ISSN: 0974-5823 Vol. 7 No. 1 January, 2022

International Journal of Mechanical Engineering

An AI Enabled Framework with Composite Metric based Feature Selection for Efficient Brain Stroke Detection

[1]A.Devaki [2]Dr.C.V.Guru Rao

Research Scholar, Department of Informatics, Osmania University, Hyderabad

Dean School of Computer Science &AI, S R University, Warangal

***Abstract* - Research on brain stroke has attracted researchers and academia due to the increasing mortality rate of the disease. With the advancements of Artificial Intelligence (AI) based approaches in the form of Machine Learning (ML), there are unprecedented possibilities. From the existing literature, it is understood that there are many efficient supervised learning methods for brain stroke detection. However, they suffer from quality of training when dataset has irrelevant and redundant features. Unless, there is an efficient mechanism to deal with identification of contributing features, the prediction models tend to be mediocre in performance. To overcome this problem, in this paper, we proposed an algorithm known as Composite Metric based Feature Selection (CMFS) which has required mathematical model to detect best features. We proposed another algorithm named Multi Model Machine Learning Approach for Brain Stroke Prediction (MMMLA-BSP) for efficient detection of brain stroke. This algorithm exploits CMFS to improve quality in training phase. It takes many prediction models in pipeline and evaluates the models with feature selection and without feature selection. The usage of the proposed CMFS leverages performance of the prediction models. It is reflected in the results obtained from the empirical study. Highest prediction accuracy is achieved by KNeighborsalong with CMFS with 96.0288%. The findings in this paper are encouraging and the proposed framework can be used in Clinical Decision Support System (CDSS) of healthcare units to diagnose brain stroke using data-driven approach.**

Keywords – Brain Stroke Detection, Machine Learning, Feature Selection, Composite Metric

1. INTRODUCTION

Brain stroke is the condition where blood supply to brain is interrupted. It is of two types known as ischemic stroke and haemorrhage stroke. The former is caused by artery blockage while the latter occurs due to bleeding. Research on brain stroke has attracted researchers and academia due to the increasing mortality rate of the disease. With the advancements of Artificial Intelligence (AI) based approaches in the form of Machine Learning (ML), there are unprecedented possibilities. However, the ML models for brain stroke detection provide deteriorated performance if the training data quality is not up to the mark. Therefore, there is need for feature selection. Feature selection is the process in which each attribute or feature in the dataset is evaluated to know whether it can contribute to prediction process. Without feature selection, when all features are considered, it causes deterioration in performance of prediction models.

Cai *et al*. [1] explored the modus operandi of feature selection in machine learning models. Bommert *et al*. [2] focused on filter methods for feature selection. They discussed about many filter methods. Sanchez-Marono *et al*. [3] also discussed about filter methods such as Relief, Correlation based Feature Selection (CFS), Fast Correlated-Based Filter (CFBF) and Interact. Khalid *et al*. [4] focused on both feature extraction methods and feature selection methods. Cherrington *et al*. [5] explored different filter methods and their challenges in feature selection. Chandrashekar and Sahin [6] studied different filter and wrapper methods and underlying metrics used. Min and Fangfang [7] proposed a hybrid method known as Filter-Wrapper Hybrid Method (FWHM) that combines both filter and wrapper approaches in the same architecture for feature selection. From the existing literature, it is understood that there are many efficient supervised learning methods for brain stroke detection. However, they suffer from quality of training when dataset has irrelevant and redundant features. Unless, there is an efficient mechanism to deal with identification of contributing features, the prediction models tend to be mediocre in performance. From the study of existing feature selection methods, it is observed the need for hybrid approaches towards enhancing feature selection efficiency. Our contributions in this paper are as follows.

1. We proposed a ML based framework that could guide the research and complete it to achieve efficient brain stroke prediction.
2. We proposed an algorithm known as Composite Metric based Feature Selection (CMFS) which has required mathematical model to detect best features.
3. We proposed another algorithm named Multi Model Machine Learning Approach for Brain Stroke Prediction (MMMLA-BSP) for efficient detection of brain stroke. This algorithm exploits CMFS to improve quality in training phase.
4. A prototype application is built using Python data science platform to evaluate the proposed framework and the underlying algorithms.

The remainder of the paper is structured as follows. Section 2 reviews literature on different ML techniques and feature selection approaches. Section 3 presents the proposed methodology and underlying algorithms for efficient brain stroke detection. Section 4 provides details of experiments and results. Section 5 draws conclusions of the research carried out and presented in this paper. It also gives possible future scope of the research.

2. RELATED WORK

This section reviews literature on the brain stroke detection methods and also feature selection approaches.

2.1 Machine Learning for Brain Stroke Detection

Machine learning is widely used for brain stroke detection. Babu *et al*. [13] proposed an effective method for brain stroke detection. It makes use of Naïve Bayes and Random Forest (RF) methods for prediction models. Lei *et al*. [14] focused on stroke research using machine learning approaches. They focused on early diagnosis of stroke and prevention of the same. They built a monitoring service based on heart rate to reduce risk of stroke. Villar *et al*. [15] proposed an intelligent recognition system for brain stroke detection early. They developed a human activity recognition system for monitoring health of patients. In the process, they built methodology for early detection of stroke. Dijck *et al*. [16] on the other hand focused on quantitative assessment of brain stroke using machine learning along with hybrid feature selection method.

