Journal of King Saud University – Computer and Information Sciences 34 (2022) 4993–5006

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Feature selection using efficient fusion of Fisher Score and greedy searching for Alzheimer's classification

Muhammed Niyas K.P., Thiyagarajan P. *

Department of Computer Science, Central University of Tamil Nadu, Thiruvarur 610005, India

ARTICLE INFO

Article history:
Received 21 September 2020
Revised 16 December 2020
Accepted 16 December 2020
Available online 1 January 2021

Keywords:
Fisher Score
Mild Cognitive Impairment
Alzheimer's
Feature selection

ABSTRACT

Alzheimer's disease is one of the deadly progressive neurodegenerative diseases among aged populations. But, the progression of the disease can be reduced by proper treatment of the disease during the early stages of cognitive impairment. The main objective of this study is to implement an efficient feature selection algorithm for the detection of Alzheimer's patients at the baseline stage itself using multimodal data. In this paper, we propose an efficient fusion of Fisher Score ranking and greedy searching heuristic as feature selection criteria for Alzheimer's prediction. The proposed algorithm provides a Balanced Classification Accuracy of 90% and 91% and Multi Area Under the Curve of 0.97, 0.98 using Support Vector Machine, K-Nearest Neighbor respectively for classifying Normal Controls, Mild Cognitive Impairment, and Alzheimer's patients on Alzheimer's Disease Neuroimaging Initiative-TADPOLE dataset at baseline visit itself. Moreover, the proposed algorithm also provides better sensitivity, specificity of 84%, 82.5% using Support Vector Machine, K-Nearest Neighbor for binary classification of Mild Cognitive Impairment, and Alzheimer's patients on the Australian Imaging and Biomarker Lifestyle dataset also. Our results indicate that the proposed methodology with efficient feature selection is promising and can outperform the state of the art methods for early detection of Alzheimer's.

1. Introduction

Alzheimer's is considered to be the most deadly neurodegenerative disease among aged populations (Alzheimer's Association, 2019; Ferrari et al., 2018; Eldholm et al., 2018). The prevalence of the disease is high among aged people in both developing and developed countries around the world (Eldholm et al., 2018; Matthews et al., 2019; Pan and Nicolazzo, 2018; Winblad et al., 2016; Tsoy et al., 2019). According to the World Dementia Report 2019, it has estimated that over 50 million people around the world are suffering from dementia, which is like the population size of South Korea or Spain (Wortmann, 2012; Lorenz et al., 2019). Moreover, it is estimated that the worldwide prevalence of Alzheimer's is expected to triple 35.5 million by 2050 (Langa, 2015).

The exact parameters responsible for the occurence of Alzheimer's is not yet known. However, researchers experimenting widely on finding out the reasons and causes of Alzheimer's are unable to generalize the parameters that are responsible for Alzheimer's Disease (AD) (Petersen et al., 1999). However, there are many local and global parameters of brain which are responsible for the cognitive functionalities as well as Cerebro Spinal Fluids (CSF) of the individual can be used as bio markers (Zimmermann et al., 2018). Such characteristics can be found out using the advanced brain image acquisition techniques like Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Functional Magnetic Resonance Imaging (fMRI). etc. Many neuropsychological assesement tests like Mini Mental State Examinations (MMSE), Clinical Dementia Rating (CDR), Alzheimer's Disease Assesement Scale Cognitive (ADAS-Cog) are also used to identify AD patients (Kurlowicz and Wallace, 1999; Salmon and Bondi, 2009). It is possible to track out the minute variations in local parts of the brain structure using these techniques. However, a doctor will face difficulty in predicting AD patients because these biomarker variations are difficult to generalize (Zimmermann et al., 2018). In such cases, a Machine Learning (ML) model trained using various multi modal features can be useful in predicting AD (Zhang et al., 2012; Zhang and Shen, 2011).

* Corresponding author.
E-mail address: thiyagu.phd@gmail.com (Thiyagarajan P.).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2020.12.009
1319-1578/

Normally, a patient visits a doctor when he or she is suspicious of their cognitive abilities. All the patients visiting the doctor may not be converted to Alzheimer's in the future. In such cases, the task of a doctor is to predict the AD patients of the future at the baseline stage itself. Such predictions can be pretty useful in reducing unnecessary health care costs. Moreover, the rate of progression of AD can be effectively reduced during the early stages of detection. In other words, if the doctor can find out the AD patients during the early stages of their disease. Then, doctors can design an effective medication strategy for them (Kira and Rendell, 1992). It is also to be noted that the treatments are effective during the early stages of predictions (somewhere during the first visit of the patient itself) (Kira and Rendell, 1992).

