Journal of Autonomous Intelligence (2024) Volume 7 Issue 5
doi: 10.32629/jai.v7i5.1667

ORIGINAL RESEARCH ARTICLE

Stroke risk prediction with 5G infrastructure based learning framework

A. Revathi[1,*], K. Santhi[2], R. Swathi[1], P.V.R.D. Prasad Rao[2]

[1] *Department of Computer Science and Engineering, Sri Venkateswara College of Engineering, Karakambadi Road, Tirupati 517502, Andhra Pradesh, India*

[2] *Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur 522 302, Andhra Pradesh, India*

*** Corresponding author:** A. Revathi, arevathi20@gmail.com

ABSTRACT

Due to the lack of an effective remedy, stroke becomes worldwide cause of death and long-term impairment. Despite the need for deep learning-based approaches on extensive and well labeled data, they possess the potential to outperform existing algorithms for predicting stroke risk. Moreover, there exists a significant disparity between the instances of favorable and unfavorable occurrences within this dataset. By using the specialized knowledge of a related field, transferable knowledge aids in resolving minor data difficulties, especially when many data sources are available. This paper introduces a novel stroke risk assessment system dubbed HDTL-SRP, which utilizes a hybrid deep transfer learning approach to leverage information from several associated resources, such as external stock data and data on chronic conditions like hypertension and diabetes. After undergoing rigorous testing in both simulated and actual scenarios, the recommended framework outperforms the most powerful stroke prediction algorithms. Additionally, it showcases the feasibility of utilizing 5G/B5G infrastructures to provide assistance for several hospitals in real-world scenarios.

Keywords: deep learning; framework; risk prediction; 5G systems

ARTICLE INFO

Received: 25 March 2024
Accepted: 28 April 2024
Available online: 11 September 2024

1. Introduction

On average, a stroke occurs approximately every forty minutes. According to the data, the mortality rate for stroke sufferers within one year is 50%[1]. The treatment and rehabilitation expenditures impose an exceedingly burdensome financial responsibility on both the survivors' family and the healthcare system. The total expenditures, including direct and indirect, of stroke incidents during 2014 and 2015 amounted to almost 45.5 million US dollars[2]. Hence, it is crucial to possess accurate stroke prediction in order to promptly adopt cost-effective interventions to mitigate the risk of stroke and postpone its onset.

Several studies have been conducted to develop stroke danger predict (SRP) models utilizing medical data, including retinal scans and electronic health records. Reports indicate that DNNs, or deep neural networks, have the best level of accuracy in predicting strokes[3]. However, a prominent drawback of this particular model is its reliance on a substantial amount of accurately labeled data. Consequently, tiny segments of the whole stroke data collection are usually spread out over many institutions. Moreover, there is a significant imbalance between the good and bad cases in stroke statistics. Hence, in practical implementation, the DNN-based Mrp systems may exhibit suboptimal performance[4] that represented in the **Figure 1**.

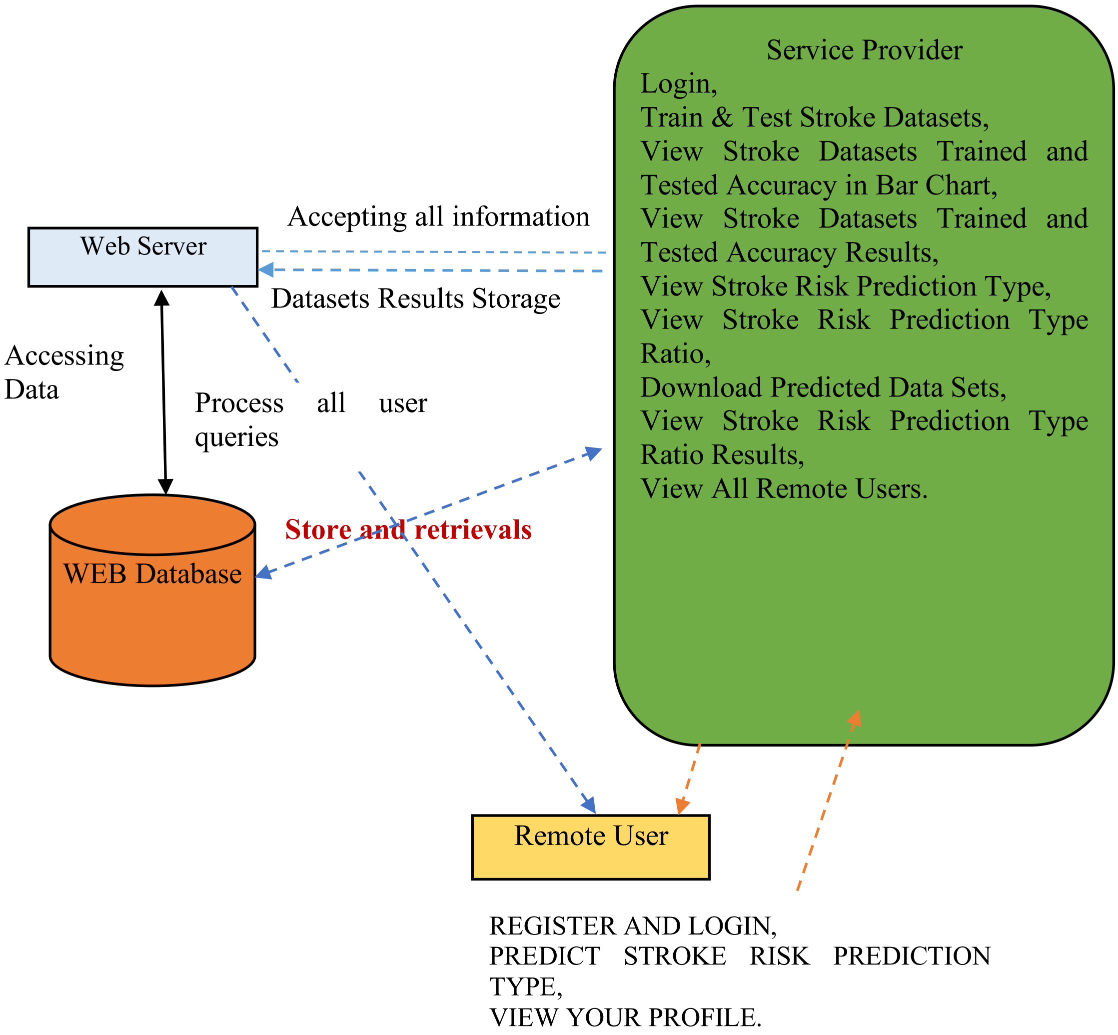

Figure 1. Architecture of the system.

