WEP Warwick Evans Publishing

Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 12
ESAC 2026

Application of Machine Learning Methods in Stroke Prediction

Jiahui Hu [1, *], Zhenzhi Shi [2]

[1] Master of Business Administration Shanghai Jiaotong University Shanghai, China

[2] Class 3, Grade 11 Nanjing Foreign Language School Nanjing, China

* Corresponding Author Email: hujiahui0720@sjtu.edu.cn

Abstract. Stroke is currently one of the largest causes of death and disability worldwide, hence it needs effective strategies to detect it early. The paper discusses the role that ML techniques may play in predicting stroke risk, along with methodologies, performance metrics, and their clinical implications. There are benefits associated with integrating ML into stroke prediction models: large dataset processing, improved accuracy, and timely risk assessment. In this paper, we have considered five distinct ML models: Linear Regression, Logistic Regression, Support Vector Machines, Random Forest, and Neural Networks. All these models provide different perspectives and insights about stroke risk analysis. In this work, vigorous analytical methods will be used that involve correlation analysis, SVM, and logistic regression. All these analyses will source a comprehensive stroke dataset from Kaggle. The results show that parameters like age, heart disease, glucose level, and hypertension are important indices that could predict the risk of stroke. Notwithstanding the odds of challenges that may result from data imbalance and bias, our study is going to implement various strategies aimed at mitigating these issues to provide precision and accuracy in our findings. We conclude from the study that, indeed, ML algorithms, especially deep neural networks, are capable of providing effective improvement in the prediction of long-term outcome of patients who have ever suffered from ischemic stroke. This current study joins other efforts in progress geared toward improving early detection and, consequently, treatment outcomes for stroke patients.

Keywords: Machine Learning; Data Visualization; Stroke; Disease Prediction.

1. Introduction

Stroke remains to be a leading cause of death and long-term disability worldwide, posing a substantial burden on healthcare systems and society at large.[1] Early detection and prevention are critical to mitigating its impacts. The emergence of machine learning (ML) has revolutionized the field of medical diagnostics by enabling the analysis of vast and complex datasets, such as electronic health records (EHRs) [2,3]. ML algorithms can identify intricate patterns and relationships among risk factors, which traditional statistical methods might overlook. This paper delves into the application of various ML techniques for stroke risk prediction, examining their methodologies, performance metrics, and practical implications [4]. The integration of ML in stroke prediction offers a number of advantages. It allows researchers to process and learn from large datasets, improve prediction accuracy through advanced algorithms, and provide timely risk assessments that can guide preventive interventions. However, the application of these techniques also poses challenges, such as ensuring data quality, managing imbalanced datasets, and addressing ethical considerations related to patient data privacy. The anticipated outcomes of this study are expected to provide valuable insights into the effectiveness of such techniques within a clinical setting and their potential for improving patient outcomes through the means of early interventions. In this paper, five machine learning models will be considered: linear regression, logistic regression, support vector machines, random forests, and neural networks. These models have their own efficacies and capabilities with respect to the analyzing and predicting of stroke risk from complex data sets.

2. Related Work

At the moment, the scientific community is focusing on developing extremely efficient tools and techniques for monitoring and predicting life-threatening diseases. Recently, research studies have been conducted which focus on the application of ML algorithms in order to evaluate stroke risk [5, 6]. This paper will review the studies that have been conducted with their results.

In the paper "Performance Analysis of Various Machine Learning Approaches in Stroke Prediction," the authors have used three ML algorithms for stroke risk prediction [5]: logistic regression, decision tree, and random forest. They wanted to assess the performance of these algorithms in predicting the occurrence of a stroke in a patient. Results obtained showed that the Random Forest algorithm performed the best, with accuracy at 99.98%. This was followed by decision trees and logistic regression with accuracies of 99.46% and 81.34%, respectively [7]. The study, therefore, goes on to expound more on the potential of machine learning in improving the current prediction of stroke risk and further underscores the need for robust algorithms and significant features toward ensuring highly accurate predictions.