As the exact parameters responsible for the AD patients are unknown, a feature selection based approach for the ML algorithm can be useful (Kira and Rendell, 1992). In this paper, a novel feature selection based on the fusion of Fisher Score (FS) ranking criteria and greedy searching strategy for classifying the patients into three categories namely Mild Cognitive Impairment (MCI), AD and Normal Controls (NC) at baseline visit is performed on both Alzheimer's Disease Neuroimaging Initiative-TADPOLE (ADNI-TADPOLE) and Australian Imaging and Biomarker Lifecycle (AIBL) datasets. For training purposes, the cross-sectional baseline data of the patients are being utilized. This paper is organized as follows: Section 1 contains the background, Section 2 contains the materials, Section 3 contains the methodology, Section 4 deals with the experimental results and discussion, Section 5 deals with limitations and future works and Section 6 deals with conclusions respectively.

2. Related works

This section contains detailed information about the related works done in the classification of AD patients. MRI, PET image features with a Support Vector Machine (SVM) are used for the classification of MCI and AD patients (Hao et al., 2020). Further, MRI and PET are also combined with a Random Forest (RF) feature selection algorithm and Gaussian classifier, reported with an accuracy of 78% in classifying MCI and NC patients (Forouzannezhad et al., 2020). The researchers also conducted the studies on multimodal data consists of MRI, cognitive tests, demographics of Open Access Series Imaging Studies (OASIS) for the prediction of dementia patients (Khan and Zubair, 2020; Battineni et al., 2019). A fusion of RF classifier with a correlation-based feature selection method is used for the classification of demented, non-demented, and non-demented to demented converters (Khan and Zubair, 2020). An SVM classifier is used for classifying demented, non-demented, and non-demented to demented converters with multimodal data (Battineni et al., 2019). A transfer learning-based deep learning model was used for the classification of AD, MCI, and healthy patients using MRI data (Farooq et al., 2017). Researchers also proposed an MRI deformation quantification model using SVM for the classification of AD and MCI (Long et al., 2017). Further, multimodal data consists of an MRI. PET and CSF are used for the classification of MCI and non-MCI using SVM (Zhang et al., 2011).

Moreover, the ADNI-TADPOLE challenge dataset is widely used by the researchers for experimenting their methods (Marinescu et al., 2018). There are mainly two types of data cleaning approaches observed from the previous works on the ADNI-TADPOLE challenge set: 1. Data Imputation by substituting values instead of missing values, 2. Deleting the samples with missing values. Researchers experimented by performing data imputation on the missing values on longitudinal data collected at various time points (Ghazi et al., 2019; Nguyen et al., 2018; Iddi et al., 2019; Vivar et al., 2019). Researchers performed imputation and training on the ADNI-TADPOLE MRI biomarker data using Long Short Term Memory Recurrent Neural Networks (LSTM RNN) (Ghazi et al., 2019). Researchers also again proposed a Recurrent Neural Network (RNN) forward filling algorithm for imputation and training on multi-modal data from ADNI-TADPOLE (Nguyen et al., 2018). A multi-graph based RNN and Convolutional Neural Network (CNN) is proposed by the researchers (Valenchon and Coates, 2019) for imputation and used SVM classifier. A two stage RF classification using cognitive, MRI, CSF features where the missing data is imputed by the mean of the corresponding features are also proposed by the researchers (Iddi et al., 2019). Further, researchers also proposed a multi graph based approach for missing data imputation on ADNI-TADPOLE classifier data with SVM classifier (Vivar et al., 2019). In short, the studies are widely conducted on various types of multimodalities of features. But, researchers face a challenging task in finding the relevant features. Thus, an efficient feature selection algorithm is important in identifying the important features. Hence, our proposed methodology is focused on feature selection algorithm to overcome this problem.

2.1. Feature selection techniques: An analysis on FS

Feature selection techniques are relevant to the creation of predictive models. Many feature ranking methods and metrics are used by the researchers for selecting the relevant features (Molina et al., 2002). Feature selection is very important for Alzheimer's prediction using multi-modal data also. The relevance or combination of what type of features are the most important in distinguishing AD from other dementia is not generalized yet (Petersen et al., 1999).

