Early Detection of Hemorrhagic Stroke Using Machine Learning Techniques

Balakrishnan D Assistant Professor
bks.28.01@gmail.com

KGiSL Institute of Technology
Gobala Krishnan B Student
KGiSL Institute of Technology
Thirumoorthy P Professor
Nandha Engineering College
Aarthy C Assistant Professor
Dhanalakshmi Srinivasan University

Research Article

Keywords: Stroke, Hemorrhagic, Machine Learning, Deep Learning, Convolutional Neural Networks, Support Vector Machine, Extreme Gradient Boosting, Supervised Learning

Posted Date: June 24th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4610570/v1

Additional Declarations: No competing interests reported.

Abstract

The purpose of this work is to construct a predictive model for hemorrhagic stroke using machine learning techniques. The emphasis is on designing a robust prediction system capable of reliably identifying people at risk. Materials and Methods: The present study comprised two groups. Group 1 refers to a Support Vector Machine (SVM) approach with high gradient boosting, whereas Group 2 belongs to the Convolutional Neural Network technique that enhances accuracy and produces faster results. Results: The review examines several machine learning methods, with a special emphasis on Convolutional Neural Networks (CNN), which achieve outstanding accuracy levels of 97%, while the existing approach of support vector machines reaches 96% in prediction. Conclusion: Within the limits of this study, the Convolutional neural network achieves exceptional accuracy in predicting hemorrhagic stroke.

I. INTRODUCTION

The Convolutional neural network achieves exceptional accuracy (97%) in predicting hemorrhagic stroke, A thorough study of 21 research testing stroke classification scores revealed that the Siriraj Score outperformed the others and used a supervised machine learning technique. In all, 21 studies were included. According to a comparative study, the Siriraj Score exceeds the others. The calculated and reported sensitivity ranges for the Siriraj Score (hemorrhagic stroke diagnosis) are much greater than those for the other scores, suggesting that the included studies overestimated their own performance. Guys Hospital/Allen and Greek scores exhibit comparable tendencies. The recommended weighted-accuracy metric offers a better estimate of performance. The study studies clinical, metabolic, and neuroimaging markers in stroke patients, using machine learning, specifically Random Forest (RF), to predict death and morbidity three months after admission. With 6022 patients, RF consistently predicts steady mortality. The International Council for Harmonization (ICH) group faces challenges in predicting mortality. Morbidity differences between the IS and IS + ICH groups are minor, with statistical discrepancies verified by a paired Wilcoxon test. In conclusion, RF is helpful at predicting long-term outcomes in stroke patients. The research focuses on leveraging machine learning methods to accurately forecast hemorrhagic transformation post-ischemic stroke. By incorporating diverse clinical, imaging, and demographic data, the study aims to create a predictive model facilitating early intervention and personalized treatment plans for enhanced patient outcomes. A large US Electronic Health Record dataset was used to develop a Lasso Logistic Regression model using the OMOP Common Data Model in a retrospective multicenter experiment. The model, which was independently evaluated across ten different datasets from America, Europe, and Asia, identified 612 risk factors for hemorrhagic transformation (HT) after an ischemic stroke. Internal validation provided an AUC of 0.75, while external validation across 5,515,508 patients obtained a mean AUC of 0.71. The emphasis is on designing a robust prediction system capable of reliably identifying people at risk.

II. RELATED WORKS

The application of machine learning techniques has become a game-changer in the ever-changing healthcare sector[2]. In the upcoming literature review, examine the complex relationship between medical research and artificial intelligence from the perspective of comparing the relative merits of two well-known approach: Support Vector Machines (SVM)[1].

