Prediction of dementia using machine learning model and performance improvement with cuckoo algorithm

Sivakani Rajayyan[1], Syed Masood Mohamed Mustafa[2]
[1]Department of Computer Science and Engineering, B. S. Abdur Rahman Crescent Institute of Science and Technology, Vandalur, India
[2]Department of Computer Applications, B. S. Abdur Rahman Crescent Institute of Science and Technology, Vandalur, India

Article Info

Article history:

Received Jun 30, 2022
Revised Sep 16, 2022
Accepted Oct 1, 2022

Keywords:

Classification
Dementia
Gaussian naïve Bayes
Linear regression
Support vector machine

ABSTRACT

Dementia is a brain disease that stays in the seventh position of death rate as per the report of the World Health Organization (WHO). Among the various types of dementia, Alzheimer's disease has more than 70% of cases of dementia. The objective is to predict dementia disease from the open access series of imaging studies (OASIS) dataset using machine learning techniques. Also, the performance of the machine learning model is analyzed to improve the performance of the model using the cuckoo algorithm. In this paper, feature engineering has been focused and the prediction of dementia has been done using the OASIS dataset with the help of data mining techniques. Feature engineering is followed by prediction using the machine learning model Gaussian naïve Bayes (NB), support vector machine, and linear regression. Also, the best prediction model has been selected and done the validation. The evaluation metrics considered for validating the models are accuracy, precision, recall, and F1-Score and the highest values are 95%, 97%, 95%, and 95%. The Gaussian NB has been given these best results. The accuracy of the machine learning models has been increased by eliminating the factors which affect the performance of the models using the cuckoo algorithm.

Corresponding Author:

Syed Masood Mohamed Mustafa
Department of Computer Applications, B. S. Abdur Rahman Crescent Institute of Science and Technology
Vandalur, Chennai, India
Email: ms.masood@crescent.education

1. INTRODUCTION

The process of converting the raw data into useful data as required is called data mining. The paper gives the contribution of feature engineering and the prediction of dementia using machine learning models [1]. The features are very important for a good prediction. The data analyzer should decide what kind of data and in which format it is required; for this, feature engineering is the focus. Feature engineering is the process of making the dataset in such a way that is required for processing using machine learning models [2], so that the prediction done using the machine learning model will be the best and most accurate. The process of creating new features from the existing raw data for processing is called feature creation [3]. The required feature for good prediction can be created using data mining tools. The process of changing the data to another required format is called transformation. Data encoding is applied to convert the text to numeric values for the instances in the dataset. The process of extracting the required feature from the dataset is called feature extraction. Each and every data will be processed and based on the importance of the data the feature will be extracted. The process of selecting the required feature is called feature selection [4]. Feature selection is an important process in feature engineering.

A machine learning model is a model constructed using the machine learning algorithm for the prediction of results using an input dataset. The machine learning model can be mainly classified into supervised and unsupervised learning models [5]. The supervised learning model trains the labeled data and predicts the result. The unsupervised learning model trains the unlabeled data and predicts the results. The electrocardiogram (ECG) beats of the arrhythmia have been classified using the machine learning classifiers decision trees, support vector machine (SVM), and random forest [6]. The novel feature engineering method was introduced using the deep learning method along with the K-nearest neighbor (KNN). A 5-fold cross-validation has been implemented for the validation and the accuracy obtained was 99.77%. The advantage is featuring extraction gives a high balance of sensitivity and specificity [7]. Feature selection was done using the chi2, information gain score, and backward elimination method. The feature selection process makes the distributed denial-of-service (DDoS) process faster. The model has been constructed using the machine learning methods KNN, SVM, naïve Bayes (NB), artificial neural network (ANN), and cross-validation has been applied to find the better performance of the algorithms [8]. A two-classifier pool has been created with four classifiers in each pool. Using the ensemble method, the best classifier has been selected in this multi-subspace elastic network (MSEN) proposed model. The missing values are generated using the KNN algorithms and the outliers are removed using the isolation forest method. Feature selection has been done using the recursive feature elimination method and the model has been validated using the real-world dataset [9]. An efficient blockchain secure healthcare system for predicting disease using fog computing was created. The data are collected using fog nodes and stored in a blockchain; the rule-based cluster method has been applied to diabetes and cardio disease data and the feature selection was done using the forward selection-adaptive neuro-fuzzy inference system (FS-ANFIS) method. The accuracy of the prediction was 81% [10]. An ensemble model has been constructed for the classification of hepatitis-C medical record patients with 75 records [11].