Although there is limited data on strokes, other common chronic conditions, such as diabetes and hypertension, have sufficient information and have shown a significant link with the occurrence of strokes in clinical investigations[5,6]. Most of the existing TL studies employ single transfer approaches, such as networking transmission[7,8], features translating[9,10], instance translating[11–13], and so on. In a recent study, a hybrid adapted-embedding strategy was proposed[14]. The study provided empirical evidence that hybrid transfer techniques outperform single transfer approaches. Alternatively, this study proposes a hybrid transfer technique that utilizes external stroke data to address the issue of label imbalance. The method incorporates the use of generative instance transfer and active selection. By ensuring the confidentiality of patient data, the process of creating instance transfer allows for the interchange of accurate and reliable fake stroke information, which can be used to train SRP models. Active choice of instance facilitates the move of the most relevant instances generated to the destination domain. Furthermore, the structure's training and inference processes are designed to be dispersed, enabling it to take advantage of the fast rate of data transfer as well as the low latency of the 5G/B5G cell phone network.

The proposed framework is better equipped to create a single responsibility principle (SRP) model. However, crucial factors for the model's effectiveness include the number of transmitted layers and the specific sequence of data from various sources domains. Due to the vastness of the search area, commonly used methods such as grid and arbitrary searches for parameter modification often prove to be unproductive. The Poisson process is commonly utilized in Bayesian optimization (BO), a technique for globally optimizing black box operations, due to its user-friendly nature and ability to create a probabilistic representation for the function in question. The optimum value of the SRP model is calculated using Bayesian optimization (BO).

2. Existing system

Lim et al.[15] used retinal images to construct an SRP model employing a deep learning methodology, which yielded satisfactory results. Nonetheless, the process of collecting retinal images is sometimes expensive and time-consuming. Based on the aforementioned research, it can be deduced that the dataset is imbalanced and contains a smaller amount of stroke data. Pereira et al.[16] created a deep neural network model to forecast strokes using a dataset for 43,400 medical records. The dataset had an imbalance issue, with only 783 incidences of stroke. The purpose of the model was to tackle this imbalance problem. However, the solutions outlined before are not effective in addressing the issue of stroke information disparity and small data issues.

Centralizing stroke data from many hospitals can enhance the stroke risk prediction (SRP) model, particularly when nearby medical facilities have little data on strokes. The notion of secure A coalition with computation (SMC) involves the utilization of multiple party data to ensure the protection of data privacy. Nevertheless, it necessitates that each participant possess knowledge only of their individual input and output, a task that proves to be difficult to do. Federated learning[17] is a recently suggested architecture that also safeguards multi-party data privacy. By using many dispersed data sets, it may provide a more precise model. Federated learning employs encryption throughout the training process to ensure that data from different places is not shared. Nevertheless, the issue of insufficient data cannot be effectively addressed by the implementation of a federated learning system. Curiously, the SRP software has not employed any of the previously listed techniques.

The crucial determinant of the HDTL-SRP model’s efficacy is the settings of the net weight transference module. Historically, the optimization of hyperparameters in neural networks has been accomplished by grid search with manual search[18]. Random search[19] outperforms grid search in finding superior models under the same time constraints by exploring a larger and less favorable configuration space. Alternatively, they rely on the expertise of the specialist or make a subjective decision to establish the suitable hyperparameter. The genetic technique (GA), an evolutionary search technique, is employed to tackle optimization issues[20,21]. Nevertheless, genetic algorithm (GA) often yields suboptimal optimization efficiency due to its reliance on a substantial number of initial sample points for optimization of hyper-parameters. Bayesian optimization (BO) is a versatile approach for optimizing hyperparameters. It has been effectively utilized in the herb spearmint system and relies on Gaussian processes. The system does not utilize hybrid deep learning algorithms for the purpose of forecasting stroke risk. The transfer of stroke data from an external source does not occur in an already present system while generating an instance.

3. Advanced system

The proposed framework possesses a greater capability to generate single responsibility principle (SRP) models. However, crucial factors that significantly impact the model’s effectiveness include the quantity of moved layers and the specific order in which separate source domains are transferred. Due to the vastness of the search space, commonly used methods such as grid or randomized searches for parameter modification often prove to be inefficient[22]. The Gaussian process is frequently employed in Bayesian optimization (BO), a method for globally optimizing Blackbox functions, due to its user-friendly interface and ability to create a probabilistic representation of the function that is being optimized[23]. The optimum parameter of the SRP model is calculated using Bayesian optimization (BO). This work has made the following contributions:

This study introduces a new framework called hybrid deep add it learning-based a stroke the danger forecasts (HDTL-SRP) to train the SRP algorithm at a nearby healthcare facility with limited and imbalanced stroke data. The empirical evidence shows that, when applied to artificial as well as practical circumstances, HDTL-SRP outperforms its counterparts and illustrated with class diagram as **Figure 2** and mentioned the use case diagram as **Figure 3**.