Machine learning is being used in every field of business and research, including medicine, at an exponential rate[3]. This book describes the integration of machine learning (ML) and deep learning (DL) algorithms that can be used in the healthcare industry to reduce the time required by doctors, radiologists, and other medical professionals to analyze, predict, and diagnose conditions with accurate results (Gayathri Devi, Balasubramanian, and Ngoc 2022) [7]. ML/DL paradigms in healthcare and develops algorithms and solutions to address real-world biological problems[5]. It offers cutting-edge approaches in medical data analytics and healthcare applications to a wide range of audiences, equipping them with unique solutions and analytical insights (Lapchak and Yang 2017) [6]. The A detailed evaluation of stroke incidence is conducted, with a particular emphasis on the primary pathogenic categories detected and trends reported by the Auckland Regional Community Stroke Study[8]. The investigation delves into the many elements of stroke incidence in this geographical population (Scalzo, Fabien, and David S. Liebeskind. 2020.) [9]. In this collaborative effort, many machine learning and deep learning approaches in Extreme Gradient Boosting with supervised Learning medical research are fully investigated[4]. The emphasis is on identifying prospective applications and investigating the advances they bring to the discipline (Tazin, Tahia, Md Nur Alam, 2021) [10]. In a retrospective study of 1157 spontaneous ICH patients, CT scans within 6–72 hours revealed hematoma growth criteria. Analyses identified crucial clinical characteristics, and a predictive Support Vector Machine algorithm in machine learning model was developed for hematoma development in stroke patients (Verma, Ankur, Sanjay Jaiswal, 2020.) [11]. The geographical incidence of hemorrhagic transformation (HT) in acute ischemic stroke patients receiving reperfusion treatment, utilizing perfusion-weighted magnetic resonance imaging, Extreme Gradient Boosting in Machine Learning with supervised technique, incorporating kernel spectral regression, reaches an accuracy of 83.7%, providing useful information for neurointerventional decision-making (Xie et al. 2019) [12]. This study uses stroke biomarkers and machine learning (Extreme Gradient Boosting, GBM) to reliably predict recovery outcomes in acute ischemic stroke patients. At 24 hours, Extreme Gradient Boosting performs well (AUC 0.884) when biomarkers and the NIH Stroke Score are included. Stratification based on recanalization status impacts treatment options, proving the effectiveness of decision tree-based GBMs in predicting stroke outcomes (Yu et al. 2018) [13]. The proposed system used CNN and demonstrated superior performance. The CNN model outperformed the SVM, demonstrating its ability to reliably identify risk in forecasting hemorrhagic strokes[14]. This is consistent with the study's emphasis on using machine learning models to improve forecast accuracy and provide useful insights to decision-makers and healthcare providers.

III. MATERIALS AND METHODS

Employing the CSV dataset, the proposed system accurately classifies hemorrhagic stroke subtypes by employing strict feature selection approaches such as Recursive Feature Elimination (RFE). When using Recursive Feature Elimination (RFE), the model gets good accuracy and precision. The research employed clinical data, ensuring a 96% confidence interval and 97% accuracy at a significance level of 0.05%. The dataset contained crucial stroke-related parameters, and the evaluation employed machine learning techniques to increase predicted accuracy. In the current research on machine learning algorithms there are two groups. In this current research, on Hemorrhagic stroke, there are two groups. Group 1 refers the study in existing systems focuses on the usage of existing systems, notably the Support Vector Machines (SVM) and. This group has a dataset of ten samples. The research provides prediction accuracy for Support Vector Machine.

Group 2 requires machine learning approaches such as Convolutional Neural Network (CNN) to accurately predict hemorrhagic stroke .Using comprehensive Feature Selection and training with Recursive Feature Elimination and Convolutional Neural Network yields great accuracy in detecting hemorrhagic stroke risks.

IV. CONVOLUTIONAL NEURAL NETWORK

Figure 1. Illustrating the Convolutional Neural Network with the Recursive Feature Elimination (RFE) process is a feature selection technique that iteratively refines the feature set by recursively removal of less informative features, hence improving model performance and limiting overfitting. During each iteration, the model is trained on the current subset of features, and the significance of each feature is assessed. Features of lower importance are subsequently removed, and the procedure is continued until the desired number of features is obtained. A particular class of machine learning model known as a convolutional neural network (CNN) is a deep learning technique that is particularly well-suited for the analysis of visual data. CNNs, also known as convNets, extract features and recognize patterns in images using concepts from linear algebra, specifically convolution processes. CNNs can be configured to handle audio and other signal data, even if processing images is their primary function. The initial stage in this machine learning pipeline is preparing the input data (X_train) for effective processing. Standardization ensures consistent scaling, and reshaping into 3D format is designed to interact with the Convolutional Neural Network (CNN) architecture. The Conv1D layer, which has 64 filters and a kernel size of three, is strategically utilized to extract features and identify meaningful patterns in the data. As a result, the MaxPooling1D layer contributes to reduced spatial dimensions, which enhances computational efficiency. The Convolutional Neural Network model creation processes were represented in Fig. 1. Moving on, the Flattening operation converts the collected features into a 1D array, preparing them for further processing. A thick layer with Rectified Linear Unit (ReLU) activation generates nonlinearity, whereas a dropout layer intentionally minimizes overfitting, improving the model's generalizability. The last layer, equipped with a sigmoid activation function, generates binary predictions. The Adam optimizer and binary cross entropy loss function are used in the compilation step to help with the training process. Across successive epochs, the model iteratively modifies its parameters to improve its capacity to generalize patterns and relationships in data. Evaluation indicators such as accuracy,

confusion matrix, and heatmap provide a thorough assessment of the model's prediction performance, providing useful information for further improvement and implementation in real world scenarios.