A new method was introduced for breast cancer detection by applying machine learning algorithms to the clinical dataset. Feature selection was done using the auto-encoder and principal component analysis (PCA) algorithm. The relief-SVM has been applied to diagnose the disease [12]. Multi-classification was done on the cleaned and selected features. Machine learning models are used for the classification and K-fold cross-validation has been applied for the evaluation of the dataset [13]. The weighted ensemble method was introduced to predict Alzheimer’s disease (AD) early stage. The feature selection methods applied were recursive feature elimination (RFE) and L1 regularization [14]. The linear discriminant analysis (LDA) model along with the RF classifier was constructed for the classification of the neurodegenerative problem. For the feature extraction, the LDA method is applied and for the classification, an RF classifier is applied [15]. A machine learning (ML) classifier model was introduced for the detection of AD. The feature selection method applied was RFE and the 10-fold cross-validation was applied for the evaluation of the dataset [16]. A machine learning model was introduced for the AD classification and for the feature selection elaboration likelihood model (ELM) method was applied and cross-validation was done on the Alzheimer's disease neuroimaging initiative (ADNI) data set [17]. Various machine learning algorithms are used for generating the missing values in the open access series of imaging studies (OASIS) Alzheimer dataset [18]. A framework has been developed for the prediction of Alzheimer’s [19]. A multipath delay commutator has been proposed for enhancing the throughput and speed [20], [21]. Cooperative routing using the fresher encounter algorithm to improve efficiency [22].

A new injury metric has been introduced for computing the risk due to the injury in the head [23]. A hybrid algorithm has been introduced using the combination of Harris Hawks optimizer (HHO) and cuckoo search (CS) and chaotic maps. Feature selection is also implemented for better performance. The CS maintains the vector’s position and HHO balances the exploitation and exploration process. Experimental and statistical analysis has been done [24]. A modified cuckoo search algorithm has been introduced along with the variational parameter and logistic map (VLCS) to solve the dimension problems [25]. Three new cuckoo search algorithms have been developed by using various parameters and the performance has been tested using the mathematical model. Finally, concluded that the cuckoo algorithm improves the performance by changing the parameters [26]. A modified cuckoo search algorithm has been used for tournament selection in robot path planning [27]. A new hybrid model has been developed using the conditional mutual information maximization algorithm and the cuckoo search algorithm for the prediction of the disease [28]. The particle swarm optimization method tolerates uncertainty and imprecision to a maximum extent [29]. ANN controller for load frequency control of a four-area interconnected power system. This controller is designed by optimal control theory to defeat the issue of load frequency control. A feed-forward neural network with multi-layers and Bayesian regularization backpropagation training function is utilized [30]. The literature shows that various machine learning methods are used for the prediction and classification of the disease but no performance improvement has been done. So, the prediction of dementia has been done using the machine learning model and the performance of the model has been improved using the cuckoo algorithm. The result of the proposed model has been compared with [31], [32], and the proposed shows the best result. In the proposed framework the Gaussian NB, SVM, and linear regression (LR) classifiers are used for the machine learning model construction. These classifiers belong to the supervised learning model. The validation of the result has been done using the hold-out cross-validation. The performance of the model has improved using the cuckoo algorithm. Outlines of the contributions are OASIS dataset for processing, feature engineering concept implementation, constructing the machine learning model, prediction of the dementia disease, validating and analyzing the results obtained, and performance improvement of the model using the cuckoo algorithm.

The prediction of dementia disease has been done using machine learning techniques and the result has been validated. The improvement in the result has been done using the cuckoo algorithm. This paper is organized with an introduction followed by the proposed framework, method, result and discussion, and conclusion.

2. PROPOSED FRAMEWORK

The proposed model has the dataset and is subjected to feature engineering, and the selected features are given to the machine learning model for prediction and the validation has been done. Figure 1 describes the proposed model; the feature engineering concept has been applied to the OASIS dataset for the prediction. The machine learning model has been constructed and the result has been validated. Finally, the performance of the machine learning model has been improved using the cuckoo algorithm. In [33]–[39], feature engineering, feature creature feature transformation, feature selection, and feature extraction process have been implemented. Gaussian NB, SVM, and LR classifiers are used for the construction of the machine learning model and the validation has been done using the hold-out validation technique.

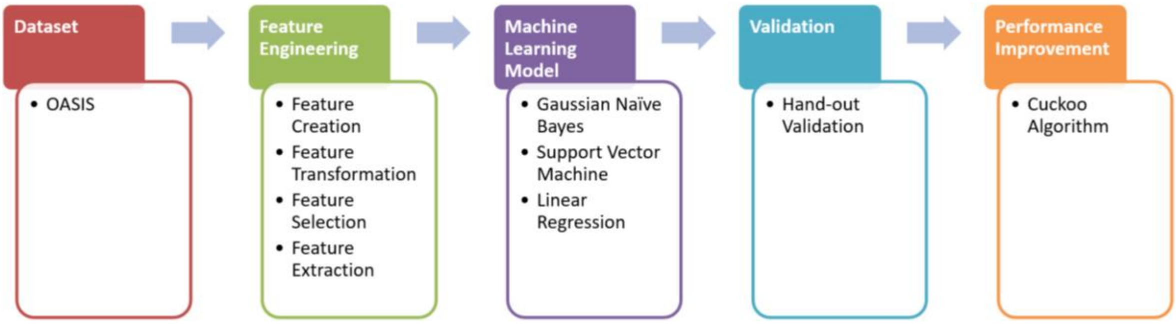

Figure 1. A framework of the proposed model

2.1. Dataset

The OASIS dataset has been taken for validation of the proposed model; the execution has been done using the WEKA and python tools. The input features considered are specified in the dataset section. The dataset has been subjected to a feature engineering process and the processed data has been given for the machine learning model and the validation has been done. The performance of the model has been improved by using the cuckoo algorithm. OASIS longitudinal dataset is a freely available dataset in the Washington University Alzheimer's Disease Research Center, the attributes and the description are given in Table 1.