Sirsat *et al*. [17] reviewed different ML methods for stroke prediction. They opined that supervised ML models are good for stroke detection. They also investigated on different hyperparameters used with ML techniques that influence the learning process and improve detection performance. They also covered some of the feature selection methods used to improve performance of prediction models. Chen *et al*. [18] proposed a smart ML model using Internet of Things (IoT) for diagnosis of brain haemorrhage. Pathanjali *et al*. [19] proposed an AI based framework for stroke detection. They explored different prediction models suing ML based approaches including fuzzy models. Aishvarya *et al*. [20] used MRI imagery and machine learning techniques for brain stroke detection. Bangare [21] does similar kind of research to know brain anomalies. Vamsi *et al*. [22] investigated on various brain stroke prediction models based on machine learning techniques. The techniques are employed on the data pertaining to family history.

2.2 Feature Selection Methods

This subsection reviews feature selection methods found in the literature. Cai *et al*. [1] explored the modus operandi of feature selection in machine learning models. They discussed about different kinds of feature selection techniques such as supervised, unsupervised and semi-supervised. Feature selection methods can be classified into filter and wrapper methods. Bommert *et al*. [2] focused on filter methods for feature selection. They discussed about many filter methods. Filter methods compute score of each feature and based on the score, a feature is either selected or discarded. Sanchez-Marono *et al*. [3] also discussed about filter methods such as Relief, Correlation based Feature Selection (CFS), Fast Correlated-Based Filter (CFBF) and Interact. Khalid *et al*. [4] focused on both feature extraction methods and feature selection methods. They also discussed about filter and wrapper approaches related to feature selection. Cherrington *et al*. [5] explored different filter methods and their challenges in feature selection.

Chandrashekar and Sahin [6] studied different filter and wrapper methods and underlying metrics used. Filter methods used different ranking techniques to identify good features while wrapper methods used predictor in order to find suitable features. Min and Fangfang [7] proposed a hybrid method known as Filter-Wrapper Hybrid Method (FWHM) that combines both filter and wrapper approaches in the same architecture for feature selection. Jovic *et al*. [8] discussed about different metrics used for feature selection. They include Chi-square, information gain, correlation, gain ratio and so on. They found that feature selection can be used in text mining, image processing, medical data analysis and industrial applications. Talavera [9] evaluated various filter and wrapper methods in clustering application. Ferreira and Figueiredo [10] proposed a filter method for feature selection using high-dimensional data. They explored both unsupervised and supervised feature selection methods on large datasets. They also investigated different measures known as similarity measures and dispersion measures.

As found in the work of Chilamkurthy *et al*. [11], the machine learning approaches suffer from scarcity of quality training (labelled) datasets. They found that the quality of training is essential for better performance of supervised learning models. Therefore, an important research gap found is that there is further need for improving feature selection so as to leverage quality in training phase to get rid of performance deterioration of prediction models. Zhang and Sejdić *et al*. [12] also distinctly identified the problem of selecting ideal features for both medical data-driven and image analysis approaches in order to use most useful contents of images in supervised learning. In fact, they opined it as most challenging problem in machine learning. From the existing literature, it is understood that there are many efficient supervised learning methods for brain stroke detection. However, they suffer from quality of training when dataset has irrelevant and redundant features. Unless, there is an efficient mechanism to deal with identification of contributing features, the prediction models tend to be mediocre in performance.

3. METHODOLOGY

This section presents the proposed methodology for stroke detection using data-driven approach. It includes the algorithms defined for feature selection and detection of brain stroke. It is supervised learning approach that learns from training dataset to gain required knowledge for prediction. The rationale behind the supervised learning approach is that it can exploit prior experiences of domain experts of physicians in the healthcare domain. It maps inputs to outputs based on the knowledge gained from learning. Supervised learning has two phases such as training and testing. In training phase, the prediction model learns from data given for training. In fact, a knowledge model is created after training phase. In the testing phase, this knowledge model is used to have predictions. A framework is proposed as shown in Figure 1 to detect brain stroke efficiently. To bring about novelty and performance enhancement, the framework includes a feature selection algorithm known as Composite Metric based Feature Selection (CMFS). The framework also uses another algorithm named Multi Model Machine Learning Approach for Brain Stroke Prediction (MMMLA-BSP) for efficient prediction of stroke.