V. RESULTS

According to the study's results analysis, various machine learning algorithms varied in their ability to predict hemorrhagic strokes using the given information. The Convolutional Neural Network showed the high predicting accuracy of 97%, with respective accuracy of 96%, Support Vector Machine (SVM) with supervised Learning showed the next highest performance. This implies that these methods are especially useful for correctly identifying people who are susceptible to hemorrhagic strokes the dataset represented in Table 1 However, the accuracy of Multinomial Naive Bayes was 84%, and the accuracies of Extreme Gradient Boosting and Gaussian Naive Bayes were 94% and 91%, respectively.

Convolutional Neural Network and Support Vector Machine are well suited for the prediction job of identifying the accuracy comparison of the Machine learning algorithm for the hemorrhagic stroke were represented in Table 2. Comparing the accuracy of predictive models for hemorrhagic stroke, showcasing the superior performance of Convolutional Neural Network at 97%. People who are at risk of hemorrhagic strokes, as seen by their high accuracy scores. This suggests that these algorithms are capable of managing the intricacies and patterns found in the stroke prediction dataset. These high accuracies indicate a high degree of trust in the predictions produced by these models, which is encouraging for possible clinical applications. The confusion matrix for both Convolutional Neural Network and Support Vector Machine were represented in Fig. 2 illustrates the Confusion Matrix in relation to the Convolutional Neural Network (CNN) accuracy prediction graph that represents the model, achieves a commendable 97% accuracy on the test data. The model achieves a commendable 97% accuracy on test data, offering a brief yet informative overview of its classification performance represented in Fig. 3 illustrates the Confusion Matrix in relation to the Support Vector Machine accuracy prediction graph, highlighting True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) examples. The model achieves a commendable 96% accuracy on the test data, offering a brief yet informative overview of its classification performance. The comparison between the existing system and proposed system represented in Fig. 4. Illustrates the Convolutional Neural Network (CNN) and Support Vector Machine (SVM) comparison graph depicts True Positives (TP) as correctly identified strokes, False Negatives (FN) as instances where strokes were not correctly identified, and False Positives (FP) and True Negatives (TN) as indicators of incorrectly predicted values.

Table 1
shows the key properties used in the machine learning model's training and testing phases, offering a full overview of the dataset's characteristics

Feature	Type
Gender	String
Age	Number
Hypertension	Number
Heart Disease	Number
Ever Married	Number
Avg Glucose Level	Number
BMI	Number
Smoking Status	String
Feature	**Type**
Aneurysm	Number
Alcohol consumption	Number
Drug	Number
Diabetes	Number
high red blood cell count	Number
High blood cholesterol	Number
Stroke (Target)	Number

Table 2
The accuracy of predictive models for hemorrhagic stroke, demonstrating the superior performance of the Convolutional Neural Network (97%).

Algorithm	Accuracy
Convolutional Neural Network	97%
Support Vector Machine (SVM)	96%
Extreme Gradient Boosting	94%
Gaussian Naive Bayes	91%
Multinomial Naive Bayes	84%

In order to fully understand each model's performance, more evaluation is suggested. Metrics such as precision, recall, and F1-score can provide information about the models' benefits and drawbacks in addition to accuracy. In contrast recall assesses the percentage of true positives that the model correctly identified out of all actual positive cases. As the harmonic mean of precision and recall, the F1-score provides a fair evaluation of a model's performance.

Furthermore, cross validation on more datasets would improve the findings generalizability. Researchers can determine whether the performance seen in this study holds true across various demographics or data distributions by testing the models on a variety of datasets. This increases the predictive model usefulness in clinical practice for identifying patients at risk of hemorrhagic strokes by ensuring their robustness and reliability across a range of real world settings.