In the study" Stroke Risk Prediction with Machine Learning Techniques," the authors employed various machine learning models and a stacking method to develop a robust framework for long-term stroke risk prediction [6].Using a dataset from Kaggle, they applied several prepro- cessing techniques and evaluated the performance of mul- tiple models. The stacking model outperformed others, achieving an AUC of 98.9%, precision of 97.4%, recall of 97.4%, and accuracy of 98%[8]. The study concluded that the stacking method is effective During the process of detecting high-risk stroke candidates, with high predictive ability. Future work will focus on incorporating.

Researchers in the referenced study aimed to collect a stroke dataset from Sugam Multispecialty Hospital in India. They utilized data mining and machine learning algorithms to classify different types of strokes [7]. Their findings revealed that combining Support Vector Machine (SVM) with ensemble-bagging techniques improved classification accuracy to 91%. However, an Artificial Neural Network (ANN) trained with the stochastic gradient descent (SGD) algorithm outstripped all others, attaining a precision level of exceeding 95%. This indicates that ANN with SGD is particularly effective for stroke classification and could significantly enhance diagnostic accuracy in medical applications.

In another study titled "Stroke Prediction with Machine Learning Methods among Older Chinese," researchers also proposed gathering a stroke dataset from Sugam Multispecialty Hospital in India to refine the classification of stroke types using diverse data mining and machine learning algorithms. Among the tested algorithms, SVM with ensemble-bagging achieved a 91% accuracy rate, but the ANN model trained with SGD surpassed this, reaching over 95% accuracy [8]. These results highlight the potential of ANN with SGD to improve diagnostic accuracy in clinical settings.

In the study "Stroke Prediction of Outcomes in Acute Stroke," researchers focused on developing an advanced machine learning model to enhance the ability to predict acute stroke patient outcomes [9], aiding in treatment decisions and prognosis. This study used data from a prospective cohort of ischemic stroke patients admitted within seven days of initial manifestation of symptoms between January 2010 and December 2014, excluding those with a pre-stroke mRS >2, missing 3-month mRS, or who underwent recanalization treatment. In the study, three machine learning algorithms were evaluated: a deep neural network (DNN), random forest, and logistic regression. Notably, the DNN models outperformed the others, achieving an AUC score that was significantly higher than the benchmark set by the reputable Acute Stroke Registry and Analysis of Lausanne (ASTRAL) score. These findings suggest that machine learning algorithms, particularly deep neural networks, could improve long-term outcome predictions for ischemic stroke patients.

In the realm of stroke risk prediction, while significant strides have been made, several gaps remain in the literature, as highlighted in our" Related Work" section. The limitations of existing datasets, often characterized by regional specificity and imbalance, pose a challenge for developing broadly ap- plicable predictive models [10]. The generalizability of machine learning algorithms, particularly in diverse patient populations, is another area that requires further exploration. Moreover, the lack of transparency in complex models like deep neural networks can hinder clinical acceptance due to concerns about model interpretability. Ethical considerations regarding data privacy and the need for robust long-term risk prediction models are also areas that have not been fully addressed [11].

Our research endeavors to bridge these gaps, especially in managing imbalanced datasets and enhancing the gen- eralizability of predictive models. By leveraging an open- source dataset from Kaggle, we have employed Synthetic Minority Over-sampling Technique (SMOTE) to address the issue of data imbalance, thereby enriching the minority class and allowing our models to discern stroke risk factors moreeffectively [2, 3]. In our analysis, we have not only evaluated individual models but also utilized ensemble methods like stacking and majority voting to synthesize predictions, thereby improving the precision and robustness of our models. By this approach, A predictive performance is enhanced while simultaneously providing a more nuanced understanding of stroke risk factors [4, 5].

Furthermore, we have focused on enhancing model inter- pretability through visualization techniques such as correlation matrices and decision boundaries, making the decision-making process of our models more transparent and understandable to medical professionals. This transparency is crucial for building trust in the predictive capabilities of our models [6,7].

In a recent study presented at the 2023 IX International Conference on Information Technology and Nanotechnology, Faskhutdinova et al. explored the application of machine learning methods for predicting stroke outcomes [11]. Their research aimed to investigate the potential of various algorithms in identifying patterns that could contribute to more accurate stroke prediction models. The findings from this study provide valuable insights into the ongoing development of diagnostic tools that leverage the power of machine learning.