Table 1. Data description

Sl. No.	Data Attributes	Data Description
1	SubjectID	Identification of the subjects
2	MRIID	Identification of the MRI
3	MF	Gender
4	Hand	Dominant hand
5	Age	Age of the patient
6	Educ	Education of the patient
7	SES	Socio-economic status of the patient
8	MMSE	Mini mental state examination
9	CDR	Clinical dementia rating
10	eTIV	Estimated total intra-cranial volume
11	nWBV	Normalize whole brain volume
12	Visit	Visiting of the patient
13	MRDelay	Time delay
14	ASF	Atlas scale factor of the patient
15	Group	Class of the patient

2.2. Feature engineering

Feature engineering has a major role in good prediction. In feature engineering, the following tasks are implemented for the prediction of dementia. Feature engineering has feature creation, feature transformation, feature selection, and feature extraction. The feature creation process of creating the features as required for the processing from the raw dataset is called feature creation. It contains two types: data creation and data imputation.

Data creation, in which data required for processing, can be created from the raw data and is called data creation. In this dataset, we have 15 data attributes for processing. The data imputation process of replacing the missing value with any other data is called data imputation. There are many types of imputation techniques. In this dataset, SES and MMSE attributes have some missing values. The mean imputation method has been applied for generating the missing values.

The process of converting the data from one format to another required format is called feature transformation. Label encoding has been implemented for converting the text to numeric format. The process of selecting the required features for the processing is called feature selection. i) correlation-based feature selection (CFS) subset evaluator and ii) info gain attribute evaluator

CFS subset evaluator is one of the feature selection techniques. CFS means based on the correlation between the attributes the features are selected along with the best search method. Info gain attributes evaluator means information gain-based attribute evaluator; the features are selected based on the entropy method along with the Ranker search method. This is one of the feature selection techniques. The required features that can be extracted from the given features for the best prediction are called feature extraction. The PCA method has been applied for the extraction of the features. Using the co-variance principle, the features are extracted in PCA.

2.3. Machine learning model construction

The machine learning model is the model which processes the data which we can feed and generate the result by learning the data. The data can be split into training data and testing data. The training data is given to train the model and for the validation testing data can be given. Various machine learning algorithms are used for constructing machine learning models.

- Gaussian NB: a type of NB classifier; also called normal distribution. In this classifier, each feature is continuously distributed.
- SVM: a supervised learning classifier where the classification can be done using the margin maximization principle.
- LR: a predictor, used for predicting the result using the relationship between the continuous data.

2.4. Validation

Validation is the process of evaluating the machine learning models and finding the error rate using the dataset. 80-20 data splitting has been done for the training and testing; this type of validating is called as hold-out validation technique. The result is predicted and the evaluation is done using the confusion matrix and selecting the best prediction model from the given models.

2.5. Performance improvement of the machine learning model

To improve the performance of the machine learning models the mathematical model has been done using the cuckoo algorithm. Generally, the performance can be improved by applying any one of the given methods such as, adding more data or handling the missing values and the outliers or proper use of feature engineering or parameter turning methods or using any algorithms. Hypothesis testing can be also done for finding a better solution. Hypothesis testing can be of two types of null hypothesis H_0 and alternate hypothesis H_A. H_0 shows that there will change in the accuracy and H_A shows that there will be alternate accuracy for the models. The pseudo-code for the proposed model is given below.

```
Input oasis longitudinal dataset
Feature Creation
        D←Dataset_creation
        D1←mean_imputationD
Feature Transformation
        D2←Label_EncodingD1
Feature Extraction
        D3←PCAD2
Feature Selection (CFS Subset Evaluator & Best First search, Info Gain AttributeEvaluator&
Ranker)
D4←Best_AttributesD3
Classification
```

```
Classification D4
        Best_ModelGaussian NB, SVM, LR
        Validation
Hold-out cross validationD4
        Output Result←dementia_predictionBest_Model
```

3. RESULTS AND DISCUSSION

In OASIS longitudinal dataset, 373 instances are available among which 146 data records come under the class demented, 190 data records come under non-demented, and 37 data records come under the converted class. Figure 2 shows the class distribution in the dataset. Figure 3 represented the missing data in the dataset. The OASIS dataset is an incomplete dataset. Among the 15 data attributes SES and MMSE attributes have incomplete data values, in the feature engineering task, the missing data has been generated and converted to a complete dataset.

