine learning Algorithms

The objective model of this project is a supervised binary classification model with the response variable of Received EVT as a binary variable. Therefore, the binary classification algorithms were applied to develop the prediction patient receive EVT. Four different supervised learning algorithms which are specialize for binary classification were applied to the preprocessed dataset: logistic regression, decision tree, random forest, and support vector machine (SVM). According to those four algorithms were suitable for simple binary classification, however, the XG boost model is more suitable for complex classification studies which is not suitable for this study.

### 3.4.1 Logistic Regression

Logistic regression is a well-known statistical technique that analyzes the relationship between multiple independent and dependent variables and estimates the probability of occurrence of result variables by fitting linear regression equation to sigmoid function to create the logistic regression model (45). The response variable for this study is in the binary value. Thus, the binary logistic regression algorithm was applied to solve the problem. Equation 1 represents the linear regression equation, and equation 2 represents the sigmoid function used for the binary logistic regression model.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_n X_n \qquad (1)$$

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots\beta_n X_n)}} \qquad (2)$$

Where the parameter $\beta$ value between 0 and 1, which is the probability of the positive class label given $X$. $X$ is the weight of each attribute of the dataset, and n represents the number of the attribute in the dataset and $p$ represents the probability of the binary outcome.

### 3.4.2 Decision Tree

The decision tree algorithm is a decision support tool that shows the various possibilities and results in a flowchart-like tree structure. The decision tree consists of the internal node, the attribute or attributes of the dataset, the branch represents a decision rule, and the leaf node represents the outcome (46) The Figure 6 depicts the structure of the decision tree algorithm.

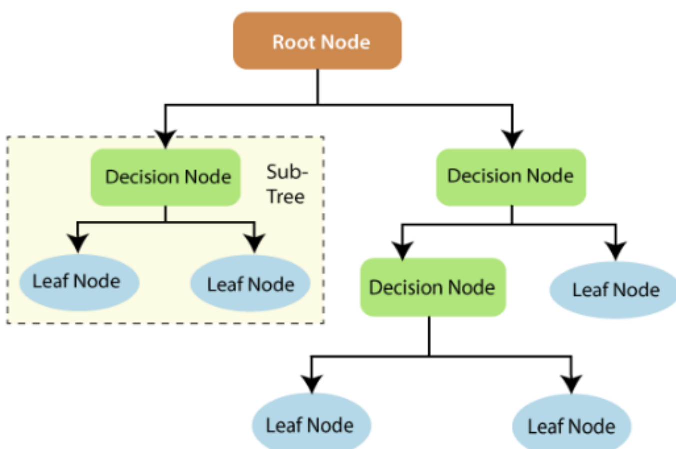


*Figure 6: Decision Tree Structure* (47)

The decision tree is based on the attribute selection measure (ASM), which provides the rank of each dataset attribute. The best score attribute will be selected as a splitting attribute. In the classification tree, the information gain, since the split point, results in the largest information gain for a given criterion. The following equation 3, represents the information gain calculation.

$$IG(D_p, f) = I(D_p) - \sum_{j=1}^{m} \frac{N_j}{N} I(D_j) \quad (3)$$

The parameters are labeled following:

$f$: Feature split on

$D_p$: Dataset of the parent node

$D_j$: datset of the jth child node where

$I$: Impurity criterion

$N$: Total number of samples

$N_j$: Number of samples at jth child node

There are two types of impurity criterion (I): entropy and Gini. Entropy measures the randomness or impurity in a dataset. The Gini index is used to measure the purity of a specific class after splitting along a particular attribute. Both entropy and Gini index are calculations of the dataset's purity, and as both values are close to 0, the information gain increases according to the equation (3) (48).

### 3.4.3 Random Forest

The random forest algorithm is an extended version of the decision tree. The general procedure is ensembled the multiple decision trees and makes a prediction collected from the votes of the individual trees (49). The following process chart depicts the process of the random forest algorithm.

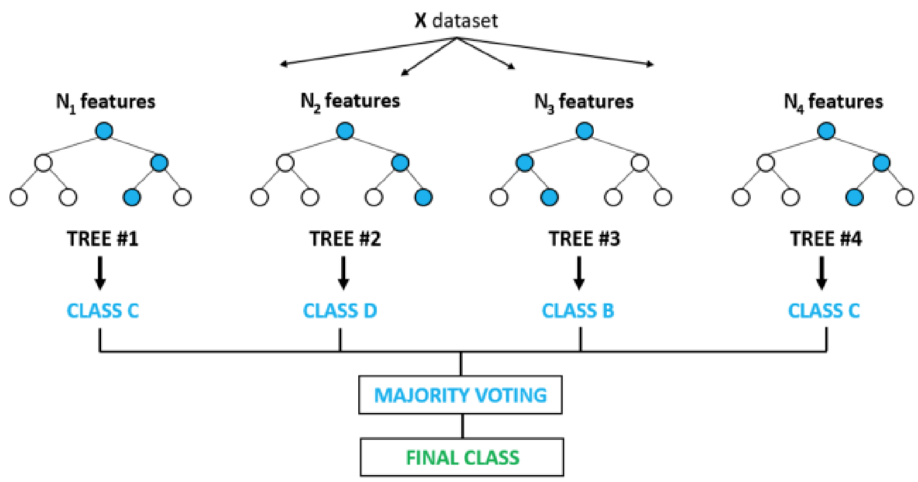


*Figure 7: Example of the Random Forest* (50)

### 3.4.4 SVM

The SVM is a commonly used kernel-based machine learning approach to classification, which creates multiple hyperplanes and then selects the one with the maximum margin. The basic process of the SVM has generated a hyperplane to separate the data into different classes, and a hyperplane is chosen to maximize the margin between classes. The margin represents the distance between the hyperplane and the nearest data point (51).

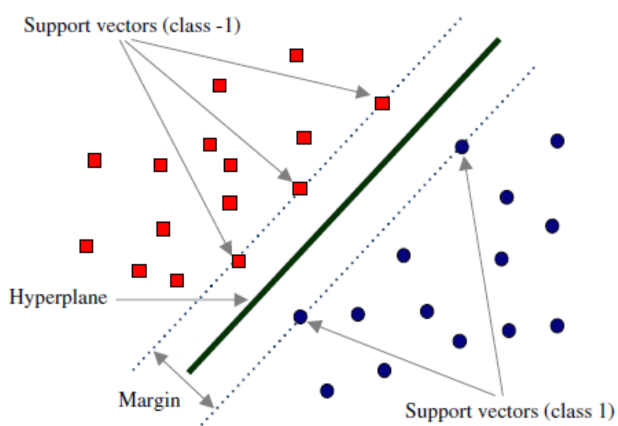


*Figure 8: SVM Representation of SVM Hyperplane* (52)

Equation 4 represents the binary classification SVM, and 5 and 6 represent the optimization of weight factor w.

$$f(x) = sgn(w^T x + b) \qquad (4)$$

$$Minimize \ \frac{||w||^2}{2} \qquad (5)$$

$$Subject\ to\ y_i(w^T x_i + b) \geq 1 \qquad \forall i = 1, \dots, n \qquad (6)$$

Where x is the input vector, w is the weight vector, b is the bias term, and f(x) represents the decision function which returns +1 if the argument is positive or zero and -1 if the argument is negative (53).

## 3.5 Model Evaluation

Three tools were used for the model evaluation to determine the best performance model. The main methods for evaluating binary classification models are the confusion matrix, classification report model, and AUC (Area Under Curve).

### 3.5.1 Confusion Matrix

The confusion matrix is a well-known evaluation method for the performance of classification models in machine learning and statistics. It summarizes the models' performance by comparing the result variables' labels with the true class labels of the dataset. For example, the binary classification model displays four different numbers, which are TP (True Positive), FP (False Positives), TN (True Negative), and FN (False Negative); all of those options are shown in the matrix on Figure 9. These four numbers are the correct and incorrect observation numbers on each label. The following figure is the template of the confusion matrix of binary classification (54).

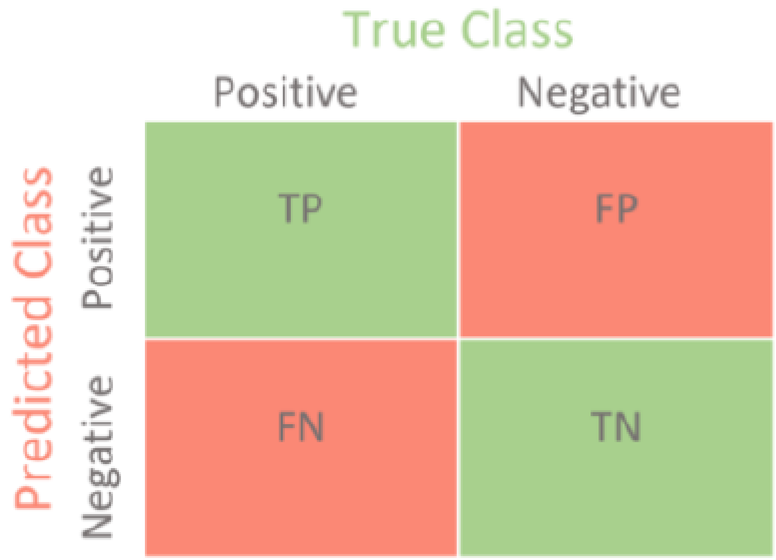


*Figure 9: Template of Confusion Matrix*

### 3.5.2 Classification Report

Use the four values provided by the confusion matrix to calculate the metrics of the evaluation:

1. Precision: The accuracy of positive predictions, which is the proportion of TP among all instances classified as positive.
2. Recall: The completeness or sensitivity of the classification, which is the proportion of TP among all actual positive instance.
3. F1-Score: Balance between precision and recall, which is weighted average of precision and recall.
4. Support: Total number of observations in each class.

These values key indicators to calculate the evaluation metrics for model performance and comparison between algorithms. The Figure 10 is the template of the classification model from Python which calculates the performance metrics.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.94 | 1.00 | 0.97 | 1444 |
| 1 | 0.00 | 0.00 | 0.00 | 89 |
| accuracy | | | 0.94 | 1533 |
| macro avg | 0.47 | 0.50 | 0.49 | 1533 |
| weighted avg | 0.89 | 0.94 | 0.91 | 1533 |

*Figure 10: Example of the binary classification report from the Python*

The classification report is the main method of measuring performance of the classification models.

### 3.5.3 AUC (Area Under Curve)

The AUC's main purpose is to compare different algorithms and optimal thresholds for making predictions to decide the best model among the various machine learning models. In addition, AUC is commonly used to evaluate binary classification performance. Area calculation under the ROC (Receiver Operating Characteristic) curve, which is a plot of the TP rate against the FP rate at various threshold values. The TP and FP rates calculation is shown in equations 7 and 8 values based on the confusion matrix.

$$TPR = \frac{TP}{TP+FN} \quad (7)$$

$$FPR = \frac{FP}{FP+TN} \quad (8)$$

The value of AUC represents the probability that a randomly chosen positive example is ranked higher than a randomly chosen negative example by the classification model. The indicator of AUC is between 0 to 1. An indication of AUC close to 1 means perfect classification performance, while close to 0 means missing out on classification (55).

### 3.6 Model Development and Evaluation

All models were developed using Python 3 on Visual Studio Code. The functions that were used the "sklearn" packages to apply the four machine learning algorithms and get the evaluation metrics variables. The coding for the models is attached in the Appendix A.

# 4.0 Results

Data was obtained for patients that were transferred from a referring stroke centre in Nova Scotia to QEII. The total number of patients records obtained was 101 and end up with 78 patient data after data preprocessing. Also, as mentioned in Section 3.3.1, new dataset was provided from NS health. The number of patients that match both datasets were 42 (merged data). The number of patients records used in model development is described in the following diagram shown in figure 11.

