International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-13 Issue-1, December 2023

A Hybrid Model for Predicting Classification Dataset based on Random Forest, Support Vector Machine and Artificial Neural Network

Priyanka Mazumder, Siddhartha Baruah

Abstract: *Machine Learning offers a rich array of algorithms, and the performance of these algorithms can vary significantly depending on the specific task. Combining these traditional algorithms can lead to the development of innovative hybrid structures that outperform individual models. One such novel hybrid model is the Hybrid Support Random Forest Neural Network (HSRFNN), which is designed to deliver enhanced performance and accuracy. HSRFNN represents a fusion of Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN) to leverage their respective strengths. This hybrid model consistently outperforms the individual models of Random Forest, SVM, and ANN. In this study, ten diverse datasets sourced from UCI and Kaggle data repositories were considered for evaluation. The accuracy of the HSRFNN model was meticulously compared with the three traditional algorithms, namely Random Forest, Support Vector Machine, and Artificial Neural Network. Various accuracy metrics, such as Correctly Classified Instances (CCI), Incorrectly Classified Instances (ICI), Accuracy (A), and Time Taken to Build Model (TTBM), were used for the comparative analysis. This research strives to demonstrate that HSRFNN, through its hybrid architecture, can offer superior accuracy and performance compared to individual algorithms. The choice of datasets from different sources enhances the generalizability of the results, making HSRFNN a promising approach for a wide range of machine learning tasks. Further exploration and fine-tuning of HSRFNN may unlock its potential for even more challenging and diverse datasets.*

Keywords: *Machine Learning, Random Forest, Support Vector Machine, Artificial Intelligence, Accuracy*

I. INTRODUCTION

Algorithms in Machine Learning are well trained to provide possible outcomes. Sometime full potentiality of a particular algorithm cannot be determined due to certain factors that lead the accuracy of prediction to be uncertain. Many studies have shown that combining machine learning algorithm provide improved result compare to individual prediction of algorithm [1][11][16][17].

Manuscript received on 04 November 2023 | Revised Manuscript received on 15 November 2023 | Manuscript Accepted on 15 December 2023 | Manuscript published on 30 December 2023.
** Correspondence Author (s)*
Priyanka Mazumder*, Department of Computer Applications, Assam Science and Technical University, Tetelia Road, Jhalukbari, Guwahati (Assam), India. E-mail: priyank.mazumder@gmail.com, ORCID ID: 0000-0002-0761-780X
Dr. Siddhartha Baruah, Department of Computer Applications, Jorhat Engineering College, Garamur, Jorhat (Assam), India. E-mail: sbaruain@gmail.com

Combining multiple machine learning algorithms, a technique known as ensemble learning, proves to be a potent strategy for reducing prediction uncertainty and elevating overall accuracy. Ensemble methods harness the collective strengths of diverse algorithms, culminating in the creation of a more resilient and dependable predictive model. This approach is rooted in the fundamental principle that the synergy of multiple algorithms compensates for their individual weaknesses while amplifying their strengths. Both ensemble and hybrid algorithms aim to enhance the performance of machine learning models. They do so by leveraging the strengths of multiple algorithms or techniques to achieve more accurate and robust predictions or decisions. The combination can be done creating Hybrid machine Learning Model (HML). HML is an approach that combines more than one machine learning algorithm together to solve a particular problem. The main goal of HML is to improve the strength of individual algorithm so that overall performance of the model can be presented. HML help to decrease the individual algorithm limitations [2][13][15][18][26][27][28]. A single algorithm sometime cannot justify the performance depending upon certain dataset. It has been noticed that many single algorithm show different performance result if dataset is somewhat altered [3][14][19]. Classification is the technique where the class values are provided with label. Classification provides various categorized algorithms depending on Supervised and Unsupervised types. Supervised algorithm built with quality of individual learning with the help of example. Supervised can further subdivided into Parametric and Non Parametric classification. Parametric Classification depends on the probability distribution of each class and Non Parametric classifications are used when the density function is unknown. Unsupervised classification is a type of algorithm which learn from the observations i.e. the unsupervised algorithm do not have class value and possible class value is assigned depending of training result the algorithm gets[7][8][20][21]. There are different algorithms can be grouped to create HML. In this study three machine learning algorithm are combined to create a classifier for better performance. Classification algorithm has advantage that in research it has been noted that Supervised Algorithm combination provides best Hybrid Model outcomes than others [4][5][6][22]. The three classifiers that create HML are Random Forest (RF), Support Vector Machine (SVM) and Artificial Neural Network (ANN). The new approach is named as Hybrid Support Random Forest Neural Network (HSRFNN). HSRFNN is the development model based on three specifications:-

1. Improve the performance of predicting dataset.
2. Apply the advantages of each algorithm to develop a strong ensemble model.
3. HSRFNN will deploy the deep predicting knowledge for creating the appropriate decision.

