IRSTI 28.23.01 https://doi.org/10.26577/jpcsit2024-02i03-03

A. Zhalgas , M. Toleubek*

Astana IT University, Astana, Kazakhstan
*e-mail: moldir.toleubek@astanait.edu.kz

A COMPARATIVE ANALYSIS OF MACHINE LEARNING CLASSIFIERS FOR STROKE PREDICTION

Abstract. Stroke, a major global health concern, is characterized by sudden neurological deficits and impaired cerebral function. Advancements in technology and the integration of medical records offer opportunities to enhance stroke care and diagnosis. By mining and analyzing electronic health records, valuable insights into the interdependencies of stroke risk factors can be gained, aiding in prediction. This research provides a comprehensive review of the application of machine learning classifiers in stroke prediction, considering various techniques, features, and performance measures utilized in previous studies. The novelty of this work is to emphasize the potential of machine learning classifiers in improving stroke prediction, with a focus on feature selection, data pre-processing, and model evaluation. The aim is to shed light on the strengths and limitations of different classifiers, including Decision Trees, AdaBoost, and Gradient Boost, and their performance metrics in stroke prediction. By achieving this goal, effective stroke risk assessment models can be developed, leading to improved patient outcomes through early intervention and targeted preventive measures. The findings reveal that machine learning classifiers, particularly AdaBoost and Gradient Boost, show promising performance in stroke prediction. These classifiers demonstrate high recall rates and balanced F1 scores, indicating their efficacy in identifying individuals at risk of stroke. This research contributes to the growing body of knowledge in stroke prediction and highlights the potential of machine learning techniques in enhancing stroke care. The integration of machine learning classifiers with stroke prediction holds great promise in improving patient outcomes. By harnessing the power of electronic health records and utilizing appropriate techniques and features, healthcare providers can enhance their ability to identify and intervene in stroke cases, ultimately leading to better preventive measures and care strategies.

Key words: classification, statistical analysis, Decision Tree, AdaBoost, Gradient Boost.

1. Introduction

Stroke is an acute cerebrovascular disease characterized by focal neurological deficits and global cerebral impairments. It develops suddenly as a result of disrupted cerebral blood circulation, lasting for at least 24 hours. The care and diagnosis of strokes could be improved with technological improvements and the integration of medical records. Caregivers can learn important information about the interdependency of risk variables for stroke prediction by methodically mining and analyzing electronic health records. Stroke accounted for roughly 5.5% of all deaths worldwide in 2019 [1], according to the Global Burden of Disease Research. Notably, in the Southeast Asian and Western Pacific areas, stroke accounted for more than 40% of all noncommunicable disease fatalities [2]. Hypertension, diabetes, smoking, obesity, physical inactivity, and a poor diet are all common risk factors for stroke. Stroke imbalances are worsened by socioeconomic variables such as low education and restricted access to healthcare [3]. For example, O'Donnell et al. [4] discovered that modifiable variables account for 90% of the global risk of stroke. Prevention techniques emphasize risk factor reduction through public health programs and community-based interventions. Stroke prevention has benefited from efforts to manage hypertension, encourage smoking cessation, and enhance nutrition and physical exercise [5]. Stroke has a substantial worldwide health impact, with regional differences in incidence, risk factors, preventative initiatives, and therapeutic techniques. Implementing effective stroke preventive programs and increasing access to high-quality care are critical to lowering the global burden of disease. In such circumstance, machine learning can play a critical role in accurately and efficiently predicting at a lesser cost. For many years, several machine learning classifiers have been utilized in medical disciplines to perform proper analyses and predict appropriate outcomes based on patterns in large, unbalanced datasets.

Machine learning techniques have showed promise in the realm of healthcare due to their ability to anticipate and detect disorders such as stroke. The purpose of this literature review is to examine the use of machine learning classifiers in stroke prediction by analyzing the various techniques, features, and performance measures used in previous researches. Several machine learning classifiers, including Logistic Regression, Support Vector Machines (SVM), Random Forests, Artificial Neural Networks (ANN), and Gradient Boosting Algorithms, have been employed in stroke prediction [6,7]. Each classifier has advantages and disadvantages that determine its appropriateness for stroke prediction.

Demographic data, medical history, clinical exams, laboratory test results, and neuroimaging findings have all been used in studies to predict stroke. Some researches [8] have also included genetic data and lifestyle variables. Model performance and generalizability are heavily reliant on feature selection and preprocessing. Machine learning classifier performance in stroke prediction is often measured using measures such as accuracy, sensitivity, specificity, area under the receiver operating characteristic curve (AUC-ROC), and F1 score. To test model generalizability, cross-validation approaches such as k-fold cross-validation are often used. External validation with separate datasets is required to validate the classifiers' robustness. Machine learning classifiers' effectiveness in stroke prediction differs among researches. Several classifiers, on the other hand, routinely obtain high accuracy and AUC-ROC values. SVM, Random Forests, and ANN, for example, have shown promising results in reliably predicting stroke risk based on various datasets and feature combinations [9]. The specific performance of each classifier is determined by the dataset properties, feature quality, and study population.