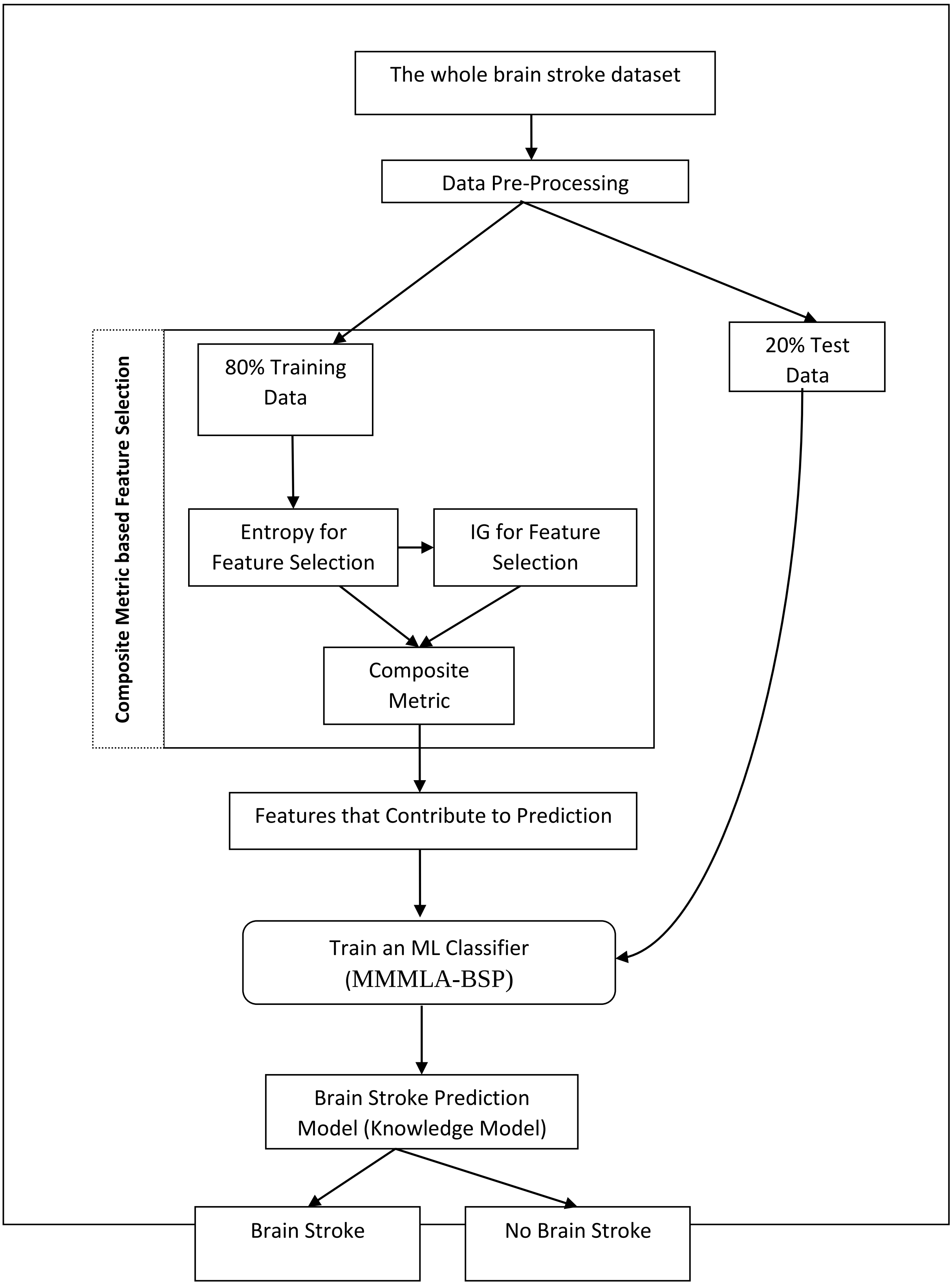

Figure 1: Supervised machine learning framework for brain stroke detection

The given brain stroke dataset is subjected to pre-processing. Pre-processing is made programmatically to make the data ready for training and testing phases of the supervised learning process. Once the data is pre-processed, the training data is given to the proposed CMFS algorithm. This algorithm is meant for identifying features that contribute more to class label prediction (brain stroke prediction). Once feature selection is completed, the features that are not relevant and the features that are redundant are eliminated. This will bring about quality in training process. The selected features along with the domain data is given to brain stroke prediction algorithms. The algorithms learn from the data and creates a brain stroke prediction knowledge model. This model is further used for testing where each instance in the test data is subjected to prediction of presence or absence of brain stroke.

3.1 Data Collection

Brain stroke data is the data of patients collected from UCI machine learning repository. This dataset contains data suitable for data-driven approach. The dataset contains attributes associated with an Electronic Health Record (EHR) of patients. Patients vitals are recorded in the form of dataset. The dataset has 5100 instances and 12 attributes. Every patient is identified with a unique ID.

Attribute	**Description**	**Possible Values**
ID	Identifies patient uniquely	Any numeric value that is unique.
GENDER	Identifies gender of patient	Male | Female
AGE	Specifies age of patient	Any positive numeric value reflecting age of the person
HYPERTENSION	Shows whether the patient has hypertension health condition	0 | 1 indicating no hypertension and hypertension respectively
HEART DISEASE	Shows whether the patient has heart disease already	0 | 1 indicating no heart disease and heart disease respectively
EVER MARRIED	Shows whether the patient has ever married	Yes | No indicating patient has ever married or not respectively
WORK TYPE	Shows the work type of the patient	Never Worked, Children, Private Job, Public Job, Self Employed
RESIDENCE TYPE	Indicates the residence type of the patient	Urban | Rural based on the residence location
AVG_GLUCOSE_LEVEL	Shows average glucose level of patient	A non-negative numeric value
BMI	Indicates body mass index of the patient	A non-negative numeric value
SMOKING	Shows whether patient has smoking habit	Unknown, Never Smoked, Formerly Smoked, Smokes
STROKE	Diagnosis column indicating whether patient has stroke disease	0 | 1 indicating no stroke disease and stroke disease respectively

Table 1: Shows attributes in the datas*et al*ong with description and possible values

As presented in Table 1, the dataset details are provided. Each attribute is supposed to have patient specific value as given by healthcare units. The last attribute is the diagnosis attribute or class label whose value is given by domain experts or doctors. That column with corresponding values constitutes the ground truth which is later used to know prediction performance of ML models.

