Motyka Similar Feature Selected Softsign Deep Neural Classification For Stroke Disease Prediction

Dr. D. Nalini

Assistant Professor, Department of Computer Science, Kurinji College of Arts and Science (Affiliated to Bharathidasan University), Trichy-2

ABSTRACT

Stroke is the main etiology of long-term disability. Data mining techniques were introduced globally for performing the accurate stroke occurrence prediction depending on the risk factors associated with patient data. Patient data comprises the number of features where the irrelevant gets discarded to improve the prediction accuracy performance. Different feature selection and classification techniques were carried out to improve the stroke disease prediction performance. But, the existing techniques failed to improve the prediction accuracy and reduce the time consumption during stroke disease prediction. In order to address these problems, Motyka Similar Feature Selected Softsign Deep Neural Classification (MSFSSDNC) Method is introduced. The key aim of MSFSSDNC Method is to perform an efficient stroke disease prediction with higher accuracy and lesser time consumption. MSFSSDNC Method comprises four layers, namely one input layer, two hidden layers and one output layer. MSFSSDNC Method collects the input patient data in the layer. MSFSSDNC Method performed Motyka Similarity Index Feature Selection in hidden layer 1. Motyka Similarity Index Feature Selection is carried out to find the relevant features (i.e., important risk factors) that influence the stroke disease with lesser time complexity. Motyka Similarity Index Feature Selection reduces the overfitting problems and increases the classification ability of MSFSSDNC Method to attain enhanced stroke disease prediction results. After that the selected features are sent to the hidden layer 2. In that layer, MSFSSDNC Method employed the softisgn activation function for classifying the patient data for performing the stroke disease diagnosis. Finally, the obtained results are transmitted to the output layer. This in turn helps to improve the stroke disease diagnosis performance. An experimental evaluation of MSFSSDNC Method is carried out using Stroke Prediction Dataset on factors such as prediction accuracy, prediction time, and false positive rate with respect to number of patient data. The experimental results depicts that the proposed MSFSSDNC Method increases the accuracy and minimizes the time of stroke disease prediction when compared to state-of-the-art works.

Keywords: Stroke, softisgn activation function, Motyka Similarity Index Feature Selection, softisgn activation function

1. INTRODUCTION

Stroke disease is considered as the second leading cause of death. According to the World Health Organization (WHO), the estimated death from cardiovascular diseases gets increased to 17.7 million people in 2017 and about 6.7 million died due to the stroke disease. A detection and prevention of stroke is an important one to avoid adverse consequences. Many techniques were introduced for performing the

stroke disease prediction. Detecting Risk Factor of Stroke Disease (DRFS) methodology was introduced in [1] to identify the symptoms related with stroke disease and preventive measures from social media content. A new architecture was constructed for grouping the tweets depending on content through spectral clustering in iterative manner. But, the prediction accuracy was not improved by DRFS methodology.

A new rough-set based technique was designed in [2] for ranking the significance of different electronic healthcare records of patient in detecting the stroke disease. The designed technique was employed on any dataset with binary feature sets. However, the time consumption was not reduced by designed technique. An integrated machine learning approach was introduced in [3] to perform subtype ischemic stroke classification. The designed approach performed the feature selection and prediction in medical data. However, the computational complexity was not minimized by designed approach.

The regularization terms were combined with the standard cross-entropy loss function in [4] to reduce the false positive and false negative predictions. The main objective was to diagnosis the stroke within one year of patient last results or medical diagnoses. But, the computational complexity was not minimized [19]. The custom regularization terms had positive effect on the training process. Intuitionistic Fuzzy Based Decision Tree was constructed in [5] to perform different stroke disease diagnosis. Hamming Distance determine the difference between values on similar variable. Intuitionistic fuzzy based decision tree presented large amount of information to stakeholders and utilized the linguistic terms for obscurity, ambiguity, and hesitation in human perception. However, the false positive rate was not minimized by Intuitionistic Fuzzy Based Decision Tree. ASPECTS correlation was carried out in [6] with mortality and morbidity in patients in acute middle cerebral artery territory infarction for determining cutoff value of ASPECTS to predict the results. But, the feature selection process was not carried out for efficient stroke disease prediction.

An intelligent system was introduced in [7] to forecast the stroke risk in rapid manner. The designed system predicted the stroke risk rate depending on fuzzy cognitive map and nonlinear Hebbian learning algorithm. However, the time complexity was not minimized by intelligent system. The machine learning techniques like logistic regression, support vector machine, random forest classifier, extreme gradient boosting (XGBoost) and fully connected deep neural network were developed in [8] for performing the prediction. The designed techniques were evaluated through area under curve of receiver operating characteristic on testing data. But, the stroke disease prediction accuracy was not improved by machine learning techniques.

A new prototype was introduced in [9] to categorize the stroke with help of text mining tools and machine learning algorithms. Machine learning was essential one in different areas such as surveillance, medicine and data management. But, the computational cost was not reduced. A stroke prediction system was introduced in [10] to identify the stroke with artificial intelligence (AI). Machine learning and deep learning techniques were introduced with collected data[20]. The essential features were extracted and prediction results were obtained depending on everyday activities. However, the time consumption for stroke prediction was not minimized at the required level.

The multiple cardiovascular and metabolic advantages were discussed in [11] with lifestyle behaviors changes like diet, physical movement, smoking cessation and alcohol consumption. Health

behavior theories were evaluated to identify the behavioral and cognitive skills for health care management. However, the error rate was not minimized during the stroke prediction process. The radiomics were combined with clinical factors in cranial computed tomography (CT) in [12] to forecast ischemic strokes in patients with silent lacunar infarction (SLI). A least absolute shrinkage and selection operator (LASSO) and cox regression analysis was carried out to choose the prognostic factors depending on ModelC with clinical factors. But, the computational complexity was not reduced.