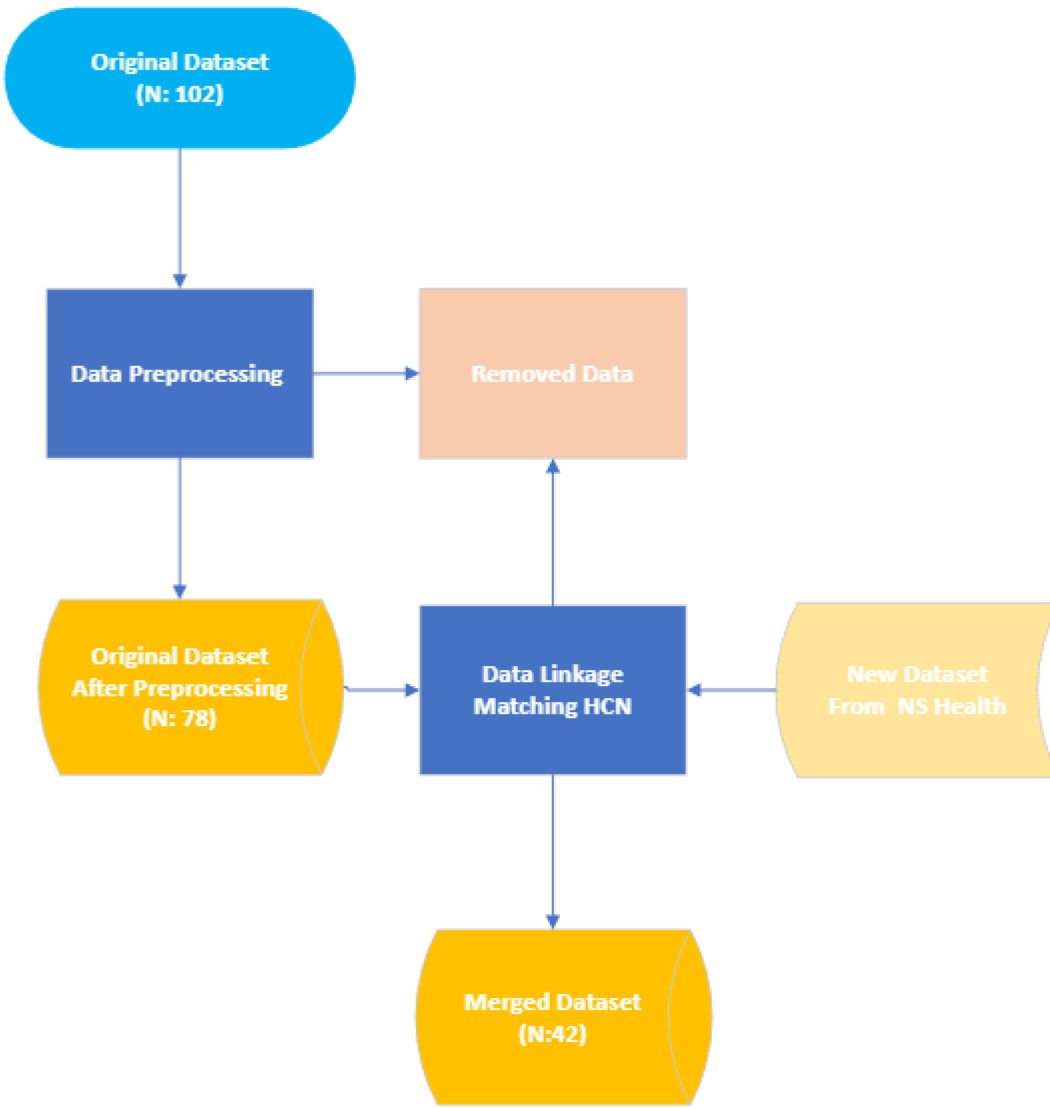


*Figure 11: Number of data change from data preprocessing and data linkage*

The result of this project is described in this chapter. Section 4.1 is described the result of data cleaning. Section 4.2. is about the data exploratory analysis of the dataset after data cleaning. Section 4.2 is about the data exploratory analysis. Section 4.3 illustrates the correlation between attributes in the dataset. Sections 4.4 to 4.7 result from the machine learning model from four different algorithms with two sets of data split 0.7/0.3 and 0.6/0.4 ratio then the section 4.8 is the model comparison. The section 4.9 is described about the false negative rate of best model.

## 4.1 Data Cleaning

The first problem detected from the given dataset was the inconsistency of the data record method and record error. For example, attributes "Occlusion location" and "Collateral status had a few spelling errors and inconsistencies in the records, such as unnecessary comments and different units of measure.

| | Number of Missing Values | Percentage of missing value (%) |
|---|---|---|
| MRN | 0 | 0 |
| Age | 0 | 0 |
| PMHx | 105 | 97.2 |
| CT time | 108 | 100 |
| Sex | 0 | 0 |
| Multiphase CT | 0 | 0 |
| CTA | 5 | 4.95 |
| ASPECTS | 1 | 1 |
| Arch Coverage | 5 | 4.95 |
| Brain Coverage | 6 | 5.94 |
| Venous Contamination | 5 | 4.95 |
| Occlusion Location | 3 | 2.97 |
| Collateral status | 10 | 9.9 |
| Patient Transferred | 0 | 0 |
| Re-imaged | 10 | 9.9 |
| Time between 1$^{st}$ and 2$^{nd}$ image | 21 | 20.8 |
| Collateral status2 | 42 | 41.6 |
| Repeat ASPECT | 33 | 32.7 |
| Thrombectomy Performed | 19 | 18.8 |

| | | |
|---|---|---|
| Onset to Recanalization Time | 56 | 55.4 |
| Date | 0 | 0 |
| mRS | 106 | 98.1 |
| Referring Center | 0 | 0 |

*Table 6: Summary of Missing Values*

## 4.2 Data Exploratory Analysis

After data preprocessing, the dataset ends up with nine attributes: age, sex, ASPECTS, occlusion location, collateral status, patient transferred, thrombectomy performed, date, and referring centre. The exploratory data analysis is proceeded to discover any characteristics of dataset.

### 4.2.1 ASPECTS

The ASPECTS score is an important standard in the decision to transfer patients for EVT and to determine EVT eligibility.

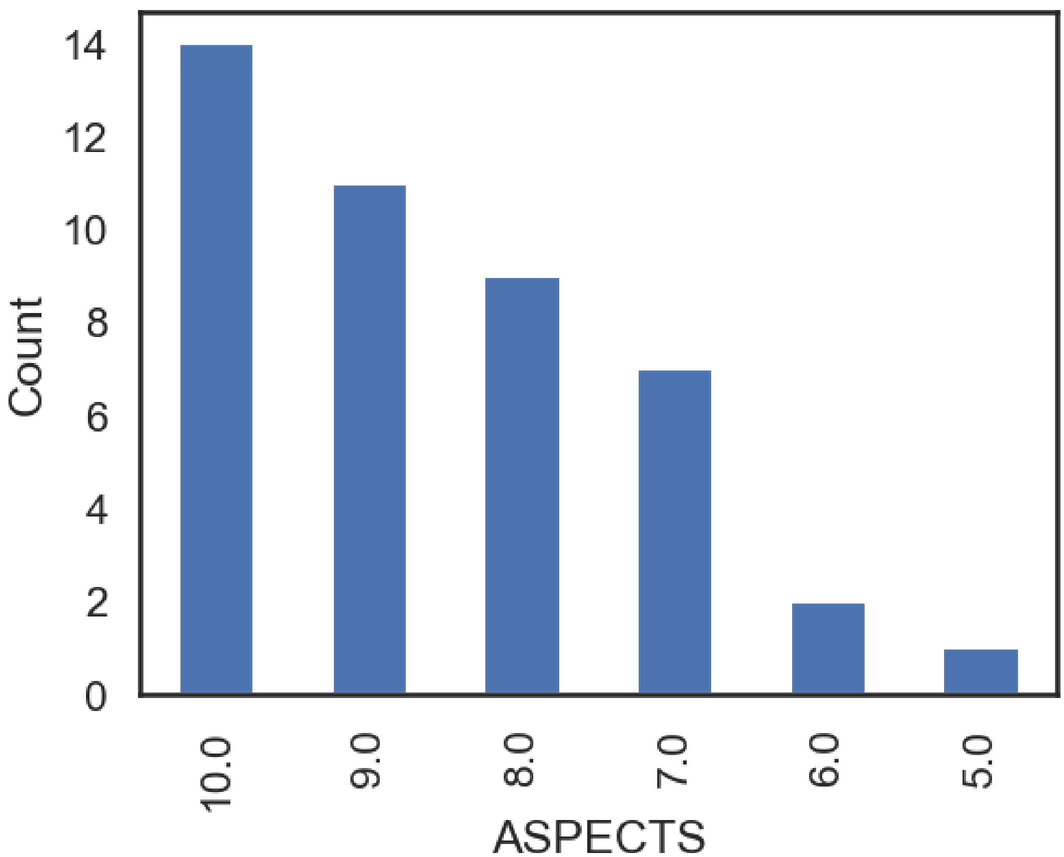


*Figure 12: Patient got EVT depends on ASPECTS.*

Figure 12 represents the number of patients who got an EVT among the transferred patients with ASPECTS. The scores under five did not have the opportunity to get the EVT, and the average value of ASPECTS was 8.56.

### 4.2.2 Occlusion Location/Collateral Status

The occlusion location and collateral status were correlated with the classification model's response variable. The location of the clot is the blood vessel in the brain where the clot persists. The location may have an effect on the changes to the ischemic core while the patient is transferred, which is why it was important to include this variable. Collateral status, for example, 60% of patients with a blood clot in the M1 region got EVT. The collateral status represents the ability of blood to flow through a collateral vessel in the brain region. The 14% of patient with less than intermediate status got EVT. The Figure 12 is the represents the patients' occlusion location and collateral status who got EVT.

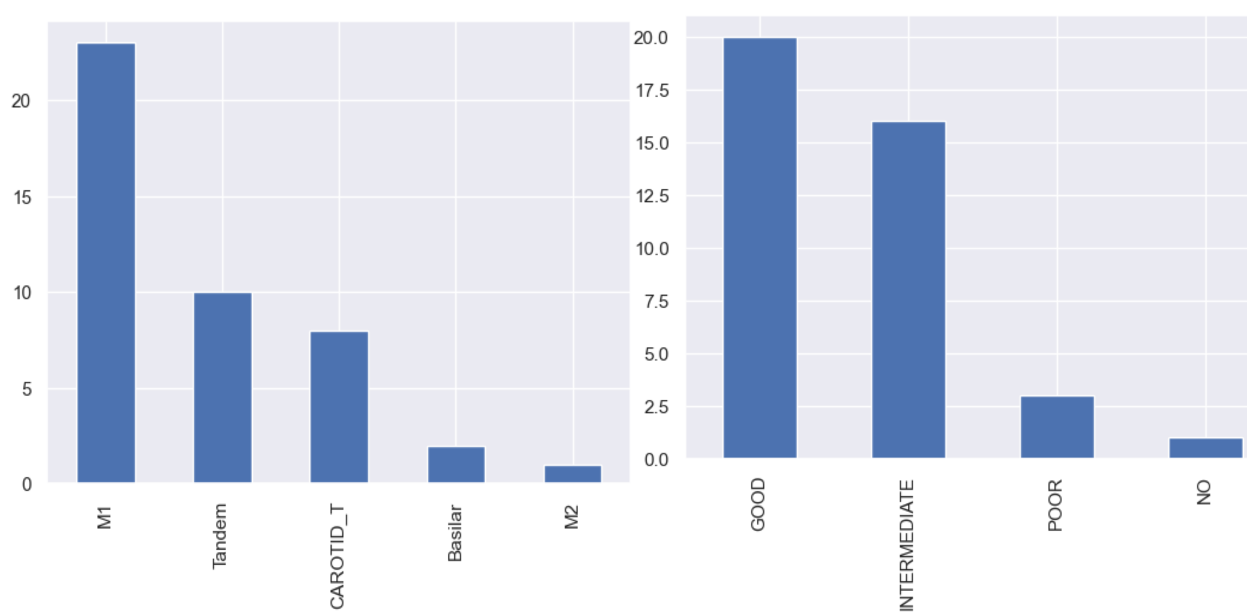


*Figure 13: Patient got EVT by Occlusion Location and Collateral Status (Occlusion: Left, Collateral: Right)*

## 4.3 Correlation Matrix

The correlation between the attributes was calculated to discover those attributes mainly affected by response variable. The following figure is the correlation table representing all attribute correlations from the original dataset.