The paper will show the development method of HSRFNN and its prediction outcome on certain dataset that are collected from UCI website and the accuracy will be collectively seen taking individual algorithm approaches. The paper will next discuss about the background knowledge of each Algorithm individually. Section 3 will discuss the proposed methodology of HSRFNN, Section 4 will discuss the performance analysis and Section 5 will follow up the conclusion.

II. BACKGROUND KNOWLEDGE

A. Random Forest

Random Forest was first introduced by Breiman(2001)[8][9] which is defined as a classifier that consist of a collection of decision tree where each treeis built by applying specific algorithm A on the training set T and an additional random vector θ. The prediction is based on majority vote over different individual predictor. There are certain approach for defining algorithm A over the Training set T. A sequence like X1,X2… where each X1is a subset of [d] of size K. Breiman(2009) presented the overall definition of random forest which can be extracted as Given an ensemble of classifiers h1(x), h2(x) . . .hK(x), and with the training set drawn at random from the distribution of the random vector Y, X, define the margin function as

$$mg\,(X, Y) = av_k\, I\,(h_k(X) = Y) - \max_{(j\neq Y)} av_k\, I\,(h_k(X) = j)$$

Where I (·) is the indicator function. The margin measures the extent to which the average number of votes at X, Y for the right class exceeds the average vote for any other class. The larger the margin, the more confidence is the classification. The generalization error is given by

$$PE = P_{X,Y}(mg(X, Y) < 0)$$

where the subscripts X, Y indicate that the probability is over the X, Y space.

In random forests, $h_k(X) = h(X,\Theta_k)$.

B. Support Vector Machine (SVM)

The SVM algorithm works to create a separator by creating large margin. SVM uses a separator which may be halfspace used in training dataset and the separated line may be called as hyperplane[23][24][25]. The main goal of SVM is to find the best hyperplane which separates the data points in two components by maximizing the margin [10][11][12][29][30]. Let S = (x1,y1)……(xm,ym)be a training set of examples, where each xi ϵ Rd and yi ϵ {±1}. We say that this training set is linearly separable, if there exists a halfspace, (w,b) such that yi = sign([w,xi] + b) for all i. Alternatively, this condition can be rewritten as:

$$\forall \epsilon\ [m],\ y_i\,((w\ ,\ x_i) + b) > 0$$

C. Artificial Neural Network (ANN)

It is defined by acyclic graph, G = (V, E) and a weight function over the edges, w: E→R. Nodes of the graph corresponds to neurons. Each single neuron is modeled as a simple scalar function, δ: R → R. Each edge in the graph links the output of some neuron to the input of another neuron. The input of the neuron is obtained by taking a weighted sum of the outputs of all the neurons connected to it, where the weighting is according to w. The ANN consists of layers-input, hidden and output. Let V0 be the input layer. It consists of n+1 neurons where n is the dimensionality of the input space. For every i ϵ [n], the output of the neuron i in Vo is simply xi. The last neuron in V0is the "constant" neuron, which always outputs 1. We denote by ʋt,ithe i th neuron of the tth layer and by ot,I (x) the output of ʋt,iwhen the network is fed with the input vector x. Therefore, for i ϵ [n] we haveot,I (x)=xiand for I = n+1 we have ot,I (x)=1. When the neural network is fed with the input vector x. Then we get,