The above mentioned existing techniques experienced many issues such as lesser feature extraction accuracy, higher time consumption, higher false positive rate, higher computational complexity, higher computational cost, lesser prediction accuracy, etc. In order to address these problems, MSFSSDNC Method is introduced in this article.

The main contribution of the work is given as:

- The objective of MSFSSDNC Method is to perform efficient stroke disease prediction with higher accuracy and lesser time consumption. MSFSSDNC Method collects the input patient data and performs Motyka Similarity Index Feature Selection.
- Motyka Similarity Index Feature Selection selects the relevant features with important risk factors that influence the stroke disease. Motyka Similarity Index Feature Selection reduces the overfitting problems and increases the classification ability of MSFSSDNC Method to attain enhanced stroke disease prediction results.
- MSFSSDNC Method employed the softisgn activation function for classifying the patient data for performing the stroke disease diagnosis. Finally, the obtained results are transmitted to the output layer. This in turn helps to improve the stroke disease diagnosis performance.

The rest of the paper is structured as: In Section 2, proposed methodology is explained with the neat architecture diagram. In Section 3, experimental settings are described and the performance result of MSFSSDNC Method is explained in Section 4. Section 5 shows the conclusion of the paper.

2. METHODOLOGY

Motyka Similar Feature Selected Softsign Deep Neural Classification (MSFSSDNC) Method is introduced for stroke disease prediction with higher accuracy and lesser time consumption. The MSFSSDNC Method combined the Motyka Similarity Index and Deep Learning Classifier. Motyka Similarity Index Feature Selection is used in MSFSSDNC Method to identify the important risk factors that influence the stroke disease with minimal time complexity. Motyka Similarity Index Feature Selection minimizes the overfitting issues and increases the classification ability of MSFSSDNC Method to obtain better stroke disease prediction results. MSFSSDNC Method used softisgn activation function to reduce the false positive rate of stroke disease diagnosis. This helps MSFSSDNC Method to attain improved softisgn activation function performance as compared to state-of-the-art works. The architecture diagram of the MSFSSDNC Method is illustrated in above Figure 1.
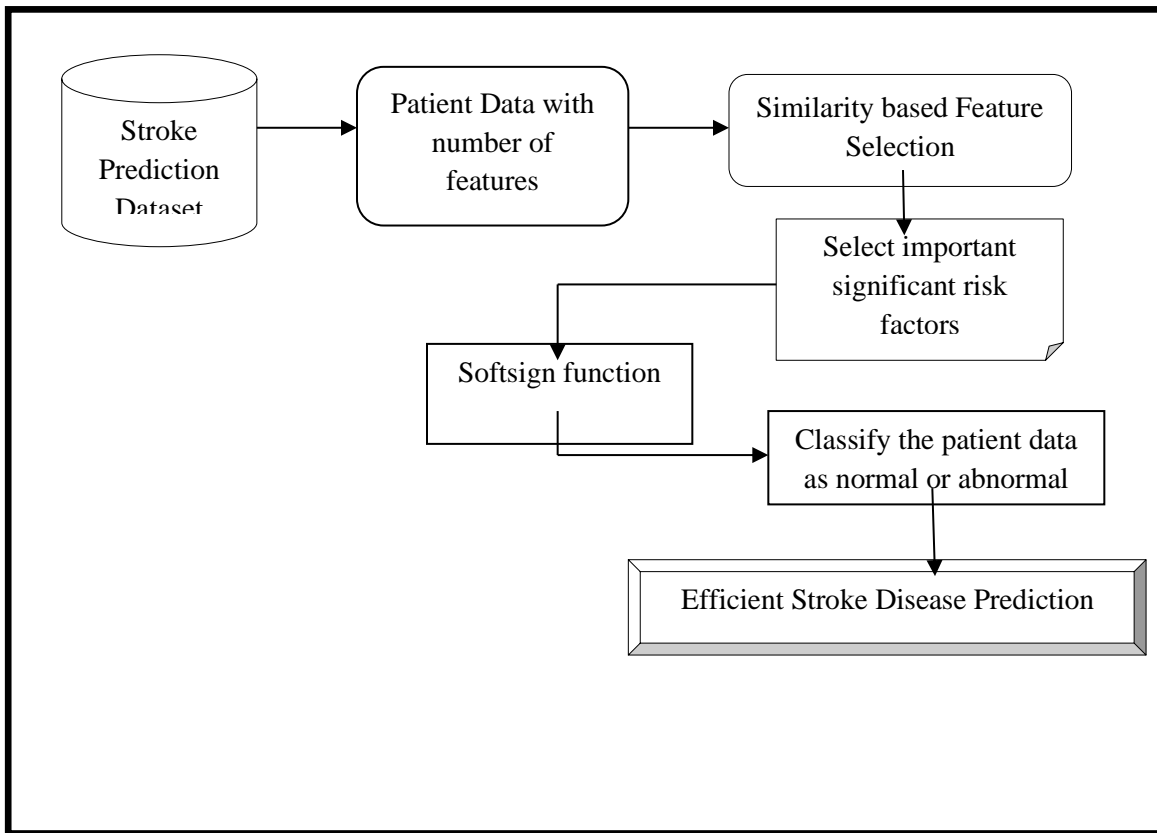
Figure 1 Architecture Diagram of the MSFSSDNC Method for Stroke Disease Prediction

Figure 1 illustrates the overall process of MSFSSDNC Method to obtain enhanced accuracy for stroke disease diagnosis with less number of risk factors. As illustrated in the above figure, the MSFSSDNC Method initially collects the medical data (i.e. Stroke prediction dataset) as input. The dataset includes the huge number of patient data '$Pad_1, Pad_2, \ldots Pad_n$' and their features '$f_1, f_2, \ldots f_m$'. The objective of MSFSSDNC Method is to identify the presence of stroke disease accurately with less number of features. Motyka Similarity Index used to select the relevant features for performing classification. Finally, Softsign activation function is used to identify stroke disease with higher accuracy. The detailed processes of MSFSSDNC Method are described in below subsections.