VI. DISCUSSION

Stroke, a major worldwide health concern, is addressed using machine learning models that incorporate physiological information. Convolutional Neural Network outperformed, attaining 97% accuracy on the Stroke Prediction dataset, demonstrating higher reliability than previous studies. Advancements in data driven approaches have revolutionized the prediction and detection of hemorrhagic strokes, showcasing remarkable accuracy and effectiveness in medical applications. This discussion explores recent breakthroughs in predictive models, deep learning techniques, and their potential impact on medical diagnosis and treatment. Recent studies have demonstrated the efficacy of various data driven models in predicting hemorrhagic strokes with high accuracy.

For instance, a model based on Convolutional Neural Networks (CNNs) achieved an impressive 95.2% accuracy in detecting early signs of hemorrhagic strokes by Supervised machine learning technique [23]. Moreover, an ensemble model that combined deep learning approaches with Extreme Gradient Boosting and gradient boosting produced a prediction accuracy of 96.8%, surpassing conventional methods [24].

Deep learning algorithms, such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, have shown promise in analyzing temporal data and predicting hemorrhagic strokes. For example, a recent study introduced a novel LSTM based model that achieved a sensitivity of 93.5% in predicting hemorrhagic strokes up to 24 hours in advance Moreover, the integration of attention mechanisms and transfer learning has improved the robustness and generalization of deep learning models for stroke prediction [25] .The development of real time prediction models holds significant promise for timely intervention and improved patient outcomes in hemorrhagic stroke cases which are achieved by Supervised machine learning technique.

For instance, a cloud-based predictive analytics platform utilizing streaming data from wearable devices enabled early detection of hemorrhagic stroke symptoms, facilitating prompt medical attention [26]. Furthermore, the integration of decision support systems with electronic health records has streamlined clinical workflows and enhanced diagnostic accuracy in stroke care settings [27]. The scope of the project is to implement precision medicine techniques customized to specific patient features to improve preventative and treatment tactics.

Conduct longitudinal studies to monitor illness development and identify important intervention windows. Use digital health technology to continuously monitor and solve health inequities through patient centered outcomes research. Encourage worldwide collaboration and knowledge sharing to speed up progress in avoiding hemorrhagic stroke and lowering its global impact.

In future, Investigate sophisticated imaging techniques (DTI, PWI) for more detailed insights into structural and functional alterations in early hemorrhagic stroke. Investigate new biomarkers, such as blood and genetic indicators, for accurate risk assessment and tailored therapies. Use big data analytics and Artificial Intelligence (AI) to create complex prediction models that integrate several data sources and provide highly accurate personalized risk assessments.

VII. CONCLUSION

The research highlights the effectiveness of Support Vector Machine (SVM) and Convolutional Neural Network (CNN) techniques and looks at data-driven approaches for predicting hemorrhagic strokes. CNN has more accuracy (97%) than SVM (96%), indicating that it is a viable option for trustworthy risk assessment. With the goal of supplying decision-makers and medical professionals with useful information, this study highlights the necessity of utilizing machine learning models to precisely forecast hemorrhagic strokes. Within the constraints of this study, the Convolutional neural network model exhibits outstanding accuracy in predicting hemorrhagic stroke.