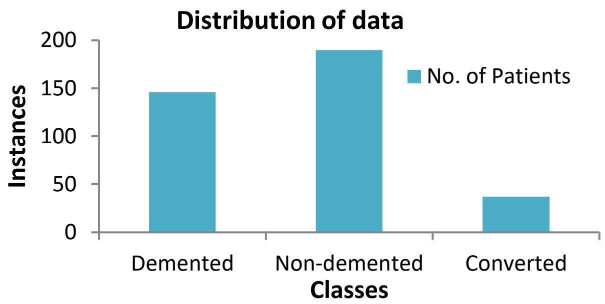

Figure 2. Graphical representation for the comparison of all the models

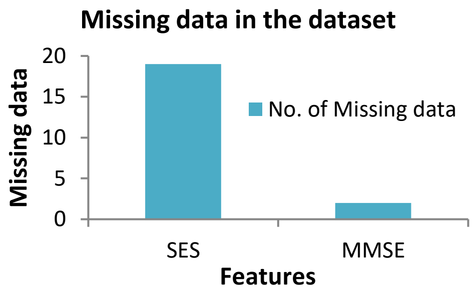

Figure 3. Graphical representation of the accuracy for all the models

The main motivation of this paper is to focus on feature engineering and to predict dementia. Also, the accuracy result is to be improved using the cuckoo algorithm. Feature engineering has been organized as feature creation followed by feature transformation, feature extraction, and feature selection. The OASIS dataset has 15 attributes and it has been analyzed that there are some missing values in SES and MMSE attributes. The mean imputation technique has been applied to generate the missing values. The missing values are represented in Figure 3. Label encoding has been applied for the SubjectID, MRIID, M/F, Hand, and Group in the dataset for converting the text to the numeric values in the data transformation task.

Feature extraction and feature selection have been implemented for reducing the dimensionality of the dataset. The CfsSubsetEval function has been applied along with the Best First search method and this function selected 12 attributes. Also, the info gain attribute evaluator along with the Ranker search method has been applied and it also selected the same 12 attributes. For the best prediction, the attributes have a major role. In the feature selection, the major attributes are selected for prediction. The machine learning model has been constructed for multi-class classification. Gaussian naive Bayes, SVM, and LR are the classifiers used for the processing. Table 2 shows the evaluation metrics for the Gaussian NB classifier.

Figure 4 represents the graphical representation of the precision, recall, and F1-score value for the Gaussian NB. Figure 5 represents the confusion matrix for the Gaussian NB. The classification of the three classes has been done and the evaluation metrics predicted have been tabulated. The performance has been analyzed and represented in graphical form. The accuracy for the NB classifier has been 95%. Table 3 shows the evaluation metrics for the SVM classifier. Figure 6 represents the graphical representation of the precision, recall, and F1-score values for the SVM. Figure 7 represents the confusion matrix for the SVM. The classification of the 3 classes has been done and the evaluation metrics predicted have been tabulated. The performance has been analyzed and represented in graphical form. The accuracy for the SVM classifier has been 89%.

Table 2. Evaluation metrics of Gaussian NB

Sl. No.	Class	Precision	Recall	F1-Score
1	Converted	1	0.40	0.57
2	Dementia	0.91	1	0.95
3	Non-Dementia	0.92	1	0.96

Table 3. Evaluation metrics of SVM

Sl. No.	Class	Precision	Recall	F1-Score
1	Converted	0.67	0.25	0.36
2	Dementia	0.96	0.93	0.95
3	Non-Dementia	0.86	1	0.93