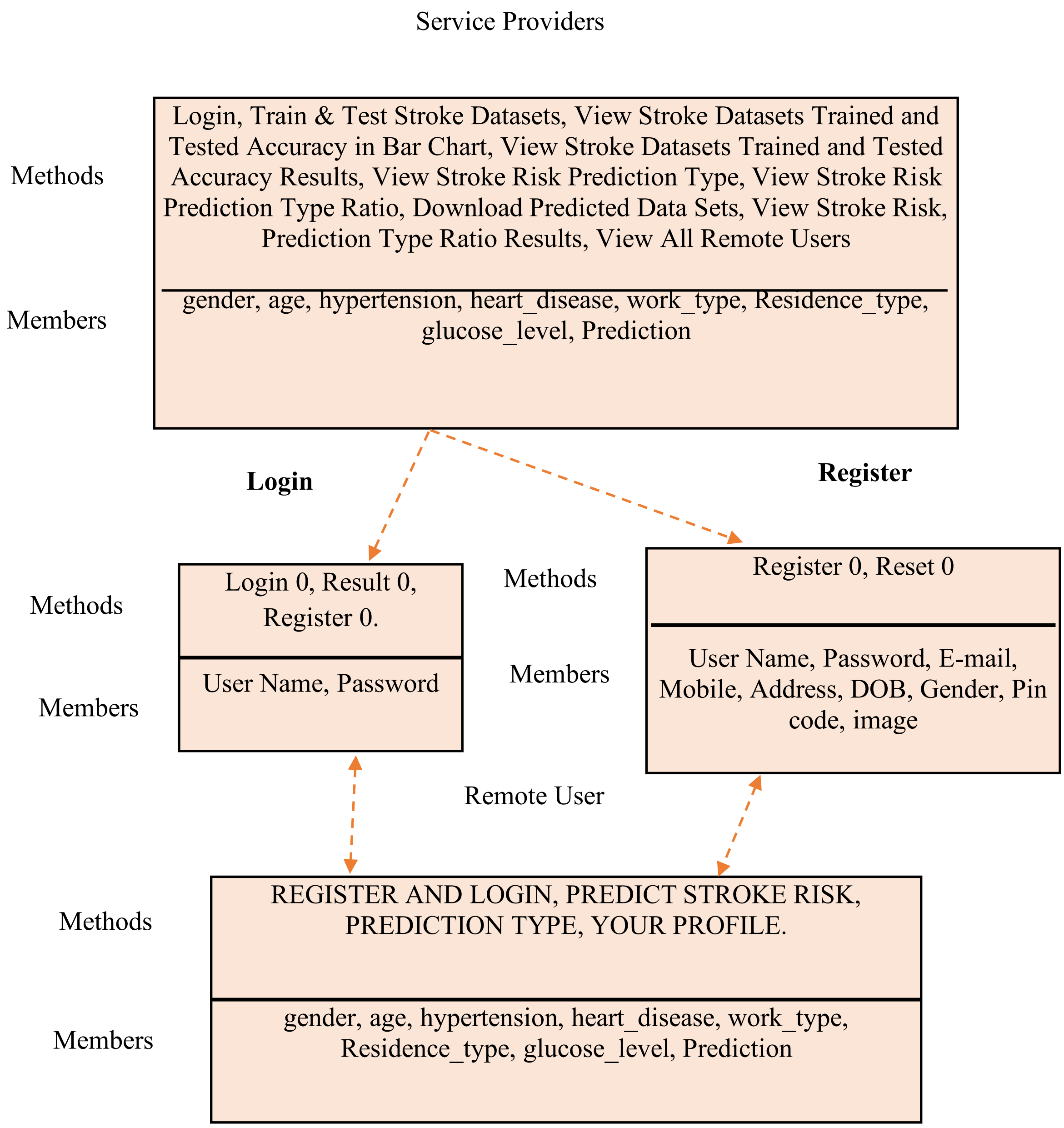

Figure 2. Class diagram.

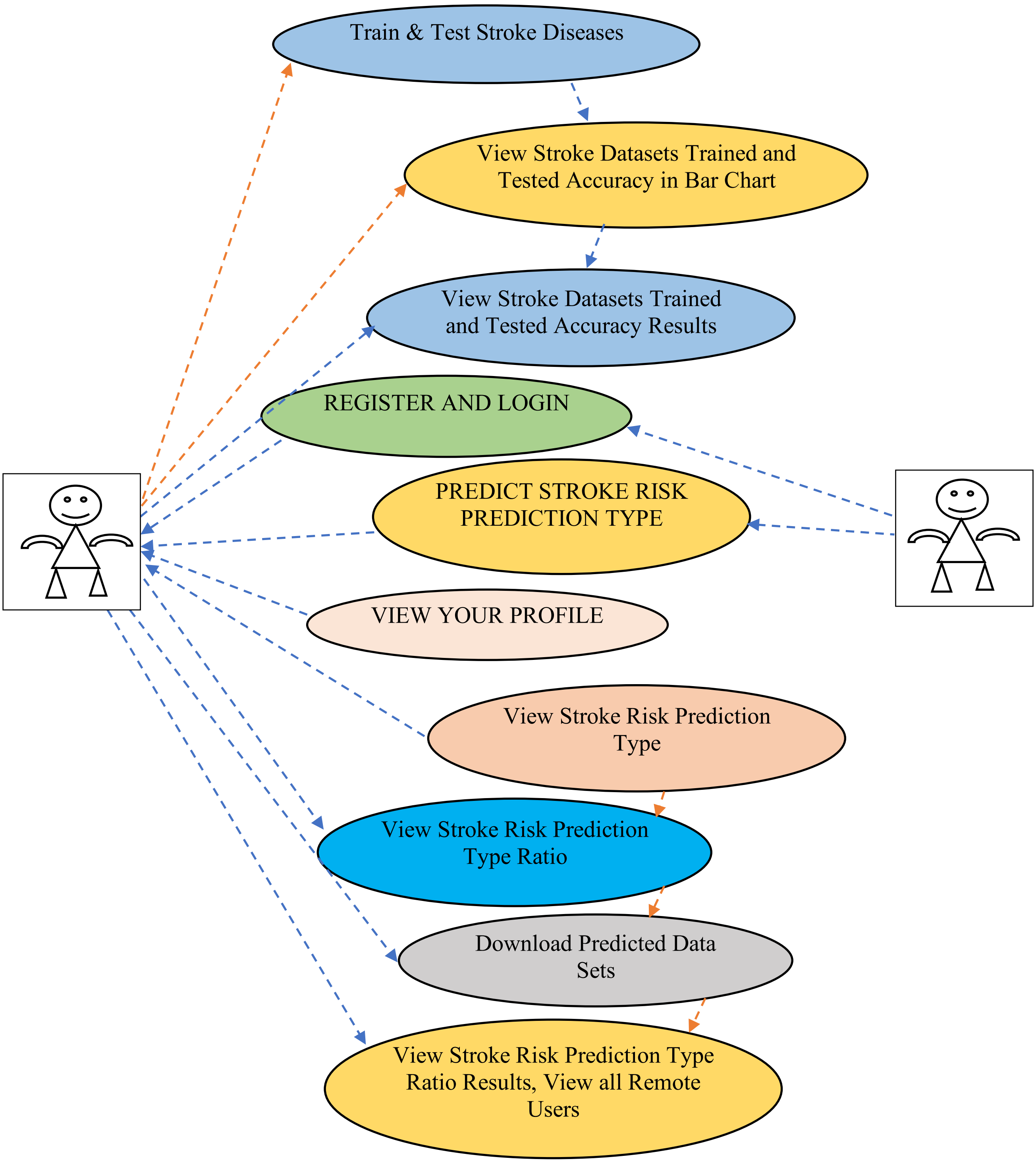

Figure 3. Use case diagram.

The HDTL-SRP architecture may be used in several hospitals to exploit the distributed medical data while upholding patient confidentiality. Moreover, the effectiveness of this distributed framework may be improved if the 5G/B5G network is available across all hospitals.

The technique aims to reproduce the data samples used to train a computerized classifier, with the goal of obtaining exact and highly accurate outcomes. The proposed method employs Bayesian optimization with artificial neural classifiers to optimize the network architecture across several sources.

Combinatorial processes refer to a set of step-by-step instructions that are followed to solve a problem or complete a task. Execution, in this context, refers to the actual carrying out of these instructions.

3.1. Classification models utilizing decision trees

Decision trees analyzers are widely employed in several areas. Their main strength comes in their capacity to derive complete decision-making insights from the given data.

The next sequential action. Consider T as a test with many potential results, labeled as $O1$, O2, & so on. Each member in set S relates to a unique result for test T, resulting in the division of S into subsets $S1$, $S2$, ..., Sn. Every component in subset Si corresponds to result Oi for test T. The variable T functions as the main node of the choosing tree represented in **Figure 4**.