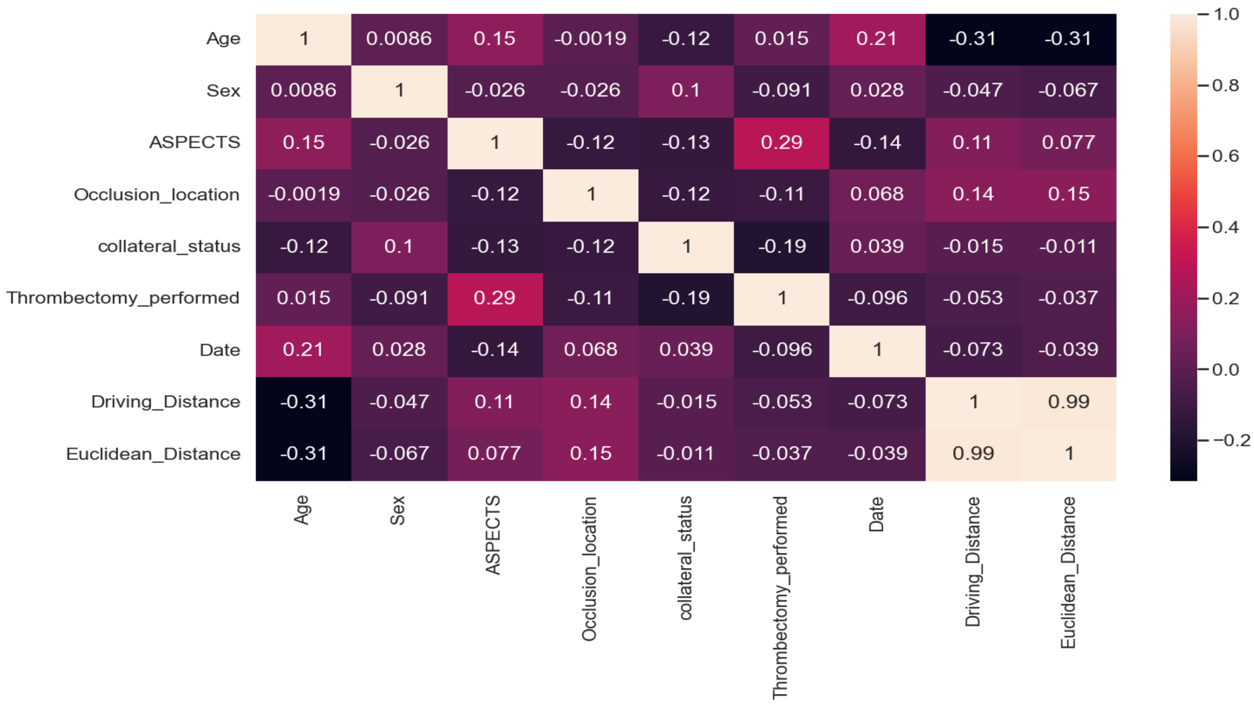


*Figure 14: Correlation Table of the Original Dataset*

Among the nine attributes, the ASPECTS was the most positively correlated, and the Occlusion Location and Collateral Status showed a high negative correlation to the response variable. The following figure shows the correlation between attributes in the merged dataset. The following table is representing the correlation coefficient for original dataset to thrombectomy performed in Table 7.

| **Variable** | **Coefficient of Correlation** |
|---|---|
| Age | 0.015 |
| Sex | -0.091 |
| ASPECTS | 0.29 |
| Occlusion Location | -0.11 |
| Collateral Status | -0.19 |

| | |
|---|---|
| Date | -0.096 |
| Driving Distance | -0.053 |
| Euclidean Distance | -0.037 |

*Table 7: Summary of correlation coefficient for original dataset*

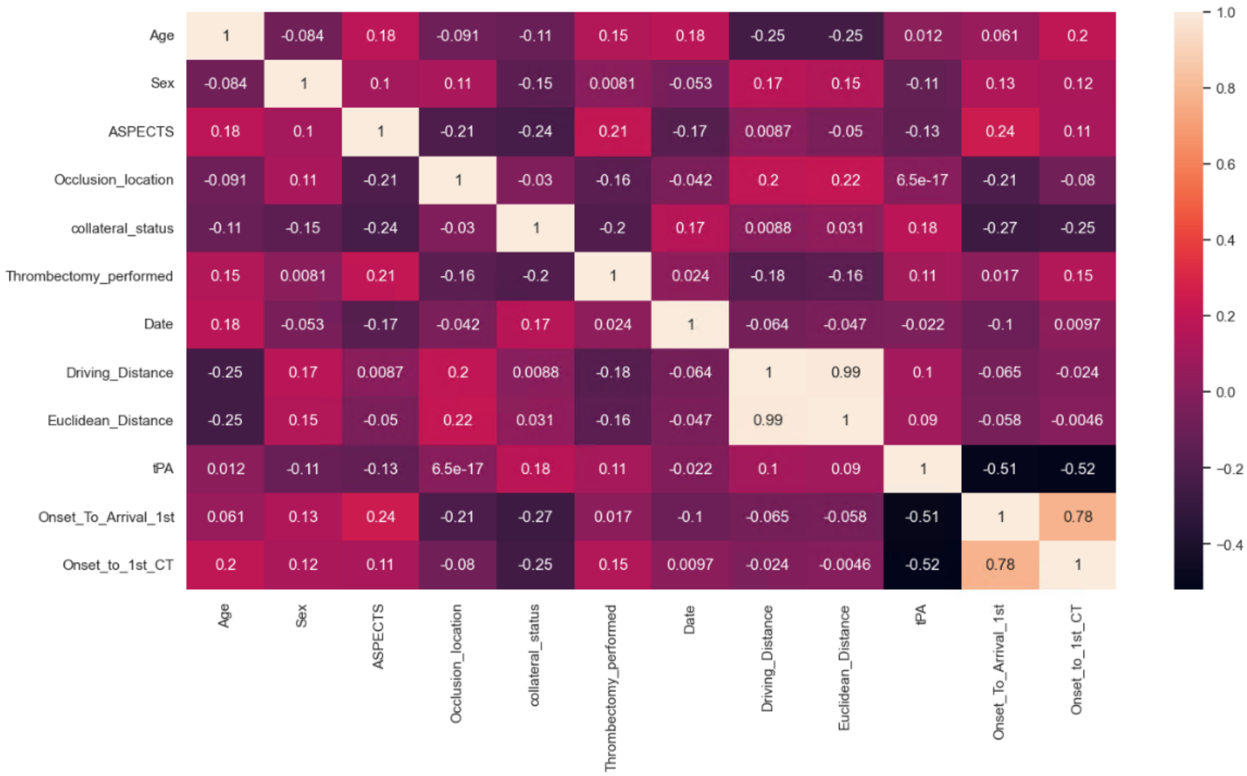


*Figure 15: Correlation Table of Merged Dataset*

The following table is the summary of correlation coefficient for merged dataset to "Thrombectomy Performed" in Table 8.

| **Variable** | **Coefficient of Correlation** |
|---|---|
| Age | 0.15 |
| Sex | 0.0081 |
| ASPECTS | 0.21 |
| Occlusion Location | -0.16 |
| Collateral Status | -0.20 |
| Date | 0.024 |
| Driving Distance | -0.18 |
| Euclidean Distance | -0.16 |
| tPA | 0.11 |
| Onset to 1$^{st}$ hospital | 0.017 |
| Onset to 1$^{st}$ CT | 0.15 |

*Table 8: Summary of correlation coefficient for merged dataset*

After adding three attributes to the dataset, correlation coefficients differ slightly from the original dataset. ASPECTS was the most positively correlated with the "Thrombectomy Performed" response variable. However, the new dataset generally increased the correlation coefficient of all other attributes to the response variable. Age also turned out to be positively correlated, and new attributes the onset to first CT time and tPA with the response variable. For the negative correlation, collateral status and occlusion location stayed negatively correlated with the response variable. The driving and Euclidean distance to the EVT center also significantly increased the negative correlation coefficient. According to the calculations of the coefficient of correlation, the merged dataset shows better potential to create better model, but it has small sample size compared to the original dataset.

## 4.4 Decision Tree Model Result

### 4.4.1 Original Dataset

The decision tree models' performance is described in the classification reports shown on Figure 15 and 16.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.31 | 0.67 | 0.42 | 6 |
| 1 | 0.75 | 0.40 | 0.52 | 15 |
| accuracy | | | 0.48 | 21 |
| macro avg | 0.53 | 0.53 | 0.47 | 21 |
| weighted avg | 0.62 | 0.48 | 0.49 | 21 |

*Figure 16: Classification Report of Decision Tree-Original Set (0.7/0.3)*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.27 | 0.40 | 0.32 | 10 |
| 1 | 0.54 | 0.39 | 0.45 | 18 |
| accuracy | | | 0.39 | 28 |
| macro avg | 0.40 | 0.39 | 0.39 | 28 |
| weighted avg | 0.44 | 0.39 | 0.40 | 28 |

*Figure 17: Classification report of Decision Tree-Original Set (0.6/0.4)*

The decision tree with 0.7/0.3 split had better performance overall—about a 9% difference in the accuracy and 7% higher in the f-1 score. Also, the AUC of the 0.7/0.3 model was 0.14 higher than the other one. The AUC of each model is shown in Figure 18.