3.2 Pre-Processing

The given data is subjected to pre-processing. In this phase, the data is observed and null values are treated with default values. Additionally, in this module, the proposed framework divides the data into two parts. The entire brain stroke data is categorized into training data and testing data. The training data has all attributes including the class label or diagnosis column while testing data has all attributes except the diagnosis column as it needs to be predicted by the algorithm.

Id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
579	Male	9	0	0	No	children	Urban	71.88	17.5	Unknown	0
5317	Female	79	0	1	Yes	Private	Urban	214.09	28.2	never smoked	1
8213	Male	78	0	1	Yes	Private	Urban	219.84	N/A	Unknown	1
12095	Female	61	0	1	Yes	Govt_job	Rural	120.46	36.8	smokes	1
12109	Female	81	1	0	Yes	Private	Rural	80.43	29.7	never smoked	1
12175	Female	54	0	0	Yes	Private	Urban	104.51	27.3	smokes	1
28048	Male	13	0	0	No	children	Urban	82.38	24.3	Unknown	0
41512	Male	57	0	0	Yes	Govt_job	Rural	76.62	28.2	never smoked	0
64520	Male	68	0	0	Yes	Self_ employed	Urban	91.68	40.8	Unknown	0
68598	Male	1.1	0	0	No	children	Rural	79.15	17.4	Unknown	0

Table 2: Shows an excerpt from the training data

The training data is shown in Table 2 where it shows the last column as diagnosis column which has ground truth. This is the basis for prediction performance evaluation. This column values are given by domain expert or doctor who has first-hand experience in dealing with brain stroke patients. The test

data will have all the attributes except the last one as it needs to be predicted by the proposed prediction models.

Notation	Description
X, Y	Random variables
p(x)	Probability density function
P(y)	Probability density function
H (X)	Entropy
H (y/x)	Conditional Entropy
SU	Symmetric uncertaity (composite metric)

Table 3: Shows notations used in the proposed feature selection algorithm

The notations used in the proposed feature selection algorithm are presented in Table 3. Section 3.3 provides the details of the proposed algorithm.

3.3 Feature Selection

Feature selection is the process in which each attribute or feature in the dataset is evaluated to know whether it can contribute to prediction process. Without feature selection, when all features are considered, it causes deterioration in performance of prediction models. Therefore, we proposed an algorithm known as Composite Metric based Feature Selection (CMFS). It is based on a composite metric that combines entropy and information gain metric in order to have a measure for determination of utility of given feature. Entropy finds uncertainty while gain finds the change in entropy. Entropy measure is computed using Eq. 1 and Eq. 2 while Eq. 3 computes information gain measure.

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \quad (1)$$

$$H(Y) = -\sum_{y \in Y} p(y) \log p(y) \quad (2)$$

$$IG = H(y) - H(y/x) \quad (3)$$

$$SU = \frac{2*IG}{H(x)+H(y)} \quad (4)$$

The Eq. 4 computes the compositive metric known as symmetric uncertainty. It is finally used in the proposed algorithm for determination of usefulness of given feature for brain stroke prediction.

Algorithm: Composite Metric based Feature Selection (CMFS)

Input: EHR dataset on brain stroke *D*, threshold *th*

Output: Useful features *F*

```
1.    Start
2.    Initialise attributes vector A
3.    A←FindAttributes(D)
4.    For each a in A
5.    Use Eq. 1 and 2 to compute entropy value
6.      Use Eq. 3 to compute information gain
7.      Use Eq. 4 to compute composite metric
8.       Add composite metric to a map M
9.    End For
10.   For each a in A
11.     Obtain composite metric from M
12.     IF feature importance satisfiesth THEN
13.        Add a to F
14.     End If
15.   End For
16.   Return F
17.   End
```

Algorithm 1: Hybrid Measures Approach for Feature Engineering

The proposed algorithm shown in Algorithm 1 takes dataset containing patient EHRs and computes importance of each feature. In the process it has two iterative processes. The first iterative process finds the importance of each feature while the second iterative process filters the features based on the importance of features and conformance with the threshold value. All the selected features are finally returned. This algorithm is only meant for finding good features. It itself cannot perform brain stroke prediction.

3.4 Brain Stroke Prediction

For brain stroke prediction, the proposed framework has the proposed Multi Model Machine Learning Approach for Brain Stroke Prediction (MMMLA-BSP) algorithm. This algorithm exploits CMFS for improving training quality of prediction models in MMMLA-BSP.

Algorithm: Multi Model Machine Learning Approach for Brain Stroke Prediction (MMMLA-BSP)

Inputs: Brain stroke dataset *D*, ML techniques *T*

Output: Brain stroke prediction results *P*

```
1.    Start
2.    Initialize results map M
3.    (T1, T2)←PreProcess(D)
4.    F←Run CMFS(T1)
5.    For each ML technique t in T
6.      Train the model t using F
7.      Fit the model t for T2
8.      Add results to P
9.      Add t and P to M
10.   End For
11.   For each map entry m in M
12.     Display confusion matrix
13.     Display P
14.   End For
15.   End
```

Algorithm 2: Multi Model Machine Learning Approach for Brain Stroke Prediction

Algorithm 2 defines MMML-BSP procedure for brain stroke detection. It has pipeline of many ML models denoted as *T*. It takes other input as brain stroke EHRs of patients denoted as *D*. In Step 3 of the algorithm, pre-processing is done to divide data into training (T1) and testing (T2) data. In Step 4, the algorithm invokes CMFS algorithm in order to find best features into the feature vector *F*. Step 5 through Step 10, there is an iterative process to have brain stroke prediction with many ML techniques denoted by

T. Each prediction model predictions are saved to *P* and the map *M* holds a map of technique and its performance. Step 11 through 14, the algorithm presents the results of all prediction models including confusion matrix (shows prediction results) and performance metrics denoted as *P*.