Researchers used FS criterion as a feature selection algorithm for the prediction of Alzheimer's (Beheshti et al., 2016; Zhou, 2016; Song et al., 2017). The proposed feature selection algorithm in Beheshti et al. (2016) finds the optimal set of features using FS and t-test scores. FS criterion ranking is similarly used in other datasets for many applications like detection of Hepatitis (Zhou, 2016), Breast Cancer (Zhou, 2016), Isochemic Stroke (Zhou, 2016). A hybrid of the Genetic Algorithm and FS is used by the researchers in Zhou (2016) for the detection of Hepatitis and Breast Cancer. Further, Isochemic Stroke detection is performed by a hybrid of FS and Mutual Information in Zhou (2016). Similarly, a hybrid of Fisher Discriminant Analysis (FDA) and FS criterion is used for multiclass classification by the researchers 9Song et al., 2017). The shortcomings of feature selection algorithms using FS is illustrated in Table 1.

However, the individual FS is a better feature ranking method despite the shortcomings as described earlier. But, finding the best combination of features using the FS is a challenging task in feature selection. In this paper, we propose a fusion-based algorithm

Table 1
Shortcomings of similar FS algorithms.

Reference	Feature Selection Algorithm	Limitation
Beheshti et al. (2016)	Hybrid (t-test and FS)	Finding threshold for optimal set of features based on ranking
Zhou (2016)	Hybrid (Genetic Algorithm + FS)	Computational time high
	Hybrid (FS + Mutual Information)	Computational time high
Song et al. (2017)	FDA + FS	Finding threshold for optimal set of features based on ranking Finding best combination of features within threshold is time consuming