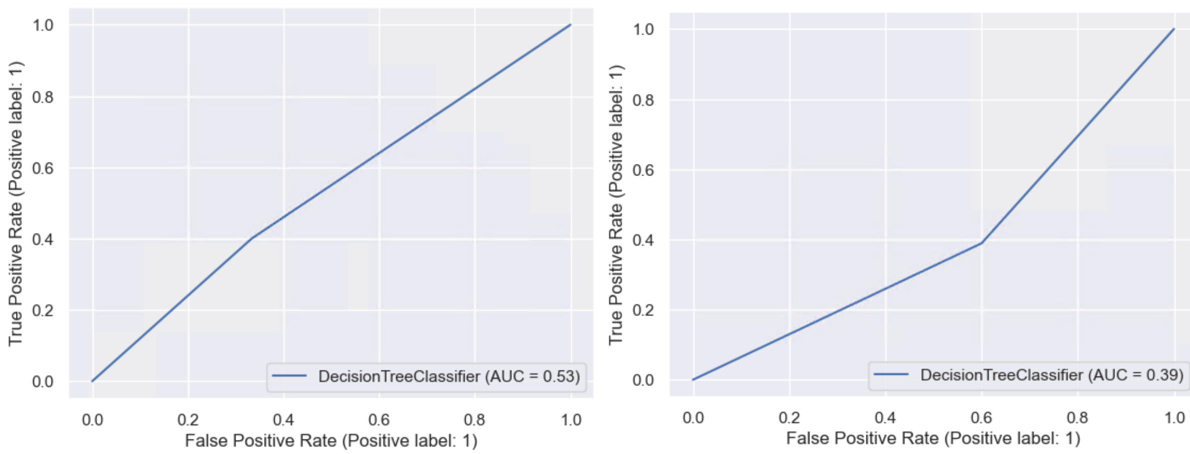


*Figure 18: AUC of Decision Tree Model-Original (Left: 0.7/0.3.Right:0.6/0.4)*

### 4.4.2 Merged Dataset

The classification reports for decision tree model for merged dataset are described in Figures 19 and 20.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.50 | 0.17 | 0.25 | 6 |
| 1 | 0.50 | 0.83 | 0.62 | 6 |
| accuracy | | | 0.50 | 12 |
| macro avg | 0.50 | 0.50 | 0.44 | 12 |
| weighted avg | 0.50 | 0.50 | 0.44 | 12 |

*Figure 19: Classification report of Decision Tree-Merged (0.7/0.3)*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.25 | 0.17 | 0.20 | 6 |
| 1 | 0.58 | 0.70 | 0.64 | 10 |
| accuracy | | | 0.50 | 16 |
| macro avg | 0.42 | 0.43 | 0.42 | 16 |
| weighted avg | 0.46 | 0.50 | 0.47 | 16 |

*Figure 20: Classification report of Decision Tree-Merged (0.6/0.4)*

In the case of the decision tree model with the merged dataset, the accuracy for each data split was equal to 50%, and the F-1 score was similar to the 0.6/0.4 one, which is 2% higher than the F-1 score. Figure 21 is the AUC of the merged dataset's decision tree models.

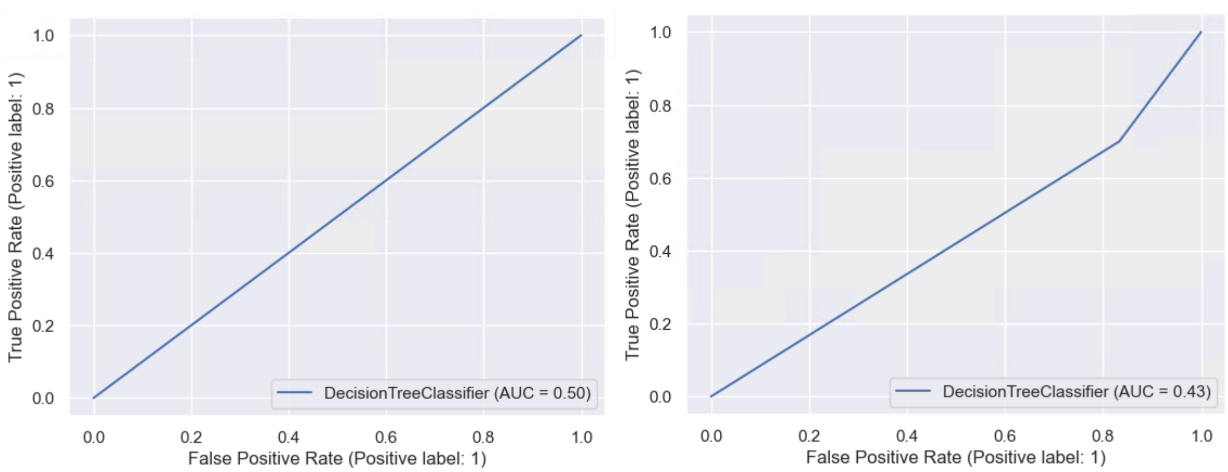


*Figure 21: AUC of Decision Tree Model-Merged (Left:0.7/0.3, Right:0.6/0.4)*

The 0.7/0.3 model showed a slightly higher AUC than the 0.6/0.4 model. In the decision tree algorithm, the merged dataset showed better overall performance when compared to the original dataset.

## 4.5 Random Forest Model Result

### 4.5.1 Original Dataset

The performance of the random forest model for the original data are described in the classification report shown on Figures 22 and 23.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 6 |
| 1 | 0.71 | 1.00 | 0.83 | 15 |
| accuracy | | | 0.71 | 21 |
| macro avg | 0.36 | 0.50 | 0.42 | 21 |
| weighted avg | 0.51 | 0.71 | 0.60 | 21 |

*Figure 22: Classification Report of Random Forest-Original (0.7/0.3)*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 10 |
| 1 | 0.64 | 1.00 | 0.78 | 18 |
| accuracy | | | 0.64 | 28 |
| macro avg | 0.32 | 0.50 | 0.39 | 28 |
| weighted avg | 0.41 | 0.64 | 0.50 | 28 |

*Figure 23: Classification Report of Random Forest-Original (0.6/0.4)*

The 0.7/0.3 split model showed better accuracy and F-1 score, about 7% and 4% higher than the other model. However, as shown in the following figure, the 0.6/0.4 split model has about 0.08 higher AUC than the 0.7/0.4 split model.

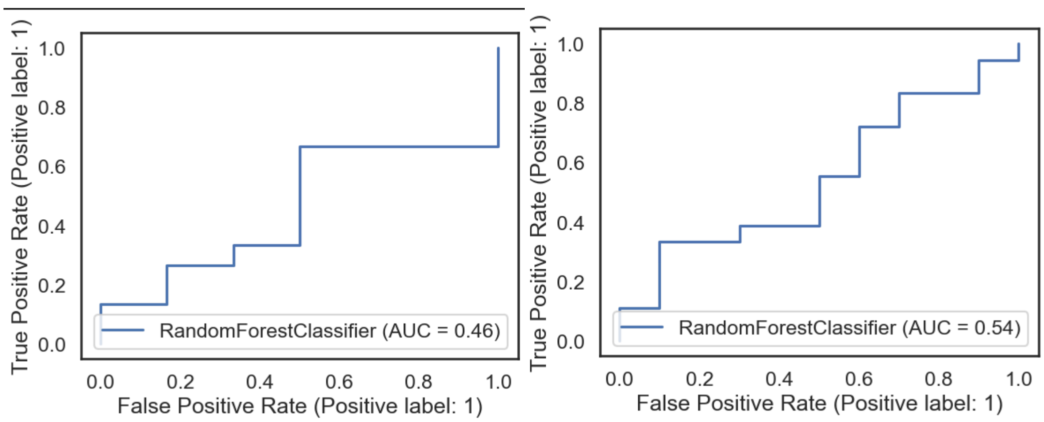


*Figure 24: AUC of Random Forest Model-Original (Left:0.7/0.3, Right:0.6/0.4)*

### 4.5.2. Merged Dataset

The classification reports shown in Figure 25 and 26 describes the model performance of the random forest model on merged dataset.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 6 |
| 1 | 0.50 | 1.00 | 0.67 | 6 |
| accuracy | | | 0.50 | 12 |
| macro avg | 0.25 | 0.50 | 0.33 | 12 |
| weighted avg | 0.25 | 0.50 | 0.33 | 12 |

*Figure 25: Classification report of Random Forest-Merged (0.7/0.3)*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 6 |
| 1 | 0.62 | 1.00 | 0.77 | 10 |
| accuracy | | | 0.62 | 16 |
| macro avg | 0.31 | 0.50 | 0.38 | 16 |
| weighted avg | 0.39 | 0.62 | 0.48 | 16 |

*Figure 26: Classification report of Random Forest-Merged (0.6/0.4)*

In the merged dataset, the split of 0.6/0.4 showed better overall performance than 0.7/0.3. The 0.6/0.4 model had a 12% higher accuracy and a 10% higher F-1 score. However, the AUC of the model had only a 0.01 difference, as shown in the following figure.

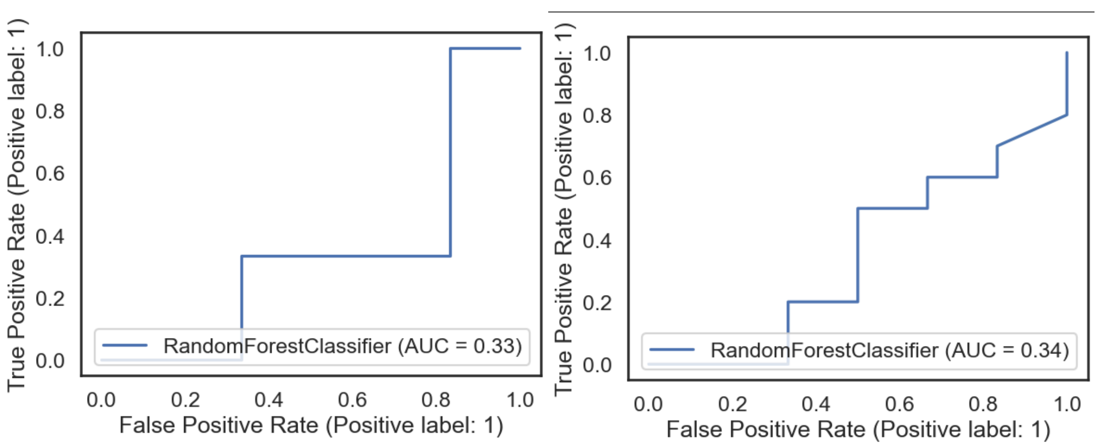


*Figure 27: AUC of Random Forest Model-Merged (Left:0.7/0.3, Right:0.6/0.4)*

In the random forest algorithm, the split of 0.7/0.3 original dataset had the best performance compared to other cases.

## 4.6 Logistic Regression Model

### 4.6.1 Original Dataset

The classification reports Figures 28 and 29 show the original dataset's logistic regression performance. The data split of 0.6/0.4 ratio got a 6% higher accuracy and 0.09 in F-1 Score.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.31 | 0.67 | 0.42 | 6 |
| 1 | 0.75 | 0.40 | 0.52 | 15 |
| accuracy | | | 0.48 | 21 |
| macro avg | 0.53 | 0.53 | 0.47 | 21 |
| weighted avg | 0.62 | 0.48 | 0.49 | 21 |

*Figure 28: Classification report of Logistic Regression-Original (0.7/0.3)*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.38 | 0.50 | 0.43 | 10 |
| 1 | 0.67 | 0.56 | 0.61 | 18 |
| accuracy | | | 0.54 | 28 |
| macro avg | 0.53 | 0.53 | 0.52 | 28 |
| weighted avg | 0.57 | 0.54 | 0.54 | 28 |

*Figure 29: Classification report of Logistic Regression-Original (0.6/0.4)*

The AUC of both models is above the 0.5 line and has a slight difference. The 0.7/0.3 split model is 0.03 higher in the AUC, as shown in the following figure.