The Decision Tree classifier is a well-known and effective machine learning technique that may be used for both regression and classification applications [10]. It is a non-parametric supervised learning approach that builds a tree-like model to predict outcomes based on a set of decision rules derived from training data. It partitions the feature space recursively based on the values of input features. Each partitioning is established by picking the most informative characteristic that divides the data the best. This procedure is continued until a stopping requirement, such as a maximum tree depth or a minimum number of samples in each leaf node, is fulfilled [11].

Adaptive Boosting (AdaBoost) is a well-known and effective ensemble learning approach for classification tasks. It is very useful in integrating weak classifiers into a strong classifier to improve their performance. Yoav Freund and Robert Schapire invented AdaBoost in 1996 [12]. The main idea of the classifier is to train several weak classifiers consecutively, with each successive classifier focusing on misclassified examples from prior classifiers. AdaBoost's ability to adaptively focus on difficult examples during training iterations can enhance stroke detection performance [13]. By combining the predictions of multiple weak classifiers, it can effectively capture complex relationships between features and the presence of stroke.

Another powerful classifier is the Gradient Boost, which is extension of AdaBoost. Gradient Boosting, as opposed to AdaBoost, works by successively adding weak classifiers to minimize a loss function by gradient descent [14]. The approach iteratively fits the weak classifiers to the loss function's negative gradient, increasing the overall model's performance with each iteration.

Our study aims to provide a comparative analysis of various stroke prediction factors and techniques by three different models as Decision Tree, AdaBoost, Gradient Boost. Above mentioned machine learning techniques will be used to look for fresh insights into stroke risk factors and develop trustworthy predictive models. Therefore, these findings contribute to the existing body of knowledge and provide new avenues for early stroke detection and prevention. By statistical analysis approach there can be visible decisions to significantly improve patient outcomes and reduce the burden of stroke on people and healthcare systems.

2. Materials and Methods

The overview of the proposed methodology is illustrated in Fig. 1. The dataset is pre-processed in the first phase. After conducting statistical analysis on a clean dataset, the subsequent findings and results provide valuable insights into the relationships and patterns within the data, emphasizing the importance of data cleanliness in ensuring the accuracy and reliability of statistical analyses. In the second phase, the pre-processed dataset is fed into multiple machine learning algorithms, enabling the exploration and application of various computational techniques to extract meaningful patterns, predict outcomes, and derive insights from the data. The output of the models is then examined using various metrics in the third phase. In the last step, the model with the best accuracy is identified.

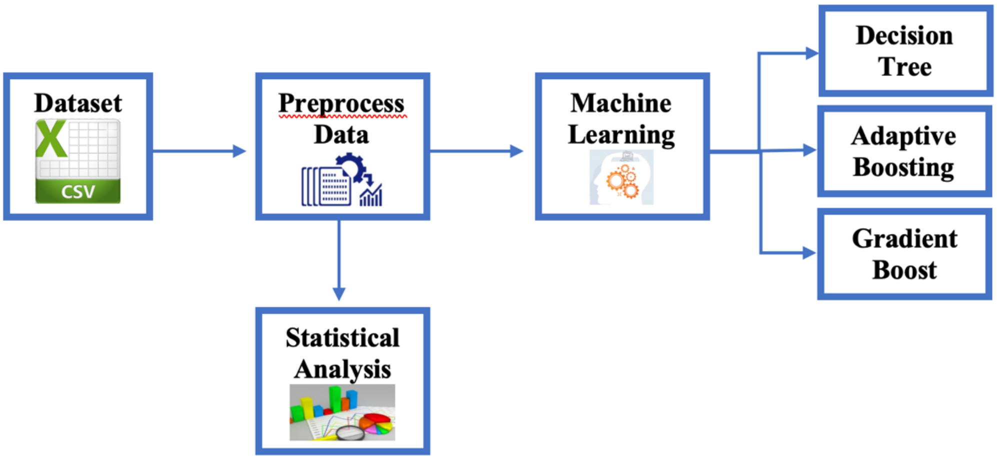
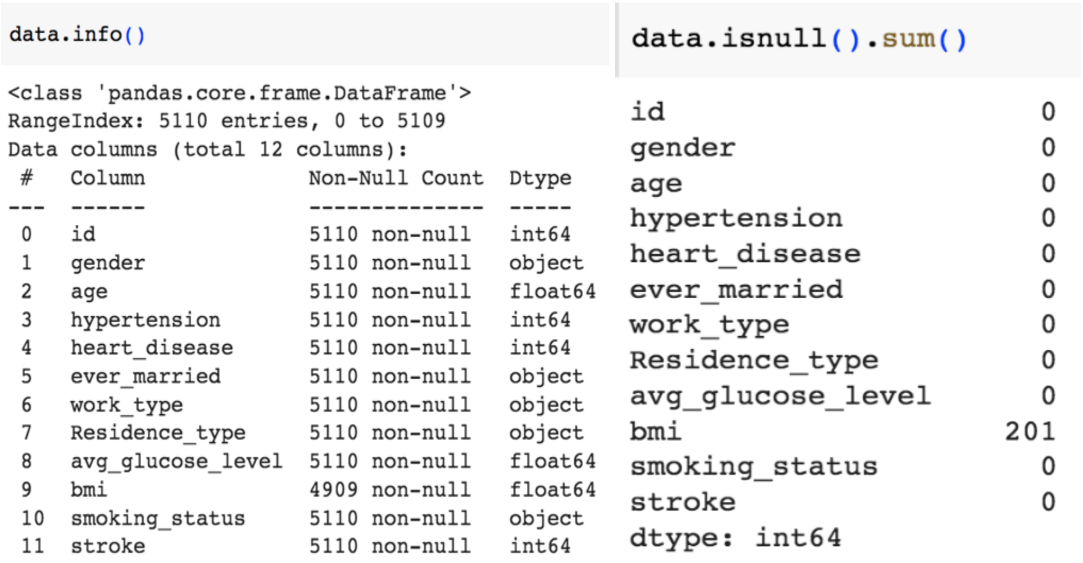