3.5 Evaluation

Different performance metrics that are widely used in the literature are known as precision, recall, accuracy and F1-score are considered for performance evaluation. These metrics are based on the computation of number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (TN).

Metric	Formula	Value range	Best Value
Precision (p)	$\frac{TP}{TP+FP}$	[0; 1]	1
Recall (r)	$\frac{TP}{TP+FN}$	[0; 1]	1
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	[0; 1]	1
F1-Score	$2*\frac{(p*\ r)}{(p+r)}$	[0; 1]	1

Table 3: Performance metrics used for evaluation

Table 3 shows the performance metrics used for evaluation of the proposed framework and underlying ML models. It shows the equation for each metric along with value range between 0 and 1 reflecting lowest and highest performance besides best value.

4. EXPERIMENTAL RESULTS

Experiments are made on the dataset described in Section 3.2 using the ML algorithms implemented using Python data science platform. The performance of the prediction models is evaluated in terms of precision, recall, F1-score and accuracy. The accuracy the prediction models are also compared with and without feature selection algorithm.

Precision Comparison	
Prediction Model	**Precision**
GaussianNB	1
BernoulliNB	1
Logistic Regression	1
Random Forest Classifier	1
Support Vector Machine	1
Decision Tree Classifier	0.058823
KNeighbors Classifier	0.029411
Gradient Boosting Classifier	1
Stochastic Gradient Descent	1
Neural Nets	0.058823

Table 4: Performance of prediction models in terms of precision

As presented in Table 4, the performance of the prediction models is provided in terms of the precision.

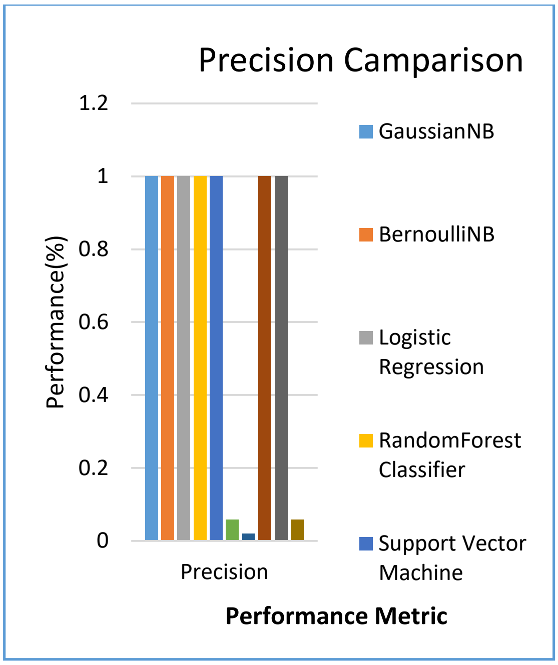

Figure 2: Performance comparison of brain stroke prediction models in terms of precision

As presented in Figure 2, the performance of brain stroke prediction models is provided. The precision performance is shown in vertical axis. The results show that there is difference in performance of the models used for brain stroke prediction in terms of precision.

Recall Comparison	
Prediction Model	**Recall**
GaussianNB	0.049853
BernoulliNB	0.962051
Logistic Regression	0.963533
Random Forest Classifier	0.964533
Support Vector Machine	0.963533
Decision Tree Classifier	0.054794
KNeighbors Classifier	1
Gradient Boosting Classifier	0.960352
Stochastic Gradient Descent	0.963533
Neural Nets	0.059712

Table 5: Performance of prediction models in terms of recall

As presented in Table 5, the performance of the different prediction models is provided in terms of the recall.

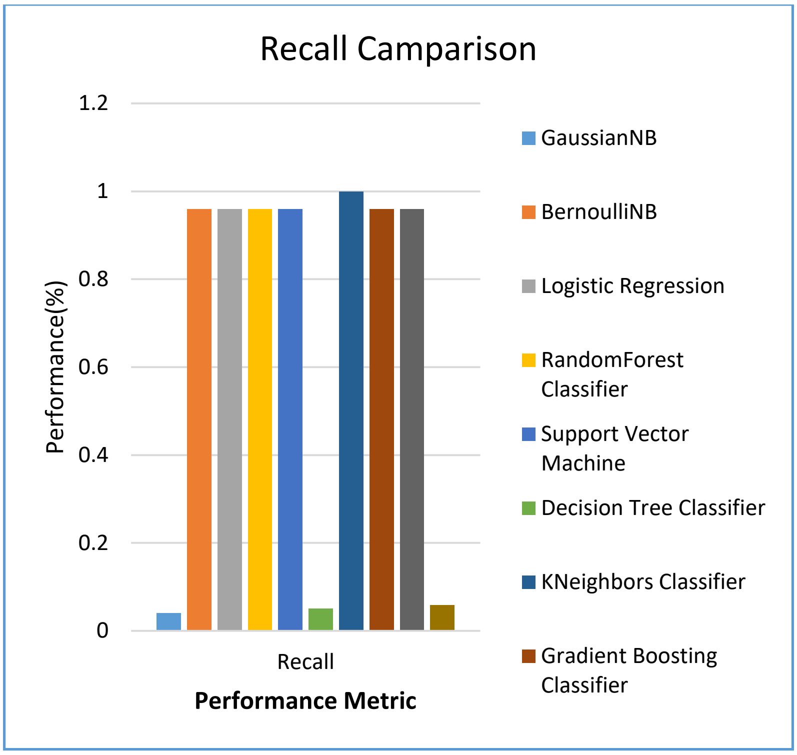

Figure 3: Performance comparison of brain stroke prediction models in terms of recall

As presented in Figure 3, the performance of brain stroke prediction models is provided. The recall performance is shown in vertical axis. The results show that there is difference in performance of the models used for brain stroke prediction in terms of recall. Highest recall is exhibited by KNeighbors classifier.