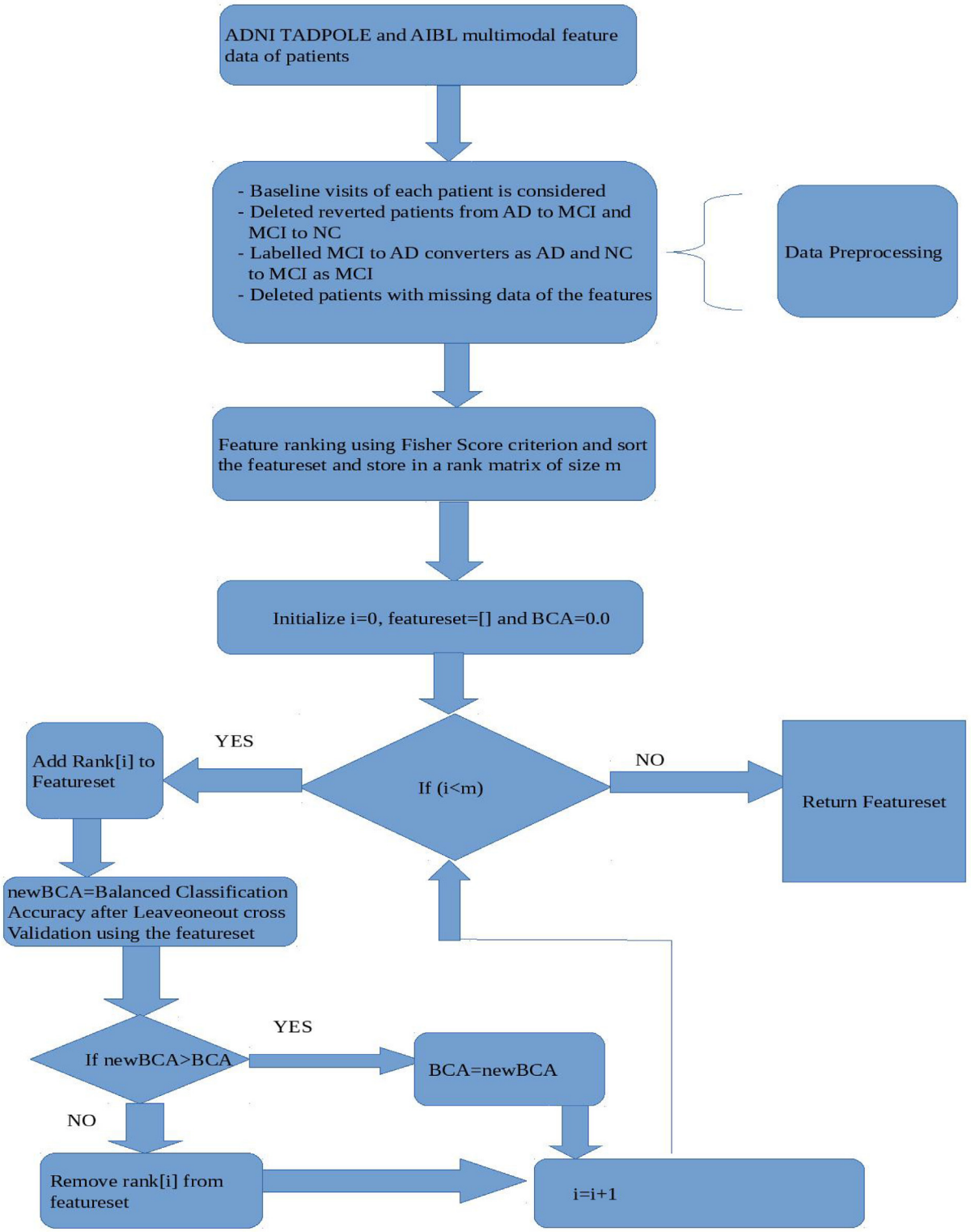


Fig. 1. The workflow of the proposed feature selection algorithm.

where the feature selection is performed using the FS and a greedy search to overcome the challenge. Initially, the features are ranked on the basis of FS and then features are added and eliminated on the basis of their performance in a greedy manner. The greedy searching heuristic in our methodology finds a better combination of features and thereby resolves the drawback of FS in considering feature combinations. The basic idea is to choose a sub-optimal set of features by searching each features in the ranking order of FS only once. Such an approach always give a sub-optimal minimal feature set because the greedy algorithm start search from the top ranked FS feature and thereafter, adds or removes features based on their performance. Hence, our algorithm solve the incapability of FS while handling multiple features through a better search heuristic. Also, the proposed feature selection algorithm

achieves a less time complexity without reducing the features also. Moreover, the approach considers all the features rather than limiting the search to some features within a threshold.

3. Proposed methodology

The main objectives of the proposed algorithm is:

- To find out a sub-optimal minimal combination of features that can maximize the performance of dimension based classifiers such as SVM and KNN.
- Designing a feature selection algorithm that is time efficient.

The description of the methodology is given below:

1. Choosing features.
2. Pre-processing of the data.
3. Ranking of features using FS.
4. Sub-optimal minimal feature subset selection using the proposed fusion of FS and greedy searching algorithm.
5. The performance evaluation of the resultant feature set which are returned by the proposed feature selection algorithm. Both Leave One Out Cross Validation (LOOCV) and stratified 10-fold cross-validation strategies are used separately for estimating the performances of the resultant feature set. The metrics such as BCA, precision, sensitivity, specificity, F1-score are used for evaluating the binary classifications. Along with all the metrics used for evaluating the performance for binary classifications, Multi Area Under Curve (MAUC) is also used for three-way classifications (AD v/s MCI v/s NC).

Fig. 1 contains the workflow of proposed feature selection algorithm.

3.1. Datasets

A description about the dataset and the features used for predicting the diagnosis is given in this section.

3.1.1. Description of dataset

The experiments are performed on ADNI-TADPOLE and AIBL dataset. The description of the datasets is as follows:

ADNI-TADPOLE. ADNI-TADPOLE dataset is published by the Alzheimer's Disease Neuroimaging Initiative (ADNI) as a competition challenge dataset for Alzheimer's. This is one of the standard freely avaialble datsets for researchers (Marinescu et al., 2018). The dataset is comprised of multimodal data of various 1737 patients collected over multiple time points. The multi-modal data includes MRI, PET, CSF and cognitive tests. The main aim of the ADNI-TADPOLE challenge dataset is to predict AD risk individuals at an early stage. The dataset is a collection of MRI, PET Region of Interests (ROI) extracted features, CSF, Genetic, and neuropsychological cognitive assessment data of individuals performed over various periods during ADNI projects such as ADNI, ADNI-1 (Marinescu et al., 2018). The participant's of the ADNI projects have given their full consent for sharing their data to the public (Marinescu et al., 2018). Hence, the dataset where the ethical concerns are already satisfied are chosen for our study. The dataset is a comprehensive collection of biomarkers from various modalities that are promising in distinguishing AD from MCI and NC (McKhann et al., 2011).

