SCHOLARLY REVIEW

The Use of Machine Learning in Stroke Detection: Performance, Medicinal Knowledge, and Future Prospects

By Devin Yue

Author Biography

Devin is a senior at St. George's School. He has a burning passion for computer science, from exploring competitive programming to learning about Machine Learning. His enthusiasm doesn't stem from coding itself but from its vast and life-changing potential in the application of CS. He will continue to innovate and design, attempting to make improvements small steps at a time.

Abstract

Impacting the health of the worldwide population, strokes pose a significant global health challenge, affecting people's lives on a massive scale. Despite its hazardous nature, strokes are largely preventable; however, due to factors such as awareness and health care access, many suffer an unfortunate fate. Many research endeavors have generated predictive models for stroke; however, less have laid out interpretable methods in the clinical context and optimized models for practicality. This research introduces clear approaches that employ interpretable methods suited to the clinical context. This paper involves engineered models that incorporate ML techniques such as dataset resampling, feature importance and selection, parameterization, and analytical techniques. The approach not only offers comprehensive methods, but also an effective model that sets up the foundations for stroke prediction in the real world. Using the Stroke Prediction Dataset, a Deep Neural Network (DNN) model and a Decision Tree (DT) model were constructed. By applying several techniques, including resampling technique SMOTE, feature selection RFE, and feature importance, both models offered insight into the influences on stroke and into the methods themselves. The multifaceted techniques enabled the DNN to achieve a 71% accuracy and the DT to achieve an astounding 95.4% accuracy, with both having only 3 input features. The high accuracy and low feature count of the DT will build a strong foundation for a cost-effective solution towards stroke prediction in the real world.

Keywords: Machine Learning, Deep learning, Stroke, Data Resampling, Feature Importance

Introduction

Cerebrovascular Accidents (CVA), more commonly known as strokes, represent a leading cause of morbidity and mortality worldwide, posing significant challenges to public health (WHO). Stroke is a condition characterized by the interruption of blood flow to a part of the brain, resulting in tissue damage and necrosis due to oxygen deprivation (John Hopkins, 2024). From causing fatality and severe disabilities to milder and recoverable conditions, stroke will affect more than a quarter of the world's population in their lifetime (The GBD 2016 Lifetime Risk of Stroke Collaborators, 2018).

However, despite its potentially life-threatening effects, 80% of strokes are preventable, according to the Centers for Disease Control, through lifestyle changes and medical assistance – such as managing dietary intake and conducting regular checkups (CDC, 2024). Nonetheless, strokes have continued to have a devastating effect on human health, being the second leading cause of death around the world and rendering millions disabled for life annually (World Stroke Organization, 2020). In order to raise standards of living and prevent millions of deaths, methods for spotting patients with high likelihoods of getting a stroke must be integrated into our healthcare systems to identify patients who require medical assistance for stroke prevention.

With artificial intelligence and data analytics advancing at incredible rates, could a machine learning (ML) model be used to identify stroke-prone patients on a cost-efficient and accurate scale? If so, the likelihood of spotting stroke-prone patients drastically increases, allowing preventional methods to be taken as early as possible (Daidone et al., 2024). Moreover, having an accessible presence of an ML stroke-detection model can increase the prevention of stroke, especially in areas with less accessible health care —such as rural districts or developing areas (Hammond et al., 2020). In this paper, algorithms and methods will be experimented with to produce an ML model that will detect stroke at a resource efficient and accurate rate.

Method

There are several medical statistics and behaviors that have been extensively, through scientific research, proven to affect stroke: age, smoking, hypertension (high blood pressure), cholesterol levels, diabetes, stress, and a variety of other factors (National Institutes of Health). The dataset that was employed included several of these features: age, hypertension, heart disease, ever-married, work type, residence type, glucose level, Body Mass Index (BMI), and smoking status (Palacios, 2020). By creating an efficient ML model, it is able to help predict the likelihood of a stroke within a patient. Prior to training the model with the dataset, preprocessing steps were taken to help maximize its accuracy and avoid inefficiency. These steps involved removing weak and inconsistent data and features that have little impact on stroke.

Data Preprocessing

First, there was weak and insufficient data within the dataset. For example, in processing the gender features, while there were three gender categories (Male, Female, and Other), the 'Other' category only had 1 entry, a case that would not only unnecessarily increase the complexity of the model, but also be an unreliable entry as there is only one singular case. Moreover, there were many instances that a feature would have 'NaN' as a value; this means that the instance should be rid of to maintain a rigid model that accounts for every feature.