F1-Measure Comparison	
Prediction Model	**F1-Measure**
GaussianNB	0.094972
BernoulliNB	0.960985
Logistic Regression	0.980854
Random Forest Classifier	0.989085
Support Vector Machine	0.980854
Decision Tree Classifier	0.057945
KNeighbors Classifier	0.054142
Gradient Boosting Classifier	0.988937
Stochastic Gradient Descent	0.989057
Neural Nets	0.057917

Table 6: Performance of prediction models in terms of F1-score

As presented in Table 6, the performance of the different prediction models is provided in terms of the F1-score.

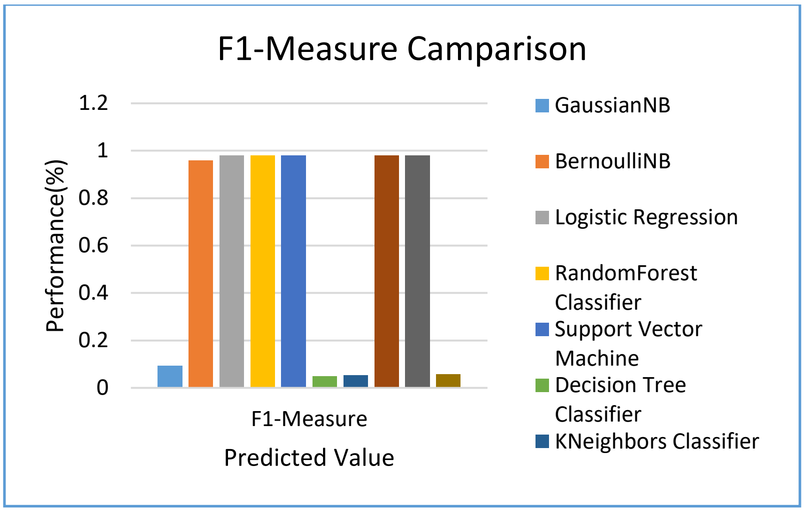

Figure 4: Performance comparison of brain stroke prediction models in terms of F1-score

As presented in Figure 4, the performance of brain stroke prediction models is provided. The F1-score performance is shown in vertical axis. The results show that there is difference in performance of the models used for brain stroke prediction in terms of F1-score. Highest performance is exhibited by Gradient Boosting classifier.

Accuracy Comparison	
Predicted Model	**Accuracy**
GaussianNB	0.220216
BernoulliNB	0.959085
Logistic Regression	0.959085
Random Forest Classifier	0.959085
Support Vector Machine	0.959084
Decision Tree Classifier	0.916967
KNeighbors Classifier	0.960288
Gradient Boosting Classifier	0.954755
Stochastic Gradient Descent	0.959085
Neural Nets	0.921708

Table 7: Performance of prediction models in terms of accuracy

As presented in Table 7, the performance of the different prediction models is provided in terms of the accuracy.

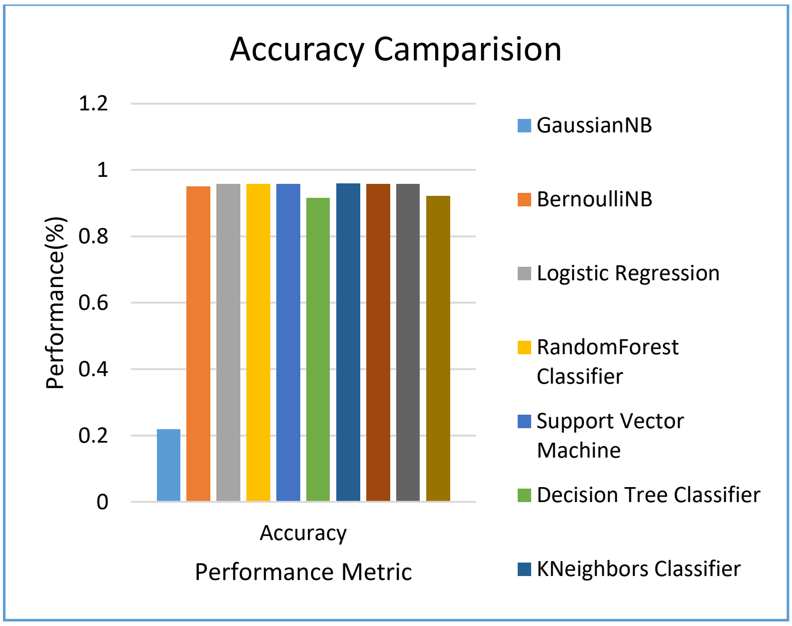

Figure 5: Performance comparison of brain stroke prediction models in terms of accuracy

As presented in Figure 5, the performance of brain stroke prediction models is provided. The accuracy performance is shown in vertical axis. The results show that there is difference in performance of the models used for brain stroke prediction in terms of accuracy. Highest performance is exhibited by KNeighbors classifier.