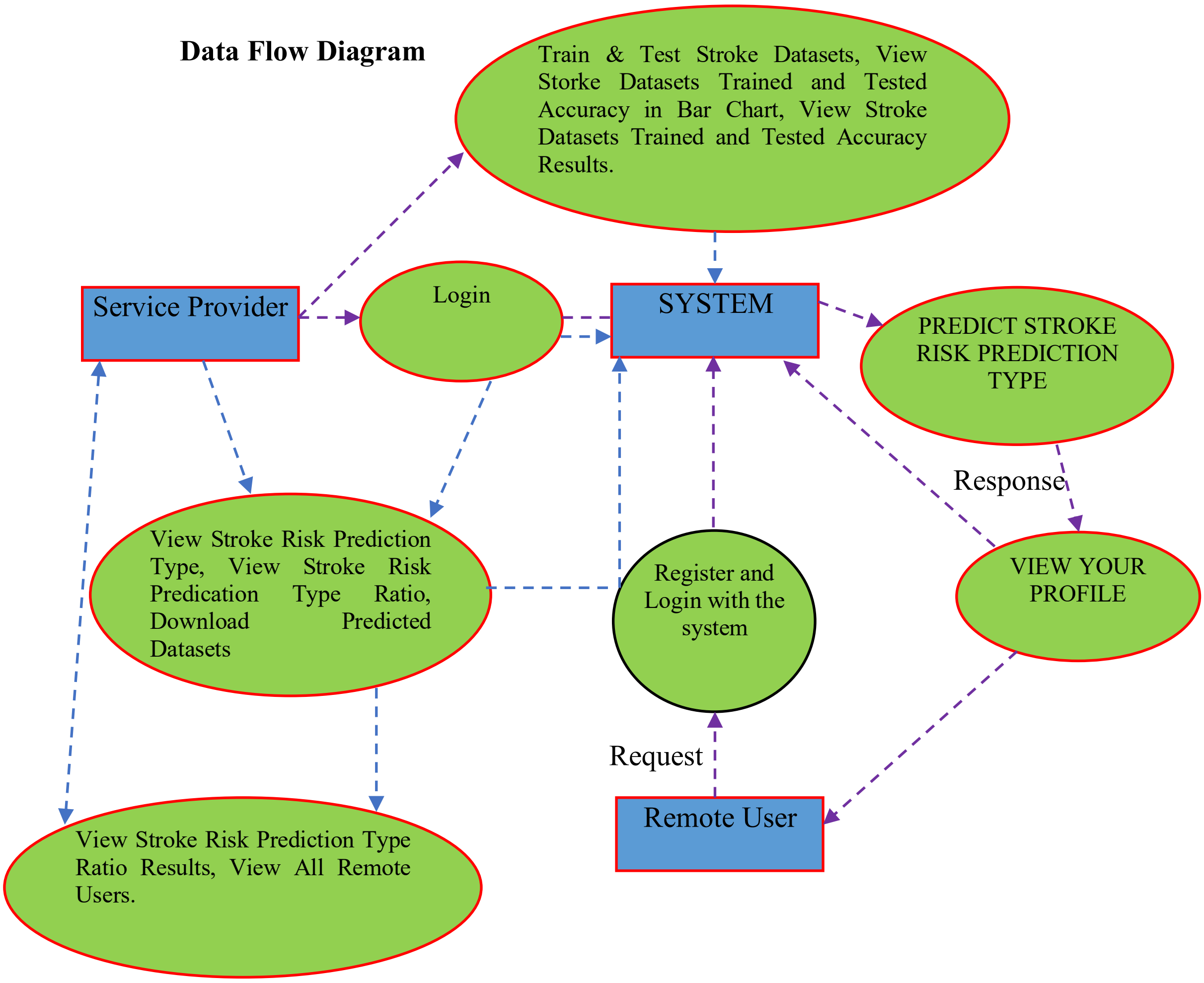

Figure 4. Data flow diagram.

3.2. Asymmetric is a method used in machine learning

The system offers a prognostic model including an assemblage of feeble prognostic models, often in the shape of decision trees. The user's text has two citations,[24] and[25]. However, it sets itself apart by allowing the optimization of any loss function that can be differentiated.

3.2.1. K-nearest neighbors (KNN)

1) This classification algorithm is simple yet extremely efficient.
2) Isolation made by measure is similar.
3) Non-parametric denotes a statistical technique that assumptions not relayed on the inherent pattern of the data.

3.2.2. Passive learning

Knowledge is not acquired until the assessment instance is presented.

Visual representation:

1) The training dataset consists of the k-nearest examples in the characteristic space.
2) The term "characteristic space" denotes a space consisting of non-metric variables that serve as classification parameters.
3) Learning is based on instances and functions in a lazy manner, since examples that are comparable to the vectors of inputs for testing or predictions takes some period be included in the already trained dataset.

3.3. Logistic regression classifiers

The method of logistic regression is used to examine the interaction between a variable that is categorical and a group of independent variables. Logistic regression pertains to the scenario when the variable under investigation is binary, meaning it can only assume two distinct values, including 0 and 1 or Yes and Nay. The logistic regression technique is used when the variable of interest has at least three distinct values, such as engaged, separated, divorcing, or widowed as well while the data utilized for a parameter that relies on numerous imputations may vary across different types of data, the practical implementation of the approach remains consistent. Logistic regression, or LR, is an alternative approach to discriminant analysis for examining variables with categorical responses. Logistic regression is often regarded by statisticians as a more adaptable and appropriate method for modeling various scenarios, in comparison to discriminant analysis. Logistic regression differs from discriminant analysis in that it by not requiring that every variable that is independent follows a normal distribution.