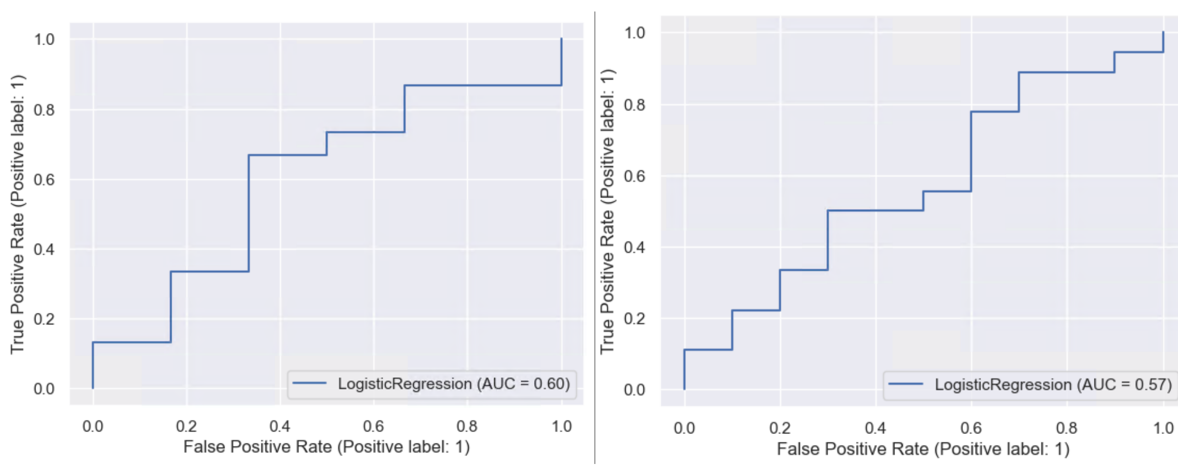


*Figure 30: AUC of Logistic Regression Model-Original (Left:0.7/0.3, Right:0.6/0.4)*

### 4.6.2 Merged Dataset

The logistic regression algorithm's performance on the merged dataset was poor compared to the original dataset. In the case of the merged dataset, the split of 0.7/0.3 ratio had 11% higher accuracy and 0.11 F-1 score. The reports are shown in Figures 31 and 32.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.40 | 0.33 | 0.36 | 6 |
| 1 | 0.43 | 0.50 | 0.46 | 6 |
| accuracy | | | 0.42 | 12 |
| macro avg | 0.41 | 0.42 | 0.41 | 12 |
| weighted avg | 0.41 | 0.42 | 0.41 | 12 |

*Figure 31: Classification report of Logistic Regression-Merged (0.7/0.3)*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.22 | 0.33 | 0.27 | 6 |
| 1 | 0.43 | 0.30 | 0.35 | 10 |
| accuracy | | | 0.31 | 16 |
| macro avg | 0.33 | 0.32 | 0.31 | 16 |
| weighted avg | 0.35 | 0.31 | 0.32 | 16 |

*Figure 32: Classification report of Logistic Regression-Merged (0.6/0.4)*

The logistic model's AUC was significantly lower compared to other algorithms. Since AUC is extremely low, defining a better model was pointless. The following figure depicts the AUC of both models.

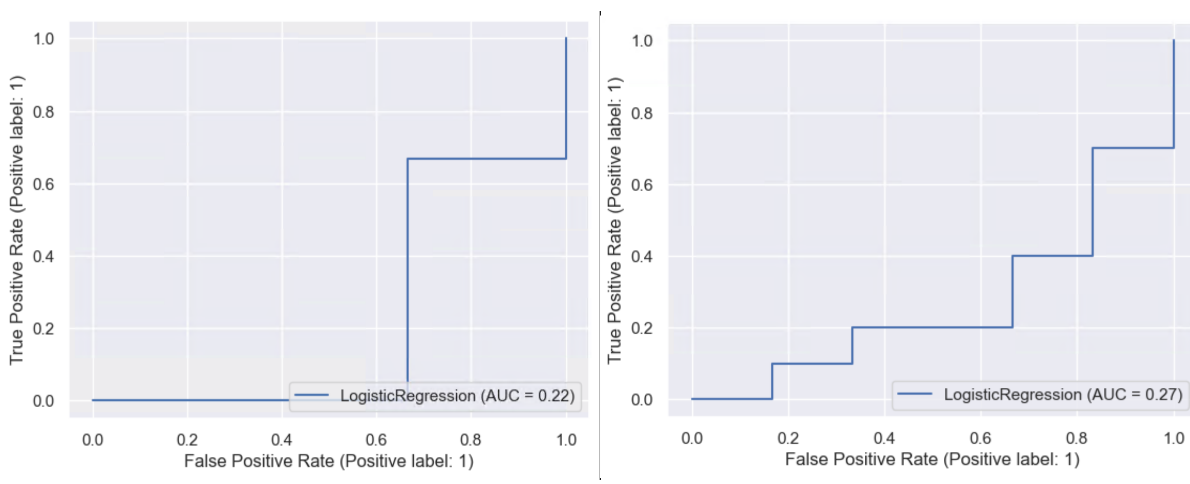


*Figure 33: AUC of Logistic Regression Model-Merged (Left:0.7/0.3, Right:0.6/0.4)*

## 4.7 Support Vector Machine

### 4.7.1 Original Dataset

The support vector machine algorithms on the original dataset performance are summarized in the classification report shown in Figures 33 and 34.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 1.00 | 0.17 | 0.29 | 6 |
| 1 | 0.75 | 1.00 | 0.86 | 15 |
| accuracy | | | 0.76 | 21 |
| macro avg | 0.88 | 0.58 | 0.57 | 21 |
| weighted avg | 0.82 | 0.76 | 0.69 | 21 |

*Figure 34: Classification report of SVM-Original (0.7/0.3)*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.67 | 0.20 | 0.31 | 10 |
| 1 | 0.68 | 0.94 | 0.79 | 18 |
| accuracy | | | 0.68 | 28 |
| macro avg | 0.67 | 0.57 | 0.55 | 28 |
| weighted avg | 0.68 | 0.68 | 0.62 | 28 |

*Figure 35: Classification report of SVM-Original (0.6/0.4)*

The result of the SVM with the original dataset made a remarkable result. Accuracy of the 0.7/0.3 split model had 76% accuracy, 8% higher than the 0.6/0.4 split model, and 0.86 F-1 score, 0.07 higher than the 0.6/0.4 split model. Also, the 0.7/0.3 split model is slightly higher in AUC than the 0.6/0.4 split model. The AUC graphs are shown in the following figure.

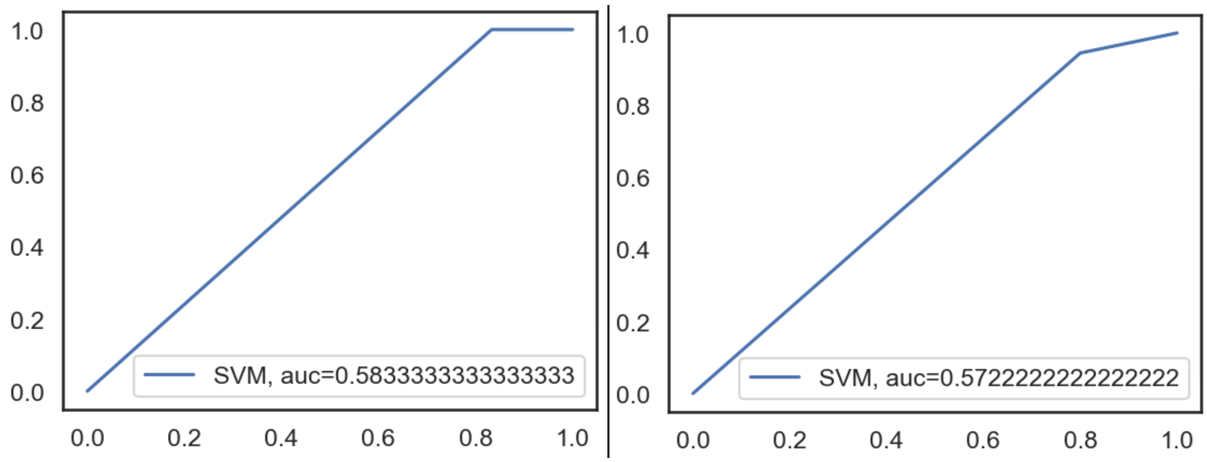


*Figure 36: AUC of SVM-Original (Left:0.7/0.3, Right:0.6/0.4)*

### 4.7.2 Merged Dataset

The performance of the SVM on the merged dataset are shown in the classification reports shown in Figure 37 and 38.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 1.00 | 0.17 | 0.29 | 6 |
| 1 | 0.55 | 1.00 | 0.71 | 6 |
| accuracy | | | 0.58 | 12 |
| macro avg | 0.77 | 0.58 | 0.50 | 12 |
| weighted avg | 0.77 | 0.58 | 0.50 | 12 |

*Figure 37: Classification report of SVM-Merged (0.7/0.3)*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 1.00 | 0.17 | 0.29 | 6 |
| 1 | 0.67 | 1.00 | 0.80 | 10 |
| accuracy | | | 0.69 | 16 |
| macro avg | 0.83 | 0.58 | 0.54 | 16 |
| weighted avg | 0.79 | 0.69 | 0.61 | 16 |

*Figure 38: Classification report of SVM-Merged (0.6/0.4)*

In the case of SVM with merged data, the data split of 0.6/0.4 ratio had a better performance with 11% higher accuracy and a 0.09 higher F-1 score than the 0.7/0.3 split model. The following figure shows that the AUC of both models was equal in this case.

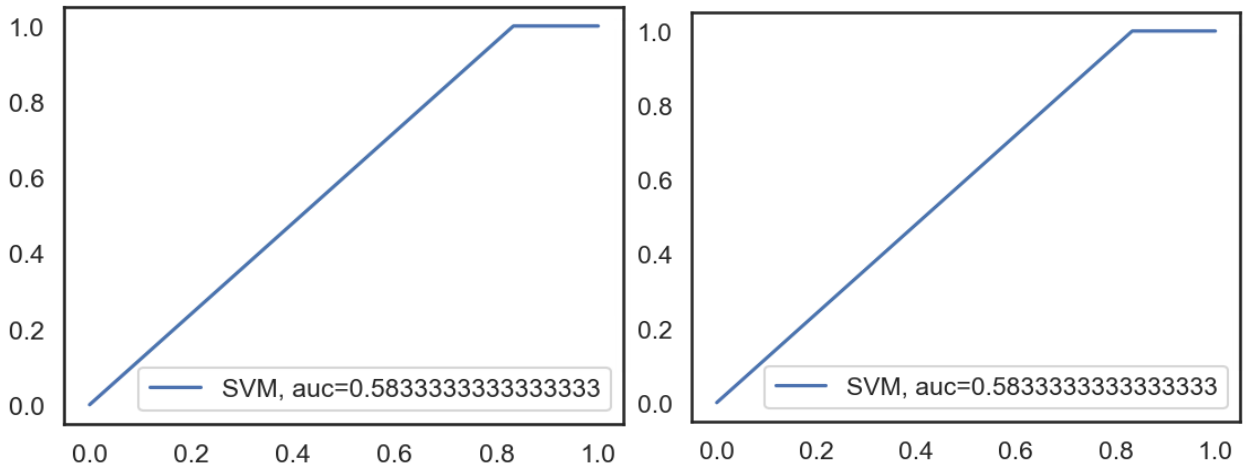


*Figure 39: AUC of SVM-Merged (Left:0.7/0.3, Right:0.6/0.4)*

## 4.8 Model Comparison

The performance was collected using two split methods in 0.7/0.3 and 0.6/0.4 ratios on four different algorithms. The result summary will be discussed in two sections: the original and merged datasets. The three-performance metrics were used for the model comparison: accuracy, AUC, and F1 score.

### 4.8.1 Original Dataset Models

The following three figures represent the comparison of model performance by the evaluation metrics mentioned in the previous section: Accuracy (Figure 40), AUC (Figure 41), and F-1 Score (Figure 42).