Figure 1 – Overview of the proposed methodology

The dataset includes medical risk variables that have been linked to stroke. These factors include, among others, hypertension, cardiac disease, smoking status, and body mass index (BMI), age, etc. The existence of missing BMI values in the dataset highlights the difficulties in data collecting and data quality. Missing data can be caused by a variety of circumstances, including incomplete survey replies, technological problems during measurements, or participant noncompliance. In this situation (Fig. 2) the dataset had 201 missing BMI values, which had to be handled in order for the analysis to be accurate. During the preprocessing stage, the missing BMI data were replaced with a value of 0. While this strategy appears to be simple and easy, it is critical to recognize the potential consequences.



```
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5110 entries, 0 to 5109
Data columns (total 12 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   id                 5110 non-null   int64
 1   gender             5110 non-null   object
 2   age                5110 non-null   float64
 3   hypertension       5110 non-null   int64
 4   heart_disease      5110 non-null   int64
 5   ever_married       5110 non-null   object
 6   work_type          5110 non-null   object
 7   Residence_type     5110 non-null   object
 8   avg_glucose_level  5110 non-null   float64
 9   bmi                4909 non-null   float64
 10  smoking_status     5110 non-null   object
 11  stroke             5110 non-null   int64
```

(a)

```
data.isnull().sum()

id                     0
gender                 0
age                    0
hypertension           0
heart_disease          0
ever_married           0
work_type              0
Residence_type         0
avg_glucose_level      0
bmi                  201
smoking_status         0
stroke                 0
dtype: int64
```

(b)

Figure 2 – (a) Information about dataset; (b) missing data

There are several steps involved for each classification method after data pre-processing. The process of constructing a decision tree involves several key steps. Firstly, feature selection is performed to identify the most relevant features, enhancing accuracy and interpretability. Next, the dataset is partitioned into a training set and a testing set for model evaluation. The decision tree is then constructed recursively, selecting the best features based on a specified criterion. Pruning techniques can be applied to address overfitting, improving the tree's predictive performance. Visualization of the decision tree provides insights into decision rules and feature hierarchy. Evaluation metrics are employed to assess the model's performance on the testing set. Hyperparameters are refined to optimize the decision tree through techniques like cross-validation or grid search. Finally, the trained and refined decision tree can be deployed for prediction on new data, employing the learned decision rules. This comprehensive process ensures the effective utilization of decision trees for classification tasks.

The AdaBoost algorithm involves a series of steps to train and utilize weak classifiers for stroke detection. Initially, weak classifiers such as decision stumps or trees are selected as base models and trained on equally weighted subsets of the data. The AdaBoost algorithm is then applied iteratively, where each weak classifier is trained on a weighted training set and evaluated based on its performance. Misclassified samples receive higher weights, while correctly classified samples receive lower weights. The predictions of multiple weak classifiers are combined through weighted voting, with each weak classifier's weight determined by its performance. The final prediction is made by aggregating the weighted predictions. The trained classifier is evaluated on a separate testing set using various evaluation metrics like accuracy, precision, recall, F1 score, and AUC-ROC. Hyperparameters, such as the number of weak classifiers and the learning rate, can be tuned using techniques like cross-validation or grid search to optimize the model's performance. Additionally, AdaBoost provides insights into feature importance through the weights assigned to weak classifiers, allowing for interpretation of the model and identification of influential features in stroke prediction.