Wang et al. conducted a systematic review in 2020, which was published in PLoS ONE, focusing on the predictive capabilities of machine learning models in the context of stroke outcomes [12]. Their comprehensive analysis of structured data demonstrated the efficacy of these models in improving the accuracy of stroke predictions. The review underscored the significance of machine learning as a transformative approach in medical diagnostics, offering a systematic evaluation of existing methodologies and their clinical implications.

In conclusion, our research not only improves the accuracy of stroke risk prediction but also enhances the generalizability and interpretability of our models, offering valuable insights for clinical applications and patient care.

3. Analyze Stroke Dataset

In this section, we embark upon an examination of a compre- hensive stroke data set, delving into its intricacies through the application of rigorous analytical techniques. Our approach encompasses Correlation Analysis, Support Vector Machine, and Logistic Regression serves as a pivotal tool, enabling us to decipher the probabilistic relationship between the attributes and the occurrence of a stroke, thereby facilitating predictive insights. Crucially, we endeavor to assess the efficacy of Logistic Regression by employing the Confusion Matrix model, which offers a quantitative framework for evaluating classification performance. This study highlights the importance of uncovering correlations within disparate information sources, as they pave the way for both medical practitioners and patients alike to anticipate potential outcomes prior to definitive diagnoses. In the past few years, machine learning (ML) algorithms were widely applied for analyzing stroke data from multiple perspectives to get an insight into the dataset [6] [7] [8] [9] [10]. While acknowledging that our research may not occupy the forefront of medical breakthroughs, we remain steadfast in our commitment to conducting an independent, thorough analysis. Our objective is to contribute, however modestly, to the body of knowledge surrounding stroke prediction, aiming for precision and accuracy in our findings. By doing so, we aspire to contribute to the ongoing efforts aimed at enhancing early detection and improving treatment outcomes for those affected by this devastating condition.

3.1. Data Source

We obtained the dataset for analyzation from https: // www. kaggle. com/datasets/fedesoriano/stroke-rediction-dataset, which is an open-source dataset. This database contains part of their legal personal information, health information, and possible disease characteristics.

3.2. Correlation Matrix

In our analysis, we initially leverage correlation computa- tions as a cornerstone to procure numerical data. This pivotal step involves calculating the correlation matrix, which shows a matrix that serves as a beacon, illuminating the intensity and directionality of the intricate relationships between variable pairs. Our primary lens is directed towards unraveling how each attribute intertwines with the occurrence of stroke.

Subsequently, we visualize this intricate web of correlations through two captivating images. The first image, a heat map, vividly portrays the correlation matrix, painting a landscape of how diverse attributes, particularly those pertinent to patients with robust health indicators, are interconnected. The color palette plays its roles, with red ones indicates strong positive correlations, while for blue color vice versa, thus conveying the strength of these relationships at a glance.

The visualization shown as Fig. 1 illustrates a clear rela- tionship between age and stroke incidence, although further refinement is needed to enhance readability. Fig. 2, a bar chart, meticulously orchestrates the correlation coefficients of each attribute in a harmonious symphony with stroke. This visual narrative underscores a compelling narrative: age emerges as the paramount factor, boasting the strongest positive corre-lation with stroke, trailed closely by heart disease status, average glucose level, and hypertension status (also marital status but it is conditionally ignored), each contributing their unique rhythm to this intricate dance.