$$a_{t+1,j}\,(x) = \Sigma\,(r: (\,v_{t,r}\,,\,v_{t+1,j}\,)\,\epsilon\,E$$

and $$o_{t+1,j}\,(x) = \sigma\,(a_{t+1,j}\,(x))$$

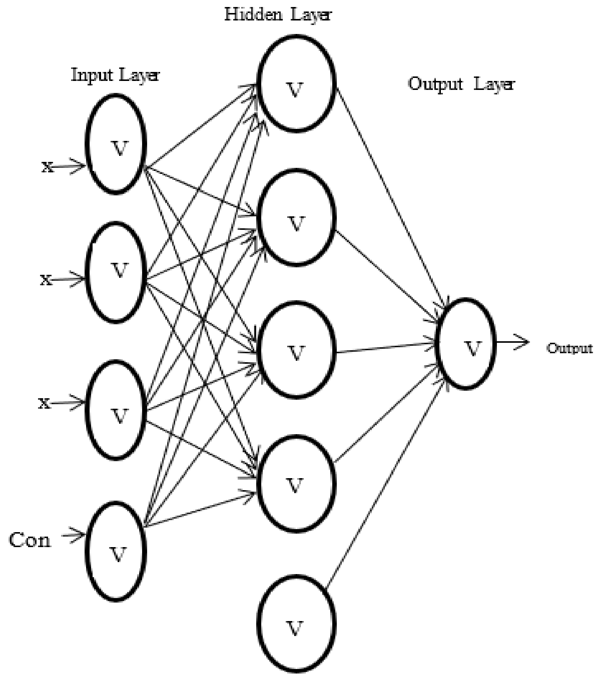

Fig. 1: Separation of Steps used in ANN

III. DATA DESCRIPTION AND PRE-PROCESSING

A. Data Description

In this research paper, a classification dataset sourced from both the UCI and Kaggle Open Data repositories was meticulously curate and systematically evaluated to assess the accuracy fluctuations of a Hybrid Model. Both UCI and Kaggle repositories are invaluable resources, offering a diverse collection of datasets for rigorous testing and validation of machine learning algorithms. The detailed dataset information is thoughtfully presented in Table 1 for reference. The UCI (University of California, Irvine) Machine Learning Repository stands as a revered haven for machine learning and data science professionals. It boasts an extensive repository of datasets, widely recognized for their recurrent application in research, experimentation, and the rigorous assessment of machine learning algorithms. Spanning diverse domains, these datasets are routinely utilized for multifaceted tasks, encompassing classification, regression, clustering, and beyond.

This rich resource is a prime destination for researchers and data enthusiasts, serving as a cornerstone for benchmarking, evaluation, and the evolution of machine learning models.

B. Data Pre-Processing

The formats of the datasets are different so they need to be converted. Some dataset has large number of missing values which need to be avoided; some instances had raw data which need to be left out. Dataset have different clusters which need to be combined and made a complete training dataset. The dataset's attributes are ranked based on some trained ranking algorithm because of which attribute selection is very well provided. The following steps shown in Figure 2 for data preprocessing are considered.

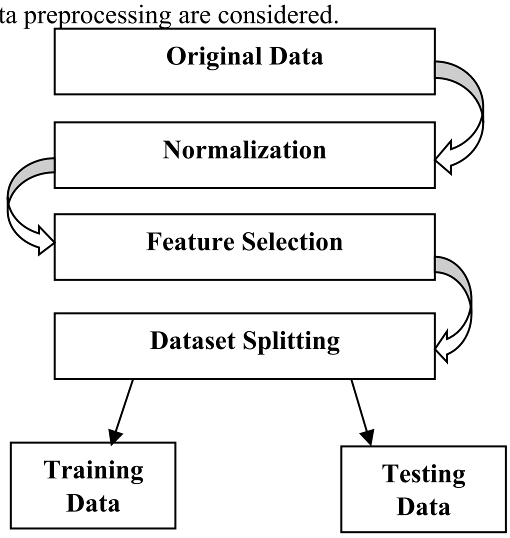

Fig. 2: Steps of Data Pre-Processing

Steps are explained as below:

1. Original Dataset: The original dataset include all the downloaded dataset from Kaggle and UCI data repository.
2. Normalization: Some datasets value ranges vary greatly. A basic minimum and maximum standard value need to be examined and used inside the dataset. Missing value instances are deleted so that the model must not get effect. The standardization can be expressed as:

$$x' = (x-\mu) / \delta$$

Where, μ represent the mean value of the feature δ represent the standard deviation of feature values

3. Feature Selection: In feature selection, the properties attribute selection is done. It is one of the most flexible and allows various search and evolution method to be combined. The process of discretization is also done, it allows a range of numeric attributes in the dataset into nominal attributes. The main advantage of Feature Selection is to find the best set of feature that allows creating useful module.
4. Dataset Splitting: Dataset is split into two parts, training data and testing data respectively. A complete 90% split is for training data and 10% for testing data. Training data is used to make model informative and testing data is used to test the class value depending on the trained information.

IV. PROPOSED METHODOLOGY

The paper developed a Hybrid Algorithm which will be predicting different dataset along with individual algorithms. The development structure of the hybrid model is discussed in Fig 3 and detailed explanation of each steps are given below. The steps are discussed below

1. UCI and Kaggle : UCI and Kaggle is Online Data repository, from the two data repository 10 datasets are taken. The dataset belongs to classification category which means all the dataset have class label.
2. Datasets: The selection of dataset depended on the well structured data collection. The 10 dataset are medical dataset. The hybrid model will test accuracy to all 8 dataset individually along with compare with other three algorithms i.e. Random Forest, Support Vector Machine and Artificial Intelligence.
3. Training data- Some data from individual dataset are separated. 90% data is kept as training and 10% data is kept as testing data. Training data will help to train the model and a last testing data will help to predict the outcome with maximum support class value.
4. Random Splitting: The training dataset will be separated randomly by using Random Forest technique i.e. shuffling of data along with separation among certain group.
5. Linear Separation: Separated individual group of data will be checked using hyper plane signification among the class value. Basically it will check either the separated and combined groups of data are splitted without collisions in hyper plane.
6. Tree Generator Selection Root Node: The ANN algorithm will check and find the best selective Root Node so that tree can be formed. It will select using neural network fundamental to generate the best root node from the separated groups.
7. Majority Voting: Depending on formation of free individual test data will be tested and result depending on data will be given.