The Gradient Boosting algorithm follows a sequential process to build an ensemble of weak classifiers. Initially, a simple model like a decision stump or shallow decision tree is initialized as the first weak classifier. The weak classifier is then trained on the training data, and its predictions are computed. Next, the residuals, which represent the differences between the predicted and actual target values, are calculated for each sample in the training set. Subsequently, additional weak classifiers are added to the ensemble one by one. Each new classifier is fitted to the negative gradient (residuals) of the loss function, focusing on the samples that were not well predicted by the previous classifiers. The new weak classifier is integrated into the ensemble by combining its predictions with the predictions of the previous weak classifiers, weighted by the learning rate. This iterative process continues for a specified number of iterations or until a stopping criterion is met. Finally, the final prediction is made by aggregating the predictions of all the weak classifiers in the ensemble.

Model evaluation for stroke prediction with 2 classes using machine supervised learning involves assessing the performance of the model on the test set using appropriate metrics [15]. Confusion matrix is a table that shows the number of true positives, true negatives, false positives, and false negatives for each class. A confusion matrix for multi-class classification with labels 0 (non-stroke) and 1 (stroke) would look like Fig. 3. True positives (TP) are the number of samples that are correctly classified as positive for a particular class. For example, the number of samples that are actually has stroke (label 1) and are correctly classified as stroke patients (predicted label 1). True negatives (TN) are the number of samples that are correctly classified as negative for a particular class. For example, the number of samples that are actually has no stroke (label 0) and are correctly classified as not being stroke patient (predicted label not 1). False positives (FP) are the number of samples that are incorrectly classified as positive for a particular class. For example, the number of samples that are actually healthy (label 0) but are misclassified as stroke patient (predicted label 1). False negatives (FN) are the number of samples that are incorrectly classified as negative for a particular class. For example, the number of samples that are actually stroke patients (label 1) but are misclassified as not having stroke (predicted label not 1).

The TP, TN, FP, and FN values can be used to calculate various performance metrics, such as precision, recall, and F1 score, for each class. Precision is the proportion of correctly predicted positive samples out of all predicted positive samples. It is useful when the cost of false positives is high. Precision calculated according to the following formula:

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

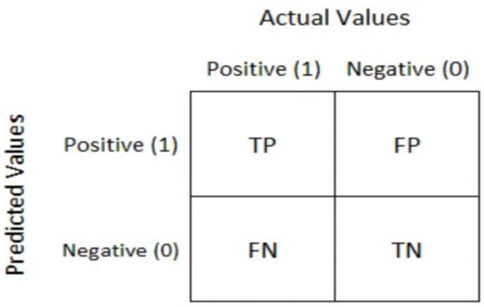


Figure 3 – Confusion Matrix

The TP, TN, FP, and FN values can be used to calculate various performance metrics, such as precision, recall, and F1 score, for each class. Precision is the proportion of correctly predicted positive samples out of all predicted positive samples.

F1 score is the harmonic mean of precision and recall and is useful for imbalanced datasets where both false positives and false negatives are important. F1 score calculated according to the following formula:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN}$$

2.1. Statistical Analysis

The qualitative and quantitative analysis were evaluated for open access stroke disease dataset which contains various medical and lifestyle features of individuals, including age, hypertension, heart disease, smoking status, and more. The data itself is a sample of big population and used to generalize the whole population. Understanding the intricate relationships between age, gender, geography, and various health factors in stroke occurrences is crucial for developing targeted preventive measures and improving patient outcomes. The findings discussed in this article provide valuable insights into these connections, enabling healthcare professionals and researchers to refine strategies that aim to reduce the burden of strokes on individuals and communities. By prioritizing stroke prevention efforts and implementing tailored interventions, we can make significant strides in minimizing the impact of strokes and improving overall public health. Correlations between features were obtained for multivariate analysis of values. As it is seen from Fig. 4 age, hypertension and average glucose level affected to disease, whereas body mass index had less contribution to stroke. Correlation obtained by heatmap function only shows linear correlation, therefore the values could be smaller than expected.