Second, there was a heavy imbalance within the dataset between the two classes: stroke and no stroke. With 207 having suffered from strokes and 2727 having none, the imbalance was around a ratio of 1:13 between the minority class (stroke) and the majority class (no stroke). In machine learning, an imbalance dataset can pose several challenges and consequences for the performance of the ML model. First, a bias towards the majority class will be formed. Since the model's performance is evaluated through accuracy, the model will develop a tendency to always pick the majority class and neglect the minority class. For example, if the model were to predict all cases as having no stroke on the imbalance dataset, it would achieve a 93% accuracy. Although nominally a successful value, it is futile in helping medical practitioners realize stroke-prone patients. To

negate this effect, the technique of Synthetic Minority Oversampling Technique (SMOTE) is employed. Until the dataset is balanced, SMOTE will be used to continuously synthesize new instances of the minority class through extrapolating on the instances of the current group (Satpathy, 2024). However, the dataset SMOTE synthesizes may not be reflective of the actual distribution, skewing the model to recognize the potentially limited information of the minority class.

Evaluation Methods

Within the dataset, there were features that, at first, seemed to have little scientific connections with stroke: marital status, work type, and residence type. In each of these cases, there is scientific evidence that demonstrates these features' effect on stroke. A meta-analysis published in BMJ Journals found that divorced and widowed individuals had a higher risk of stroke, being 35% and 16% higher, respectively, than those currently married (Wong et al., 2019). In a study published in Lancet that was conducted on more than 600,000 employees, it was discovered that those who worked longer than 55 hours a week were 33% more likely to experience a stroke than those who worked 40 hours weekly (Johnson, 2021). The American Heart Association has also stated how the lower standards of living in rural areas, involving education, food insecurity, living conditions, and medical access, has an effect on rural residents' stroke chance (American Heart Association, 2023). However, within the dataset, the marital status and work type features have very ambiguous inputs. The 'marital status' feature only shows whether or not the patient has married before; it has no inputs regarding divorce or the current state of marriage. The 'work type feature' also only has three unique inputs: 'Private', 'Self-employed', and 'Government Job'. These have quite a vague reference to how stressful and how long individuals have to work. Nonetheless, to ensure the accuracy of the model, feature selection will be used to test whether or not the inclusion of these features, along with the rest, can have a beneficial effect on the model.

ML is ubiquitous in every corner of the methods and the results. A variety of methods were used to both build the model from the thousands of data entries and to evaluate the success of the model. There are four scoring metrics that were looked into to help evaluate the performance of the model: recall, precision, accuracy, and F1. Based on the prediction of the model and the actual answer, four classes emerge, True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN), as shown in Figure 1 below:

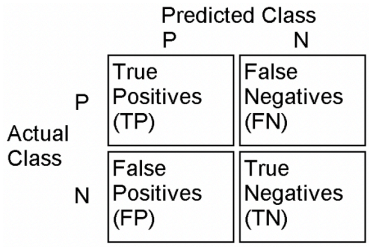

Figure 1: A table that outlines the basis of the four terms based on predicted class and actual class, a.k.a. A Confusion Matrix. Image from mlxtend.

In the context of stroke prediction, a true positive would represent the model predicting a patient to have a stroke when the patient did suffer from stroke. A false positive would represent the model predicting a patient to have a stroke when the patient didn't suffer from stroke. A false negative would represent the model predicting a patient to not have a stroke when the patient did have one. A true negative would represent the model predicting a patient to not have a stroke when the patient didn't.

From the confusion matrix, four evaluation methods can be used: recall, precision, F1 score and accuracy. The four scores may be evaluated through several formulas, as shown below:

$$\frac{TP}{TP+FN}$$

Recall Formula

$$2\,x\,\frac{Recall \times Precision}{Recall + Precision}$$

F1 Formula

$$\frac{TP}{TP+FP}$$

Precision Formula

$$\frac{TP+TN}{All\ Instances}$$

Accuracy Formula

Recall measures, out of the actual positives, how many did the model find; a high recall means the model is proficient at identifying all the patients with stroke. Precision measures, out of the predicted positives, how many were actually correct; a high precision means the model is proficient at separating

those with strokes and those without, rather than just guessing everything as having a stroke. The F1 score is the method of looking at both recall and precision as a harmonic whole. Accuracy measures the percentage of correct answers the model was able to predict, generating a holistic measure.