AIBL. AIBL dataset is published in November 14, 2016, for identifying which biomarkers, cognitive characteristics, lifestyle factors help in identifying AD from MCI patients. Similar to ADNI-TADPOLE, AIBL also one of the standard freely avaialble datasets for researchers (Ellis et al., 2009). The participant's of the AIBL projects have also given their full consent for sharing their data to the public (Ellis et al., 2009). Hence, the dataset is selected for our study. The cohort for the study is selected from Australia. There over 1000 participants data consists of various factors such as lab data, medical history data, lifestyle data, cognitive data is collected over various time intervals. The dataset consists of 4.5 years longitudinal cohort study examined in 6 month intervals (Ellis et al., 2009).

3.2. Features description

The selection of the features are done after reviewing the previous related works. Two datasets: ADNI-TADPOLE and AIBL are considered for the study. The following features are selected from the ADNI-TADPOLE and AIBL datasets.

3.2.1. ADNI-TADPOLE

The experiments are performed on ADNI-TADPOLE D1, D2 merged file dataset. We have chosen 21 features from this dataset. The following features from various multi modalities are chosen for the experiments. The Table 2 contains a detailed information about the features and their explanation.

- MRI: Hippocampus, Ventricles, Wholebrain, Entorhinal, Fusiform, Middle Temporal Gyrus, Intra Cranial Volumes as obtained from Freesurfer software.
- PET: Fluorodeoxyglucose (FDG), AV-45 Florbetpir measurements captured by PET image acquisition technique.
- Genetic Factors: Apolipoprotein E4 (APOE4) which is a genetic factor for AD.
- Cognitive Tests: Clinical Dementia Rating Scale Box (CDRSB), Mini Mental State Examination (MMSE), Rey Audital Verbal Learning Test (RAVLT), Alzheimer's Disease Cognitive Assesement Scale (ADAS), Functional Activities Questionnaire (FAQ).
- Education: Education level of the patient.
- Age: Age of the patient at baseline (in years).

The Table 2 contains a detailed information about the features of ADNI-TADPOLE and their explanation.

3.2.2. AIBL

18 multi-modal features are chosen from AIBL dataset. The following features from various multi modalities are chosen for the experiments using AIBL.

- Cognitive Tests: Mini Mental State Examination (MMSCORE),[1] Clinical Dementia Rating Global (CDGLOBAL),[2] Total number of story units recalled -Logical Memory Immediate Recall (LIMMTOTAL), Total number of story units recalled-Partial Score of LM test (LDELTOTAL).
- Lab Data: Thyroid Stimulate Hormone, Vitamin 12, Red Blood Cell Count, White Blood Cell Count, Platelets, Hemoglobin, Mean Cell Hemoglobin, Mean Cell Hemoglobin Concentrate, Urea Nitrogen, Serum Glucose, Cholestrol and Creatinine.

The Table 3 contains a detailed information about the features of AIBL and their explanation.

It is observed that out of the features used for the study from both the datasets, there are only two common features namely CDRSB and MMSE. Consequently, a study by merging the two datasets are impossible. Hence, the experiments are conducted sepa-

[1] Same as the MMSE, cognitive test in the ADNI-TADPOLE.
[2] Same as the CDRSB, cognitive test in the ADNI-TADPOLE.

Table 2
Description about the features in ADNI-TADPOLE as explained in https://github.com/swhustla/pycon2017-alzheimers-hack/blob/master/docs/tadpole_data_dictionary.csv AC-Anterior Cingulate, PC-Parietal Cortex.

Feature	Description
Ventricles	Ventricles Volume
Hippocampus	Hippocampus Volume
WholeBrain	WholeBrain Volume
Entorhinal	Entorhinal Volume
Fusiform	Fusiform Volume
MidTemp	Med Temp Volume
ICV	Intracranial Volume
CDRSB	Clinical Dementia Rating Scale Box
MMSE	Mini-Mental State Examination
RAVLT_learning	Rey Auditory Verbal Learning Test
RAVLT_immediate	Rey Auditory Verbal Learning Test (5 sum)
RAVLT_Forgetting	Rey Auditory Verbal Learning Test Forgetting
RAVLT_Perc_Forgetting_bl	Rey Auditory Verbal Learning Test Percentile Forgetting
ADAS11	Alzheimer's Disease Assesement Cognition Scale 11
ADAS13	Alzheimer's Disease Assesement Cognition Scale 13
FAQ	Functional Activities Questionnaire
FDG	Average FDG-PET of angular, temporal, and posterior cingulate
AV45	Average AV45 SUVR of frontal, AC, precuneus, and parietal cortex relative to the cerebellum
APOE4	Apolipoprotein E4
Age	Age at baseline
PTEDUCAT	Education
DX	Diagnosis Status

Table 3
Description about the features in AIBL as explained in http://adni.loni.usc.edu/data-dictionary-search/.