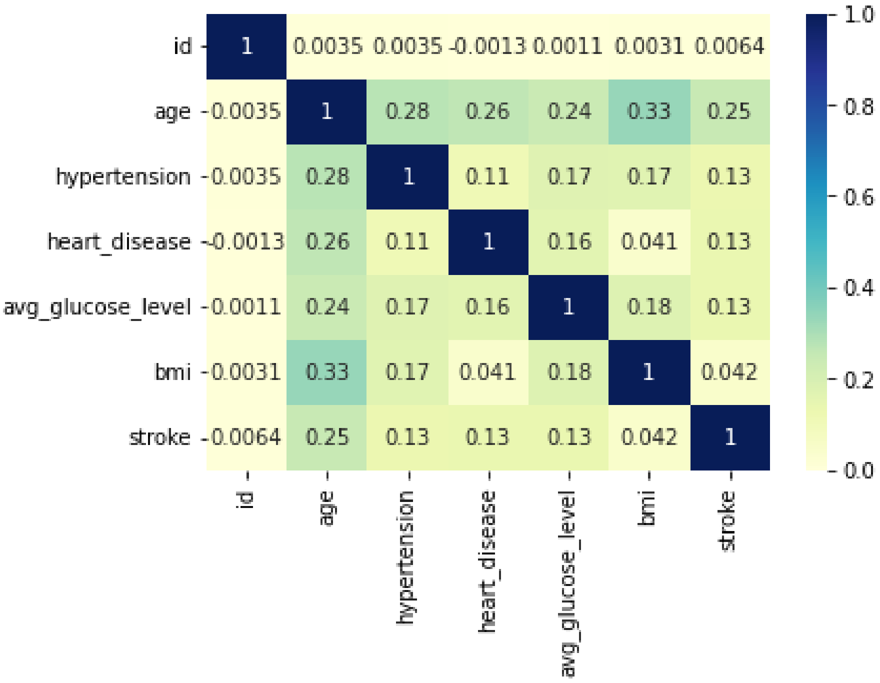


Figure 4 – A Correlation Heatmap

Individuals who have had a stroke are frequently beyond the age of 40, with the average age being 67.73, according to the statistics shown in the figure. Furthermore, a close analysis of the right and left graphs reveals a continuous pattern showing that the risk of stroke increases with age. Notably, those aged 70-80 had the largest number of stroke cases, accounting for 88 people, or roughly 35% of all stroke cases. Fig. 5 shows a strong relationship between age and the chance of having a stroke. As you become older, your chances of having a stroke increase. This research emphasizes the significance of age as a risk factor for stroke. With a mean age of 67.73 years for stroke patients, it is clear that strokes mostly impact people in their late 60s and beyond. This information is useful for healthcare professionals since it highlights the need for specific preventative actions and medications for older people.

Recognizing the significant influence of age on stroke prevalence is critical for healthcare practitioners and policymakers. Understanding the constant pattern depicted in the image allows for the development of age-specific measures to reduce the risk of stroke among the elderly. Implementing preventative measures, such as promoting healthy lifestyles and frequent medical check-ups, may be particularly beneficial for those in their 70s and 80s, who account for the majority of stroke occasions.

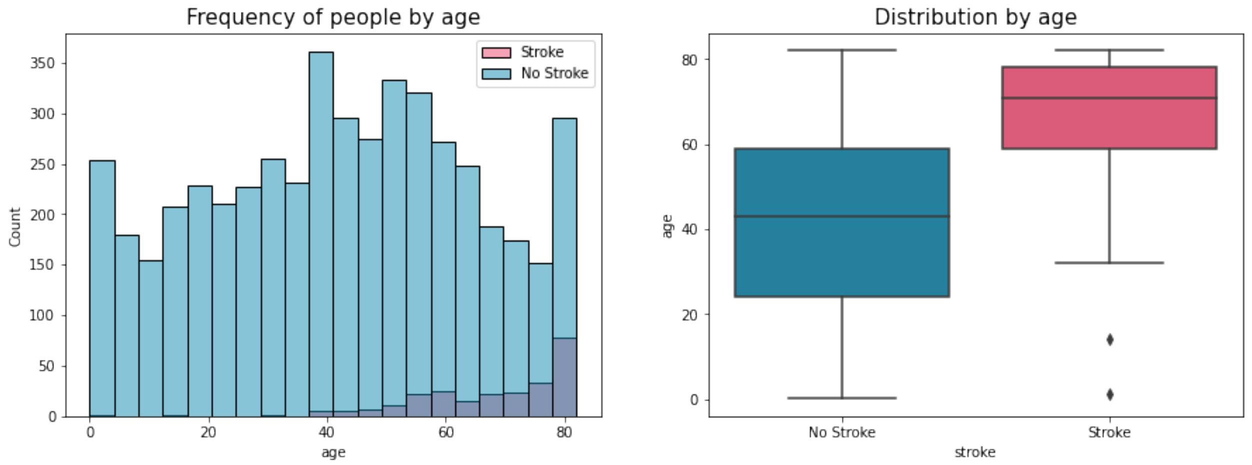


Figure 5 – Relation between age and stroke

Fig. 6 demonstrates an intriguing gender-based finding about the age at which people are at danger of having a stroke. According to the research, the majority of stroke cases in men occur after the age of 40. Women, on the other hand, have had strokes before the age of 40 in certain circumstances (specifically, 8 occurrences).