Two types of models were tested to realize key insights into stroke: Deep Neural Networks (DNN) and Decision Trees (DT). DNNs are essentially neural networks that have hidden layers between the input and the output layer, making them 'deep' and different from linear models like logistic regression, as shown in Figure 2 below:

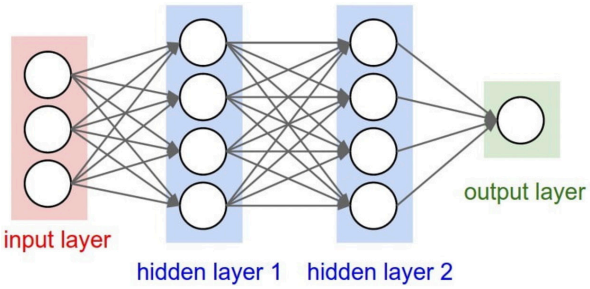

Figure 2: Visualization of a Deep Neural Network. Figure from https://www.bmc.com/blogs/deep-neural-network/.

A simple explanation of DNNs is how each layer will pass on information to the next through the weighted lines. These passed-on values will go through an activation function —in this case, it will be a ReLU (Rectified Linear Unit) function. This will occur throughout the hidden layers until the output layer is reached and a continuous output is displayed. These hidden layers introduce non-linearity to the model, allowing the model to make more complicated connections between the inputs and outputs. Since the cause of stroke is a result of multiple factors that have many influences, it is a complex condition that cannot be predicted as accurately with linear models compared to non-linear models with deeper patterns. DTs are another type of non-linear models that are able to spot complex relationships within a dataset. DTs, based on the feature values of instances, partition the dataset into multiple sub-nodes, aiming for a high homogeneity –the degree to which the instances in the subnodes are of the same class– within those subnodes. The tree will continuously do this to create a model that aims for a high homogeneity in its final state, separating the two groups efficiently.

For these two models, it is important to address two error evaluation methods employed. Binary Cross Entropy Error (BCE) , an error evaluation method typically used for binary classification, was used with the DNN (Sorokin, 2024). The formula for the BCE is shown below, with y representing the actual result and y^ representing the predicted result:

$$BCE(y, y\hat{}) = -[ylog(y\hat{}) + (1 - y)log(1 - y\hat{})]$$

Binary Cross Entropy Error Formula

Through the BCE formula, the difference between the predicted value and actual value is calculated for the entire result, with a score closer to 0 representing near-perfection. In addition to BCE, Gini Impurity was used to evaluate the error within the DT model. The formula of

$$G = 1 - \sum_{i=i}^{N} (p_i)^2$$

Gini Impurity Formula

In the Gini Impurity formula, N represents the number of classes and pi represents the homogeneity of the class in that subnode (Kaushik, 2023). Homogeneity would be calculated with the ratio between the number of instances of a specific class to the number of instances in an entire subnode. With this, the model would continue to modify the partitions to minimize the Gini Impurity within the model. With these models and statistical approaches, we were able to optimize the model's performance with our dataset.

Feature Selection

Feature selection and feature importance were implemented on these models to analyze trends lying within stroke, providing correlational knowledge that reinforces medical practices. Moreover, these two methods also help optimize the model, reducing overfitting and improving performance.

Feature selection is a process that chooses a subset of the existing features to create a more optimal model. There are three different types of methods for feature selection: Filter Methods, Wrapper Methods, and Embedded Methods. For the DNN

model, Recursive Feature Elimination (RFE), a type of Wrapper Method, is employed to find an optimal feature subset. RFE works by recursively removing the least important feature and then retraining the model; this is repeated for a predefined number of times or when no features are remaining. Then, RFE is able to select the best performing subset based on accuracy scores. By eliminating the less important features, RFE decreases the model's complexity, allowing for faster computation, and reduces noise –irrelevant data that doesn't contribute to an underlying pattern. From a medical standpoint, this means being able to discover stroke-prone patients with a higher accuracy and with less data —which costs time and money.

Feature importance is a technique that determines how relevant a feature is in determining the model's predictions. In the DT, feature importance was determined through the Gini Impurity Reduction. Gini Impurity Reduction is the decrease in Gini Impurity at each split of the tree when a feature is used to split the data into more homogenous groups. By summing up all the Gini Impurity Reductions within each split of the DT, the importance of each feature in the DT is found on a relativistic scale to each other, allowing us to compare the effects of each feature on the final output. In the DNN, feature importance was determined with a method called Permutation Feature Importance (PFI), which is typically used. To evaluate a feature's effects on the model's output, PFI randomly shuffles all the values within the feature, breaking the feature's effects on the model. If the feature was very important, a strong decline in the model's performance may be seen, as shown in Figure 3 below:

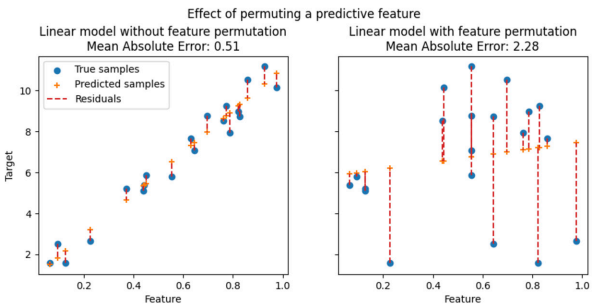

Figure 3: The above charts show how PFI would affect a feature that plays a significant role in determining the predictive outcome in a linear model. Image from ScikitLearn.

Prior to the PFI, the feature values were in a relatively linear relationship with the predicted outputs. However, after the PFI, the original established relationship is no longer applicable to the feature; thus, all the contributions of the feature towards the model are now misleading and inaccurate, leading to poorer performance. On the other hand, if the model's performance remained the same, then it can be inferred that the feature played little role in the model's predictions, as shown in Figure 4 below:

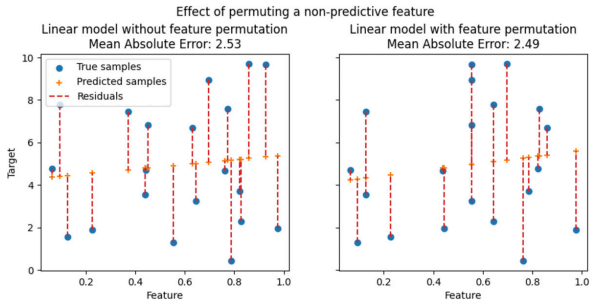

Figure 4: The above charts show how PFI would affect a feature that plays little role in determining the predictive outcome in a linear model. Image from ScikitLearn.

Prior to the PFI, the feature values were all over the place and didn't have a linear relationship. Thus, after the PFI, no real impact was made because they continued to not follow the linear trend. Although the trend isn't linear within a DNN, after shuffling, a similar effect will be seen as data become randomized within the feature. PFI is especially useful when it comes to deep learning models, where the patterns in the hidden layers may be opaque. With PFI, the importance of each feature can be evaluated, resulting in similar benefits as feature selection: lower complexity, reduces noise, and optimizes the model. By optimizing the model and analyzing it with feature selection and importance, an efficient and effective tool can be created to assist medical practitioners in discovering stroke-prone patients, mitigating the consequences of stroke.

Results

DNN #0:

To demonstrate the effects of an imbalance dataset, a model was constructed and its performance was evaluated. As illustrated, it can be seen that the model quickly came into an optimized position at around 40 steps, which is quite low for a deep learning model, as shown in Figure 5 below:

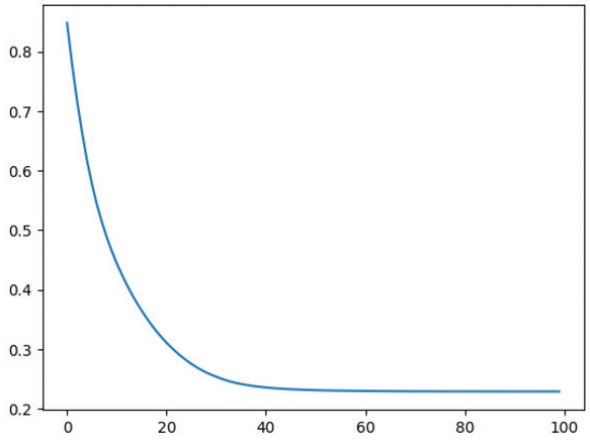

***Figure 5**: BCE loss graph for DNN #0.*

Values	Precision	Recall	F1 Score	Accuracy
0 (No Stroke)	0.93	1.00	0.96	0.93
1 (Stroke)	0.00	0.00	0.00	

***Table 1**: Performance scores of the model (DNN #0) with an imbalanced dataset.*

This would mean the model had a relatively low complexity. From this, along with the zero F1 score, as shown in Table 1, we can conclude that the model had simply predicted all cases to have no stroke in order to gain a high accuracy. To prevent this from occurring in future models, SMOTE was performed on the dataset, allowing for the creation of a more practical and realistic model.

DNN #1:

DNN #1 was a model with one hidden layer that had 2 nodes, a rather simple construction. The model's complexity is higher than DNN #0 because the graph settles down after around 150 steps, as shown in Figure 6 below. Although the model's F1 score is non-zero, which demonstrates the effectiveness of SMOTE, it has only achieved an accuracy of 69% (shown in Table 2 below), a low success rate in the field of medical practice.