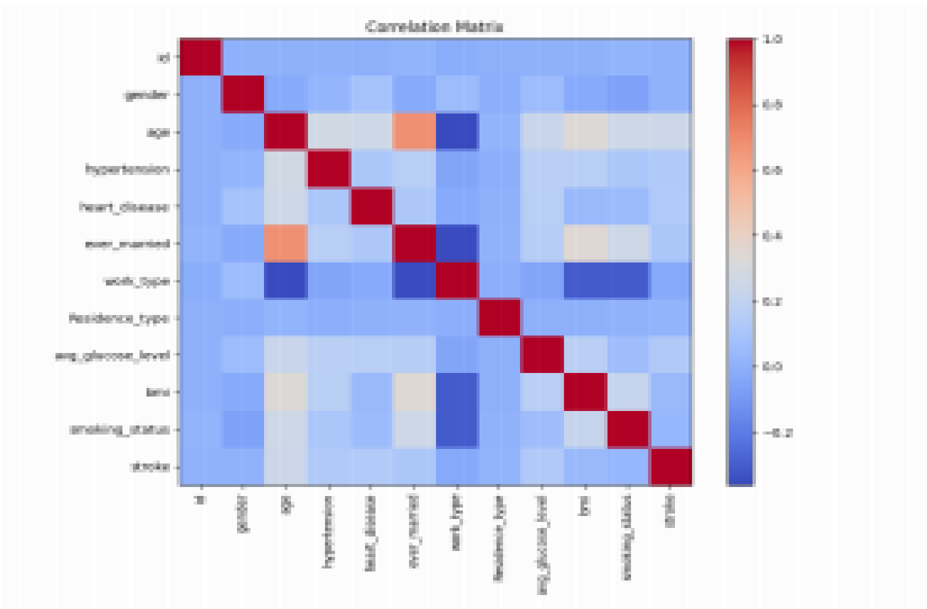

Fig 1. Image of Correlation Matrix.

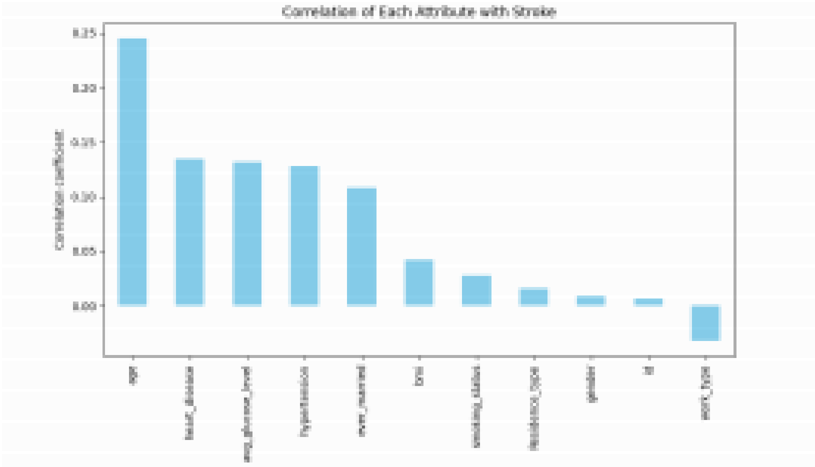

Fig. 2. Image of Correlation Bar Chart of stroke and other attributes.

Though from these images, we worked out some relation- ship between stroke and other health fact, but obviously the result is not up to our satisfactory. In further research, we try to find more crucial evidence between these facts and stroke. Also, we realized that due to the moderate strong relationship between age and marriage, we might ignore one of those attributes in our study.

3.3. SVM Decision Boundary

The Support Vector Machine (SVM) model is a frequently used math-based machine learning algorithm in realms including but not limited to disease prediction for pattern recognition, classification, and regression tasks, etc. SVM achieves strong classification performance and good generalization ability, making it widely applicable in machine learning. We start by using the pandas library, we meticulously parse through our CSV file, extracting the vital information it holds. To streamline the visualization process, we meticulously select two pivotal features: age, as a proxy for physiological maturity and vulnerability, and average glucose level, potentially indicative of certain health indicators.

Subsequently, we embark on a strategic division of our dataset, leveraging the train-test split method to segregate it into two distinct components – a training set comprising eighty percent of the data, meticulously crafted to nurture our model's learning, and a test set, accounting for the remaining twenty percent, designed to rigorously evaluate its predictive prowess.

With our dataset thus prepared, we proceed to harness the power of the SVM, specifically employing the linear kernel for its inherent simplicity and interpretability, to train our model on the training set. This process enables the model to discern the intricacies within the data, learning to distinguish between the patterns associated with stroke occurrence and those devoid of such events.

Ultimately, we bring our analysis to life through visual- ization, depicting the decision boundary of the SVM model alongside the data points it has classified. In this visual nar- rative, blue dots serve as poignant reminders of samples that are currently unencumbered by stroke, while red dots starkly contrast, representing instances where stroke has indeed oc- curred. This visualization not only offers a compelling visual representation of our model's predictions but also underscores the intricate interplay between age, health indicators, and the risk of stroke.