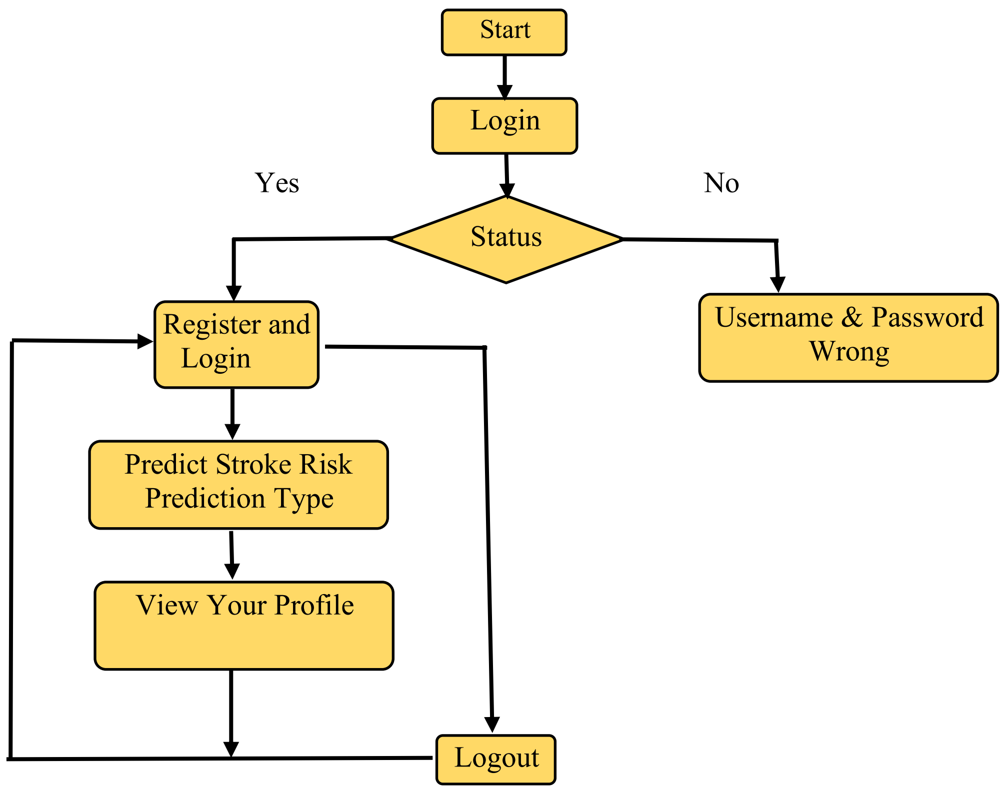

Figure 5. Flow chart of remote user.

This program does binary, logistic, or linear multinomial regression by utilizing independent variables that can be either numeric or categorical. The report provides comprehensive information on the model of regression, quality of fit, risks ratios, intervals of trust, probability, and deviance. The investigation comprises a thorough examination of residuals, which includes analyzing diagnostic reports and images. The software independently looks for a selected group of variables that are autonomous to identify the optimal regression structure with the minimum number of variables that are independently determined and represented in **Figure 5**. This tool offers confidence intervals for expected values and showcases ROC curves to aid in identifying the optimal threshold for categorization. It simplifies result confirmation by proactively categorizing unnecessary rows throughout the inquiry.

3.4. Naïve Bayes

The Naïve Bayes technique is a form of supervised learning that assumes independence between the presence or absence of a particular feature within a class and the presence or absence of any other attributes.

However, in spite of this reality, it showcases resilience and effectiveness. This technology has comparable outcomes to other supervised education methods. Different justifications have been put out in the scholarly literature. The purpose of this lesson is to clarify the concept is representational bias by providing a specific and detailed explanation. The ignorant Bayes classifier, basic classifiers model, logistical regression (LR), and nonlinear support vector machine model (SVM) are all examples of linear classifier. The difference is in the approach used to determine the characteristics of the classifier, often known as the learning bias technique.

Naive Bayes classification strategy

Although a Naive Bayes classification strategy is commonly employed in studies, it is not typically adopted by practitioners seeking practical and valuable outcomes. Researchers have demonstrated that this particular method is highly versatile and practical for execution. The requirements of the system in question may be easily approximated, and it exhibits rapid learning capability even when handling extensive databases. Furthermore, its accuracy is significantly superior when compared to alternative methods. However, the end customers are not given an example that is easily comprehensible and actionable, and as a result, they struggle to fully comprehend the significance of this strategy.

Therefore, we offer a novel method for obtaining information. The classifier is more comprehensible, and its use is likewise simplified. The initial segment of this training is devoted to elucidating the theoretical facets of the novice's Bayes classifier. Afterwards, we apply the method on a dataset using Tanagra. In order to evaluate the findings acquired from the model, we contrast them to the results produced by other linear approaches, such as logistic regression, linear discriminant analysis, and linear vector machines (SVM). The results exhibit a substantial level of coherence. This is the primary aspect that contributes to the outstanding success of the strategy in comparison to other ways. In the second section, we employ other software tools on the identical dataset, including Weka programme 3.6.0, R 2.9.2, the Knime program programme 2.1.1, Tangerine 2.0b, and RapidMiner, also 4.6.0. The primary goal is to comprehend the acquired outcomes mentioned in the **Figure 6**.