V. PROPOSED HSRFNN

Load your datasets into memory, which are labeled as D1, D2, D3, ..., D8.

1. Load your datasets into memory, which are labeled as D1, D2, D3, ..., D8. 2.Apply data preprocessing techniques and feature selection to clean and optimize the datasets. This step ensures that the data is in a suitable format for testing. 3. Split each preprocessed dataset into training (90%) and testing (10%) subsets, maintaining separate subsets for each dataset (D1_train, D1_test, D2_train, D2_test, etc.).

For each dataset D_i, repeat the following steps:

4.1 Randomly divide the training data (e.g., D1_train) into individual groups [D1 (d1, d2, d3 …), D2 (d1,d2,d3…..)……..D8(d1,d2,d3)]

4.2 Train a Support Vector Machine (SVM) model on each individual group.

4.3 Check for data collisions or inconsistencies among the predictions from the individual groups. If any collisions are detected (e.g., disagreements between group predictions), go to the next step.

5. If a collision is detected during training, re-shuffle the data within the individual groups for that dataset to reduce the likelihood of collisions.

Retrieval Number: 100.1/ijitee.A97571213123
DOI: 10.35940/ijitee.A9757.1213123
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

5.1 Return to Step 4 and retrain the SVM models on the updated individual groups.

5.2 Continue this process until no collisions are detected or until a predefined number of iterations is reached.

6. After resolving collisions, update the individual groups and retrain the SVM models.

7. Build the root node and generate tree.

8. Repeat Steps 4-6 for each dataset (D1, D2, D3, ..., D8) to ensure that each dataset is properly handled and collisions are minimized.