Prediction Model	**Accuracy (without FS)**	**Accuracy (with FS)**
GaussianNB	0.2092052	0.220216
BernoulliNB	0.91113075	0.959085
Logistic Regression	0.91113075	0.959085
Random Forest Classifier	0.91113075	0.959085
Support Vector Machine	0.9111298	0.959084
Decision Tree Classifier	0.87111865	0.916967
KNeighbors Classifier	0.9122736	0.960288
Gradient Boosting Classifier	0.90701725	0.954755
Stochastic Gradient Descent	0.91113075	0.959085
Neural Nets	0.8756226	0.921708

Table 8: Performance of prediction models with and without feature selection in terms of accuracy

As presented in Table 8, the performance of the different prediction models with and without feature selection is provided in terms of the accuracy.

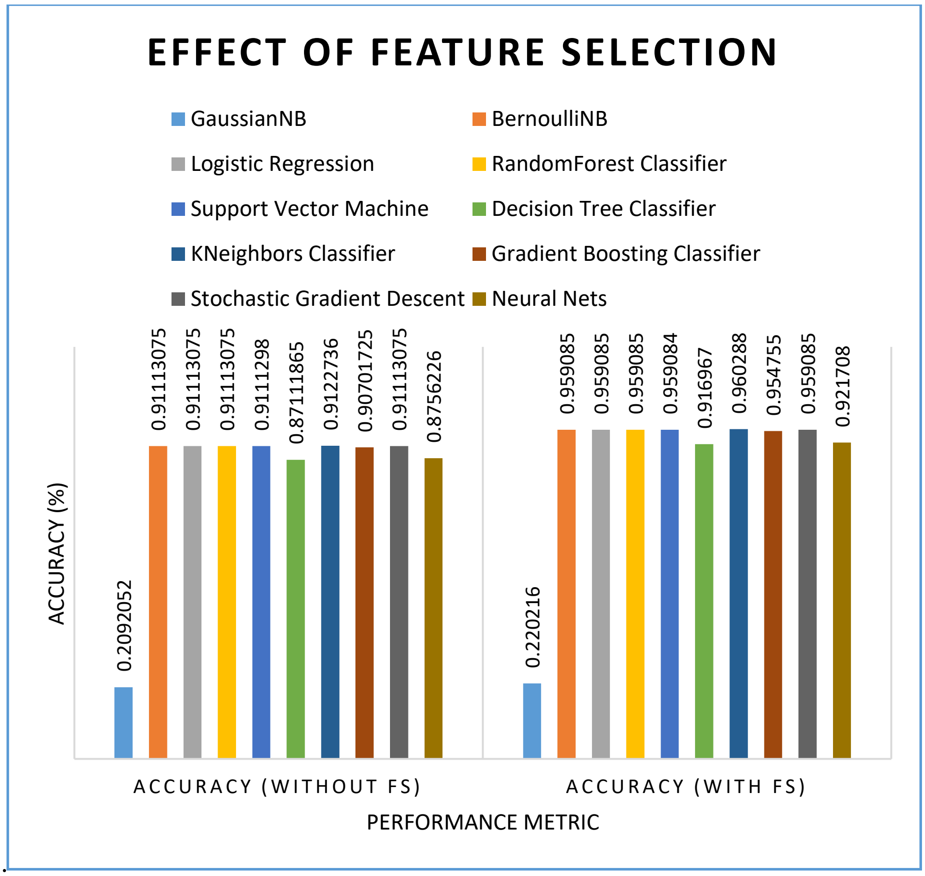

Figure 6: Performance comparison of brain stroke prediction models with and without feature selection in terms of accuracy

As presented in Figure 6, the performance of brain stroke prediction models is provided. The accuracy performance is shown in vertical axis. The results show that there is difference in performance of the models used for brain stroke prediction in terms of accuracy. Another important observation is that the proposed feature selection algorithm is able to improve performance of prediction models. The highest accuracy with feature selection is exhibited by KNeighbours classifier with 96.0288%. The performance of the same prediction model without feature selection is 91.22736%. Therefore, it is evident that there is clear performance improvement with the proposed CMFS algorithm.

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed an algorithm known as Composite Metric based Feature Selection (CMFS) which has required mathematical model to detect best features. We proposed another algorithm named Multi Model Machine Learning Approach for Brain Stroke Prediction (MMMLA-BSP) for efficient detection of brain stroke. This algorithm exploits CMFS to improve quality in training phase. It takes many prediction models in pipeline and evaluates the models with feature selection and without feature selection. The usage of the proposed CMFS leverages performance of the prediction models. It is reflected in the results obtained from the empirical study. Highest prediction accuracy is achieved by KNeighbors along with CMFS with 96.0288%. The findings in this paper are encouraging and the proposed framework can be used in Clinical Decision Support System (CDSS) of healthcare units to diagnose brain stroke using data-driven approach. In future, we intend to improve the performance of the framework further with the introduction of ensemble approach.

References

[1] Cai, Jie; Luo, Jiawei; Wang, Shulin; Yang, Sheng (2018). Feature selection in machine learning: a new perspective. Neurocomputing, p1-13.