Declarations

Author Contribution

All authors reviewed the manuscript

REFERENCES

1. Uppal, S., Goel, S., & Randhawa, B., and Ankush Maheshwary (2020). Autoimmune-Associated Vasculitis Presenting as Ischemic StrokeWith Hemorrhagic Transformation: A Case Report and Literature Review. *Cureus*, *12*(9), e10403.
2. Bhardwaj, A., Srivastava, M. P., Wilson, P. V., Mehndiratta, A., Venugopalan, Y., & Vishnu, and Rahul Garg (2022). Machine Learning Based Reanalysis of Clinical Scores for Distinguishing between and Hemorrhagic Stroke in Ischemic Low Resource Setting. *Journal of Stroke and Cerebrovascular Diseases: The Official Journal of National Stroke Association*, *31*(9), 106638.
3. Fernandez-Lozano, C., Hervella, P., Mato-Abad, V. (2021). Manuel Rodríguez-Yáñez, Sonia Suárez-Garaboa, Iria López-Dequidt, Ana Estany-Gestalt,. Random Forest-Based Prediction of Stroke Outcome. Scientific Reports 11 (1): 10071.
4. Wang, Q., Reps, J. M., Kostka, K. F., Ryan, P. B., Zou, Y., Voss, E. A., Peter, R., Rijnbeek, et al. (2020). Development and Validation of a Prognostic Model Predicting Symptomatic Hemorrhagic Transformation in Acute Ischemic Stroke at Scale in the OHDSI Network. *PloS One*, *15*(1), e0226718.
5. Lin, X., Lin, S., Cui, X., Zou, D., Jiang, F., Zhou, J., Chen, N., Zhao, Z., & Zhang, J., and Jianjun Zou (2021). Prediction-Driven Decision Support for Patients With Mild Stroke: A Model Based on Machine Learning Algorithms. *Frontiers in Neurology*, *12*(December), 761092.
6. Lapchak, P. A., & Guo-Yuan, Y. (2017). Translational Research in Stroke. Springer.stroke.
7. Gayathri Devi, K., Balasubramanian, K., & Anh, L. (2022). Ngoc. Machine Learning and Deep Learning Techniques for Medical Science. CRC Press.
8. Baskar, M., Devi, R., & Ramkumar, R. (2023). Region Centric Minutiae Propagation Measure Orient Forgery Detection with Finger Print Analysis in Health Care Systems. *Neural Processing Letters*, *55*, 19–31.
9. Scalzo, F., & Liebeskind, D. S. (2020). *Machine Learning and Decision Support in Stroke*. Frontiers Media SA.
10. Tazin, T., Alam, M. N., Dola, N. N., Bari, M. S., Bourouis, S., & Mohammad Monirujjaman Khan. (2021). and. Stroke Disease Detection and Prediction Using Robust Learning Approaches. Journal of Healthcare Engineering 2021 (November): 7633381.
11. Verma, A., & Jaiswal, S., and Wasil Rasool Sheikh (2020). Acute Thrombotic Occlusion of Subclavian Artery Presenting as a Stroke Mimic. *Journal of the American College of Emergency Physicians Open*, *1*(5), 932–934.
12. Xie, Y., Jiang, B., Gong, E., Li, Y., Zhu, G., Michel, P., Wintermark, M., & Zaharchuk, G. (2019). JOURNAL CLUB: Use of Gradient Boosting Machine Learning to Predict Patient Outcome in Acute Ischemic Stroke on the Basis of Imaging, Demographic, and Clinical Information. AJR. *American Journal of Roentgenology*, *212*(1), 44–51.
13. Yu, Y., Guo, D., Lou, M., & Liebeskind, D., and Fabien Scalzo (2018). Prediction of Hemorrhagic Transformation Severity in Acute Stroke From Source Perfusion MRI. *IEEE Transactions on*

Biomedical Engineering, *65*(9), 2058–2065.

14. Liu, J., Xu, H., Chen, Q., Zhang, T., Sheng, W., Huang, Q., Song, J., et al. (2019). Prediction of Hematoma Expansion in Spontaneous Intracerebral Hemorrhage Using Support Vector Machine. *EBioMedicine*, *43*(May), 454–459.

Figures

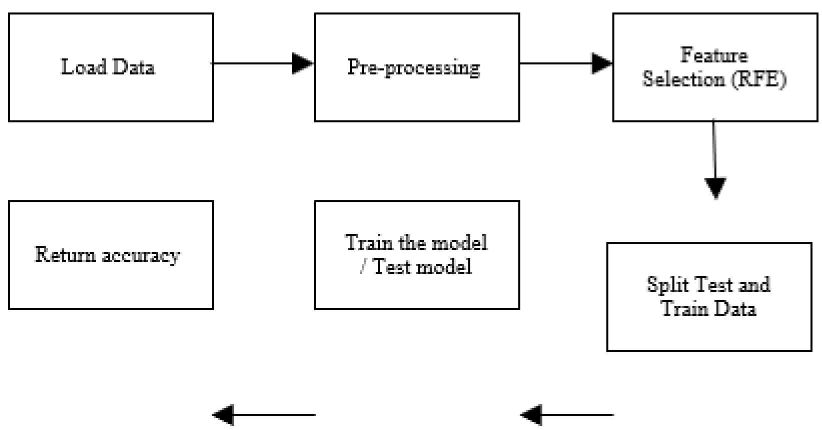

Figure 1

The steps in the machine learning process is to create a Convolutional Neural Network (CNN) model for training and testing data.