9. Finally Maximum Tree generates by selecting best Root Node(R) using ANN algorithm specifications.

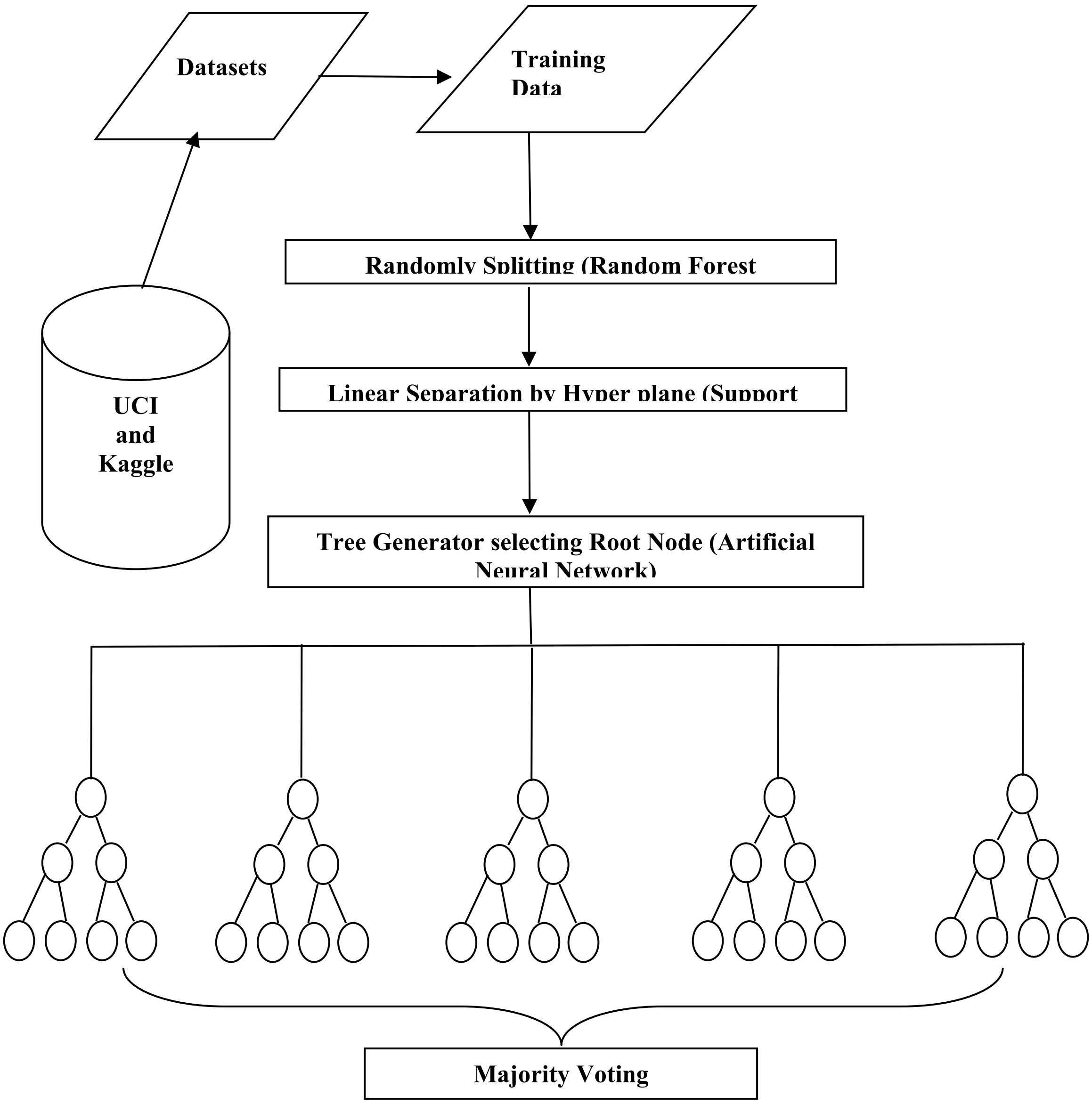

Fig. 3: Hybrid Algorithm Structure Flow

Table 1 Dataset Descriptions

Dataset Name	Characteristics	Instances	Attributes	Attribute Type
Iris	Multivariate	150	4	Real
Heart Disease	Multivariate	303	13	Categorical, Integer, Real
Diabetes	Multivariate, Time-Series	500	20	Categorical, Integer, Real
Breast Cancer	Multivariate	569	30	Real
Lung Cancer	Multivariate	697	19	Integer
Liver Disorder	Multivariate	345	5	Categorical, Integer, Real
Thyroid Disease	Multivariate, Domain-Theory	5100	5	Categorical, Real
Arrhythmia	Multivariate	452	279	Categorical, Integer, Real

Retrieval Number: 100.1/ijitee.A97571213123
DOI: 10.35940/ijitee.A9757.1213123
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Table 2: The Accuracy Depending on Correctly Classified Instances (CCI), Incorrectly Classified Instances (ICI), Accuracy (A) and Time Taken to Built Model (TTBM). Shown all the Comparison between Algorithms with HSRFNN

Dataset Name	Algorithm	C.C.I	I.C.I	A(%)	TTBM
Iris	RF	100	50	66	2.5
	SVM	90	60	60	3.93
	ANN	110	40	73	3.23
	HSRFNN	140	10	93	2.73
Heart Disease	RF	230	73	75	5.25
	SVM	215	88	70	3.21
	ANN	190	113	62	3
	HSRFNN	280	23	92	3.75
Diabetes	RF	415	85	83	7.1
	SVM	317	183	63	6.23
	ANN	132	368	26.4	5.23
	HSRFNN	459	41	91.8	4.23
Breast Cancer	RF	435	134	76	3.13
	SVM	414	155	72	2
	ANN	301	268	52	3.11
	HSRFNN	515	54	90	2.15
Lung Cancer	RF	515	182	73	2.12
	SVM	441	256	63	3.31
	ANN	400	297	57	4.23
	HSRFNN	655	42	93	3.22
Liver Disorder	RF	290	55	84	2.33
	SVM	313	32	90	1.09
	ANN	145	200	42	2.12
	HSRFNN	330	15	95.6	1.14
Thyroid Disease	RF	4123	977	80	2.23
	SVM	4900	200	96	2.12
	ANN	4950	150	97	2.44
	HSRFNN	5075	25	99	2.48
Arrhythmia	RF	390	62	86	1.12
	SVM	400	52	88	1.05
	ANN	313	139	69	1.68
	HSRFNN	430	22	95	1.78