[2] Bommert, Andrea; Sun, Xudong; Bischl, Bernd; Rahnenführer, Jörg; Lang, Michel (2020). Benchmark for filter methods for feature selection in high-dimensional classification data. Computational Statistics & Data Analysis, 143, 106839, p1-19.

[3] Yin, Hujun; Tino, Peter; Corchado, Emilio; Byrne, Will; Yao, Xin (2007). [Lecture Notes in Computer Science] Intelligent Data Engineering and Automated Learning - IDEAL 2007 Volume 4881 || Filter Methods for Feature Selection – A Comparative Study. , (Chapter 19), p178–187.

[4] Khalid, Samina; Khalil, Tehmina; Nasreen, Shamila (2014). [IEEE 2014 Science and Information Conference (SAI) - London, UK (2014.8.27-2014.8.29)] 2014 Science and Information Conference - A survey of feature selection and feature extraction techniques in machine learning, p372–378.

[5] Cherrington, Marianne; Thabtah, Fadi; Lu, Joan; Xu, Qiang (2019). [IEEE 2019 International Conference on Computer and Information Sciences (ICCIS) - Sakaka, Saudi Arabia (2019.4.3-2019.4.4)] 2019 International Conference on Computer and Information Sciences (ICCIS) - Feature Selection: Filter Methods Performance Challenges. , p1–4

[6] Chandrashekar, Girish; Sahin, Ferat (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), p16–28.

[7] Min, Hu; Fangfang, Wu (2010). [IEEE 2010 Second Global Congress on Intelligent Systems (GCIS) - Wuhan, Hubei, China (2010.12.16-2010.12.17)] 2010 Second WRI Global Congress on Intelligent Systems - Filter-Wrapper Hybrid Method on Feature Selection. , p98–101.

[8] Jovic, A.; Brkic, K.; Bogunovic, N. (2015). [IEEE 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) - Opatija, Croatia (2015.5.25-2015.5.29)] 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) - A review of feature selection methods with applications. , p1200–1205.

[9] Talavera, L. (2005). An Evaluation of Filter and Wrapper Methods for Feature Selection in Categorical Clustering. Advances in Intelligent Data Analysis VI, p440–451.

[10]Artur J. Ferreira; Mário A.T. Figueiredo (2012). Efficient feature selection filters for high-dimensional data. , 33(13), p1-11.

[11]Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N. G., Venugopal, V. K., … Warier, P. (2018). *Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. The Lancet. P2388-2396.*

[12]Zhang, Z., & Sejdić, E. (2019). *Radiological images and machine learning: Trends, perspectives, and prospects. Computers in Biology and Medicine. p1-35.*

[13] Sunil Babu, Melingi; Vijayalakshmi, V. (2019). An Effective Approach for Sub-acute Ischemic Stroke Lesion Segmentation by Adopting Meta-Heuristics Feature Selection Technique Along with Hybrid Naive Bayes and Sample-Weighted Random Forest Classification. Sensing and Imaging, 20(1), 7, p1-24.

[14] Lei, N., Kareem, M., Moon, S. K., Ciaccio, E. J., Acharya, U. R., & Faust, O. (2021). Hybrid Decision Support to Monitor Atrial Fibrillation for Stroke Prevention. International Journal of Environmental Research and Public Health, 18(2), 813, p1-19.

[15] Villar, José R.; Chira, Camelia; Sedano, Javier; González, Silvia; Trejo, José M. (2015). A hybrid intelligent recognition system for the early detection of strokes. Integrated Computer-Aided Engineering, 22(3), p215–227.

[16] Van Dijck, Gert; Van Hulle, Marc M.; Van Vaerenbergh, Jo (2006). 2006 International Conference of the IEEE Engineering in Medicine and Biology Society - Hybrid Feature Subset Selection for the Quantitative Assessment of Skills of Stroke Patients in Activity of Daily Living Tasks. , p5699–5703.

[17] Sirsat, M. S., Fermé, E., &Câmara, J. (2020). Machine Learning for Brain Stroke: A Review. Journal of Stroke and Cerebrovascular Diseases, 29(10), 105162.

[18] Chen, H., Khan, S., Kou, B., Nazir, S., Liu, W., & Hussain, A. (2020). A Smart Machine Learning Model for the Detection of Brain Haemorrhage Diagnosis Based Internet of Things in Smart Cities. Complexity, 2020, 1–10.

[19] Pathanjali C, Priya T, Monisha G, SamyukthaBhaskar and RuchitaSudarshan K. (2020). Machine Learning for Predicting Ischemic Stroke. International Journal of Engineering Research & Technology (IJERT). 9 (5), p1334-1337.

[20] R Aishvarya, RJ Anand, BatchaliVasundhara, Dr S Sudha. (2020). Early Detection of Brain Stroke using MRI Images. International Research Journal of Engineering and Technology (IRJET). 7 (9), p2258-2261.

[21] Dr. Sunil L. Bangare. (2021). Brain Tumor Detection Using Machine Learning Approach. Design Engineering, 7557-7566.

[22] Vamsi B., Bhattacharyya D., Midhunchakkaravarthy D. (2021) Detection of Brain Stroke Based on the Family History Using Machine Learning Techniques. In: Saha S.K., Pang P.S., Bhattacharyya D. (eds) Smart Technologies in Data Science and Communication. Lecture Notes in Networks and Systems, vol 210. Springer, Singapore. https://doi.org/10.1007/978-981-16-1773-7_2.