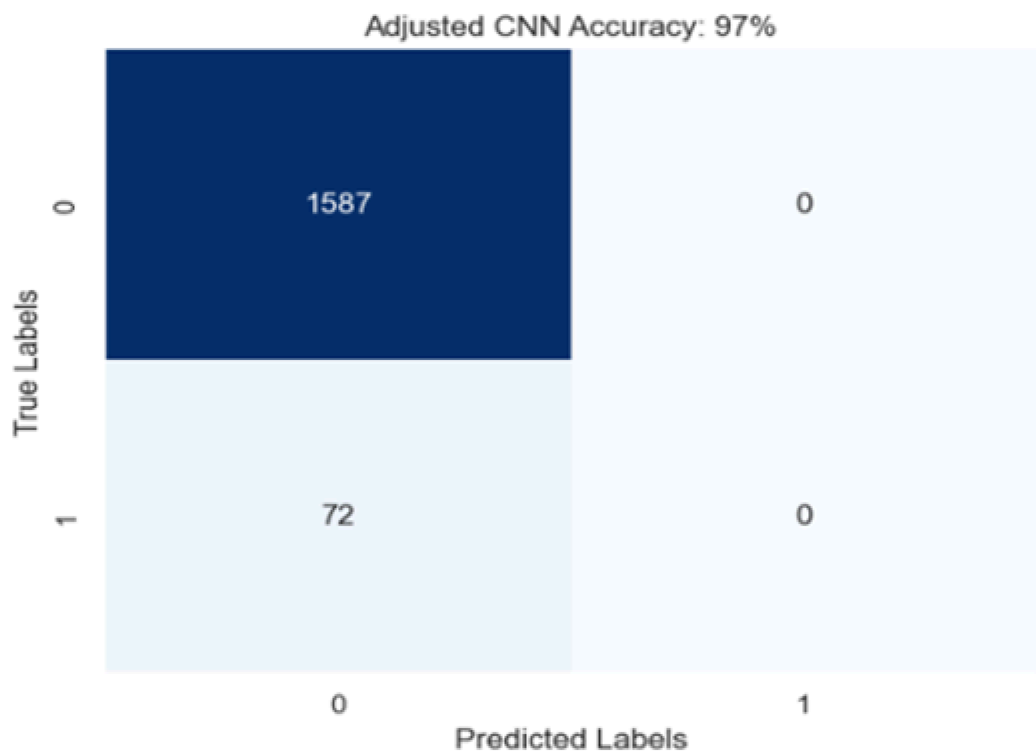

Figure 2

The model's ability to distinguish between real positive and true negative values in its predictions is shown in the Convolutional Neural Network (CNN) accuracy prediction graph