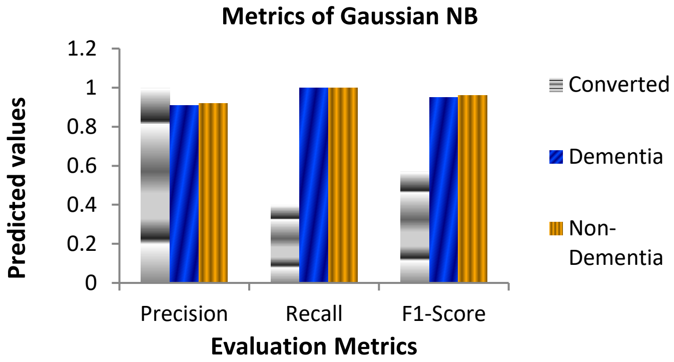

Figure 4. Graphical representation for the evaluation metrics for Gaussian NB

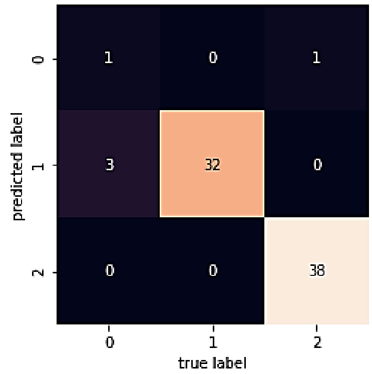

Figure 5. Confusion matrix for Gaussian NB

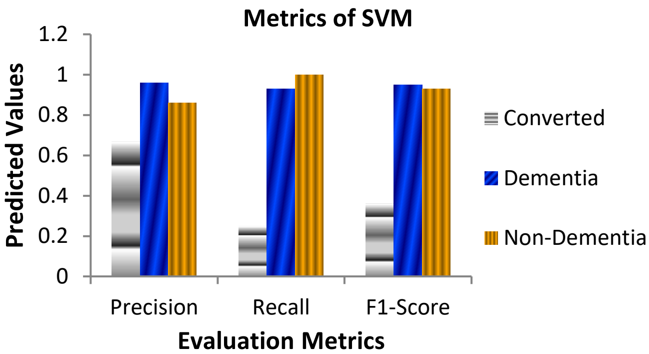

Figure 6. Graphical representation of evaluation metrics for SVM

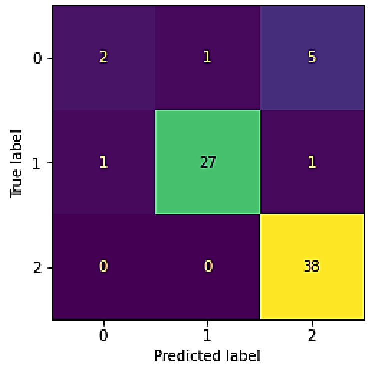

Figure 7. Confusion matrix for SVM

Figure 8 represents the graphical representation of the precision, recall, and F1-score values for the LR. Figure 9 represents the confusion matrix for the LR. The classification of the 3 classes has been done and the evaluation metrics predicted have been tabulated. The performance has been analyzed and represented in graphical form. The accuracy for the LR classifier has been 63%. Table 4 shows the evaluation metrics for the LR classifier. Table 5 shows the comparison of all the models and Figure 10 shows the graphical representation of the comparison chart. Table 6 shows the comparison of the accuracy of all the classifiers. Table 6 shows the comparison of the accuracy of all the models, and Figure 11, shows the graphical representation of the accuracy comparison chart. The Gaussian NB, SVM, and LR machine learning models have been used for the prediction of dementia. The validation has been done using the hold-out cross-validation technique. By comparing the evaluation metrics and the accuracy it is identified that the Gaussian NB classifier is the best model among these models.

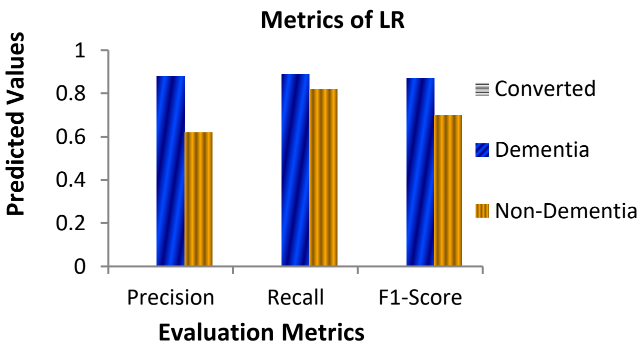

Figure 8. Graphical representation for the evaluation metrics for LR

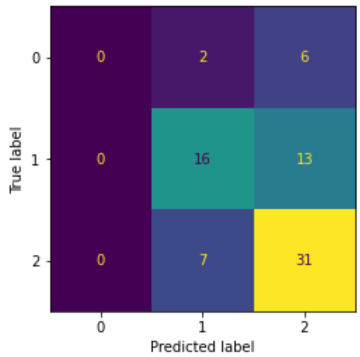

Figure 9. Confusion matrix for LR

Table 4. Evaluation metrics of LR

Sl. No.	Class	Precision	Recall	F1-Score
1	Converted	0.00	0.00	0.00
2	Dementia	0.88	0.89	0.87
3	Non-Dementia	0.62	0.82	0.70

Table 5. Comparison of the classifiers

Sl. No.	Class	Precision	Recall	F1-Score
1	Gaussian NB	0.93	0.92	0.90
2	SVM	0.64	0.55	0.59
3	LR	0.56	0.59	0.75

Table 6. Comparison of the accuracy for the classifiers

Sl. No.	Classifiers	Accuracy %
1	Gaussian NB	95
2	SVM	89
3	LR	63

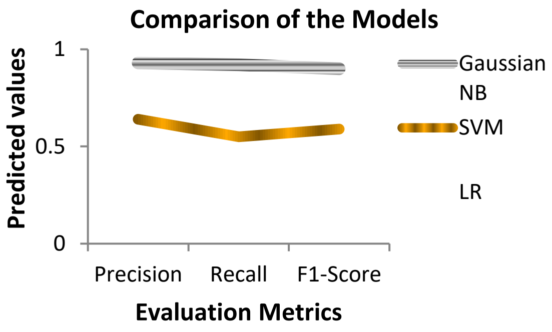

Figure 10. Graphical representation for the comparison of all the models

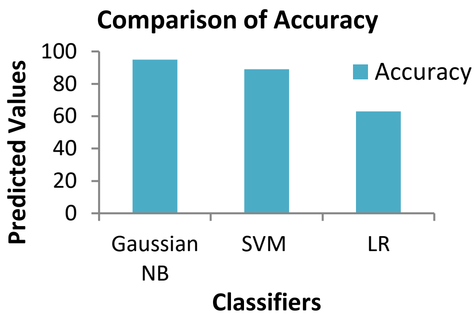

Figure 11. Graphical representation of the accuracy for all the models

3.1. Performance improvement

The performance improvement has been done using the cuckoo algorithm. The LR model has an accuracy of 63%; it is very less when compared to the other two models, so the performance of this model is to be improved. This is done using the mathematical model, along with this cuckoo algorithm is implemented to show better performance.