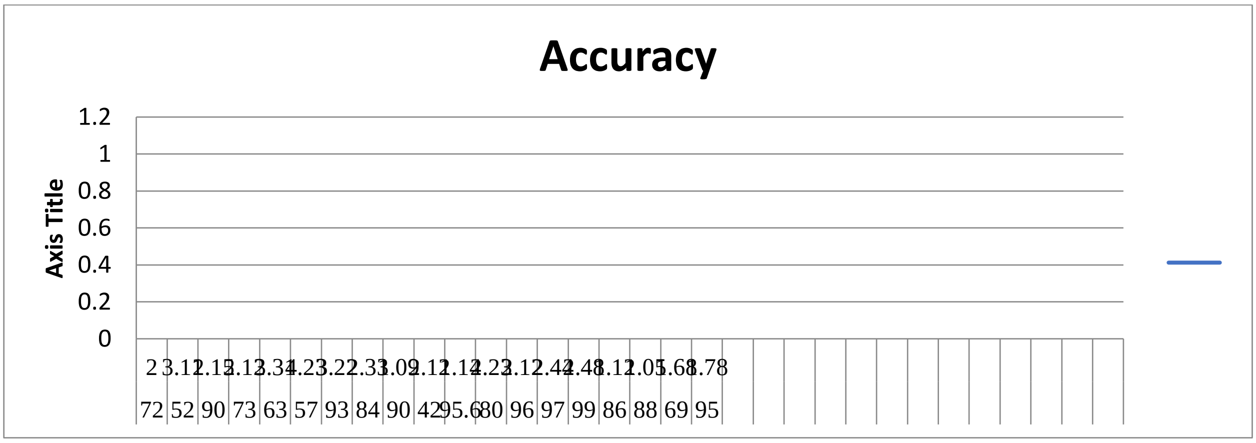

Fig. 4: Accuracy of HSRFNN

VI. RESULT AND DISCUSSION

The hybrid algorithm HSRFNN is constantly provided the best accuracy in all the dataset properties. The algorithm seems to be better than any other individual algorithm which means the HSRFNN fitted better compare to Random Forest, Support Vector Machine and Artificial Neural Network. The hybrid algorithm provided stable result among all datasets. The paper tried to provide a hybrid algorithm which perform or provide considerable accuracy by taking classification dataset. The Hybrid model shows higher accuracy compare to other traditional methods. The comparison is shown in Table 2. The consistent performance of changes in dataset is not affected while performing accuracy by HSRFNN. The main goal of the study is to provide a combined new hybrid model that must not get affected by the change of the dataset. Datasets are pre-processed by various techniques like generalization, re-sampling of dataset, feature selection and ranking of feature.

Retrieval Number: 100.1/ijitee.A97571213123
DOI: 10.35940/ijitee.A9757.1213123
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Ranking of feature allow choosing maximum ranking scored attributes so that the traditional algorithm and hybrid model must not get affected due to dataset. Hybrid model provided consistent accuracy by testing all 8 dataset. Cross Validation techniques is also incorporated to maximize the performance of the algorithm. The individual comparison for accuracy graph is shown in Figure 4.

VII. CONCLUSIONS

Hybrid Algorithm is the development of new method from constructive methods. The algorithm HSRFNN goal is to provide best outcomes by combining strength of each algorithm together. These lead improved performance, robustness and efficiency in providing accuracy. The goal of this paper is to combine nature of different algorithm to achieve best result than using single algorithm alone. In future there will testing of algorithm depending on more large dataset. HSRFNN provides suitable result for classification dataset, in future the different types of dataset will be considered for better performance. A Hybrid Support Random Forest Neural Network (HSRFNN) is a novel algorithm that integrates the strengths of both Random Forest and Neural Networks to improve model performance. This approach combines the ensemble learning capabilities of Random Forest with the deep learning capabilities of Neural Networks. The goal of HSRFNN is to capitalize on the advantages of each method and create a more robust and accurate model. This hybrid approach can potentially provide better results than using either Random Forest or Neural Networks in isolation. In the future, HSRFNN may be further tested and optimized, particularly on larger datasets, to assess its potential in various applications.

DECLARATION STATEMENT

Funding	No, I did not receive.
Conflicts of Interest	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material	Not relevant.
Authors Contributions	All authors have equal participation in this article.