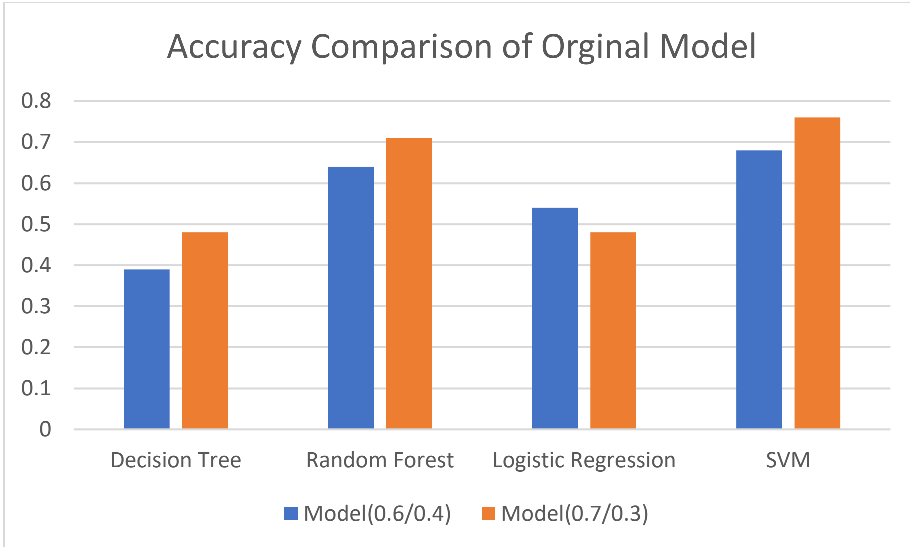


*Figure 40: Accuracy Comparison of Models (Original Dataset)*

From the accuracy perspective, SVM performed best on the original dataset with 76% accuracy on 0.7/0.3 data split ratio and 68% on 0.6/0.4 data split ratio. On the other hand, the decision tree model had the lowest accuracy in the case of the original dataset at 48% on the 0.7/0.3 data split ratio and 39% on the 0.6/0.4 data split ratio.

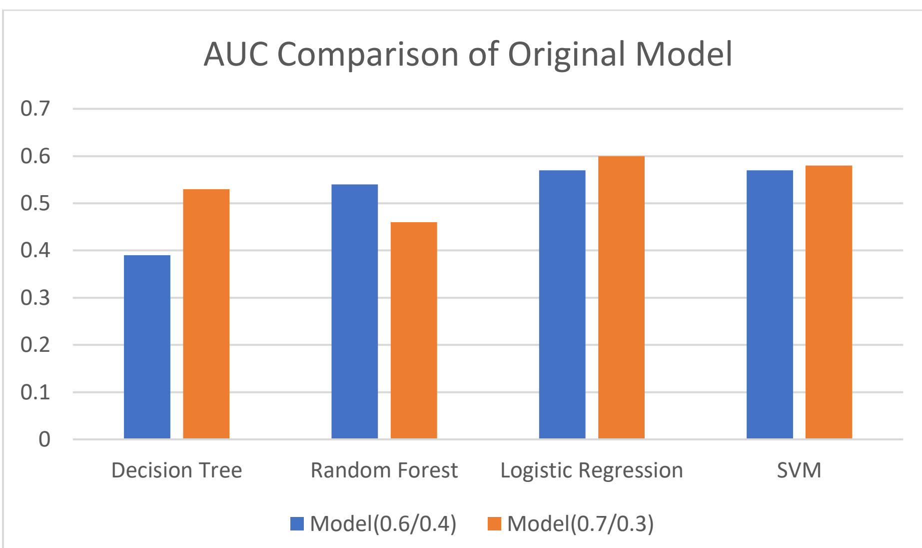


*Figure 41: AUC Comparison of Models (Original)*

The model's prediction performance can be measured through the AUC comparison. On the original dataset, the logistic regression model with a split data ratio of 0.7/0.3 had the highest AUC of 0.6. On the other hand, with a divided data ratio of 0.6/0.4, the SVM had the highest AUC of 0.58.

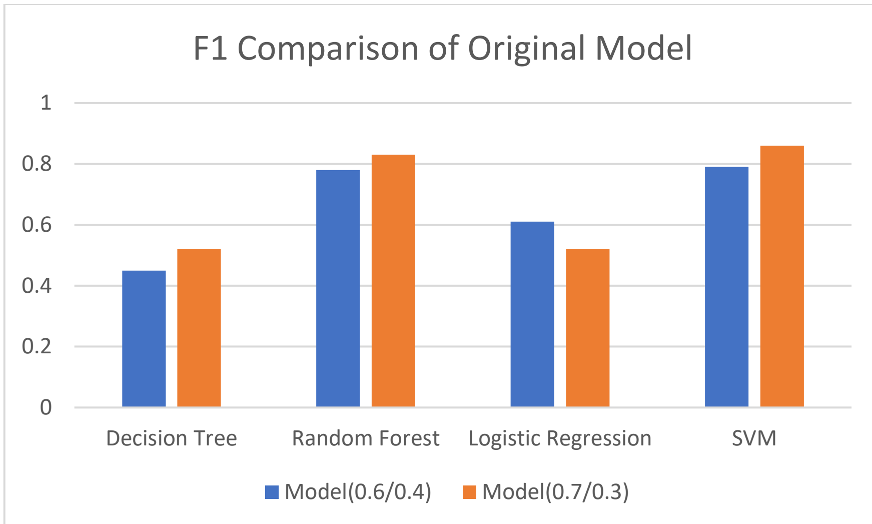


*Figure 42: F1 Score Comparison of Models (Original)*

Since the model is a binary classification and based on a small dataset, the F1 score should considered for the model performance metrics. The accuracy is looking for the correct prediction out of all predictions, while the F1 score measures the balance between precision and recall. Figure 42 shows that the SVM had the highest F1 score compared to other algorithms, 0.86 in the data split of 0.7/0.3 ratio and 0.71 in the 0.6/0.4 ratio.

### 4.8.2 Merged Dataset Models

The same metrics from the previous section were applied to the model comparison for the merged dataset. First, the accuracy of each model was compared. As shown in the following figure, the SVM had the highest accuracy for the merged dataset and both cases of the split data method: 58% in 0.7/0.3 split and 69% in 0.6/0.4 split. However, on the merged dataset, the logistic regression model had the lowest accuracy for both cases.

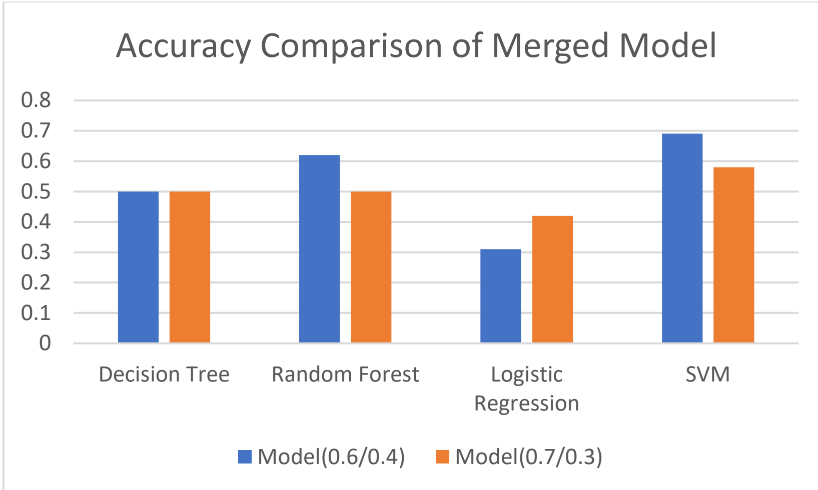


*Figure 43: Accuracy Comparison of Models (Merged)*

In comparing the AUC of the model, the SVM also had the highest AUC among the four algorithms. For both cases of split ratio, SVM had 0.58 in AUC. Also, the logistic regression model performed worse in AUC, as shown in the following figure.

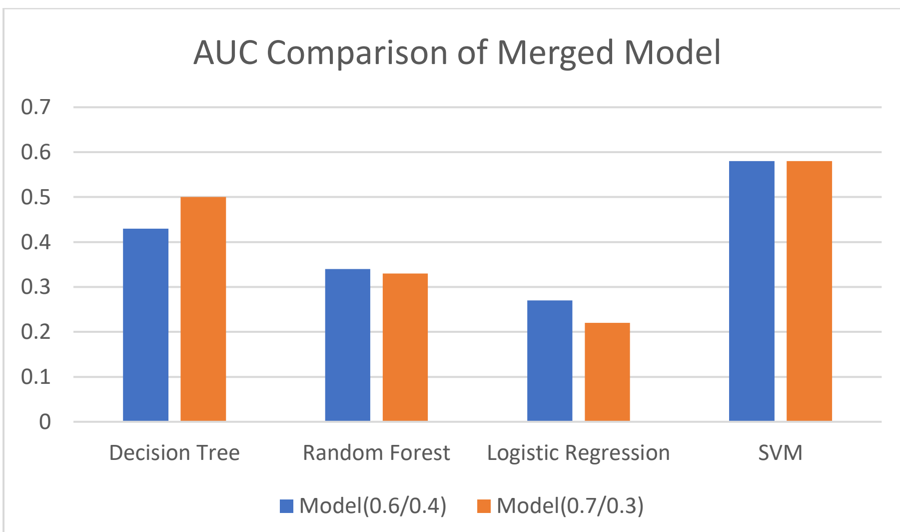


*Figure 44: AUC Comparison of Models (Merged)*

Then F1 score was also compared for the models from the merged dataset. As shown in the following figure, the SVM made the highest F1 score over the other models, with 0.8 in the data split of 0.6/0.4 ratio and 0.71 in the data split of 0.7/0.3.

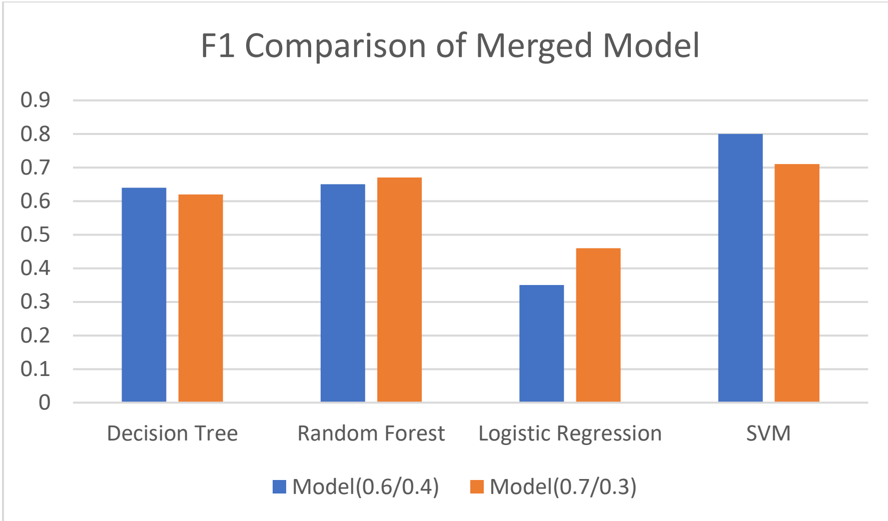


*Figure 45: F1 Comparison of Models (Merged)*

## 4.9 False Negative

The machine learning model in the healthcare system, the false negative rate is a critical metric. It represents the model fail on negative classification. In this study the false negative is the case of don't transport patient that should be transport to EVT centre which can cause the serious consequences. According to the previous section the SVM had a best performance. However, to be apply into the healthcare filed the false negative rate should be checked. The Table 9 is the false negative rate for each case of SVM model.

| Original Dataset-0.7/0.3 | Original Dataset-0.6/0.4 | Merged Dataset-0.7/0.3 | Merged Dataset-0.6/0.4 |
|---|---|---|---|
| 0 | 28.4% | 0 | 0 |

*Table 9: False Negative Rate of SVM model*

Among the four cases of SVM, three types had 0 case of false negative except the SVM from original dataset with 0.6/0.4 data split.