Fig. 3 essentially demonstrates that the SVM model identifies a correlation existing between glucose level, age, and the occurrence of stroke. The red dots are clearly scattered at the right part of the image, in which the upper part has greater red dot density. The higher the glucose level is and the older the age is, the bigger the possibility of getting a stroke is. Also, we can see that age is the more decisive factor, while the glucose level as another factor is not quite obvious related, and only after analyzing scatter density can certain conclusions be drawn. This is probably caused by the imbalanced data, but the model generally shows a expected correlation.

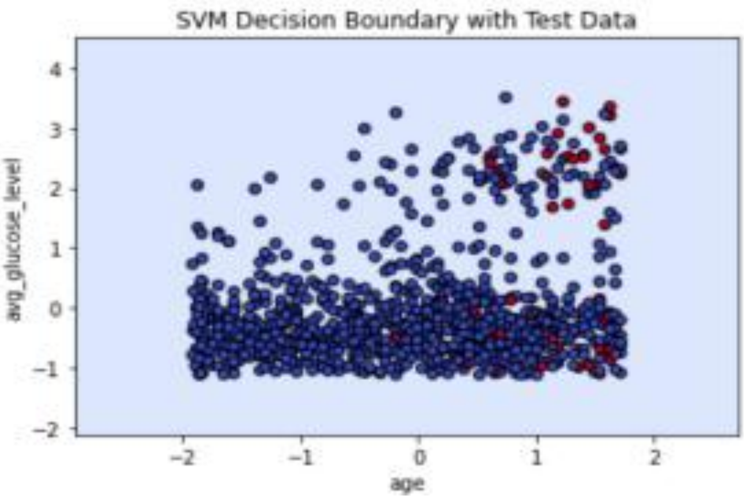

Fig. 3. Image of SVM Decision Boundary with test data.

3.4. Logistic Regression and Evaluation

3.4.1. Data Processing and Logistic Regression:

We include logistic regression in our data processing and analysis toolkit, as it is a powerful method for predicting binary outcomes. Just as with SVM, we begin by employing the pandas library to load and prepare our stroke-related data. However, the journey to a clean and ready-to-use dataset is multifaceted, requiring a meticulous approach to data pre-processing. One initial challenge is converting string data to numerical format for compatibility with machine learning algorithms. This step is crucial as it enables us to effectively process and analyze the data. Additionally, we scrutinize the file info, as shown in Table I, paying close attention to the non-NULL or missing data counts, particularly in columns like BMI Index, where the absence of data necessitates a decision on how to handle these gaps. In this case, we opt to delete rows containing unacceptable or incomplete data, ensuring a clean dataset for further analysis.

With a clean dataset in hand, as Fig. 4 we proceed to sum- marize the statistics of each column, gaining valuable insights into the distribution of values and identifying potential outliers or skewness. For instance, the average glucose level reveals a right-skewed distribution, whereas The BMI index exhibits a distribution that is approximately normal with a slight positive skew. However, the stroke status column, containing only zeros and ones, highlights an imbalance in our target variable, which is a common challenge in binary classification tasks.

Table I. Table Of Data Statistics.

Statistical information of data				
Number	***Data Attributes***	***Count***	***Non-NULL***	***Data Type***
1	ID	5110	non-null	int64
2	Gender	5110	non-null	object
3	Age	5110	non-null	float64
4	Hypertension	5110	non-null	int64
5	Heart disease	5110	non-null	int64
6	Ever married	5110	non-null	object
7	Work type	5110	non-null	object
8	Residence type	5110	non-null	int64
9	Avg glucose level	5110	non-null	object
10	BMI	4909	non-null	float64
11	Smoking status	5110	non-null	object
12	Stroke	5110	non-null	int64