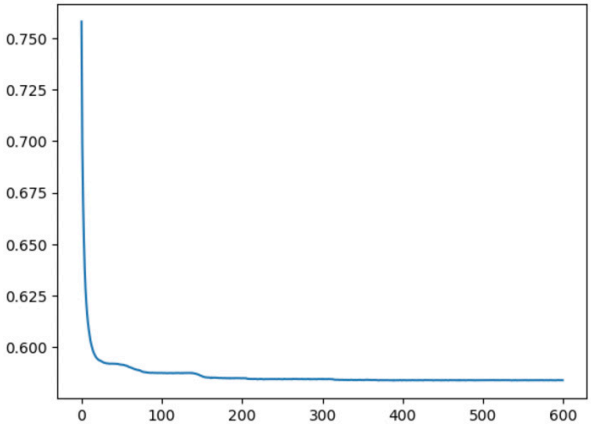

***Figure 6**: BCE loss graph of DNN #1.*

Values	Precision	Recall	F1 Score	Accuracy
0 (No Stroke)	0.70	0.69	0.69	0.69
1 (Stroke)	0.68	0.68	0.68	

***Table 2**: Performance scores of DNN #1.*

DNN #2:

The second model, a more complicated model, had 2 hidden layers with 12 nodes each. The first iteration was trained with 200 steps, resulting in a 67% accuracy. While the second iteration was trained with 600 steps obtaining a 71% accuracy. The model is far more complex, only settling down after 600 steps, as shown in figure 7 below:

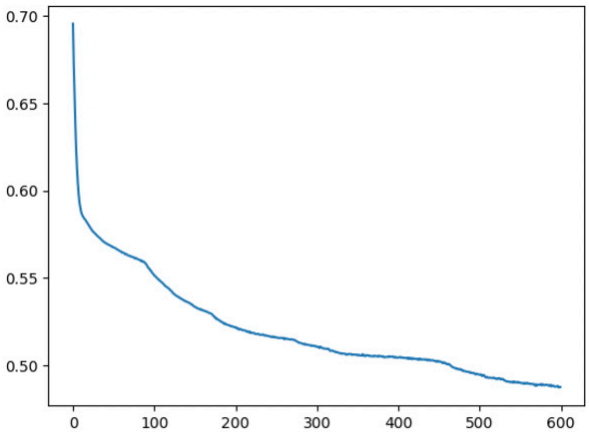

***Figure 7**: BCE loss graph for DNN #2.*

Values	Precision	Recall	F1 Score	Accuracy
0 (No Stroke)	0.72	0.70	0.71	0.71
1 (Stroke)	0.70	0.72	0.71	

***Table 3**: Performance scores of DNN #2.*

DNN #3:

DNN #3 was the product of conducting Recursive Feature Elimination (RFE) on DNN #2. The three selection features, ranked in order, are glucose level, BMI, and age. The eliminated features, ranked in order of importance, are gender, hypertension, unknown, smokes, never smoked, heart disease, residence type, marital status, and formerly smoked. The performance of DNN #3, which was the model constructed on the selected features, can be seen in Figure 8 and Table 4 below:

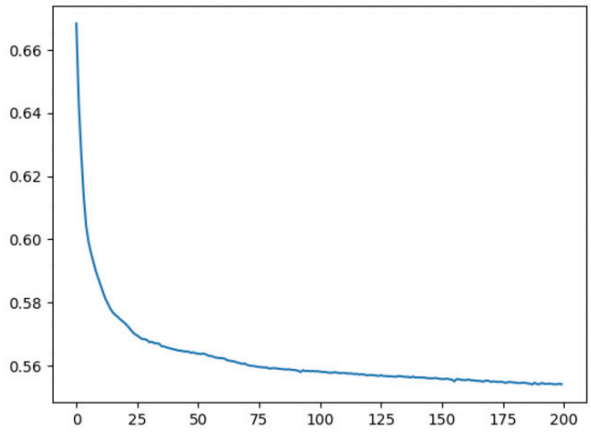

***Figure 8**: BCE loss graph of DNN #3.*

Values	Precision	Recall	F1 Score	Accuracy
0 (No Stroke)	0.72	0.72	0.72	0.71
1 (Stroke)	0.68	0.69	0.68	

***Table 4**: Performance scores of DNN #3.*

By looking at the loss graph, it can be seen that the model was able to reach a relatively stable point after 200 steps compared to DNN #2 that required 600+ steps, meaning that DNN#3 was less complex than DNN #2. Although feature selection did not produce better performances — with both models at a 71% accuracy—, it proved to be more efficient, achieving the same performance with less information, having 9 less features.