Motyka Similar Feature Selected Softsign Deep Neural Classification Method

Motyka Similar Feature Selected Softsign Deep Neural Classification (MSFSSDNC) Method is introduced for stroke disease prediction with higher accuracy and lesser time consumption. GFLDFSSDNC Method comprises the four layers, namely one input layer, two hidden layers and one output layer for performing stroke disease prediction process. The structure of deep learning classifier is illustrated in figure 2.
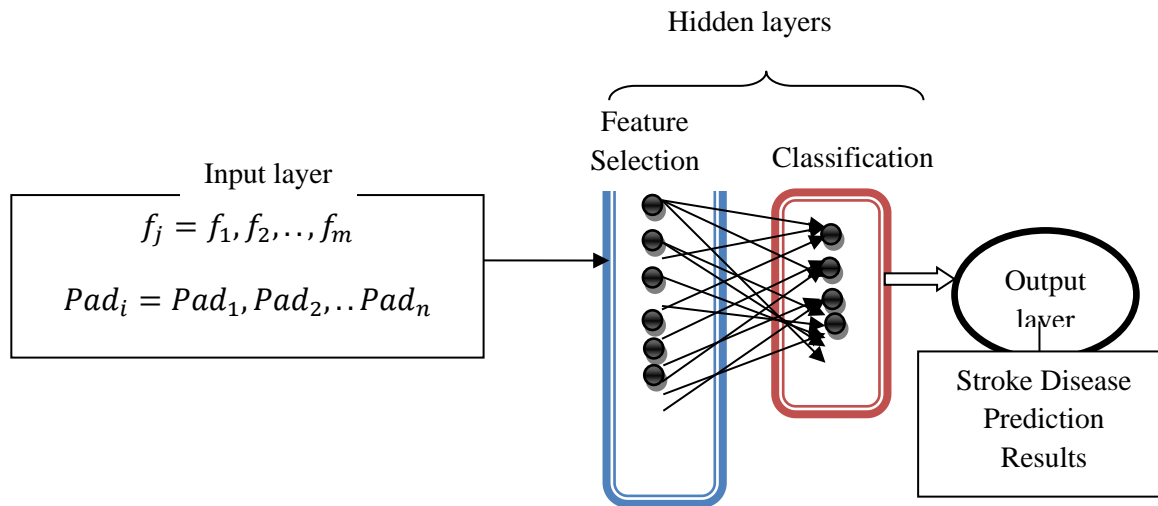
Hidden layers

Feature
Selection

Classification

Input layer

$$f_j = f_1, f_2, .., f_m$$

$$Pad_i = Pad_1, Pad_2, .. Pad_n$$

Output
layer

Stroke Disease
Prediction
Results

Figure 2 Structure of Deep Learning Classifier

The structure comprises multiple layers and the neurons like the nodes as shown in figure 2. In the Connectedness neural network, the nodes in one layer are connected to another layer and form the entire network. Let us consider the relevant features of patient data are considered as input and given to the input layer. The input of MSFSSDNC Method is denoted as '$Pad_1, Pad_2, .. Pad_n$' where 'n' denotes the total number of patient data. In input layer, the weight and bias are employed in MSFSSDNC Method. At starting, the weight is initialized and gets updated for every training error. The weights are represented by 'w_i, w_{ih}, w_{hh} and w_{ho}'. A bias neuron is used to identify the decision boundary to partition into normal data and abnormal data based on input patient data. Bias trained the algorithm faster with better quality. The input layer result is attained as,