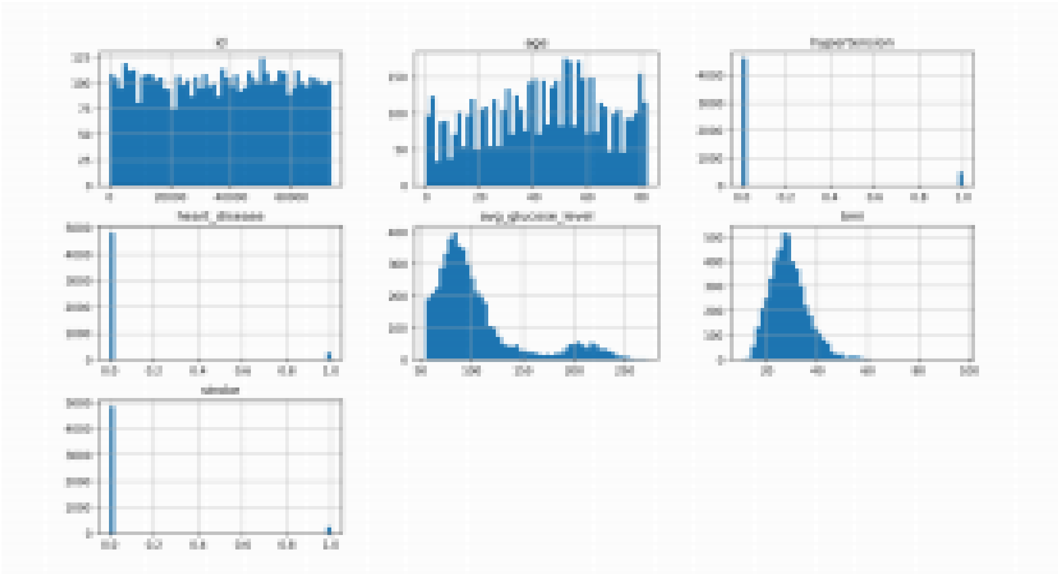

Fig. 4. Image of data spread.

To address this imbalance and prepare the data for modeling, we engage in a series of pre-processing steps. We commence by dropping unnecessary columns such as IDs, which do not contribute to the predictive power of our model. Subsequently, we meticulously remove rows with missing data, ensuring a complete dataset. Furthermore, we convert categorical vari-ables into numerical representations, as required by most machine learning models. This process not only simplifies data processing but also augments the model's capability to discern patterns within the data. Recognizing the importance of preserving the original data, we also process the data into a new file, maintaining two versions of the dataset: the raw and the processed. This approach ensures that we can always refer back to the original data if needed.

Finally, we reshape the data into a format that is acceptable by our model, storing the target variable (stroke status) in a separate y-array and the feature variables in a new file. With our data ready, we proceed to split it into training and testing sets, selecting a ratio of 80:20 to ensure a robust assessment of our model's capability.

Using the logistic regression model, we generate Fig.5 In the context of logistic regression, we acknowledge the chal- lenge posed by the imbalanced dataset and employ techniques such as balancing the weights of the classes to give equal importance to both zeros and ones. This approach helps to address the bias that can arise from an unequal distribution of target values, leading to a more balanced and accurate model. We then fit the pre-processed data to our logistic regression model, embarking on the final stage of our data analysis journey, as the image generated.

```
intercept: [-4.86518362]
coef: [[-0.11422339  0.07239545  0.63837752  0.35693811 -0.18401733 -0.12108939
   0.10552638  0.00375776  0.00747827  0.25020545]]

p_pred:
[[0.71330006 0.28669994]
 [0.73133895 0.26866105]
 [0.21265076 0.78734924]
 ...
 [0.95531469 0.04468531]
 [0.93585107 0.06414893]
 [0.74279233 0.25720767]]
```

Fig. 5. Image of Logistic Regression.

3.4.2. Confusion Matrix and Model Evaluation:

After the model training, we can predict the usability using the model, and we can evaluate the performance of each attribute by the following index, for instance, intercept coefficient and probability. We can also print out the actual result of its prediction: the score, or the confusion matrix according to the intercept and coefficient is the parameter of the actual model. Prediction is the probability of each prediction it makes and will convert it into zeros and ones.