Permutation feature importance

Using the PFI on the optimized DNN #2, it was discovered that age, glucose level, and heart disease were the 3 most influential factors in the model's predictions, with age having a very significant impact compared to others —lowering the model's performance by more than 15%. The top three features, along with residence type, smoking, and hypertension, all impacted the model by more than 1%, with the remaining features having seemingly little impact, as shown in Figure 9 and Table 5 below:

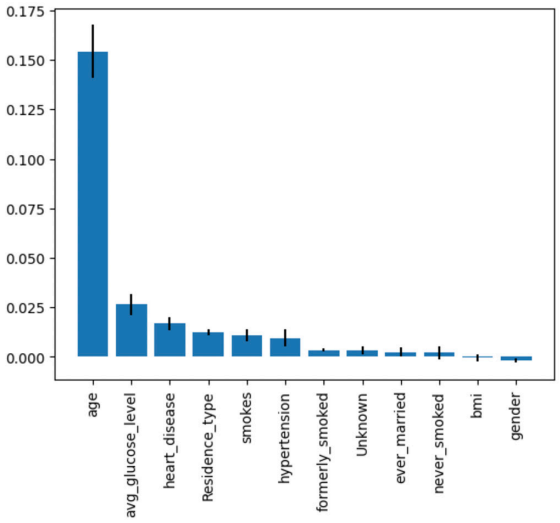

***Figure 8**: A bar graph of how PFI of each feature impacted mode performance.*

Feature Name	Performance Drop(%)	Error
Age	15.4	1.3
Glucose Level	2.6	0.5
Heart Disease	1.7	0.3
Residence Type	1.2	0.2
Smokes	1.1	0.3
Hypertension	1.0	0.4
Formerly Smoked	0.4	0.1
Unknown	0.3	0.2
Ever Married	0.3	0.2
Never Smoked	0.2	0.3
BMI	0.0	0.2
Gender	-0.2	0.1

***Table 5**: Performance drops of model evaluated through PFI on each feature.*

After evaluating feature importances on DNNs, Decision Trees were tested and optimized.

DT #1:

The DT model's accuracy far outperformed that of the DNN, achieving an outstanding 94.6% accuracy, as shown in Table 6 below, compared to the 71% of the DNN model.

Values	Precision	Recall	F1 Score	Accuracy
0 (No Stroke)	1.00	0.89	0.94	0.946
1 (Stroke)	0.90	1.00	0.95	

***Table 6**: Performance scores of DT #1.*

Then, by analyzing the Gini Impurity Reduction, the feature importances of the DT model was found. Glucose level, age, and BMI are seen to be the most significant factors, with remaining features having smaller impacts, as shown in Figure 9 and Table 7 below:

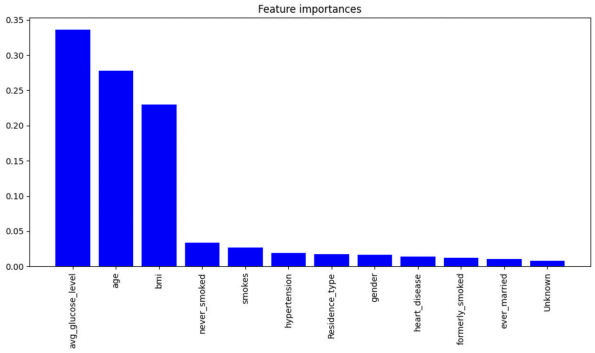

***Figure 9**: A bar graph of feature importances from DT #1. The scores add up to 1 and the top three features have significant differences.*

Feature Name	Importance Scores
Glucose Level	0.336
Age	0.277
BMI	0.230
Never Smoked	0.033
Smokes	0.027
Hypertension	0.019
Residence Type	0.017
Gender	0.016
Heart Disease	0.014
Formerly Smoked	0.012
Ever Married	0.010
Unknown	0.008

***Table 7**: Importance Scores of each feature from DT #1. Scores add up to 1.*

DT #2:

The top 3 features (age, glucose level, and BMI) from DT #1 were used to train DT #2 to evaluate the effects of feature selection. The performance surpassed that of the previous model by 0.8% in accuracy and better results in both precision and recall, as shown in Table 8 below:

Values	Precision	Recall	F1 Score	Accuracy
0 (No Stroke)	1.00	0.91	0.95	0.954
1 (Stroke)	0.92	1.00	0.96	

***Table 8**: Performance scores of DT #2.*

Discussion

Both models proved to be successful in bettering the chances of spotting stroke-prone patients. However, the DT model outperformed the DNN model by a significant margin, boasting a 94.6% accuracy compared to a 71% accuracy. Despite both being non-linear models, how was the DT model able to far exceed the accuracy of the DNN model?