REFERENCES

1. Shahhosseini, Mohsen, and Guiping Hu. "Improved weighted random forest for classification problems." Progress in Intelligent Decision Science: Proceeding of IDS 2020. Springer International Publishing, 2021.
2. Sudiana, D.; Lestari, A.I.; Riyanto, I.; Rizkinia, M.; Arief, R.; Prabuwono, A.S.; Sri Sumantyo, J.T. A Hybrid Convolutional Neural Network and Random Forest for Burned Area Identification with Optical and Synthetic Aperture Radar (SAR) Data. *Remote Sens.* **2023**, *15*, 728. https://doi.org/10.3390/rs15030728
3. Togunwa Taofeeq Oluwatosin, Babatunde Abdulhammed Opeyemi, Abdullah Khalil-ur-Rahman "Deep hybrid model for maternal health risk classification in pregnancy: synergy of ANN and random forest" Frontiers in Artificial Intelligence, VOL-6 (2023) ,https://www.frontiersin.org/articles/10.3389/frai.2023.1213436 , 10.3389/frai.2023.1213436, ISSN:2624-8212
4. Aruna Kumari G.L. , Dr Padmaja P. , Dr Jaya Suma G. , "Logistic regression and Random forest-based hybrid classifier with recursive feature elimination technique for diabetes classification", International Journal of Advanced Trends in Computer Science and Engineering, Volume 9, No.4,(2020)http://www.warse.org/IJATCSE/static/pdf/file/ijatcse379942020.pdf https://doi.org/10.30534/ijatcse/2020/379942020
5. E. Syed Mohamed, Tawseef Ahmad Naqishbandi, Syed Ahmad Chan Bukhari, Insha Rauf, Vilas Sawrikar, Arshad Hussain, "A hybrid mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms", Healthcare Analytics,Volume 3,(2023), ISSN 2772-4425, https://doi.org/10.1016/j.health.2023.100185. (https://www.sciencedirect.com/science/article/pii/S2772442523000527)
6. M Hemalatha, "A hybrid random forest deep learning classifier empowered edge cloud architecture for COVID-19 and pneumonia detection", Expert Systems with Applications, Volume 210, (2022), ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2022.118227. (https://www.sciencedirect.com/science/article/pii/S0957417422013781)
7. P. Mazumder and S. Baruah, "A Community Based Study for Early Detection of Postpartum Depression using Improved Data Mining Techniques," *2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, Bangalore, India, 2021, pp. 1-7, doi: 10.1109/CSITSS54238.2021.9682941.
8. Breiman, Leo(2001), "Random Forests", Machine Learning,45(1),5-32 https://doi.org/10.1023/A:1010933404324.
9. Breiman, L. (1996). "Bagging predictors", Machine Learning, 24(2), 123-140.
10. Cortes, Corinna, Vapnik, Vladimir (1995), "Support-vector networks",Machine Learning, 273- 297, 20(3), https://doi.org/10.1007/BF00994018
11. Yassine Al Amrani, Mohamed Lazaar, Kamal Eddine El Kadiri, "Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis", Procedia Computer Science, Volume 127, (2018), 511-520,ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.01.150.
12. L.A. Demidova, I.A. Klyueva, A.N. Pylkin, "Hybrid Approach to Improving the Results of the SVM Classification Using the Random Forest Algorithm", Procedia Computer Science,Volume 150, (2019), 455-461, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2019.02.077.
13. Admassu Tsehay, Subhashni Rajkumar, Napa, Komal Kumar, Prasath, Jijendira Duraisamy, Pradeep, Engidaye, Minychil (2022), "Random forest and support vector machine based hybrid liver disease detection", Bulletin of Electrical Engineering and Informatics ,VL - 11,10.11591/eei.v11i3.3787
14. Hasan Mehedi A. M., Nasser M., Pal B., Ahmad S., "Support Vector Machine and Random Forest Modeling for Intrusion Detection System", Journal of Intelligent Learning Systems and Applications, (2014), Vol 6(1)
15. H. Xu and M. Ma, "An Improved Hybrid Model base on SVM and Random Forest for the Prediction of Corporate Taxation," *2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, Changsha, China, 2021, pp. 1035-1038, doi: 10.1109/ICCASIT53235.2021.9633659.
16. Tun W, Wong JK, Ling SH. Hybrid Random Forest and Support Vector Machine Modeling for HVAC Fault Detection and Diagnosis. Sensors (Basel). 2021 Dec 7;21(24):8163. doi: 10.3390/s21248163. PMID: 34960257; PMCID: PMC8704049.
17. Lilhore, U.K., Manoharan, P., Sandhu, J.K. *et al.* "Hybrid model for precise hepatitis-C classification using improved random forest and SVM method". *Sci Rep* **13**, 12473 (2023). https://doi.org/10.1038/s41598-023-36605-3.
18. Jorge Alexander Ángeles Rojas, Hugo D. Calderón Vilcn,Ernesto N. Tumi Figueroa,Kent, Jhunior Cuadros Ramos, Steve S. Matos Manguinuri, Edwin F. Calderón Vilca, "Hybrid Model of Convolutional Neural Network and Support Vector Machine to Classify Basal Cell Carcinoma", Comp. y Sist. vol.25 no.1 Ciudad de México ene./mar. 2021 Epub 13-Sep-2021, https://doi.org/10.13053/cys-25-1-3431
19. Fatih Bal, Fatih Kayaal, "A Novel Deep Learning-Based Hybrid Method for the Determination of Productivity of Agricultural Products: Apple Case Study" ,*IEEE Access,* 25 January 2023, Digital Object Identifier 10.1109/ACCESS.2023.3238570
20. Jackins, V., Vimal, S., Kaliappan, M. *et al.* AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *J Supercomput* **77**, 5198–5219 (2021). https://doi.org/10.1007/s11227-020-03481-x