A confusion matrix will help define the goodness of a model used for classification. It is basically a table used in supervised learning, wherein each column shows the number of predicted instances against every row, which denotes the number of actual instances. In binary classification, a confusion matrix is a 2x2 table that displays four key metrics: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). Logistic Regression and Confusion Matrix are related in terms of how logistic regression predictions should feed a confusion matrix. Such a matrix is basic for the evaluation of model performance. You can further use the confusion matrix to determine a number of performance metrics by detailed parameters.

Our Logistic Regression Model's Confusion Matrix (Fig. 6) and detailed parameter shown in Table II indicates that it has performed reasonably well with the processed data, demon- strating a good ability to classify the data into the correct cat- egories. However, despite these positive results, it's clear that there are still some biases present in the model's predictions that we haven't fully eliminated.

To take the performance of our model to the next level, it will become imperative that we address the issue of label imbalance in the dataset. This biasing due to class imbalance, skewed towards the majority class, results in inaccurate predictions for the underrepresented classes. Thus, in this respect, dataset balancing becomes essential, and this could be effectively done through either oversampling of the minority class or under-sampling of the majority class, or techniques such as SMOTE, which creates artificially developed examples from the minority class.

Table II. Confusion Matrix Parameters

Reported Items	Parameters			
	Precision	***Recall***	***F1-score***	***Support***
0.0	0.99	0.71	0.83	939
1.0	0.12	0.84	0.21	43
Micro-Avg	0.55	0.78	0.52	982
Weighted-Avg	0.95	0.72	0.80	982

*The overall f1-score(accuracy) is 0.71996.

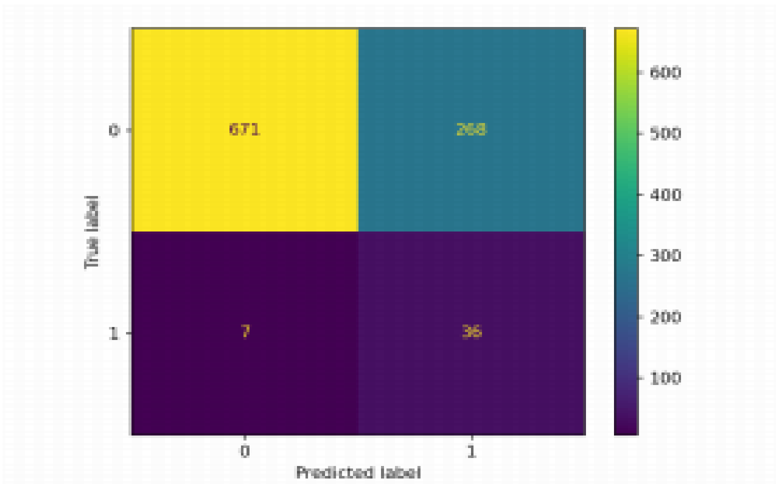

Fig. 6. Image of Confusion Matrix.

3.5. Discuss

In This section, we conducted a concise comparison of the outcomes derived from multiple experiments. These experi- ments unanimously revealed a pronounced correlation between age, glucose levels, and hypertension with the incidence of heart disease and stroke. Consequently, these factors can be designated as primary indicators for predicting the risk of stroke. Conversely, the Body Mass Index (BMI) and smoking status were not found to exhibit a significant association with these conditions, thereby not being considered primary predictive markers.

Additionally, we provided a concise overview of the application scenarios for several models. The Correlation Matrix efficiently summarizes the relationships between various indicators, facilitating the identification of valuable research directions. The SVM model visually represents large datasets, enabling straightforward analysis. The logistic regression model offers a straightforward approach to predicting binary classification probabilities. Lastly, the Confusion Matrix is utilized to assess the performance of logistic regression models, providing valuable insights into their accuracy and reliability.

4. Conclusion

In this essay, we delve into the intricate relationship between stroke occurrence and its potential predisposing factors, employing a diverse array of data analysis models as our in-vestigative tool. Our endeavor has illuminated the pivotal role played by age, heart disease, glucose levels, and hypertension in forecasting the likelihood of a stroke, underscoring their significance as key considerations in risk assessment. However, the path to unraveling these correlations has not been without challenges. We encountered formidable obstacles in the form of data imbalance and bias. Recognizing the gravity of these issues, we implemented rigorous processes and optimization strategies aimed at mitigating their influence.