There may be several reasons that can explain this phenomenon. First, DTs are typically better than DNNs on datasets that have less than 10,000 cases because DNNs typically need huge datasets

to generalize well and to avoid overfitting. In this dataset, which had around 3000 cases, a DT would typically perform better than a DNN because of its simplicity — thus, being less prone to overfitting. Second, DTs require much less tuning than DNNs. Aside from feature selection, the DNN model wasn't fine tuned with regularization, learning rates, and data augmentation. Thus, it was far more prone to overfitting and wasn't fully optimized. On the other hand, the DT, on a structural scale, is simpler and more compatible with the small dataset. Thus, even though the DT wasn't pruned or modified with any ensemble methods, it still achieved an outstanding performance. Lastly, the optimization process for the DT model was more straightforward than that of the DNN. The interpretability of a DT is much better than that of the DNN; its decision paths are clear mathematical boundaries that can be seen. However, the DNN is a black box; there is no clear way to interpret what is going on inside and between each hidden layer. Information can only be gained through manipulating the inputs, outputs, and certain methods, such as permutation feature importance. Thus, it is much harder to discover and solve issues that cause the model to underperform. Due to the DT model's better compatibility with the dataset, it was able to achieve significant success compared to the DNN.

The results obtained from performing feature importance and feature selection in both DNNs and DTs hold significance in medical information and practicality.

In the DNN models, the performance of DNN #2 —the original model— was the same as the performance of DNN #3 —the original model after performing RFE. Both DNNs achieved an accuracy of 71%. Despite DNN #3 only having 3 features — being 'age', 'glucose level', and 'BMI' — for input, it achieved the same performance as DNN #2, which had 12 features. This implies that a significant number of features were noise and provided no additional insight into the assessing the likelihood of stroke in a patient. When this is applied to real-life practical models, it would mean that less data would need to be collected, reducing the resources required and making everything more convenient.

However, after PFI was applied on DNN #2, a different feature ranking was achieved than that of the RFE. In the PFI, it was found that age was a very significant influence (15.6%) with the next three features being glucose level (2.6%), heart disease (1.7%), and residence type (1.2%). This, however, is different from the top 4 features that were found through RFE —which were age, BMI, glucose level, and gender. With age and glucose level being ranked top features in both assessments, it shows how the model is dependent on these two features, showing their importance. However, the rankings of the remaining features were conflicting between the two methods. Although heart disease and residence type was ranked third and fourth with PFI, they were ranked ninth and tenth in RFE, respectively. Although BMI and gender were ranked second and fourth in RFE, they were ranked second last and last in PFI. The incongruence between the two rankings demonstrate that both results may be inconclusive. A possible explanation could be how the model itself is underperforming. With a 71% accuracy, the model's behaviors and assessed feature rankings may not be closely aligned with that of real life. Thus, the underperformance of the model, in conjunction with the dissimilar rankings, implies that the top ranked features may be inaccurate and would require further assessment.

Contrary to the DNN, the DT model had performed exceptionally well, with the base model (DT #1) achieving a 94.6% accuracy. Thus, the feature importance rankings generated from DT #1 should be more reliable than the ones from the DNN. The feature importance ranking showed three significant features: glucose level (0.336), age (0.277), and BMI (0.230). Due to the interpretability of DTs and this model's accurate performance, the results from this feature importance is promising and demonstrates important medical values: how age, glucose level, and BMI are very important aspects in influencing stroke. In addition, this ranking list is quite similar to that from DNN #2 with RFE; this would mean that the ranking generated from PFI was less reliable and inaccurate, potentially implying the inefficacy of PFI. However, since DNN #2 underperformed, the accuracy of the rankings may have been arbitrary, meaning that no conclusive comparisons may be made between PFI and RFE.

With the information of the top three features from the DT #1, DT #2 was built with only the top three features and achieved a higher performance than DT #1 with an accuracy of 95.4%. This aligns

with the expectations for feature selection: lowering model complexity and improving model performance. DT #2 cut down 9 features and was able to perform 0.8% better than DT #1, demonstrating the effects of feature selection. In addition, the improved performance implies that the nine removed features were noise; although other features, in reality, may influence stroke, they are likely less significant and more random, creating noise and causing the model to underperform.