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n \quad (1)$$

The LR equation is represented by (1), where *Y* is the predicted output, $\beta 0, \beta 1, \dots, \beta n$ are the weighted input and are considered as the co-efficient of the x terms and $x, x_2, \dots, x_n$ are the inputs. The parameter which affects the performance of the machine learning classifiers is considered as the noise and specified as given in (2),

$$Y_n = fX_n + N \quad (2)$$

where fX_n is the function of input and *N* is the noise. In this, the parameters considered as noise are missing values MV and feature encoding FE (3). Using mean imputation and feature transformation both problems are solved here (4).

$$Y_n = f(X_n) + MV + FE \quad (3)$$

$$Y_n = f(X_n) + imputation + FeatureTransformation \quad (4)$$

Also, the cuckoo algorithm is used to remove the noise. The concept behind this algorithm is the cuckoo bird lays the egg by choosing the other bird's nest. Sometimes the other birds identify the egg, and they will drop the egg so that the count of the cuckoo in the next generation will be reduced. If it does not identify, the egg will be hatched and the cuckoo in the next generation will be increased. In this case, there is a chance of an increase or decrease in the cuckoo in the next generation. The same concept is implemented here to reduce the noise. As the birds drop the egg the count of the cuckoo birds will be decreased similarly the parameters which affect the performance of the machine learning models will be dropped and better performance will be given by the models. Figure 12 represents the steps involved in the performance improvement in the machine learning model. The performance improvement has been done in three steps; they are the machine learning model, the cuckoo algorithm implementation, and the performance improvement. Using the concept of the cuckoo algorithm the performance has been increased for the machine learning models by eliminating the parameters which affect the performance of the models.

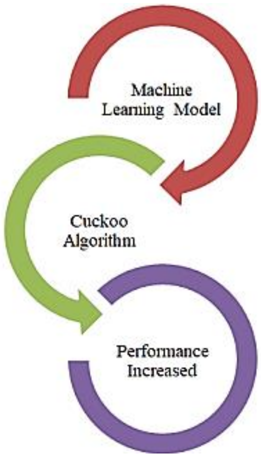

Figure 12. Steps in the performance improvement

4. CONCLUSION

Dementia is a brain disorder disease. This paper focused on feature engineering and the prediction of dementia disease using the machine learning models Gaussian NB, SVM, and LR. The feature engineering focus tasks are feature creation, feature transformation, feature selection, and feature extraction. In feature creation, data creation and data imputation have been done; mean imputation has been applied for generating the missing values. Label encoding has been done for feature transformation. CFS Subset Evaluator along with the best first search method and the info gain attribute evaluator along with the ranker method has been applied for selecting the best attributes. Initially, there are 15 features in the oasis dataset and 12 features are selected as the best attributes for the prediction. The evaluation parameters considered for validating the models are precision, recall, F1-score, and accuracy; for all the metrics the Gaussian NB was given the highest values as given 95%, 97%, 95%, and 95%. So, from these metrics, it is concluded that the Gaussian NB is the best classifier. Hold-out cross-validation has been done for validating the models. The accuracy of the algorithms has been improved by eliminating the parameters which affect the performance of the machine learning models using the cuckoo algorithm. In the future, the plan is to predict dementia using other machine learning models.

REFERENCES

[1] T. Khatibi and N. Rabinezhadsadatmahaleh, "Proposing feature engineering method based on deep learning and K-NNs for ECG beat classification and arrhythmia detection," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 1, pp. 49–68, Mar. 2020, doi: 10.1007/s13246-019-00814-w.
[2] M. Aamir and S. M. A. Zaidi, "DDoS attack detection with feature engineering and machine learning: the framework and performance evaluation," *International Journal of Information Security*, vol. 18, no. 6, pp. 761–785, Dec. 2019, doi: 10.1007/s10207-019-00434-1.
[3] J. Dhar, "Multistage ensemble learning model with weighted voting and genetic algorithm optimization strategy for detecting chronic obstructive pulmonary disease," *IEEE Access*, vol. 9, pp. 48640–48657, 2021, doi: 10.1109/ACCESS.2021.3067949.
[4] P. G. Shynu, V. G. Menon, R. L. Kumar, S. Kadry, and Y. Nam, "Blockchain-based secure healthcare application for diabetic-cardio disease prediction in fog computing," *IEEE Access*, vol. 9, pp. 45706–45720, 2021, doi: 10.1109/ACCESS.2021.3065440.
[5] D. Chicco and G. Jurman, "An ensemble learning approach for enhanced classification of patients with hepatitis and cirrhosis," *IEEE Access*, vol. 9, pp. 24485–24498, 2021, doi: 10.1109/ACCESS.2021.3057196.
[6] A. U. Haq *et al.*, "Detection of breast cancer through clinical data using supervised and unsupervised feature selection techniques," *IEEE Access*, vol. 9, pp. 22090–22105, 2021, doi: 10.1109/ACCESS.2021.3055806.
[7] V. K. Gupta, A. Gupta, D. Kumar, and A. Sardana, "Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model," *Big Data Mining and Analytics*, vol. 4, no. 2, pp. 116–123, Jun. 2021, doi: 10.26599/BDMA.2020.9020016.
[8] A. H. Syed, T. Khan, A. Hassan, N. A. Alromema, M. Binsawad, and A. O. Alsayed, "An ensemble-learning based application to predict the earlier stages of Alzheimer's disease (AD)," *IEEE Access*, vol. 8, pp. 222126–222143, 2020, doi: 10.1109/ACCESS.2020.3043715.
[9] C.-H. Lin, S.-I. Chiu, T.-F. Chen, J.-S. R. Jang, and M.-J. Chiu, "Classifications of neurodegenerative disorders using a multiplex blood biomarkers-based machine learning model," *International Journal of Molecular Sciences*, vol. 21, no. 18, Sep. 2020, doi: 10.3390/ijms21186914.