$$Input = \sum_{i=1}^{n} pad_i * w_i + bias \qquad (1)$$

From (1), 'pad_i' denotes the patient data with the relevant feature. 'w_i' symbolizes initial weight allocated at the input layer. After that, the input layer result is transmitted to the hidden layer 1. Consequently, the MSFSSDNC Method chooses the subset of relevant features from input medical dataset. The Motyka Similarity Index Feature Selection minimizes data dimensionality through removing the redundant features for stroke disease prediction with higher accuracy. In MSFSSDNC Method, Motyka Similarity Index is used to identify the correlation between features and stroke symptoms[18]. The diagrammatic representation of Motyka Similarity Index Feature Selection is illustrated in figure 3.
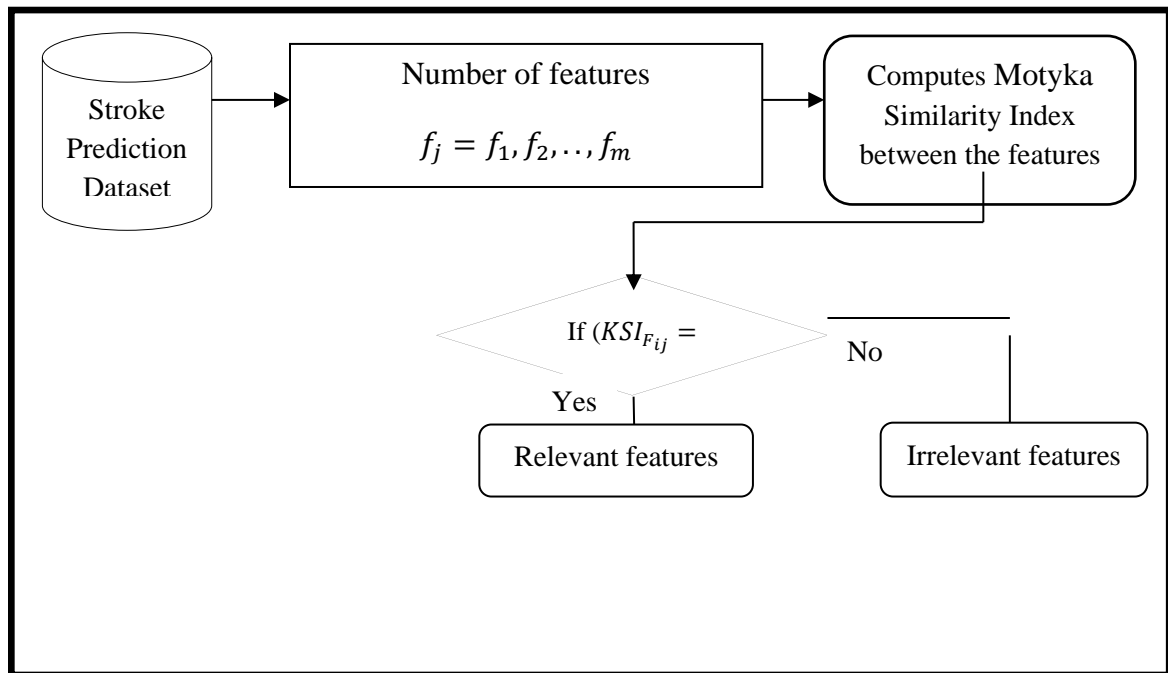
Figure 3 Motyka Similarity Index Feature Selection Process

From figure 3, the MSFSSDNC Method initially measures the Motyka Similarity Index for each medical feature in a given dataset. From that, the Motyka Similarity Index between features and stroke symptoms is determined as,

$$MSI(Pad_i, Sym_i) = \frac{\sum_{i=1}^{n} min\ (x_i, y_i)}{\sum_{i=1}^{n} x_i + y_i} \qquad (2)$$

From (2) 'Pad_i' represent the patient data with a number of features '$f_j = f_1, f_2, .., f_m$' whereas 'Sym_i' denotes the number of symptom features related to tumor disease. 'x' symbolize the number of features in both 'Pad_i' and 'Sym_i'. 'y' denote the number of features in patient data 'Pad_i' but not takes place in 'Sym_i'. The output of Motyka Similarity Index ranges from 0 to 1. When the similarity values lies between '0.5 to 1', it is considered as the relevant features. Otherwise, it is considered as the irrelevant features. From the result value, the features with higher Motyka Similarity Index value are considered as more relevant to perform the stroke disease prediction. After that, the relevant features are sent to the hidden layer 2. In hidden layer 2, a softsign activation function is used to identify the classification results for patient transaction data. The softsign activation function 'SAF' is given by,

$$SAF = \frac{MSI(Pad_i, Sym_i)}{1 + |MSI(Pad_i, Sym_i)|} \qquad (3)$$

From (3), the 'SAF' varies from the value -1 to $+1$. When the 'SAF' value lies between '0 to 1', then the patient data is considered as the normal patient data. When the 'SAF' value lies between '-1 to 0', then the patient data is considered as the abnormal patient data. The result obtained at the hidden layer is given as,

$$0 < SAF < 1 \rightarrow \text{Normal Patient data}$$
$$-1 < SAF < 0 \rightarrow \text{Abnormal Patient data}$$

$$\text{Hidden Layer} = \sum_{i=1}^{n} \text{Input} * w_i + [w_{ih} * MSI(Pad_i, Sym_i) + [w_{hh} * SAF] \tag{4}$$

$$\text{Hidden Layer} = \begin{cases} 0 < SAF < 1 \rightarrow \text{Normal Patient data} \\ -1 < SAF < 0 \rightarrow \text{Abnormal Patient data} \end{cases} \tag{5}$$

From (4) and (5), 'w_{ih}' denotes the weight allocated between input layer and hidden layer. The hidden layer results are sent to the output layer. An output layer displays the classification output for every input patient data. The output layer result is given as,

$$\text{Output layer} = w_{ho} * \text{Hidden layer} \tag{6}$$

From (6), 'w_{ho}' denotes the weight assigned between the hidden layer and output layer. The algorithmic description of Motyka Similar Feature Selected Softsign Deep Neural Classification for stroke disease prediction is given below,

---

\\ **Motyka Similar Feature Selected Softsign Deep Neural Classification Algorithm**
Input: Patient data '$Pad_1, Pad_2, .. Pad_n$' with their features '$f_1, f_2, .. f_m$'
Output: Improved stroke disease prediction performance
Step 1: Begin
Step 2:   Initialize number of patient data with their features
Step 3:   **For** each patient data 'Pad_i' at input layer
Step 4:   The input layer transmits patient data to the hidden layer 1
Step 5:   Hidden layer 1 selects the relevant features and transmit to the hidden layer 2
Step 6:   Hidden layer 2 computes softsign activation function with selected features
Step 7:   The output layer displays the patient data is normal patient data or abnormal patient data
Step 8:   **end for**
Step 9: end

Algorithm 1 Motyka Similar Feature Selected Softsign Deep Neural Classification Algorithm

Algorithm 1 describes the algorithmic process of Motyka Similar Feature Selected Softsign Deep Neural Classification for stroke disease diagnosis with higher accuracy and lesser time consumption. The number of patient data is taken as input at the input layer. The input layer transmits the patient data to the hidden layer 1. Hidden layer 1 selects the relevant features and transmitted to the hidden layer 2. After that, hidden layer 2 determines the softsign activation function with the selected features. Finally in output layer, the stroke disease prediction result (i.e., normal data or abnormal data) is displayed. By this way, MSFSSDNC Method efficiently identifies the stroke disease with higher accuracy and lesser time consumption.