## 5.0 Discussion

By comparing the three metrics, the effective ratio for the data split on each dataset and the best algorithm could be selected with the highest potentiality. The original dataset's split data ratio of 0.7/0.3 better performance than 0.6/0.4 data split. However, in the merged dataset, the data split of 0.6/0.4 ratio generally showed a better result than the 0.7/0.3 split since it had a small sample size. In accuracy metrics comparison, the SVM dominated all other algorithms in the case of both datasets. The accuracy of SVM models was usually 10~30% higher than the lowest accuracy models.

Furthermore, the AUC evaluation was different for each dataset. The logistic regression models from the original dataset had the highest AUC and SVM second. But the merged dataset logistic regression case had the lowest AUC, and SVM had the highest AUC. Finally, the metrics comparison of the F-1 Score also shows that the SVM performed best in all cases. Therefore, overall evaluation selects the SVM as the best-fit algorithm for predicting transfer patient for EVT and the best potential for future study for further benefits.

Furthermore, in the correlation table from Chapter 4, variable attributes were discovered that help decide on the EVT eligibility. The ASPECTS and age were the main elements in deciding the transfer of patients to the EVT center. However, the correlation table shows occlusion location, collateral status, driving distances to the EVT center, and onset to 1st CT time are important variable to consider prior to transferring patients for EVT. ASPECTS still had the highest correlation with the EVT eligibilities, but the previously mentioned attributes had higher coefficients of correlation than the age of the patient.

## 5.1 Limitations

The most significant limitation of this study was the given dataset. The dataset was collected from various neuroradiology residents by hand. Therefore, the dataset had some inconsistencies in the data recording method; even if it were standardized through data cleaning, some data may still be inaccurate. Furthermore, the small data size was this study's most critical limitation, which can cause overfitting. Since the given dataset was 101 patients and through the data preprocessing, only 78 of them were used for model development, and the merged dataset with additional data attributes resulted in only 42 patients.

The next limitation of the project was the inconsistent data record. Since, the dataset was recorded from different hospitals across Nova Scotia and recorded by various neuroradiology residents. The patient data recording methods were different for each resident. Thus, the inconsistency of the data was substituted with the standardized values, but some inaccuracy could be caused. Furthermore, among the 21 given attributes, more than five of the attributes were deleted due to missing values. Among the deleted attributes, some of the attributes could be highly affected by the received EVT prediction.

## 5.2 Future Study

The first future study that could be initiated would be the same study to be conducted on a larger dataset. As mentioned in the previous section, fewer data can cause many problems, which results in a low-quality model. If the dataset is more than several thousands of data samples will generate significantly improved performance and stability compared to this study's result. However, the number of stroke patients annually in Nova Scotia that receive EVT is approximately 100 per year, therefore, the study would have to use national data to achieve this desired number. Furthermore, to improve the model's overall performance, new algorithms could be applied such as the ensembled model can be applied to the study, which is a combination of multiple algorithms to create stronger model.

The subsequent possible future study from would be to create a clinical decision support tool. In the current state, only 30~50% of patients are available to get the EVT when the patient is transferred to the EVT center. However, the prediction accuracy of the machine learning model is more than 95% of accuracy. In that case, the machine learning model can be developed into the clinical assistant to support decision-making for transferring patients for EVT, helping to lower the time for a decision. As the decision-making time decreases, the patient will have a higher chance of getting an EVT. Furthermore, the current system's futile transfer rate is 70~50% and the machine learning model have higher accuracy on decision making on EVT patient transfer. The application of the machine learning models decrease in futile transfer rate will bring many benefits by saving the cost incurred from unnecessary transfer costs.

# 6.0 Conclusion

The project's main objective was to determine the adaptability of the machine learning algorithm predicts transfer patient EVT. The overall performance of each model showed effective results and proved that machine learning algorithms were adaptive to the prediction of patient transfer for EVT. The support vector machine algorithms performed best among the four different algorithms. Furthermore, this study discovers attributes of patient information correlated that support with EVT eligibility decision. The most correlated attribute was the ASPECTS. The following attributes are also related to the prediction variable: ages, occlusion locations, collateral status, driving distance, Euclidean distances, tPA, and onset to first CT time. This study discovered many potentials for the machine learning model's future studies, and its application to stroke system optimization through improved decision making. First, this study has a certain opportunity to improve by applying the larger dataset and boost algorithms. Secondly, the clinical assistant tool can be created based on the machine learning model that can improve the current state of the EVT provision to ischemic stroke patients. The current failure rate of getting EVT is 70~50% after ischemic stroke patients are transferred. The failure rate can be decreased by creating the clinical assistant tool. The clinical assistant tool will bring many benefits by saving the costs for the patient transferring and shortening the processing time for decision-making, which can reduce the delay in patient transportation and the fatality of ischemic stroke patients.

## References


1. Learn about stroke [Internet]. World Stroke Organization. [cited 2021Mar25]. Available from: https://www.world-stroke.org/world-stroke-day-campaign/why-stroke-matters/learn-about-stroke

2. About stroke [Internet]. Centers for Disease Control and Prevention. Centers for Disease Control and Prevention; 2022 [cited 2023Mar25]. Available from: https://www.cdc.gov/stroke/about.htm#print

3. Kissela BM, Jane Khoury MC, Alwell K, Charles Moomaw BJ, Woo D, Opeolu Adeoye M, et al. Age at stroke Temporal trends in stroke incidence in a large, biracial population [Internet]. 2012. Available from: www.neurology.org

4. Canada PHAof. Government of Canada [Internet]. , 2017 - Canada.ca. / Gouvernement du Canada; 2019 [cited 2023Mar25]. Available from: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/stroke-canada-fact-sheet.html.

5. Tissue plasminogen activator for acute ischemic stroke. New England Journal of Medicine. 1995;333(24):1581–8.

6. TISSUE PLASMINOGEN ACTIVATOR FOR ACUTE ISCHEMIC STROKE T HE N ATIONAL I NSTITUTE OF N EUROLOGICAL D ISORDERS AND S TROKE rt-PA S TROKE S TUDY G ROUP *. 1995.

7. Malhotra K, Gornbein J, Saver JL. Ischemic strokes due to large-vessel occlusions contribute disproportionately to stroke-related dependence and death: A review. Vol. 8, Frontiers in Neurology. Frontiers Media S.A.; 2017.

8. Endovascular treatment of stroke [Internet]. CIRSE. 2018 [cited 2023Mar25]. Available from: https://www.cirse.org/patients/ir-procedures/endovascular-treatment-of-stroke/

9. Goyal M, Menon BK, Van Zwam WH, Dippel DWJ, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: A meta-analysis of individual patient data from five randomised trials. The Lancet. 2016 Apr 23;387(10029):1723–31.

10. Kurmann CC, Kaesmacher J, Pilgram-Pastor S, Piechowiak EI, Scutelnic A, Heldner MR, et al. Correlation of Collateral Scores Derived from Whole-Brain Time-Resolved Flat Panel Detector Imaging in Acute Ischemic Stroke. American Journal of Neuroradiology. 2022 Nov 1;43(11):1627–32.

11. Gao MM, Wang JZ, Liao J, Reiter SD, Vyas M V., Linkewich E, et al. Delayed thrombectomy center arrival is associated with decreased treatment probability. Canadian Journal of Neurological Sciences. 2020 Nov 1;47(6):770–4.

12. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.

13. Maier O, Schröder C, Forkert ND, Martinetz T, Handels H. Classifiers for ischemic stroke lesion segmentation: A comparison study. PLoS One. 2015 Dec 1;10(12).

14. Grosser M, Gellißen S, Borchert P, Sedlacik J, Nawabi J, Fiehler J, et al. Improved multi-parametric prediction of tissue outcome in acute ischemic stroke patients using spatial features. PLoS One. 2020 Jan 1;15(1).

15. Subudhi A, Dash M, Sabut S. Automated segmentation and classification of brain stroke using expectation-maximization and random forest classifier. Biocybern Biomed Eng. 2020 Jan 1;40(1):277–89.

16. Lee H, Lee EJ, Ham S, Lee H Bin, Lee JS, Kwon SU, et al. Machine learning approach to identify stroke within 4.5 hours. Stroke. 2020;860–6.

17. You J, Tsang ACO, Yu PLH, Tsui ELH, Woo PPS, Lui CSM, et al. Automated Hierarchy Evaluation System of Large Vessel Occlusion in Acute Ischemia Stroke. Front Neuroinform. 2020 Mar 24;14.

18. Allen M, James C, Frost J, Liabo K, Pearn K, Monks T, et al. Use of Clinical Pathway Simulation and Machine Learning to Identify Key Levers for Maximizing the Benefit of Intravenous Thrombolysis in Acute Stroke. Stroke [Internet]. 2022 Jul 15; Available from: https://www.ahajournals.org/doi/10.1161/STROKEAHA.121.038454

19. Bandi V, Bhattacharyya D, Midhunchakkravarthy D. Prediction of brain stroke severity using machine learning. Revue d'Intelligence Artificielle. 2020 Dec 31;34(6):753–61.

20. Adam SY, Yousif A, Bashir MB. Classification of Ischemic Stroke using Machine Learning Algorithms. Vol. 149, International Journal of Computer Applications. 2016.

21. Hayashi Y, Shimada T, Hattori N, Shimazui T, Yoshida Y, Miura RE, et al. A prehospital diagnostic algorithm for strokes using machine learning: a prospective observational study. Sci Rep. 2021 Dec 1;11(1).

22. Abedi V, Avula V, Chaudhary D, Shahjouei S, Khan A, Griessenauer CJ, et al. Prediction of long-term stroke recurrence using machine learning models. J Clin Med. 2021 Mar 2;10(6):1–16.

23. Hung LC, Sung SF, Hu YH. A machine learning approach to predicting readmission or mortality in patients hospitalized for stroke or transient ischemic attack. Applied Sciences (Switzerland). 2020 Sep 1;10(18).

24. Ntaios G, Sagris D, Kallipolitis A, Karagkiozi E, Korompoki E, Manios E, et al. Machine-Learning-Derived Model for the Stratification of Cardiovascular risk in Patients with Ischemic Stroke. Journal of Stroke and Cerebrovascular Diseases. 2021 Oct 1;30(10).

25. Kim C, Lee SH, Lim JS, Kim Y, Jang MU, Oh MS, et al. Impact of 25-Hydroxyvitamin D on the Prognosis of Acute Ischemic Stroke: Machine Learning Approach. Front Neurol. 2020 Jan 31;11.

26. Mitra J, Bourgeat P, Fripp J, Ghose S, Rose S, Salvado O, et al. Lesion segmentation from multimodal MRI using random forest following ischemic stroke. Neuroimage. 2014;98:324–35.

27. Monteiro M, Fonseca AC, Freitas AT, Pinho E Melo T, Francisco AP, Ferro JM, et al. Using Machine Learning to Improve the Prediction of Functional Outcome in Ischemic Stroke Patients. IEEE/ACM Trans Comput Biol Bioinform. 2018 Nov 1;15(6):1953–9.

28. Meier R, Burri M, Med M, Fischer S, Mckinley R, Jung S, et al. Beyond Accuracy: Investigating the Potential Clinical Utility of Predicting Functional Dependency and Severe Disability or Death in Successfully Reperfused Patients using Machine Learning. Available from: https://doi.org/10.1101/2020.11.17.20232280

29. Jiang B, Zhu G, Xie Y, Heit JJ, Chen H, Li Y, et al. Prediction of clinical outcome in patients with large-vessel acute ischemic stroke: Performance of machine learning versus span-100. American Journal of Neuroradiology. 2021 Feb 1;42(2):240–6.