As we embark on future research endeavors, optimizing our approach to age as a variable becomes paramount. Rather than simply acknowledging its influence, we must strive to disentangle its effects from those of other, potentially more actionable, factors. Such imbalance problems also exist in data of glucose level. This necessitates a nuanced and sophisticated analysis that leverages techniques such as SMOTE (Synthetic Minority Over-sampling Technique) to mitigate data imbal- ance. By minimizing the confounding influence of age through SMOTE and other strategies, we can gain a clearer and more focused view of the intricate interplay that gives rise to stroke. SMOTE enables the generation of additional minority class samples, balancing the dataset with respect to age distribution, thereby enhancing the accuracy and solidity of our analytical outcomes.

This study exemplifies the successful conversion of intricate machine learning models into clinically applicable risk assessment instruments, thereby facilitating the integration of artificial intelligence within the realm of medical practice. By identifying and validating pivotal stroke risk factors, it establishes a more exhaustive foundation for risk evaluation in clinical settings. The introduction of a straightforward, non-invasive, and efficient tool for stroke risk assessment empowers healthcare professionals to promptly pinpoint high-risk patient populations and initiate preventive measures in a timely and effective manner. This not only advances the precision of risk stratification but also underscores the potential of AI in enhancing patient outcomes and reducing the burden of stroke.

Authors Contribution

All the authors contributed equally, and their names were listed in alphabetical order.

References

[1] J.Heo, J. G. Yoon, H. Park, Y. Kim, H. Nam, and J. Heo, "Machine Learning–Based Model for Prediction of Outcomes in Acute Stroke," Stroke, vol. 50, 2019, pp. 1263-1265.

[2] A.Faskhutdinova, D. Grigorieva, B. Garafutdinov, and V. Mokshin, "Investigation of Machine Learning Methods for Stroke Prediction," in Proc. 2023 IX Int. Conf. Information Technology and Nanotechnology (ITNT), 2023, pp. 1-6.

[3] W.Wang et al., "A systematic review of machine learning models for predicting outcomes of stroke with structured data," PLoS ONE, vol. 15, 2020, p. e0229980.

[4] J.Kim, Y. J. Choo, and M. Chang, "Prediction of Motor Function in Stroke Patients Using Machine Learning Algorithm: Development of Practical Models," J. Stroke Cerebrovasc. Dis., vol. 30, 2021, p. 105856.

[5] Asadi, R. Dowling, B. Yan, and P. Mitchell, "Machine Learning for Outcome Prediction of Acute Ischemic Stroke Post Intra-Arterial Therapy," PLoS ONE, vol. 9, 2014, p. e88225.

[6] M.S. Sirsat, E. Fermé, and J. Câmara, "Machine Learning for Brain Stroke: A Review," J. Stroke Cerebrovasc. Dis., vol. 29, 2020, p. 105162.

[7] F. Abdel-Gawad, S. El-Sayed, and M. M. Ismail, "From Data to Diagnosis: Applied Machine Learning for Stroke Prediction in Computational Healthcare," J. Artif. Intell. Metaheuristics, in press.

[8] S. Mainali, M. E. Darsie, and K. Smetana, "Machine Learning in Action: Stroke Diagnosis and Outcome Prediction," Front. Neurol., vol. 12, 2021, p. 743789.

[9] R. Jeena and S. Kumar, "Machine intelligence in stroke prediction," Int. J. Bioinform. Res. Appl., vol. 14, pp. 29-48, 2018.

[10] M.U. Emon et al., "Performance Analysis of Machine Learning Approaches in Stroke Prediction," in Proc. 2020 4th Int. Conf. Electronics, Communication and Aerospace Technology (ICECA), 2020, pp. 1464-1469.

[11] A. Faskhutdinova, D. Grigorieva, B. Garafutdinov, and V. Mokshin, "Investigation of Machine Learning Methods for Stroke Prediction," in Proc. 2023 IX Int. Conf. Information Technology and Nanotechnology (ITNT), 2023, pp. 1-6.

[12] W. Wang et al., "A systematic review of machine learning models for predicting outcomes of stroke with structured data," PLoS ONE, vol. 15, 2020, p. e0229980.