FLOW CHART: Service Provider

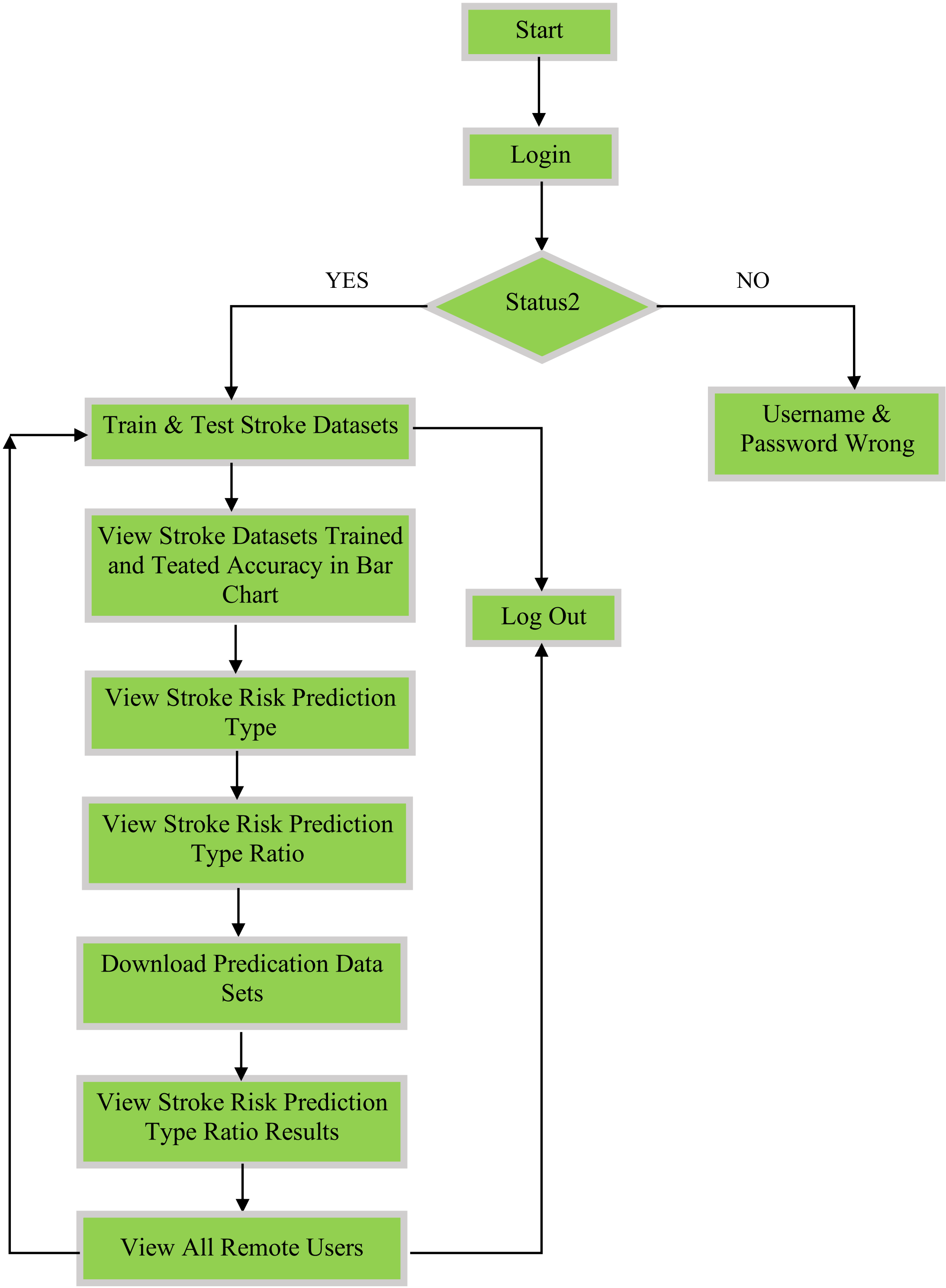

Figure 6. Flow chart of service provider.

3.5. Random forest

Random forest algorithms, also known as random choice forests, are a type of ensemble learning technique used for tasks involving classification and regression. The process involves creating multiple decision trees all through the training phase. The algorithm known as random forest assigns a class to a particular job, such as categorization, by picking the category that is most commonly chosen by a majority of the resulting trees. The outcome for tasks involving regression is the average estimate of the different trees. Randomised decision forests address the issue of decision trees' tendency to overfit with their training data. Random forests often exhibit superior performance compared to trees of choice, while their precision is lower than that of gradients boosted trees. Nevertheless, the system's performance can be influenced by the attributes of the data.

Tin Kam Ho developed the fundamental methodology for random forests of choices in 1995. This software utilized the random domain approach, which is a method for applying Eugene Kleinberg's "stochastic discriminate" classification method.

Leo Breiman, Jr. and Emilie Cutler created an enhanced version of the procedure, which they formally named "Unplanned Forests" in 2006. In 2019, Minitab, Inc. acquired ownership of the trademark. The update merges Breiman's "bagging" method with a random choice of features, initially suggested to produce a set of decision trees having regulated variance.

Random trees are commonly employed as "Blackbox" models in business environments due to their ability to produce reliable predictions across many datasets with minimal preparation. Support vector machine (SVM) models

Discriminant neural network approaches seek to identify discriminant functions that can accurately predict labels for fresh instances in classification tasks. This is accomplished by analyzing a training dataset that is both independently and evenly distributed (iid). In contrast to generative machine learning methods that involve the computation of dependent probability payments, a discriminant classification function straightforwardly assigns an individual data point x to one of the multiple groups in the sorting task. Discrimination techniques, albeit less powerful than generative algorithms, are more appropriate for detecting outliers in predicting tasks. Their advantage is in their capacity to employ minimal computational resources and training data, particularly in scenarios where the feature space is highly dimensional and just posterior probabilities are needed. Geometrically, the process of learning a classifier entail identifying the equation that precisely describes a surface in several dimensions. This equation enables efficient discrimination between various classes within the feature space.

The support vector machine (SVM) is a mathematical technique that addresses the convex optimization issue using analytical means. As a result, it consistently produces the same ideal hyperplane parameter. Genetic algorithms (GAs) and Perceptrons are commonly employed in machine learning for classification tasks, however they do not offer the same degree of dependability as the technique under consideration. The selection of beginning and finishing criteria has a substantial influence on the outputs of perceptual neurons. The SVM model variables are uniquely determined for a specific training set when training a particular kernel that maps information from the input domain to the feature space. However, the perceptron/GA classification models demonstrate variability every time the learning process is initialized. The main objective of neural networks (GAs) and perceptual neurons is to reduce the error during the training phase, resulting in the convergence of multiple hyperplanes that satisfy this criterion.

Research based on observation and experimentation:

Table 1 represents the data utilized to evaluate the proposed methodology is obtained from electronic healthcare records (EHR) systems at three medical centers situated within the same city. In collaboration with neurology specialists at the school, we have chosen 23 distinct characteristics, which encompass the use of the output portion of a DNN (deep neural network) utilizing the sigma function. The chosen instance will be included into the desired domain to systematically provide a unified compilation of stroke data. The study examined the crucial medical data and blood test results of participants between January 2014 and December 2018 is illustrated in **Figure 7**.

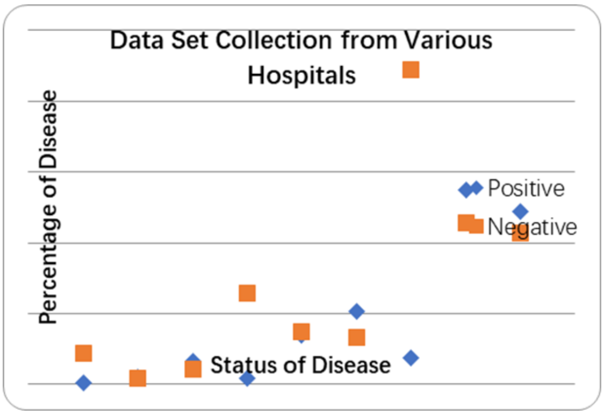

Figure 7. Details of data collections from various hospitals.

Table 1. Data collection sets.

Disease	Doman	Positive	Negative
Stroke	Target	128	2159
Hypertension	Source	529	469
Diabetes	Source	1638	1056
Stroke	Source	415	6477
Hypertension	Source	3502	3752
Diabetes	Source	5168	3318
Stroke	Source	1883	22,207
Hypertension	Source	13,769	11,414
Diabetes	Source	12,196	10,711

The assessment of correctness (Acc) is employed to scrutinize the effectiveness of detection. Nevertheless, there was a conspicuous disparity in the stroke-related datasets, with a greater abundance of death data in contrast to non-survival data. Consequently, we included a further three metrics such as sensitive (Sn), specific (Sp), and positive prediction value (PPV). The Sn metric quantifies the likelihood of accurately identifying cases of non-survival, while the Sp meter measures the chance of properly detecting cases of survival. PPV denotes the likelihood of accurately detecting the absence of survival. The term "real positive" (TP) refers to the accurate identification of stroke occurrences, while "true negative" (TN) indicates the exact prediction of the absence of strokes. The FP (false affirmative) and FN (false positive) rates represent the occurrence of inaccurate forecasts for strokes and non-strokes, correspondingly. The equations utilized to compute sensitivities (Sn), particularity (Sp), positivity predictive value (a PPV), and reliability (Acc) were as follows: The sensitivity (Sn) may be calculated using the formula Sn = TP/ (TP + FN), where TP represents the number of true positives and FN represents the number of false negatives. The amount of specificity (Sp) is determined by dividing the number of genuine negatives (TN) by the total of true positive and false positives (TN + FN). The positive predicting value (PPV) may be calculated using the formula TP/ (TP + FP), where TP represents the number of true positive cases and FP represents the number of false positive cases. The precision of the model used for classification is determined by the formula Acc = (TP + TN)/ (TP + FN + FP + TN).