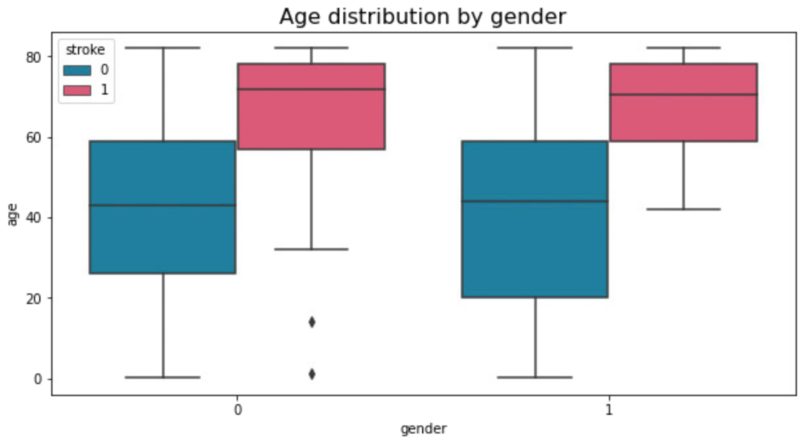


Figure 6 – Age distribution by gender for stroke and non-stroke patients

A comparable structure emerges in Fig. 7, which represents average BMI levels. Individuals with and without a history of stroke had comparable fluctuation patterns. The most common BMI range recorded in both groups is between 25 and 35. This shows that people in this age group are more common in this study's population, without regard for stroke incidence.

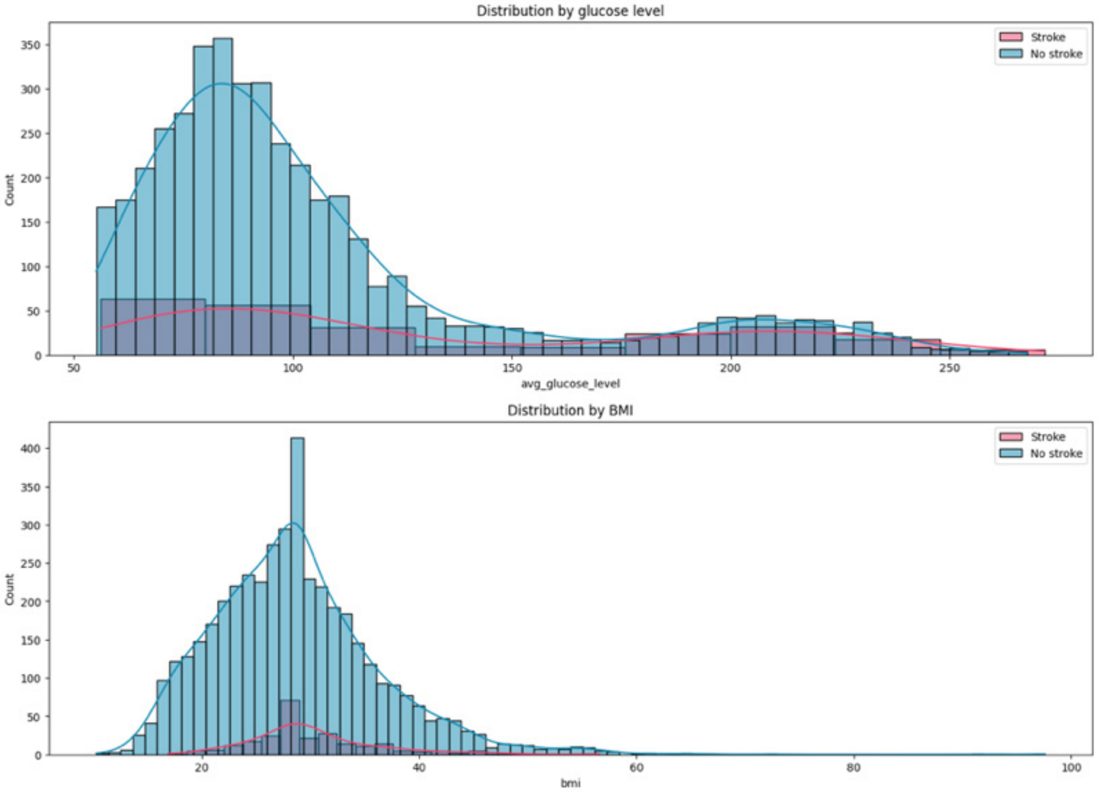


Figure 7 – Distribution of stroke and non-stroke patients by glucose level and BMI

3. Results

The dataset was divided into training and testing sets by 75%/25% ratio to evaluate the classifiers' performance.