Feature	Description
AXT117	Thyroid Stimulate Hormone
BAT126	Vitamin 12
HMT3	Red Blood Cell Count
HMT7	White Blood Cell Count
HMT13	Platelets
HMT40	Hemoglobin
HMT100	Mean Cell Hemoglobin
HMT102	Mean Cell Hemoglobin Concentrate
RCT6	Urea Nitrogen
RCT11	Serum Glucose
RCT20	Cholestrol
RCT392	Creatinine
MMSCORE	Mini Mental State Examination Score
LIMMTOTAL	Total number of story units recalled-Logical Memory Immediate Recall
LDELTOTAL	Total number of story units recalled-Partial Score of Logical Memory test
CDGLOBAL	Clinical Dementia Rating Global
DXCURREN	Diagnosis status

rately on the two datasets. By doing so, the main objective is to find: Are there any common distinguishing features in the resulting minimal feature set after executing the proposed feature selection algorithm?.

3.3. Pre-processing

In both datasets, the cross-sectional data of the patients at the baseline stage is only considered for the experiments. This step is performed to find out the effectiveness of using baseline training data for predicting AD at the baseline stage itself. As far as a doctor is concerned, the predictions using baseline data alone help in many ways. A doctor need not wait for further follow-up data for making their predictions. Thus, a doctor can predict AD within fewer data in a short period. The patients who are reverted from AD to MCI and MCI to NC are eliminated from the study. Further, the patients converted from MCI to AD, NC to MCI are considered as AD and MCI respectively. In the final step, the patients whose any of the selected feature data is missing are eliminated from the study. There is no data imputation performed on the missing data.

After the above preprocessing steps, the number of CN, MCI and AD patients obtained on ADNI-TADPOLE are 231, 358, and 94 respectively. The number of CN, MCI and AD on AIBL are 609, 143, and 105 respectively.

The demographic description of the ADNI-TADPOLE data is given in Table 4.

The demographic description of the AIBL data is given in Table 5.

3.3.1. Proposed fusion of FS and greedy searching feature selection algorithm

The algorithm initially ranks the features based on FS and then iteratively adds the ranked features one by one based on their performance. The detailed explanation of the algorithm is given below:

FS. Initially, the model ranks features based on the FS criterion. FS criterion is expected to find out features that maximize Between Class Distance (BCD) and minimizes Within Class Distance (WCD). Those features which have high FS can improve the performance of classifiers like SVM and KNN (Song et al., 2017). The FS of a feature is based on the ratio of scattering between class distances to the within-class sample distance (Gu et al., 2012). The BCD and WCD of a feature 'i' is given by the Eqs. (1) and (2) respectively. The overall FS is calculated by Eq. (3).

$$BCD_i = \sum_{i=1}^{L} (Overallmean - Meanlabel_i)^2 \quad (1)$$

$$WCD_i = \sum_{i=1}^{L} variancelabel_i \quad (2)$$

Table 4
Summary of demographic statistics of ADNI-TADPOLE dataset with mean, minimum, maximum and standard deviation.

	Statistic	NC	MCI	AD
Age (in years)	Mean	72.78	71.08	73.19
	Maximum	89.0	90.3	88.0
	Minimum	59.0	55.6	55.0
	Standard Deviation	5.807	8.230	7.239
MMSE	Mean	29.04	28.11	23.06
	Maximum	30.0	26.0	30.0
	Minimum	24.0	19.0	23.0
	Standard Deviation	1.2642	2.041	1.67

Table 5
Summary of demographic statistics of AIBL dataset with mean, minimum, maximum and standard deviation. NR-Not Reported.