21. Huang, F., Shen, J., Guo, Q. *et al.* eRFSVM: a hybrid classifier to predict enhancers-integrating random forests with support vector machines. *Hereditas* **153**, 6 (2016). https://doi.org/10.1186/s41065-016-0012-2.
22. Shen, J., Shafiq, M.O. Short-term stock market price trend prediction using a comprehensive deep learning system. *J Big Data* **7**, 66 (2020). https://doi.org/10.1186/s40537-020-00333-6
23. Chen, RC., Dewi, C., Huang, SW. *et al.* Selecting critical features for data classification based on machine learning methods. *J Big Data* **7**, 52 (2020). https://doi.org/10.1186/s40537-020-00327-4
24. Kunihiro Kamataki, Hirohi Ohtomo, Naho Itagaki, Chawarambawa Fadzai Lesly, Daisuke Yamashita, Takamasa Okumura, Naoto Yamashita, Kazunori Koga, Masaharu Shiratani; Prediction by a hybrid machine learning model for high-mobility amorphous In_2O_3: Sn films fabricated by RF plasma sputtering deposition using a nitrogen-mediated amorphization method. *J. Appl. Phys.* 28 October 2023; 134 (16): 163301. https://doi.org/10.1063/5.0160228
25. Sneha Bobde, Sharvari Role, Lokesh Khadke , Tejas Shirude , Ms. Shital Kakad, "Email Spam Detection Using Hybridization ofSVM and Random Forest", International Journal of Research in Engineering and Science (IJRES) ISSN (Online): 2320-9364, ISSN (Print): 2320-9356, Volume 11 Issue 7, July 2023 ǁ PP. 188-193
26. Wardhana, M. H., Basari, Prof. Dr. A. S. H., Mohd Jaya, Dr. A. S., Afandi, Prof. Dr. dr. D., & Dzakiyullah, N. R. (2019). A Hybrid Model using Artificial Neural Network and Genetic Algorithm for Degree of Injury Determination. In International Journal of Innovative Technology and Exploring Engineering (Vol. 9, Issue 2, pp. 1357–1365). https://doi.org/10.35940/ijitee.b6169.129219
27. Jebamalar, J. A., & Kumar, Dr. A. S. (2019). PM2.5 Prediction using Machine Learning Hybrid Model for Smart Health. In International Journal of Engineering and Advanced Technology (Vol. 9, Issue 1, pp. 6500–6505). https://doi.org/10.35940/ijeat.a1187.109119
28. Behera*, D. K., Das, M., & Swetanisha, S. (2019). A Research on Collaborative Filtering Based Movie Recommendations: From Neighborhood to Deep Learning Based System. In International Journal of Recent Technology and Engineering (IJRTE) (Vol. 8, Issue 4, pp. 10809–10814). https://doi.org/10.35940/ijrte.d4362.118419
29. Muthukrishnan, Dr. R., & Prakash, N. U. (2023). Validate Model Endorsed for Support Vector Machine Alignment with Kernel Function and Depth Concept to Get Superlative Accurateness. In International Journal of Basic Sciences and Applied Computing (Vol. 9, Issue 7, pp. 1–5). https://doi.org/10.35940/ijbsac.g0486.039723
30. Sistla, S. (2022). Predicting Diabetes u sing SVM Implemented by Machine Learning. In International Journal of Soft Computing and Engineering (Vol. 12, Issue 2, pp. 16–18). https://doi.org/10.35940/ijsce.b3557.0512222

AUTHORS PROFILE

Priyanka Mazumder received the BCA degree from Punjab Technical University, Punjab, India in 2014, received MCA degree from Center For Computer Sciences and Applications in Dibrugarh University, Dibrugarh, India in 2018. She is also a Ph.D degree candidate in Jorhat Engineering College under Assam Science and Technology University, Assam, India. She is currently serving as Assistant Professor in Department of Information Technology under School of Computing Sciences, Assam Kaziranga University, India. She has around 4+ years of teaching experience.Her research interests include Machine Learning, Data Mining, Deep Learning and Artificial Intellegence.

Dr. Siddhartha Baruah received the B.Sc. degrees in Physics from Science College Jorhat(Currently known as Jorhat Engineering College), India in 1987, the MCA degree in Computer Application from Jorhat Engineering College, India in 1990, and the Ph.D degree in Computer Science from Guwahati University, India in 2010. He is currently working as a Professor and HOD in Department of Computer Application, Jorhat Engineering College, India. He has around 28 years of teaching experiences and 12 years of research experience. He has completed several projects in Modernization and Removal of Obsolesces (MODROBS) approved by All India Council of Technical Education (AICTE). He has published several papers as well as attended different international conferences in India and Abroad. His intereset include Embedded System, IOT, Data Mining, Machine Learning.