The Decision Tree classifier achieved 0.95 precision, 0.96 recall, and 0.95 F1 score. AdaBoost and Gradient Boost, on the other hand, generated better recall scores of 0.99 but maintained precision at 0.95. Both AdaBoost and Gradient Boost had an F1 score of 0.97, showing a balanced performance between precision and recall.

Although effective with excellent precision and recall, the Decision Tree classifier falls somewhat short of AdaBoost and Gradient Boost. Both AdaBoost and Gradient Boost had excellent recall rates of 0.99, indicating their capacity to properly detect stroke patients. This is particularly crucial in stroke prediction, where the objective is to decrease false negatives and identify potential dangers as early as possible.

In Fig. 8 examined the performance of AdaBoost, Gradient Boosting, and Decision Tree classifiers using precision-recall curve analysis. AdaBoost and Gradient Boosting consistently outperformed, with comparable patterns in the precision-recall curves. The Decision Tree classifier, on the reverse, reacted differently, preferring stronger recall at the expense of poorer precision. The findings may be used to help them choose a classifier, taking into account the trade-offs between precision and recall according to the unique needs of their classification tasks.

Table 1 – Performance of proposed model

Model	Precision	Recall	F1 score
Decision Tree	0.95	0.96	0.95
AdaBoost	0.95	0.99	0.97
Gradient Boost	0.95	0.99	0.97

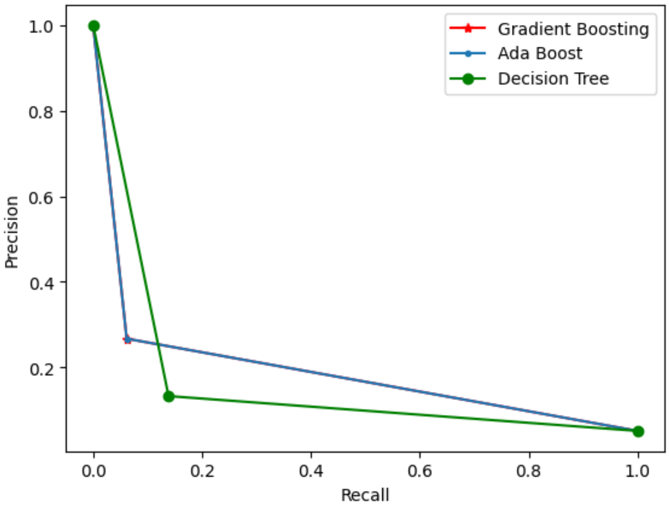


Figure 8 – Precision-Recall curves for all classifiers

4. Discussion

Furthermore, AdaBoost and Gradient Boost's F1 scores represent a harmonic combination of accuracy and recall, suggesting their capacity to produce correct predictions while minimizing both false positives and false negatives. This balanced performance is critical in stroke prediction models since both underestimation and overestimation of risk can have disastrous consequences.

Addressing problems and confirming the efficacy of these classifiers in real-world clinical situations, on the other hand, is critical. Continued research and development in this sector have the potential to transform stroke risk assessment and patient outcomes.

5. Conclusions

This research work based on stroke dataset provided a comprehensive findings according to statistical analysis and classification task. During the correlation analysis it was observed that age, hypertension, heart disease has more affection on stroke. On top of that, age and BMI demonstrated remarkable dependence. By exploratory data analysis, the average age of patients having stroke is 67-68 ears. As the analysis showed, AdaBoost and Gradient Boosting classifiers examined higher performance in contrast to Decision Tree classifier. As it is known, Gradient Boosting is modified version of AdaBoost, therefore their similar results are understandable. Machine learning classifiers have demon-strated promise in stroke prediction by combining various features and powerful methods. They can help with stroke risk assessment, allowing for early intervention and preventative actions.

Funding

This research received no external funding.

Author Contributions

Conceptualization, Methodology, Investigation,

Writing Original Draft Preparation, Project Administration, Funding Acquisition, A.Z.; Validation, Data Curation, Visualization, M.T.; Software, Resources, Supervision, Review & Editing, M.T., A.Z.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. GBD 2019 Diseases and Injuries Collaborators. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Neurology*, 396(10258), 1204–1222. https://doi.org/10.1016/s0140-6736(20)30925-9

2. Feigin, V. L., Roth, G. A., Naghavi, M., Parmar, P., Krishnamurthi, R., Chugh, S. S., Mensah, G. A., Norrving, B., Shiue, I., Ng, M., Estep, K., Cercy, K., Murray, C. J. L., & Forouzanfar, M. H. (2016). Global burden of stroke and risk factors in 188 countries, during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet Neurology*, *15*(9), 913–924. https://doi.org/10.1016/s1474-4422(16)30073-4.