	Statistic	NC	MCI	AD
Age (in years)	Mean	NR	NR	NR
	Maximum	NR	NR	NR
	Minimum	NR	NR	NR
	Standard Deviation	NR	NR	NR
MMSE	Mean	28.70	26.944	20.28
	Maximum	30.0	26.0	30.0
	Minimum	24.0	19.0	23.0
	Standard Deviation	1.2642	2.041	1.67

$$FS_i = BCD_i/WCD_i \quad (3)$$

where Overallmean is the overallmean of the feature, meanlabel is the mean of the feature corresponding to the individual labels, variancelabel is the variance of the feature corresponding to the individual labels and L is the number of classes. In short, the FS metric depends on the mean and variance of a feature corresponding to a label. That is why, our proposed algorithm that contains FS also helps in supervised learning. The proposed model as mentioned in Algorithm 1 is implemented on the FS ranked features.

The features are ranked on the increasing order of FS (the highest FS feature as rank1, the second-most highest FS score feature as rank2, and so on). The fusion of FS and greedy searching feature selection algorithm selects a sub-optimal minimal feature set after traversing through each feature in their rank wise order only once. Initially, only the highest-ranked feature is added to the required minimal feature subset (current feature set) and the performance is evaluated also. Then, the next ranked feature is added to the current feature set and the performance is evaluated for the new feature set. The newly added feature is retained in the current feature set if the performance of the classifier is increased. Otherwise, it is discarded. The performance of the classifiers is evaluated using the LOOCV on the selected cohort's data. BCA is the metric used for assessing the performance of the classifiers. Likewise, the performance of each feature is assessed by the algorithm. The features are retained or discarded based on their performances. Thus, the proposed fusion feature selection algorithm finds a sub-optimal minimal feature set by traversing through each feature only once. The pseudo-code of the proposed algorithm is given in Algorithm 1:

Algorithm 1. Fusion of FS and Greedy Searching Feature Selection Algorithm.

Input: *The set of all multi-modal features*
Output: *The minimal set of features for classifying AD*
Method:

- Find the FS for all Features
- Sort the features in decreasing order of FS (rank the top-most FS feature as 1 and so on).

```
Initialize rank = Features arranged on basis of ranks
Initialize BCA = 0.0
Initialize Featureset=[]
For i = ranks[1] to ranks[last], do
Featureset.append(ranks[i]) ie, Adding the feature to the featureset
Perform LOOCV on the dataset using featureset and evaluate the BCA
if (BCA is not increased), do item Featureset.remove (ranks[i]) ie, removing the feature from the set
Return the resultant feature set
```

Classifiers. The experiments are performed on two classifiers namely SVM and KNN for the study. Both SVM and KNN are one of the most popular supervised ML algorithms used in the medical field (Kaucha et al., 2017; Verma et al., 2017; Xing and Bei, 2019). Moreover, the experiments are not conducted on a large feature set also. Therefore, the choice of selection of SVM and KNN is suitable for our classification study. Classifiers involving neural networks are required if there are large feature sets and requirement for training large data that are computationally expensive. Hence, the classifiers involving neural networks are not considered for our study. As the proposed feature selection algorithm finds out a minimal feature subset, the typical ML classifiers such as SVM and KNN are required for the study. Moreover, they are computationally inexpensive also.

The built-in functions of Sci-kit learn package containing SVM and KNN in python are used for classifications. Our study used the Radial Basis Function (RBF) kernel of SVM and 1-Nearest Neighbor of Sci-kit learn package for our experiment. The regularization parameter, c = 10 and gamma = 0.02 are chosen as the parameters for the RBF kernel in the SVM. The choice of selection for regularization parameter and gamma value of the RBF are made after performing a grid search for various combinations of c = 10, 20, 30, 40, 50 and gamma = 0.01,0.02,0.03,0.04,0.05,0.06,0.07,0.08,0.09,0.10.

The following metrics are used to evaluate the performance of the algorithm:

- BCA:
  BCA is a metric used for measuring the accuracy of imbalanced datasets. BCA of the total system is the sum of accuracies of individual classes. The BCA of class 'i' is given in Eq. (4) and the overall BCA is given in Eq. (5):

$$BCA_i = 1/2 * [(TP/TP + FN) + (TN/TN + FP)] \quad (4)$$

  Then, the overall BCA for all the classes is mean of the individual BCA of each classes as given in the Eq. (4):

$$BCA = 1/L * \sum_{i=1}^{L} BCA_i \quad (5)$$

- Precision:
  Precision for class 'i' is given in Eq. (6):