In order to enhance the discriminator, we calculate the stochastic gradients with regard to Θd and incrementally update it. The phrase is equivalent to the summation of 1 split by the square of the root of i, where i begins at 1. The equation is the addition of the logarithmic of $D(Xi)$ and the logarithmic of 1 minus $D(G(z))$.

Where can I locate authentic occurrences of strokes?

The generator refines its settings by iterative updates using the random gradient descent process, specifically tailored to minimize the engine's loss function.

The statement is equivalent to the summation of the logarithms of the difference between one and the values generated by applying the function *G* on the input *Z*.

Therefore, the evaluation of risk for stroke prediction relies on examining the distribution of relevant health data in a graphical style, which exposes an unequal distribution. Our effort aims to tackle the difficult task of predicting attacks at smaller medical facilities that face limitations due to a lack of sufficient and imbalanced stroke data. However, there is valuable information that may be obtained from alternative medical establishments.

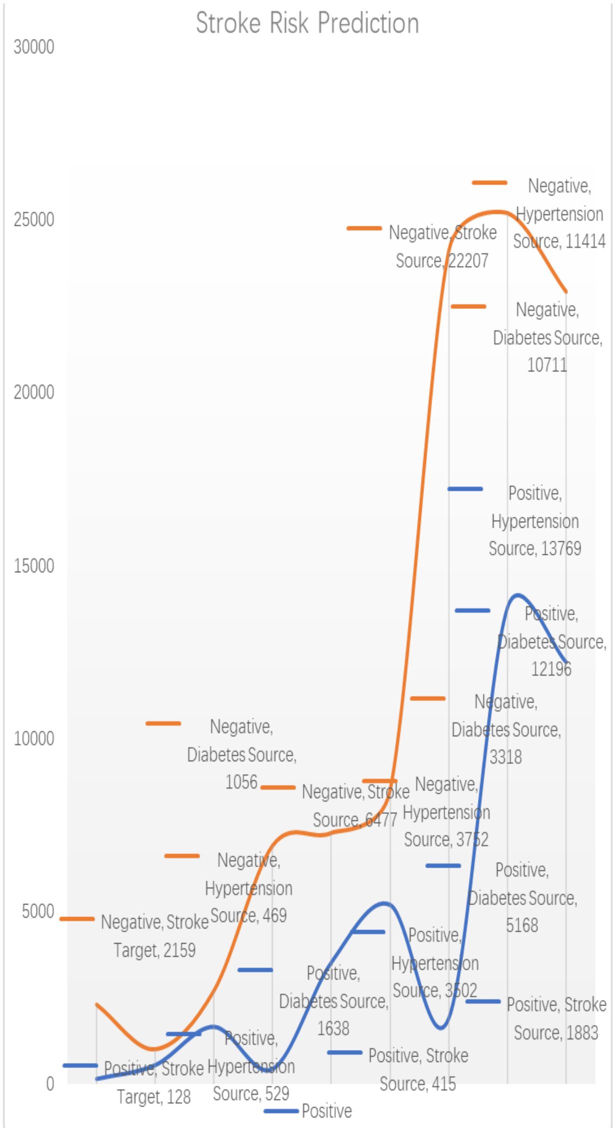

Figure 8. Stroke risk prediction analysis.

Figure 8 mentioned the performance metrics utilized for evaluating predictions are obtained by comparing projected outcomes to actual values. To be more specific, we may calculate the precise true positive (TP), true negative (TN), false positive (FP), and false negative (FN), among other figures for the test results set. Subsequently, the effectiveness of the suggested method is assessed based on four metrics: (1) Accuracy is calculated by dividing the total number of true positives and true negatives by the sum of actual positives, misleading negatives, erroneous positives, and the number of false negatives. (2) Recall is calculated by dividing the number of true positives by the sum of true positives and false negatives. The *F*1 rating is a metric that integrates accuracy and recall. The area under the curve of ROC (AUC) is a measure utilized to assess the effectiveness of a model used for classification. The numbers of 0 and 1 represent the lowest and highest efficiency, each, for all the criteria provided. In order to assess the effectiveness of BO, we employ the verification error, which measures the degree of misclassification in the validation set represented in **Table 2**.

SVM optimizes the "margin" between either positive or negative observations in a conceptual fashion. The stroke forecasting issue may be defined as the job of predicting the probability of a stroke happening within a specific time frame is represented in **Figures 9** and **10**. The problem at hand may be formulated as a task involving binary classification and can be efficiently tackled using supporting vector machines (SVM).

Table 2. Performance of network weight transfer learning versus Imbalanced stroke data.

SRP method	Accuracy	Recall	*F*1 score	AUC
SVM	0.695 ± 0.021	0.585 ± 0.017	0.555 ± 0.005	0.701 ± 0.020
DT	0.512 ± 0.015	0.563 ±0.018	0.702 ± 0.019	0.711 ± 0.015
RF	0.675 ± 0.013	0.332 ± 0.024	0.504 ± 0.027	0.685 ± 0.010
DNN	0.719 ± 0.016	0.468 ± 0.028	0.611 ± 0.026	0.576 ± 0.023
DNN+NWT	0.629 ± 0.019	0.491 ± 0.026	0.625 ± 0.026	0.781 ± 0.027

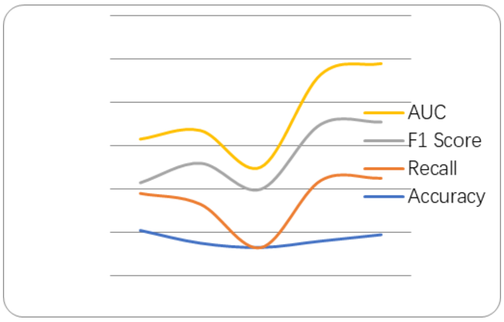

Figure 9. Imbalanced stroke data.

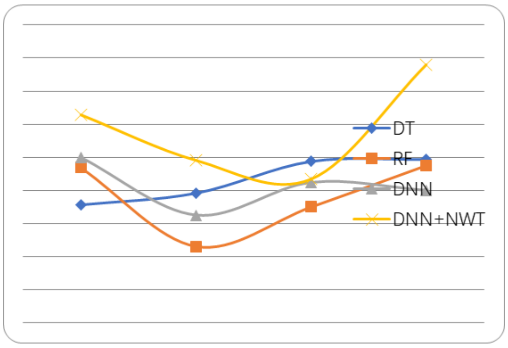

Figure 10. Balancing stroke information.