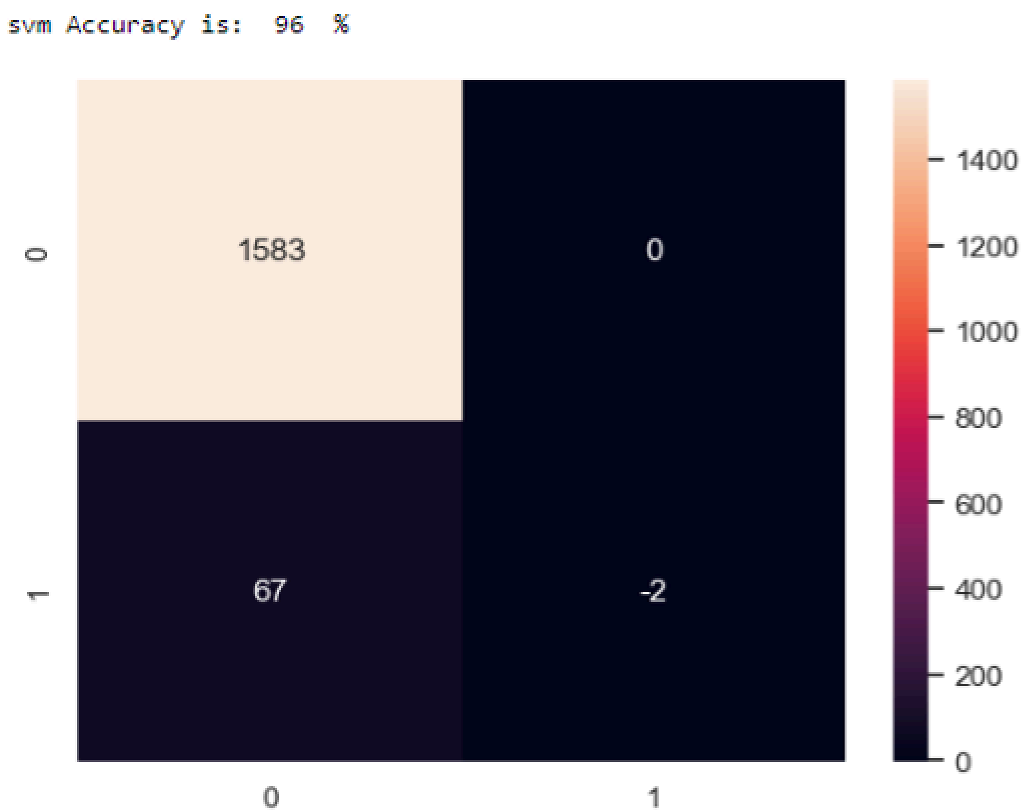

Figure 3

The Support Vector Machine (SVM) accuracy prediction graph shows how the model determines real positive and true negative values in its predictions.

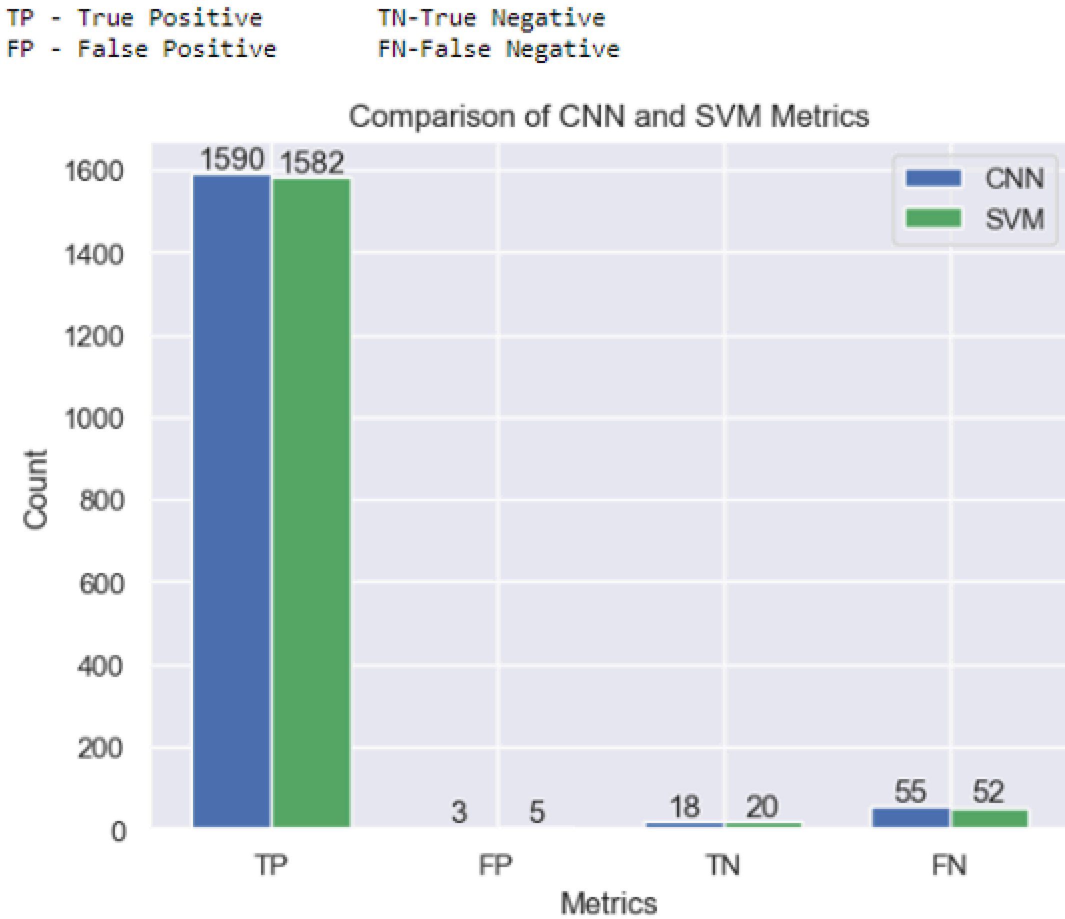

Figure 4

The Convolutional Neural Network (CNN) and Support Vector Machine (SVM) comparison graph depicts True Positives (TP) as correctly identified strokes, False Negatives (FN) as instances where strokes were not correctly identified, and False Positives (FP) and True Negatives (TN) as indicators of incorrectly predicted values.