3. EXPERIMENTAL SETTINGS

In order to evaluate the performance results of proposed MSFSSDNC Method is implemented in Java language with aid of Stroke Prediction Dataset. Stroke Prediction Dataset [13] is obtained from Kaggle to conduct the experimental process. Stroke Prediction Dataset comprises 12 attributes and 5110 patient data. The attribute information of the input dataset are id, gender, age, hypertension, heart_disease, ever_married, work_type, residence_type, avg_glucose_level, bmi, smoking_status and stroke presence or absense For experimental process, the MSFSSDNC Method considers the different number of patient data in the range of 500-5000 from dataset [17]. The performance of the MSFSSDNC Method is measured in terms of prediction accuracy, prediction time and false positive rate. MSFSSDNC Method is compared with existing Detecting Risk Factor of Stroke Disease (DRFS) methodology [1] and novel rough-set based technique [2]. The experimental processes of MSFSSDNC Method are carried out with ten different instances for diverse number of patient data and results are shown in below table and graph.

4. RESULT AND DISCUSSIONS

In this part, the performance result of the proposed MSFSSDNC Method and existing techniques are discussed. The effectiveness of the MSFSSDNC Method is explained with table and graph using parameters such as prediction accuracy, prediction time and false positive rate.

4.1 Performance Measure of Prediction Accuracy

Prediction accuracy 'Pre_{Acc}' is defined as the ratio of number of patient data that accurately classified as normal patient data or abnormal patient data to the total number of patient data. The prediction accuracy is computed as,

$$Pre_{Acc} = \frac{Number\ of\ patient\ data\ that\ accurately\ classified}{Total\ number\ of\ patient\ data} * 100 \qquad (7)$$

From (7), the prediction accuracy is computed. The prediction accuracy is measured in terms of percentages (%). The experimental values of prediction accuracy are shown in table 1 and figure 5.

Table 1 Tabulation for Prediction Accuracy

Number of patient data (Number)	Prediction Accuracy (%)		
	Detecting Risk Factor of Stroke Disease (DRFS) methodology	Novel Rough-Set Based Technique	Proposed MSFSSDNC Method
50	70	78	84
100	74	81	88
150	73	81	88
200	81	87	91
250	80	87	92
300	82	86	89
350	89	92	96

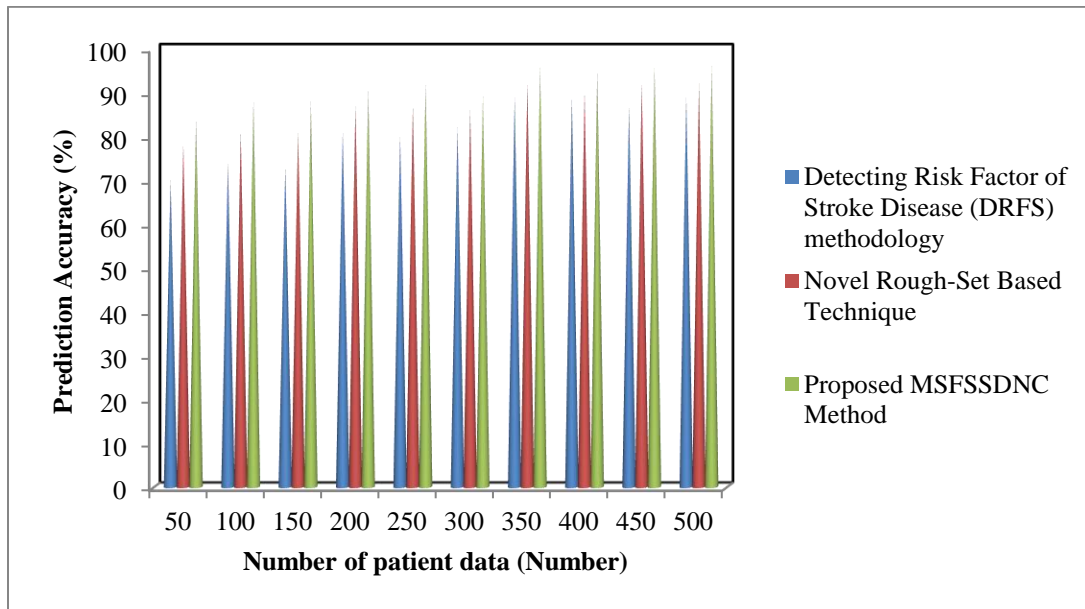| 400 | 89 | 90 | 95 |
| 450 | 87 | 92 | 96 |
| 500 | 89 | 92 | 96 |

Figure 4 Measurement of Prediction Time

Table 1 and Figure 4 shows the impact of prediction accuracy for different number of patient data in the range of 50-500 using three methods namely existing Detecting Risk Factor of Stroke Disease (DRFS) methodology [1] and novel rough-set based technique [2] and proposed MSFSSDNC Method. As discussed in the figure, the prediction accuracy of proposed MSFSSDNC Method is higher when compared to Detecting Risk Factor of Stroke Disease (DRFS) methodology [1] and novel rough-set based technique [2]. This is due to the application of Motyka Similarity Index and Softsign activation function in proposed MSFSSDNC Method on the contrary to existing works. By using Motyka Similarity Index, proposed MSFSSDNC Method selects the relevant features from the patient data in input dataset with enhanced accuracy [16]. Then, MSFSSDNC Method classified the patient data through using softsign activation function. This helps the proposed MSFSSDNC Method to improve the ratio of number of patient data that are correctly classified as normal patient data when compared to conventional works. Hence, the proposed MSFSSDNC Method attain improved prediction accuracy by 12% when compared to Detecting Risk Factor of Stroke Disease (DRFS) methodology [1] and 6% when compared to novel rough-set based technique [2].