4. Conclusion

This work has primarily aimed to address the difficulties related to stroke detection rate (SRP) by utilizing a restricted and unequal dataset of stroke cases. We have developed a cutting-edge method known as hybrid direct transfers learning-based vascular risks prediction (HDTLSRP), which consists of three fundamental elements: (1) The term "generative" refers to something that has the ability to produce or create. A case transfers global information tracker (GIT) is a method that involves the external sharing of stroke data around different healthcare facilities while ensuring anonymity. We employed the data sets to build a model based on deep learning that integrates scaled PCA. This algorithm is specifically built to autonomously forecast occurrences of strokes by analyzing the medical use history and health behaviors of the participants.

The model did not include any variables that were based on personal opinions or feelings. Our research facilitates the early detection of individuals at a heightened risk of assault, necessitating additional evaluations and appropriate treatment to prevent the deterioration of their condition. Our method eliminates the necessity for manual variable selection by humans. Due to the straightforward nature of the information being provided, we utilized a hybrid deep learning with transfer learning methodology to analyze the factors of interest. The sensitiveness, specificity, & AUC (area under the curve) values for our technique were 65.32%, 87.56%, and 85.48%, respective. Our system is not only suitable for predicting ischemic stroke with limited data, nevertheless also for detecting other diseases.

Author contributions

Conceptualization, AR and KS; methodology, AR; software, AR; validation, AR, KS and RS; formal analysis, PVRDPR; investigation, KS; resources, RS; data curation, PVRDPR; writing—original draft preparation, AR; writing—review and editing, RS; visualization, KS; supervision, AR; project administration, PVRDPR. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. Koton S, Schneider ALC, Rosamond WD, et al. Stroke Incidence and Mortality Trends in US Communities, 1987 to 2011. JAMA. 2014; 312(3): 259. doi: 10.1001/jama.2014.7692
2. Benjamin EJ, Muntner P, Alonso A, et al. Heart Disease and Stroke Statistics—2019 Update: A Report From the American Heart Association. Circulation. 2019; 139(10). doi: 10.1161/cir.0000000000000659
3. Cheon S, Kim J, Lim J. The use of deep learning to predict stroke patient mortality. International Journal of Environmental Research and Public Health. 2019; 16(11): 1876. doi: 10.3390/ijerph16111876
4. Sun C, Shrivastava A, Singh S, et al. Revisiting unreasonable effectiveness of data in deep learning era. In: Proceedings of the IEEE International Conference on Computer Vision; 2017. pp. 843-852.
5. O'Brien AA, Rajkumar C, Bulpitt CJ. Blood pressure lowering for the primary and secondary prevention of stroke: treatment of hyper tension reduces the risk of stroke. Journal of Cardiovascular Risk. 1999; 6(4): 203-205. doi: 10.1177/204748739900600402
6. Manson JE, Colditz GA, Stampfer MJ, et al. A prospective study of maturity-onset diabetes mellitus and risk of coronary heart disease and stroke in women. Archives of Internal Medicine. 1991; 151(6): 1141. doi: 10.1001/archinte.1991.00400060077013
7. Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks? Advances in Neural Information Processing Systems. 2014; 27: 3320-3328.
8. Huang JT, Li J, Yu D, et al. Cross language knowledge transfer using multilingual deep neural network with shared hidden layers. In: Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing; 2013. pp. 7304-7308.
9. Raina R, Battle A, Lee H, et al. Self-taught learning: transfer learning from unlabeled data. In: Proceedings of the 24th International Conference on Machine Learning; 2007. pp. 759-766.
10. Dai W, Xue GR, Yang Q, et al. Co-clustering based classification for out-of-domain documents. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2007. pp. 210-219.
11. Cheplygina V, Pena IP, Pedersen JH, et al. Transfer learning for multicenter classification of chronic obstructive pulmonary disease. IEEE Journal of Biomedical and Health Informatics. 2018; 22(5): 1486-1496. doi: 10.1109/jbhi.2017.2769800
12. Xu Y, Pan SJ, Xiong H, et al. A unified framework for metric transfer learning. IEEE Transactions on Knowledge and Data Engineering. 2017; 29(6): 1158-1171. doi: 10.1109/tkde.2017.2669193
13. Wan C, Pan R, Li J. Bi-weighting domain adaptation for cross-language text classification. In: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence; 2011. pp. 1535-1540.
14. Scott T, Ridgeway K, Mozer MC. Adapted deep embeddings: A synthesis of methods for k-shot inductive transfer learning. In: Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018); 2018; Montréal, Canada. pp. 76-85.
15. Lim G, Lim ZW, Xu D, et al. Feature isolation for hypo thesis testing in retinal imaging: an ischemic stroke prediction case study. Proceedings of the AAAI Conference on Artificial Intelligence. 2019; 33(1): 9510-9515.

10.1609/aaai. v33i01.33019510.
16. Pereira DR, Reboucas Filho PP, de Rosa GH, et al. Stroke lesion detection using convolutional neural networks. In: Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN); 2018. pp. 1-6.
17. Teoh D. Towards stroke prediction using electronic health records. BMC Medical Informatics and Decision Making. 2018; 18(1). doi: 10.1186/s12911-018-0702-y
18. Liu T, Fan W, Wu C. A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. Artificial Intelligence in Medicine. 2019; 101: 101723. doi: 10.1016/j.artmed.2019.101723
19. Wang F, Casalino LP, Khullar D. Deep learning in medicine—promise, progress, and challenges. JAMA Internal Medicine. 2019; 179(3): 293. doi: 10.1001/jamainternmed.2018.7117
20. Lee J, Sattigeri P, Wornell G. Learning newtricks from old dogs: Multi-source transfer learning from pre-trained networks. Advances in Neural Information Processing Systems. 2019; 32: 4372-4382.
21. Chen H, Ni D, Qin J, et al. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE Journal of Biomedical and Health Informatics. 2015; 19(5): 1627-1636. doi: 10.1109/jbhi.2015.2425041
22. Benjamin EJ, Blaha, MJ, Chiuve, SE, et al. Heart disease and stroke statistics—2017 update: A report from the American Heart Association. Circulation. 2017; 135(10): e146-e603.
23. Khosla A, Cao Y, Lin CCY, et al. An integrated machine learning approach to stroke prediction. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining; 2010. pp. 183-192.
24. Monteiro M, Fonseca AC, Freitas AT, et al. Using machine learning to improve the prediction of functional outcome in ischemic stroke patients. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2018; 15(6): 1953-1959. doi: 10.1109/tcbb.2018.2811471
25. Sung SF, Lin CY, Hu YH. Emr-based phenotyping of is chemic stroke using supervised machine learning and text mining techniques. IEEE Journal of Biomedical and Health Informatics. 2020; 24(10): 2922-2931. doi: 10.1109/jbhi.2020.2976931