Conclusion

By manipulating and fine tuning the models, two optimized models, a DNN and a DT, were produced. The methods of feature selection and importance on both models proved to grant benefits: reducing model complexity and improving model performance. On the DNN, the feature input lowered from 12 to 3 while maintaining the same accuracy of 71%. On the DT, the feature input lowered from 12 to 3 while improving the accuracy from 94.6% to 95.4%. Not only did feature selection improve model importance, it provided important insight into the medical aspect of stroke. The results, from both feature rankings from the DNN with RFE and the DT with Gini Impurity Reduction, support that age, glucose level, and BMI are the major influencers of stroke, in comparison with the remaining features within the dataset; This information is also supported by scientific research. The results from this study may build groundwork in the real world in several beneficial ways. The combination of a stroke detection model to nutrition or health tracking apps allows for the add-on to existing products, promoting both client safety and app utility. In addition, the results of feature importance from this research informs the medical forefront of the essential features, allowing researchers and data collectors to save resources on gathering information on less significant factors. Moreover, the model can serve as groundwork for the creation of a model that can detect risks of all kinds of illnesses and conditions from statistics; this lays potential to expand medical attention towards areas without healthcare access by introducing an app or unmanned medical stations. The potential for expansion on this project is boundless. With more work to be done, increasing medical access is within grasp.

References

Centers for Disease Control and Prevention. (n.d.). Preventing stroke. Centers for Disease Control and Prevention. https://www.cdc.gov/stroke/prevention/index.html

Daidone, M., Ferrantelli, S., & Tuttolomondo, A. (2023). Machine learning applications in stroke medicine: Advancements, challenges, and future prospectives. Neural Regeneration Research, 19(4), 769–773. https://doi.org/10.4103/1673-5374.382228

Global, regional, and country-specific lifetime risks of stroke, 1990 and 2016. (2018). New England Journal of Medicine, 379(25), 2429–2437. https://doi.org/10.1056/nejmoa1804492

Hammond, G., Luke, A. A., Elson, L., Towfighi, A., & Joynt Maddox, K. E. (2020). Urban-rural inequities in acute stroke care and in-hospital mortality. Stroke, 51(7), 2131–2138. https://doi.org/10.1161/strokeaha.120.029318

Health disparities among the many unique challenges for people in Rural America. American Heart Association. (n.d.). https://newsroom.heart.org/news/health-disparities-among-the-many-unique-challenges-for-people-in-rural-america

John Hopkins. (n.d.). Stroke. Johns Hopkins Medicine. https://www.hopkinsmedicine.org/health/conditions-and-diseases/stroke

Kaushik, A. (2023, June 17). Gini impurity and entropy for decision tree. Medium. https://medium.com/@arpita.k20/gini-impurity-and-entropy-for-decision-tree-68eb139274d1

Palacios, F. S. (2021, January 26). Stroke prediction dataset. Kaggle. https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset/data

Satpathy, S. (2024, July 5). Smote for imbalanced classification with python. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/10/overcoming-class-imbalance-using-smote-techniques/

Sorokin, M. (2024, January 17). Binary cross-entropy: Mathematical insights and python implementation. Medium. https://medium.com/@vergotten/binary-cross-entropy-mathematical-insights-and-python-implementation-31e5a4df78f3

U.S. Department of Health and Human Services. (n.d.). Causes and risk factors. National Heart Lung and Blood Institute. https://www.nhlbi.nih.gov/health/stroke/causes

Wong, C. W., Kwok, C. S., Narain, A., Gulati, M., Mihalidou, A. S., Wu, P., Alasnag, M., Myint, P. K., & Mamas, M. A. (2018). Marital status and risk of cardiovascular diseases: A systematic review and meta-analysis. Heart, 104(23), 1937–1948. https://doi.org/10.1136/heartjnl-2018-313005

World Health Organization. (n.d.-a). Long working hours increasing deaths from heart disease and stroke: Who, ilo. World Health Organization. https://www.who.int/news/item/17-05-2021-long-working-hours-increasing-deaths-from-heart-disease-and-stroke-who-ilo

World Health Organization. (n.d.-b). Stroke, Cerebrovascular accident. World Health Organization. https://www.emro.who.int/health-topics/stroke-cerebrovascular-accident/index.html

World Stroke Organization. (n.d.). Impact of stroke. World Stroke Organization. https://www.world-stroke.org/world-stroke-day-campaign/about-stroke/impact-of-stroke