30. Li X, Pan XD, Jiang CL, Wu MR, Liu YK, Wang FS, et al. Predicting 6-Month Unfavorable Outcome of Acute Ischemic Stroke Using Machine Learning. Front Neurol. 2020 Nov 19;11.

31. Su PY, Wei YC, Luo H, Liu CH, Huang WY, Chen KF, et al. Machine Learning Models for Predicting Influential Factors of Early Outcomes in Acute Ischemic Stroke: Registry-Based Study. JMIR Med Inform. 2022 Mar 1;10(3).

32. Chiu IM, Zeng WH, Cheng CY, Chen SH, Lin CHR. Using a multiclass machine learning model to predict the outcome of acute ischemic stroke requiring reperfusion therapy. Diagnostics. 2021 Jan 1;11(1).

33. Kappelhof N, Ramos LA, Kappelhof M, van Os HJA, Chalos V, van Kranendonk KR, et al. Evolutionary algorithms and decision trees for predicting poor outcome after endovascular treatment for acute ischemic stroke. Comput Biol Med. 2021 Jun 1;133.

34. Ramos LA, Kappelhof M, van Os HJA, Chalos V, Van Kranendonk K, Kruyt ND, et al. Predicting Poor Outcome Before Endovascular Treatment in Patients With Acute Ischemic Stroke. Front Neurol. 2020 Oct 15;11.

35. Van Os HJA, Ramos LA, Hilbert A, Van Leeuwen M, Van Walderveen MAA, Kruyt ND, et al. Predicting outcome of endovascular treatment for acute ischemic stroke: Potential value of machine learning algorithms. Front Neurol. 2018 Sep 25;9(SEP).

36. Brugnara G, Neuberger U, Mahmutoglu MA, Foltyn M, Herweh C, Nagel S, et al. Multimodal Predictive Modeling of Endovascular Treatment Outcome for Acute Ischemic Stroke Using Machine-Learning. Stroke. 2020;3541–51.

37. Fernandez-Lozano C, Hervella P, Mato-Abad V, Rodríguez-Yáñez M, Suárez-Garaboa S, López-Dequidt I, et al. Random forest-based prediction of stroke outcome. Sci Rep. 2021 Dec 1;11(1).

38. Alqudah AM, Al-Hashem M, Alqudah A. Reduced Number of Parameters for Predicting Post-Stroke Activities of Daily Living Using Machine Learning Algorithms on Initiating Rehabilitation. Informatica (Slovenia). 2021;45(4):571–81.

39. Rui C, Zhang S, Li J, Guo D, Zhang W, Tian D, et al. A Study on Predicting the Length of Hospital Stay for Chinese Patients With Ischemic Stroke Based on the XGBoost Algorithm. 2021; Available from: https://doi.org/10.21203/rs.3.rs-344337/v1

40. Uddin Emon M, Sultana Keya M, Islam Meghla T, Mahfujur Rahman M, Shamim Al Mamun M, Shamim Kaiser M. Performance Analysis of Machine Learning Approaches in Stroke Prediction.

41. RVS Technical Campus, Institute of Electrical and Electronics Engineers. Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA 2020) : 5-7, November 2020.

42. Ferdib-Al-Islam, Ghosh M. An Enhanced Stroke Prediction Scheme Using Smote and Machine Learning Techniques. In: 2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021. Institute of Electrical and Electronics Engineers Inc.; 2021.

43. Yu J, Park S, Lee H, Pyo CS, Lee YS. An elderly health monitoring system using machine learning and in-depth analysis techniques on the NIH stroke scale. Mathematics. 2020 Jul 1;8(7).

44. Lin SY, Law KM, Yeh YC, Wu KC, Lai JH, Lin CH, et al. Applying Machine Learning to Carotid Sonographic Features for Recurrent Stroke in Patients With Acute Stroke. Front Cardiovasc Med. 2022 Jan 28;9.

45. Park HA. An introduction to logistic regression: From basic concepts to interpretation with particular attention to nursing domain. J Korean Acad Nurs. 2013;43(2):154–64.

46. Sukhchandan. Decision Tree. Lecture presented at: Lecture 7; Dalhousie University; 2023 Jan 15.

47. Charbuty B, Abdulazeez A. Classification Based on Decision Tree Algorithm for Machine Learning. Journal of Applied Science and Technology Trends. 2021 Mar 24;2(01):20–8.

48. Tangirala S. Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm* [Internet]. Vol. 11, IJACSA) International Journal of Advanced Computer Science and Applications. 2020. Available from: www.ijacsa.thesai.org

49. Banerjee P, Preissner R. Bitter sweet forest: A Random Forest based binary classifier to predict bitterness and sweetness of chemical compounds. Front Chem. 2018 Apr 1;6(APR).

50. Kirasich K;, Smith T;, Sadler B. Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets [Internet]. Vol. 1, SMU Data Science Review. Available from: https://scholar.smu.edu/datasciencereviewAvailableat:https://scholar.smu.edu/datasciencereview/vol1/iss3/9http://digitalrepository.smu.edu.

51. Burges CJC. A Tutorial on Support Vector Machines for Pattern Recognition. Vol. 2, Data Mining and Knowledge Discovery. 1998.

52. Sukhchandan. Support Vector Machine. Lecture presented at: Lecture 8; Dalhousie University; 2023 Jan 15.

53. Shao YH, Chen WJ, Deng NY. Nonparallel hyperplane support vector machine for binary classification problems. Inf Sci (N Y). 2014 Apr 1;263:22–35.

54. Sukhchandan. Basic Concepts of Machine Learning Part B. Lecture presented at: Lecture 2; Dalhousie University; 2023 Jan 15.

55. Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett. 2006 Jun;27(8):861–74.

# Appendix A – Python Codes

## Decision Tree

Decision Tree Classifier

```python
from sklearn.tree import DecisionTreeClassifier
```

[381] Python

```python
dtree_model=DecisionTreeClassifier()
dtree_model.fit(X_train,y_train)
```

[382] Python

```
DecisionTreeClassifier()
```

```python
dtree_predictions = dtree_model.predict(X_test)
print(classification_report(y_test,dtree_predictions))
```

[383] Python

```
              precision    recall  f1-score   support

           0       0.31      0.67      0.42         6
           1       0.75      0.40      0.52        15

    accuracy                           0.48        21
   macro avg       0.53      0.53      0.47        21
weighted avg       0.62      0.48      0.49        21
```

```python
print(confusion_matrix(y_test,dtree_predictions))
```

[384] Python

```
[[4 2]
 [9 6]]
```

## Random Forest

```python
weight_range = [int(x) for x in np.logspace(1, 2.5, num=5)] # Weights for imbalanced data
min_leaf_range = [int(x) for x in np.linspace(8, 20, num=5)] #
best_score = 0
best_combination = (10, 5)

for w, leaf_s in [(x, y) for x in weight_range for y in min_leaf_range]:
    rf_model = RandomForestClassifier(n_estimators=150, criterion='gini',
                                      class_weight = {0: 1, 1: w}, min_samples_leaf=leaf_s,
                                      max_features = None)
    rf_model.fit(X_train, np.ravel(y_train))
    y_pred = rf_model.predict_proba(X_test)
    current_score = f1_score(y_test,np.around(y_pred[:, 1]))
    print("Weight:", w,"\tMin leaf sample:", leaf_s, "\tF1-Score:", current_score)
    if current_score > best_score:
        best_score = current_score
        best_combination = (w, leaf_s)

w, leaf_s = best_combination
rf_model = RandomForestClassifier(n_estimators=150, criterion='gini',
                                  class_weight = {0: 1, 1: w}, min_samples_leaf=leaf_s,
                                  max_features = None)
rf_model.fit(X_train, np.ravel(y_train))
```

[403] Python

```python
y_pred = rf_model.predict_proba(X_train)
print('Score on the training set:')
print(classification_report(y_train, np.around(y_pred[:, 1])))
print('roc_auc score: ', end='')
print(roc_auc_score(y_train, y_pred[:, 1]))
print('f1 score:', f1_score(y_train,np.around(y_pred[:, 1])), end='\n\n')

y_pred = rf_model.predict_proba(X_test)
print('Score on the dev set:')
print(classification_report(y_test, np.around(y_pred[:, 1])))
print('roc_auc score: ', end='')
print(roc_auc_score(y_test, y_pred[:, 1]))
print('f1 score:', f1_score(y_test,np.around(y_pred[:, 1])), end='\n\n')
```

[404] Python

## Logistic Regression

Logistic Regression

```
[387]
from sklearn.linear_model import LogisticRegression
```
Python

```
[388]
std_scaler=StandardScaler()
std_scaler.fit(X_train)
X_train_Scaled=std_scaler.transform(X_train)
X_test_Scaled=std_scaler.transform(X_test)
```
Python

```
[389]
logistic_model = LogisticRegression(random_state=101)
logistic_model.fit(X_train_Scaled,y_train)
```
Python

```
LogisticRegression(random_state=101)
```

```
[390]
logistic_predictions=logistic_model.predict(X_test_Scaled)
```
Python

```
[391]
print(classification_report(y_test,logistic_predictions))
```
Python

```
              precision    recall  f1-score   support

           0       0.31      0.67      0.42         6
           1       0.75      0.40      0.52        15

    accuracy                           0.48        21
   macro avg       0.53      0.53      0.47        21
weighted avg       0.62      0.48      0.49        21
```

```
[404]
y_pred = rf_model.predict_proba(X_train)
print('Score on the training set:')
print(classification_report(y_train, np.around(y_pred[:, 1])))
print('roc_auc score: ', end='')
print(roc_auc_score(y_train, y_pred[:, 1]))
print('f1 score:', f1_score(y_train,np.around(y_pred[:, 1])), end='\n\n')

y_pred = rf_model.predict_proba(X_test)
print('Score on the dev set:')
print(classification_report(y_test, np.around(y_pred[:, 1])))
print('roc_auc score: ', end='')
print(roc_auc_score(y_test, y_pred[:, 1]))
print('f1 score:', f1_score(y_test,np.around(y_pred[:, 1])), end='\n\n')
```
Python

## Support Vector Machine

SVM

```
[406]
from sklearn.svm import SVC
from sklearn.model_selection import GridSearchCV
import random as rd
kernels = ['rbf','linear','poly','sigmoid']

svc = SVC()
hyperParam = [{'kernel':kernels}]

gsv = GridSearchCV(svc,hyperParam,cv=5,verbose=1)
best_model = gsv.fit(X_train, y_train)
svc_pred = best_model.best_estimator_.predict(X_test)

print("Best HyperParameter: ",gsv.best_params_)
print("Best Accuracy :",best_model.score(X_test, y_test))
```
Python

```
Fitting 5 folds for each of 4 candidates, totalling 20 fits
Best HyperParameter:  {'kernel': 'sigmoid'}
Best Accuracy : 0.6190476190476191
```

```
[407]
print(classification_report(y_test,svc_pred))
```
Python

```
              precision    recall  f1-score   support

           0       0.40      0.67      0.50         6
           1       0.82      0.60      0.69        15

    accuracy                           0.62        21
   macro avg       0.61      0.63      0.60        21
weighted avg       0.70      0.62      0.64        21
```

```python
y_pred = svc.predict(X_train)
print('Score on the training set:')
print(classification_report(y_train, y_pred))
print('roc_auc score: ', end='')
print(roc_auc_score(y_train, y_pred))
print('f1 score:', f1_score(y_train, y_pred), end='\n\n')

y_pred = svc.predict(X_test)
print('Score on the dev set:')
print(classification_report(y_test, y_pred))
print('roc_auc score: ', end='')
print(roc_auc_score(y_test,  y_pred))
print('f1 score:', f1_score(y_test,  y_pred), end='\n\n')
```

[409] Python