4.2 Performance Measure of Prediction Time
Prediction Time 'Pre_{Time}' is defined as the amount of time taken for predicting the stroke disease. It is defined as the product of number of patient data and time consumed for predicting one patient data. Consequently, the prediction time is formulated as,

$$Pre_{Time} = n * \text{time for predicting one data} \qquad (8)$$

From (8), 'n' represent the number of data. The prediction time is measured in terms of milliseconds (ms). The experimental values of prediction time are shown in table 2 and figure 5.

Table 2 Tabulation for Prediction Time

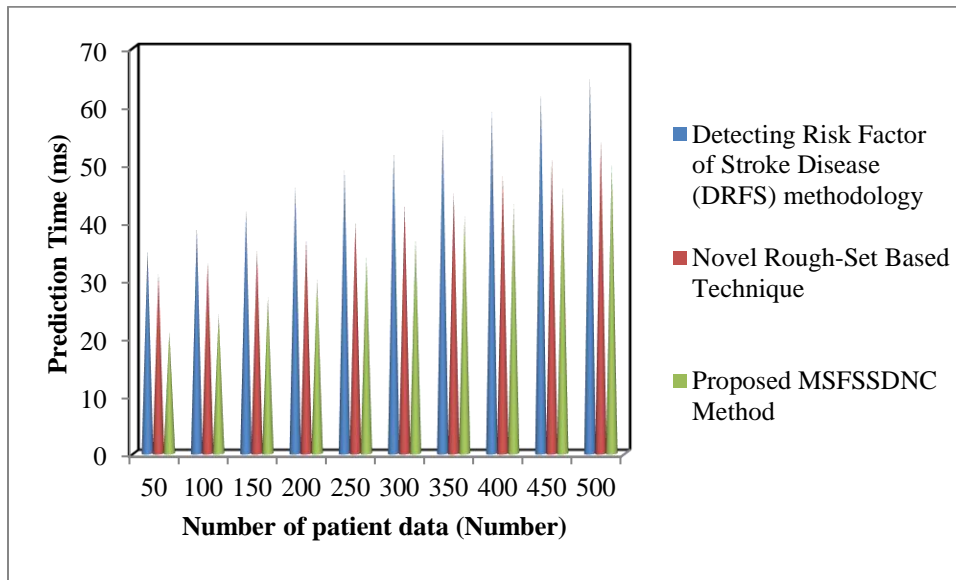| Number of patient data (Number)	Prediction Time (ms)		
	Detecting Risk Factor of Stroke Disease (DRFS) methodology	Novel Rough-Set Based Technique	Proposed MSFSSDNC Method
50	35	31	21
100	39	33	24
150	42	35	27
200	46	37	30
250	49	40	34
300	52	43	37
350	56	45	41
400	59	48	43
450	62	51	46
500	65	54	50

Figure 5 Measurement of Prediction Time

Table 2 and Figure 5 illustrates the impact of prediction time for various number of patient data in range of 50-500 using three methods namely existing Detecting Risk Factor of Stroke Disease (DRFS) methodology [1] and novel rough-set based technique [2] and proposed MSFSSDNC Method. As illustrated in the figure, proposed MSFSSDNC Method consumes higher prediction time while increasing the number of patient data when compared to Detecting Risk Factor of Stroke Disease (DRFS) methodology [1] and novel rough-

set based technique [2]. In addition, the proposed MSFSSDNC Method consumed lesser time than other conventional methods. This is due to the application of Motyka Similarity Index and Softsign activation function in proposed MSFSSDNC Method. By using the Motyka Similarity Index, MSFSSDNC Method chooses the relevant features from patient data. After that, MSFSSDNC Method categorized the patient data by softsign activation function. By this way, proposed MSFSSDNC Method reduces the time consumption for correctly classifying the patient data as normal patient data when compared to conventional works. Hence, the proposed MSFSSDNC Method consumed lesser prediction time by 31% when compared to Detecting Risk Factor of Stroke Disease (DRFS) methodology [1] and 17% when compared to novel rough-set based technique [2].

4.3 Performance Measure of False Positive Rate

False positive rate is defined as the ratio of number of patient data that incorrectly classified as normal patient data or abnormal patient data to the total number of patient data. It is formulated as,

$$\text{False Positive}_{Rate} = \frac{\text{Number of patient data that accurately classified}}{\text{Total number of patient data}} * 100 \quad (9)$$

From (9), 'False Positive$_{Rate}$' represent the false positive rate. It is measured in terms of percentage (%).