[10] J. Zhou and O. G. Troyanskaya, “Predicting effects of noncoding variants with deep learning–based sequence model,” *Nature Methods*, vol. 12, no. 10, pp. 931–934, Oct. 2015, doi: 10.1038/nmeth.3547.
[11] N. Takahashi, N. Goswami, and Y. Mitsufuji, “Mmdenselstm: An efficient combination of convolutional and recurrent neural networks for audio source separation,” in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep. 2018, pp. 106–110. doi: 10.1109/IWAENC.2018.8521383.
[12] S. Shataee, S. Kalbi, A. Fallah, and D. Pelz, “Forest attribute imputation using machine-learning methods and ASTER data: comparison of kNN, SVR and random forest regression algorithms,” *International Journal of Remote Sensing*, vol. 33, no. 19, pp. 6254–6280, Oct. 2012, doi: 10.1080/01431161.2012.682661.
[13] R. Sivakani and G. A. Ansari, “Machine learning framework for implementing Alzheimer’s disease,” in *2020 International Conference on Communication and Signal Processing (ICCSP)*, Jul. 2020, pp. 0588–0592. doi: 10.1109/ICCSP48568.2020.9182220.
[14] S. Rowson and S. M. Duma, “Brain injury prediction: Assessing the combined probability of concussion using linear and rotational head acceleration,” *Annals of Biomedical Engineering*, vol. 41, no. 5, pp. 873–882, May 2013, doi: 10.1007/s10439-012-0731-0.
[15] E. H. Houssein, M. E. Hosney, M. Elhoseny, D. Oliva, W. M. Mohamed, and M. Hassaballah, “Hybrid Harris hawks optimization with cuckoo search for drug design and discovery in chemoinformatics,” *Scientific Reports*, vol. 10, no. 1, Sep. 2020, doi: 10.1038/s41598-020-71502-z.
[16] L. Liu, X. Liu, N. Wang, and P. Zou, “Modified cuckoo search algorithm with variational parameters and logistic map,” *Algorithms*, vol. 11, no. 3, Mar. 2018, doi: 10.3390/a11030030.
[17] M. Mareli and B. Twala, “An adaptive cuckoo search algorithm for optimisation,” *Applied Computing and Informatics*, vol. 14, no. 2, pp. 107–115, Jul. 2018, doi: 10.1016/j.aci.2017.09.001.
[18] A. S. Joshi, O. Kulkarni, G. M. Kakandikar, and V. M. Nandedkar, “Cuckoo search optimization- a review,” *Materials Today: Proceedings*, vol. 4, no. 8, pp. 7262–7269, 2017, doi: 10.1016/j.matpr.2017.07.055.
[19] M. Phogat and D. Kumar, “Classification of complex diseases using an improved binary cuckoo search and conditional mutual information maximization,” *Computación y Sistemas*, vol. 24, no. 3, Sep. 2020, doi: 10.13053/cys-24-3-3354.
[20] S. Murugan, G. Babu T R, and S. C, “Underwater object recognition using KNN classifier,” *International Journal of MC Square Scientific Research*, vol. 9, no. 3, Dec. 2017, doi: 10.20894/IJMSR.117.009.003.007.
[21] M. M. Ismail, M. Subbiah, and S. Chelliah, “Design of pipelined radix-2, 4 and 8 based multipath delay commutator (MDC) FFt,” *Indian Journal of Public Health Research & Development*, vol. 9, no. 3, 2018, doi: 10.5958/0976-5506.2018.00380.7.
[22] A. Unnikrishnan and V. Das, “Cooperative routing for improving the lifetime of wireless ad-hoc networks,” *International Journal of Advances in Signal and Image Sciences*, vol. 8, no. 1, pp. 17–24, Jan. 2022, doi: 10.29284/IJASIS.8.1.2022.17-24.
[23] T. Bock, “What is feature engineering?,” *Displayr*. https //www.displayr.com/what-is-feature-engineering (accessed January 18, 2022)
[24] Y. Xiong and Y. Lu, “Deep feature extraction from the vocal vectors using sparse autoencoders for Parkinson’s classification,” *IEEE Access*, vol. 8, pp. 27821–27830, 2020, doi: 10.1109/ACCESS.2020.2968177.
[25] D. C. Yadav and S. Pal, “Prediction of heart disease using feature selection and Random forest ensemble method,” *International Journal of Pharmaceutical Research*, vol. 12, no. 04, pp. 56–66, Jun. 2020, doi: 10.31838/ijpr/2020.12.04.013.
[26] H. A. Helaly, M. Badawy, and A. Y. Haikal, “Deep learning approach for early detection of Alzheimer’s disease,” *Cognitive Computation*, vol. 14, no. 5, pp. 1711–1727, Sep. 2022, doi: 10.1007/s12559-021-09946-2.
[27] M. Akhtaruzzaman, M. K. Hasan, S. R. Kabir, S. N. H. S. Abdullah, M. J. Sadeq, and E. Hossain, “HSIC bottleneck based distributed deep learning model for load forecasting in smart grid with a comprehensive survey,” *IEEE Access*, vol. 8, pp. 222977–223008, 2020, doi: 10.1109/ACCESS.2020.3040083.