3. Johnson, C. L., Nguyen, M. T., Roth, G. A., Nichols, E., Alam, T., Abate, D., Abd-Allah, F., Abdelalim, A. A., Abraha, H. N., Abu-Rmeileh, N. M. E., Adebayo, O., Adeoye, A. M., Agarwal, G., Agrawal, S., Aichour, A. N., Aichour, I., Aichour, M. T. E., Alahdab, F., Ali, R., . . . Javanbakht, M. (2019). Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurology*, *18*(5), 439–458. https://doi.org/10.1016/s1474-4422(19)30034-1

4. O'Donnell, M., Chin, S. L., Rangarajan, S., Xavier, D., Liu, L., Zhang, H., Rao-Melacini, P., Zhang, X., Pais, P., Agapay, S., Lopez-Jaramillo, P., Damasceno, A., Langhorne, P., McQueen, M. J., Rosengren, A., Dehghan, M., Hankey, G. J., Dans, A. L., Elsayed, A., . . . Yusuf, S. (2016). Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. *The Lancet*, 388(10046), 761–775. https://doi.org/10.1016/s0140-6736(16)30506-2

5. Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., De Ferranti, S. D., Després, J., Fullerton, H. J., Howard, V. J., Huffman, M. D., Judd, S. E., Kissela, B. M., Lackland, D. T., Lichtman, J. H., Lisabeth, L. D., Liu, S., Mackey, R. H., Matchar, D. B., Turner, M. (2015). Heart Disease and Stroke Statistics—2015 Update. *Circulation*, 131(4). https://doi.org/10.1161/cir.0000000000000152

6. Zeng, Y., & Liu, X. (2019). Machine learning-based stroke prediction models: A systematic review. *Journal of Healthcare Engineering*, 2019, 8582735. doi:10.1155/2019/8582735

7. Kim, M., & Kim, J. H. (2020). Machine learning-based stroke prediction models in a general population: A systematic review. *International Journal of Environmental Research and Public Health*, 17(12), 4556. doi:10.3390/ijerph17124556

8. Fang, G., Huang, Z., & Wang, Z. (2022). Predicting Ischemic Stroke Outcome Using Deep Learning Approaches. Frontiers in Genetics, 12. https://doi.org/10.3389/fgene.2021.827522

9. Dritsas, E., & Trigka, M. (2022). Stroke Risk Prediction with Machine Learning Techniques. Sensors, 22(13), 4670. https://doi.org/10.3390/s22134670

10. Dritsas, E., & Trigka, M. (2022). Stroke Risk Prediction with Machine Learning Techniques. Sensors, 22(13), 4670. https://doi.org/10.3390/s22134670

11. Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106. doi:10.1007/BF00116251

12. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. CRC Press.

13. Freund, Y., & Schapire, R. E. (1996). *Experiments with a new boosting algorithm*. In Proceedings of the Thirteenth International Conference on Machine Learning (pp. 148-156). Morgan Kaufmann Publishers.

14. Schapire, R. E., & Freund, Y. (2012). *Boosting: Foundations and algorithms*. MIT Press.

15. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). doi:10.1145/2939672.2939785

Information About Authors

Zhalgas Aidana, MSc, Senior-lecturer of Computational and Data Science and Vice-Dean at Astana IT University, Astana, Kazakhstan. Email: aidana.zhalgas@astanait.edu.kz. She received her MSc in Mechanical Engineering from Nazarbayev University in 2018. Aidana involved in research in biomechanics, machine learning, computational mathematics. Her work experience includes positions Research Assistant(Nazarbayev University), Senior-lecturer, Acting Director of the Department Computational and Data Science (Astana IT University).

ORCID: https://orcid.org/0000-0003-1091-8483.

Toleubek Moldir, MSc, Senior-lecturer of Computational and Data Science at Astana IT University, Astana, Kazakhstan. Email: moldir.toleubek@astanait.edu.kz. She received her MSc in Applied Mathematics from Nazarbayev University in 2020. Professional Career has started at Nazarbayev University in positions of Research Assistant and Teacher Assistant in 2019 and followed by position of lecturer at Astana IT University. Her research interests include neural networks, deep learning applications, computational mathematics. Weekly attended online seminars of Professor Kharin S. in 2022-2023 and did research on Stefan type problems by using Neural Networks together with colleagues from Astana IT University. ORCID: https://orcid.org/0000-0002-8449-1251.

Submission received: 17 September, 2024.
Revised: 23 September, 2024.
Accepted: 23 September, 2024.