Table 3 Tabulation for False positive rate

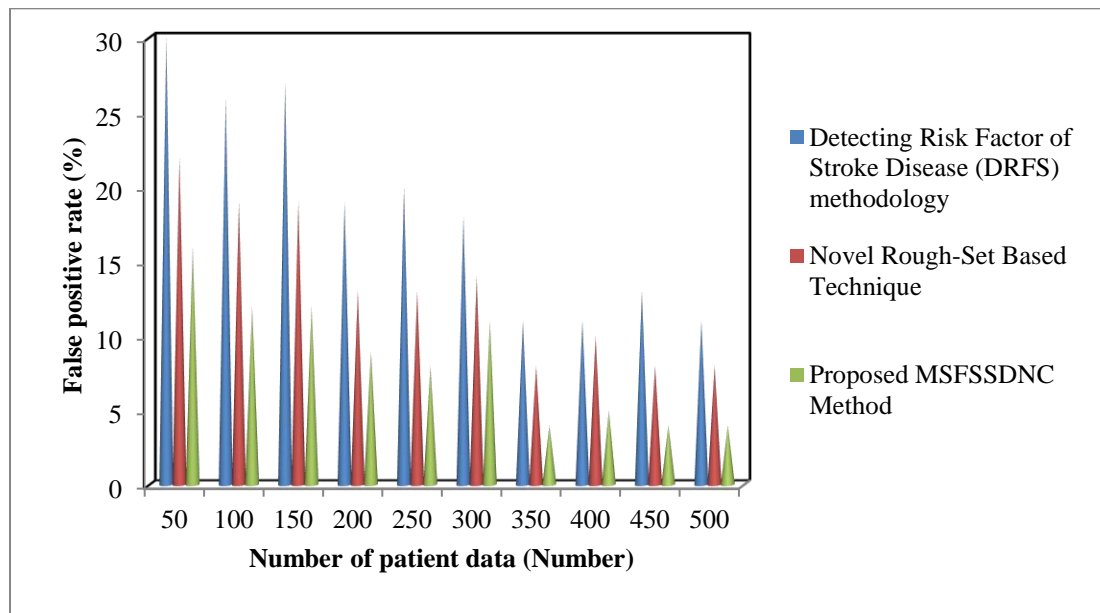| Number of patient data (Number)	False positive rate (%)		
	Detecting Risk Factor of Stroke Disease (DRFS) methodology	**Novel Rough-Set Based Technique**	**Proposed MSFSSDNC Method**
50	30	22	16
100	26	19	12
150	27	19	12
200	19	13	9
250	20	13	8
300	18	14	11
350	11	8	4
400	11	10	5
450	13	8	4
500	11	8	4

Figure 6 Measurement of False Positive Rate

Table 3 and Figure 6 shows the impact of false positive rate for different number of patient data in range of 50-500 using three techniques namely existing Detecting Risk Factor of Stroke Disease (DRFS) methodology [1] and novel rough-set based technique [2] and proposed MSFSSDNC Method. As described in figure, the proposed MSFSSDNC Method consumed lesser false positive rate when compared to other conventional methods. This is due to the use of Motyka Similarity Index and Softsign activation function in proposed MSFSSDNC Method. With Motyka Similarity Index, MSFSSDNC Method selects the relevant features from patient data. After that, MSFSSDNC Method classified the patient data by softsign activation function in accurate manner. This in turn helps MSFSSDNC Method to minimize the false positive rate during stroke disease prediction when compared to conventional works. Therefore, the proposed MSFSSDNC Method attained lesser false positive rate by 56% when compared to Detecting Risk Factor of Stroke Disease (DRFS) methodology [1] and 39% when compared to novel rough-set based technique [2;15].

5. CONCLUSION

The MSFSSDNC Method is designed with the aim of improving the stroke disease prediction performance with higher accuracy and lesser time consumption when considering patient data as input. The aim of MSFSSDNC Method is attained with the support of Motyka Similarity Index and Softsign activation function [14]. The proposed MSFSSDNC Method increases the ratio of a number of patient data that correctly predict the stroke disease with help of motyka similarity index feature selection as compared to conventional works. In addition, the proposed MSFSSDNC Method reduces the false positive rate with the help of Softsign activation function. Therefore, MSFSSDNC Method gives better performance in terms of stroke disease prediction accuracy, prediction time and false positive rate as compared to traditional algorithms. The experimental result shows that MSFSSDNC Method presents better performance with an enhancement of prediction accuracy and reduction of prediction time as compared to state-of-the-art works.

REFERENCES

[1]   S. Pradeepa, K. R. Manjula, S. Vimal, Mohammad S. Khan, Naveen Chilamkurti and Ashish Kr. Luhach, "DRFS: Detecting Risk Factor of Stroke Disease from Social Media Using Machine Learning Techniques", Neural Processing Letters, Springer, 2020, Pages 1-19

[2]   Muhammad Salman Pathan, Zhang Jianbiao, Deepu John, Avishek Nag and Soumyabrata Dev, "Identifying Stroke Indicators Using Rough Sets", IEEE Access, Volume 8, November 2020, Pages 210318 – 210327

[3]   Gang Fang, Peng Xu and Wenbin Liu, "Automated Ischemic Stroke Subtyping Based on Machine Learning Approach", IEEE Access, Volume 8, June 2020, Pages 118426 – 118432

[4]   Douglas Teoh, "Towards stroke prediction using electronic health records", BMC Medical Informatics and Decision Making, Springer, Volume 18, Issue 127, 2018, Pages 1-11

[5]   Taufik Djatna, Medria Kusuma Dewi Hardhienata and Anis Fitri Nur Masruriyah, "An intuitionistic fuzzy diagnosis analytics for stroke disease", Journal of Big Data, Springer, Volume 5, Issue 35, 2018, Pages 1-14

[6]   Ahmed Esmael, Mohammed Elsherief and Khaled Eltoukhy, "Predictive Value of the Alberta Stroke Program Early CT Score (ASPECTS) in the Outcome of the Acute Ischemic Stroke and Its Correlation with Stroke Subtypes, NIHSS, and Cognitive Impairment", Stroke Research and Treatment, Hindawi Publishing Coorporation, Volume 2021, 2021, Pages 1-10