[28] R. Mazloumi, S. R. Abazari, F. Nafarieh, A. Aghsami, and F. Jolai, “Statistical analysis of blood characteristics of COVID-19 patients and their survival or death prediction using machine learning algorithms,” *Neural Computing and Applications*, vol. 34, no. 17, pp. 14729–14743, Sep. 2022, doi: 10.1007/s00521-022-07325-y.
[29] K. Kalegowda, A. D. Iyengar Srinivasan, and N. Chinnamadha, “Particle swarm optimization and Taguchi algorithm-based power system stabilizer-effect of light loading condition,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 5, pp. 4672–4679, Oct. 2022, doi: 10.11591/ijece.v12i5.pp4672-4679.
[30] B. S. Rameshappa and N. M. Shadaksharappa, “An optimal artificial neural network controller for load frequency control of a four-area interconnected power system,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 5, pp. 4700–4711, Oct. 2022, doi: 10.11591/ijece.v12i5.pp4700-4711.
[31] G. Castellazzi *et al.*, “A machine learning approach for the differential diagnosis of Alzheimer and vascular dementia Fed by MRI selected features,” *Frontiers in Neuroinformatics*, vol. 14, no. 25, Jun. 2020, doi: 10.3389/fninf.2020.00025.
[32] H. Wang *et al.*, “Develop a diagnostic tool for dementia using machine learning and non-imaging features,” *Frontiers in Aging Neuroscience*, vol. 14, Aug. 2022, doi: 10.3389/fnagi.2022.945274.
[33] K. Sivanantham, “Deep learning-based convolutional neural network with cuckoo search optimization for MRI brain tumour segmentation,” in *Computational Intelligence Techniques for Green Smart Cities*, 2022, pp. 149–168. doi: 10.1007/978-3-030-96429-0_7.
[34] M. Sudharsan and G. Thailambal, “A hybrid learning approach for early-stage prediction and classification of Alzheimer’s disease using multi-features,” in *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, Apr. 2022, pp. 1373–1380. doi: 10.1109/ICOEI53556.2022.9776843.
[35] C. Kavitha, V. Mani, S. R. Srividhya, O. I. Khalaf, and C. A. Tavera Romero, “Early-stage Alzheimer’s disease prediction using machine learning models,” *Frontiers in Public Health*, vol. 10, Mar. 2022, doi: 10.3389/fpubh.2022.853294.
[36] J. Sheng, Y. Xin, Q. Zhang, L. Wang, Z. Yang, and J. Yin, “Predictive classification of Alzheimer’s disease using brain imaging and genetic data,” *Scientific Reports*, vol. 12, no. 1, Feb. 2022, doi: 10.1038/s41598-022-06444-9.
[37] K. Balasubramanian, N. Ananthamoorthy, and K. Ramya, “Prediction of neuro-degenerative disorders using sunflower optimisation algorithm and Kernel extreme learning machine: A case-study with Parkinson’s and Alzheimer’s disease,” *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 236, no. 3, pp. 438–453, Mar. 2022, doi: 10.1177/09544119211060989.
[38] F. Zhang, M. Petersen, L. Johnson, J. Hall, and S. E. O’Bryant, “Accelerating hyperparameter tuning in machine learning for Alzheimer’s disease with high performance computing,” *Frontiers in Artificial Intelligence*, vol. 4, Dec. 2021, doi: 10.3389/frai.2021.798962.
[39] Q. Pan, K. Ding, and D. Chen, “Multi-classification prediction of Alzheimer’s disease based on fusing multi-modal features,” in *2021 IEEE International Conference on Data Mining (ICDM)*, Dec. 2021, pp. 1270–1275. doi: 10.1109/ICDM51629.2021.00156.

BIOGRAPHIES OF AUTHORS

Sivakani Rajayyan received her B.Tech. degree from Anna university in Information Technology. She also received her M.E. degree from Sathyabama University in computer science and engineering. She is currently pursuing her Ph.D. degree in B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, Tamilnadu, India. She can be contacted at sivakani13@gmail.com.

Syed Masood Mohamed Mustafa is an associate professor at B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, Tamilnadu, India. He received his M.E., from Anna University in Computer Science and Engineering. He received his Ph.D. degree from B.S. Abdur Rahman Crescent Institute of Science. He has more than 25 years of experience in teaching. He has contributed more than 25 research articles in reputed journals and conferences. His research areas are computer networks, software engineering, information system, management information system, decision support system, cyber security, artificial intelligence, machine learning, data mining, and data security. He can be contacted at ms.masood@crescent.education.