[7]   Mahsa Khodadadi, Heidarali Shayanfar, Keivan Maghooli and Amir Hooshang Mazinan, "Prediction of stroke probability occurrence based on fuzzy cognitive maps", Journal for Control, Measurement, Electronics, Computing and Communications, Taylors and Francis, Volume 60, Issue 4, 2019, Pages 385-392

[8]   X. Li, M. Wu, C. Sun, Z. Zhao, F. Wang, X. Zheng, W. Ge, J. Zhou and J. Zou, "Using machine learning to predict stroke-associated pneumonia in Chinese acute ischaemic stroke patients", European Journal of Neurology, Wiley Publications, Volume 27, Issue 8, August 2020, Pages 1656-1663

[9]   Priya Govindarajan, Ravichandran Kattur Soundarapandian, Amir H Gandomi and Rizwan Patan, "Classification of stroke disease using machine learning algorithms", Neural Computing and Applications, Volume 32, Issue 1, February 2020, Pages 1-12

[10]   Jaehak Yu, Sejin Park, Soon-Hyun Kwon, Chee Meng Benjamin Ho, Cheol-Sig Pyo and Hansung Lee, "AI-Based Stroke Disease Prediction System using Real-Time Electromyography Signals", Applied Sciences, Volume 10, Issue 6791, 2020, Pages 1-19

[11]   Ryan R. Bailey, "Lifestyle Modification for Secondary Stroke Prevention", American Journal of Lifestyle Medicine, Volume 12, Issue 2, 2018, Pages 1-15

[12]   Jie-hua Su, Ling-wei Meng, Di Dong, Wen-yan Zhuo, Jian-ming Wang, Li-bin Liu, Yi Qin, Ye Tian, Jie Tian and Zhao-hui Li "Noninvasive model for predicting future ischemic strokes in patients with silent lacunar infarction using radiomics", BMC Medical Imaging, Volume 20, Issue 77, 2020, Pages 1-11

[13]   https://www.kaggle.com/fedesoriano/stroke-prediction-dataset

[14]   Periyasamy, R., and Nalini, D. (2015). Binary back propagation based lift association mining for heart disease and stroke identification. International journal of applied engineering research, 10(6), pp. 16071-16087.

[15]   Periyasamy, R., and Nalini, D. (2015). Lloyd Minkowski based K -Means clustering for effective diagnosis of heart disease and stroke. International review on computers and software, 10(6).

[16]   Periyasamy, R., and Nalini, D. (2015). Pairwise proximity clustering for medical disease identification. International journal of advanced computer science and technology, 5(1), pp. 1-10.

[17]   Periyasamy, R., and Nalini, D. (2015). A clinical heart disease decision supportive optimized mining method for effective disease diagnosis. International journal of computer applications, 126 (9).

[18]   Subhashini, M., & Gopinath, R. (2020). Mapreduce Methodology for Elliptical Curve Discrete Logarithmic Problems – Securing Telecom Networks, International Journal of Electrical Engineering and Technology, 11(9), 261-273.

[19]   Upendran, V., & Gopinath, R. (2020). Feature Selection Based on Multi criteria Decision Making for Intrusion Detection System. International Journal of Electrical Engineering and Technology, 11(5), 217-226.

[20]   Upendran, V., & Gopinath, R. (2020). Optimization Based Classification Technique for Intrusion Detection System. International Journal of Advanced Research in Engineering and Technology, 11(9), 1255-1262.

[21]   Jayasimman, L., Geetha Dhanalakshmi, V. (2020). Design of Enhanced Dynamic Resource Allocation Framework for Heterogeneous Cloud Environment, IJITEE, 9(3), 1563-1568.

[22]   Jayasimman, L., Geetha Dhanalakshmi, V. (2018). A Study on Spatial-Temporal Load Balancing Approach in Cloud Computing, JCSE, 6(11).

[23]   Geetha Dhanalakshmi, V., Jayasimman, L. (2018). A Review on Dynamic Resource Allocation strategies for Cloud Computing, IJSRCSAMS, 7(4).

[24]   Geetha Dhanalakshmi, V., Jayasimman, L. (2018). Resource Provisioning for Ensuring QoS in Virtualized Environments, JCSE, 6(4).

[25]   James Manoharan, J., Hari Ganesh, S. (2016). A framework for enhancing the efficiency of k-means clustering algorithm to avoid formation of empty clusters. Middle-East J. Sci. Res (MEJSR).

[26]   James Manoharan, J., Hari Ganesh, S., Sathiaseelan, JGR. (2016). Outlier detection using enhanced k-means clustering algorithm and weight-based center approach, Int. J. Comput. Sci. Mobile Comput, 5(4), 453-464.

[27]   James Manoharan, J., Hari Ganesh, S. (2016). INITIALIZATION OF OPTIMIZED K-MEANS CENTROIDS USING DIVIDE-AND-CONQUER METHOD, ARPN Journal of Engineering and Applied Sciences, 11(2), 1086-1091.

[28]   James Manoharan, J., Ganesh, S. H., Felciah, M. L. P., & Shafreenbanu, A. K. (2014, February). Discovering Students' Academic Performance Based on GPA Using K-Means Clustering Algorithm. In 2014 World Congress on Computing and Communication Technologies (pp. 200-202). IEEE.

[29]   Vijilesh, VG., Hari Ganesh, S., James Manoharan, J. (2018). An Enhancing the Performance of High Utility Itemset Mining using Utility Information Record, International Journal of Pure and Applied Mathematics, 118(17), 257-272.

[30]   Selvaramalakshmi, P., Hari Ganesh, S., James Manoharan, J. (2017). A Novel PSS Stemmer for String Similarity Joins, 2017 World Congress on Computing and Communication Technologies (WCCCT), 147-150, IEEE.