$$Precision_i = TP/(TP + FP) \quad (6)$$

  where TP is the True Positives and FP is the False Positives. The Eq. (6) calculates the precision for a single class by finding the ratio between the True Positives to the sum of the True Positives and False Positives in the confusion matrix for that class.
  The overall precision is the mean of individual precisions of each classes as given in Eq. (7):

$$Precision = 1/L * \sum_{i=1}^{L} Precision_i \quad (7)$$

- Sensitivity or True Positive Rate:
  Sensitivity for class 'i' is given in Eq. (8):

$$Sensitivity_i = TP/(TP + FN) \quad (8)$$

  where TP is the True Positives and FN is the False Negatives. The Eq. (8) calculates the sensitivity for a single class by finding the ratio between the True Positives to the sum of the True Positives and False Negatives in the confusion matrix for that class.
  The overall sensitivity is the mean of individual sensitivity of each classes as given in Eq. (9):

$$Sensitivity = 1/L * \sum_{i=1}^{L} sensitivity_i \quad (9)$$

- Specificity or True Negative Rate:
  Specificity for class 'i' is given in Eq. (10):

$$Specificity_i = TN/(TN + FP) \quad (10)$$

  where TN is the True Negative and FP is the False Positive. The Eq. (10) calculates the specificity for a single class by finding the ratio between the True Negatives to the sum of the True Negatives and False Positives in the confusion matrix for that class.
  The overall specificity is the mean of individual specificity of each classes as given in Eq. (11):

$$Specificity = 1/L * \sum_{i=1}^{L} Specificity_i \quad (11)$$

- MAUC:
  The overall MAUC is given in the Eq. (12).

$$MAUC = 1/L(L-1) * \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} (1/n_i * n_k) * [SR_i - n_i * (n_i+1)/2] + [SR_k - n_k * (n_k+1)/2] \quad (12)$$

where L is the number of distinct classes, n_i is the number of available points in the class 'i', n_k is the number of available points in the class 'k', SR_i is the sum of the ranks of class 'i' test points after ranking all the class 'i' and 'k' test points in the ascending order of posterior probabilities of belonging to class 'i' (Hand and Till, 2001).

4. Experimental results and discussions

The experiments are conducted on the Intel Core i5 processor CPU @2.40 GHz Lenovo machine. The proposed fusion of the FS and greedy searching (see Algorithm 1) feature selection algorithm returns the sub-optimal minimal features that are of high importance. The next step is to evaluate the performance of the resultant feature set. Hence, the performance evaluation of the sub-optimal features is performed using both LOOCV and stratified 10-fold cross-validation strategies (separately) on the selected cohort's data. The LOOCV is used for evaluation because of the small size of the sub-optimal minimal feature set and selected cohort's data of both the datasets (n $<$ 1000). Moreover, the computationally cheaper classifiers such as SVM and KNN are used in the study. Thus, the LOOCV strategy is used for utilizing all the selected cohort's data for both training and testing. However, the LOOCV strategy is performed on the imbalanced dataset of our study. As a result, the stratified K fold cross-validation is also used for getting more insights on the performance of the sub-optimal minimal feature set. As stratified K fold cross-validation evaluations can maintain an equal proportion of representation of the labels during each fold which is a good strategy for overcoming the drawbacks occurring in the training of imbalanced datasets (Pedregosa et al., 2011). A 10-fold stratified cross-validation is performed throughout the experiments in the paper. It is observed that the recent literatures especially in the medical field used stratified 10-fold cross-validation with good performances (Zhang et al., 2017; Marques et al., 2020; Sarawgi et al., 2020; Khagi et al., 2019). The performance metrics such as BCA, sensitivity, specificity, F1-score, and MAUC are measured every 10 folds. The average of the performance metrics in each fold is calculated to give the final performance measures. LOOCV and stratified 10-fold cross-validations are used separately while evaluating the performances using the sub-optimal minimal feature set for both three-way and binary classifications.

The performance of the proposed feature selection method is evaluated on the ADNI-TADPOLE and AIBL datasets. The experimental results reported for both ADNI-TADPOLE and AIBL along with discussions are explained in this section.

4.1. Results on ADNI-TADPOLE

The FS ranking of features on ADNI-TADPOLE dataset in descending order is given as follows: CDRSB, MMSE, ADAS-13, ADAS-11, FAQ, RAVLT_Immediate, FDG, RAVLT_perc_forgetting, RAVLT_learning, Hippocampus, AV45, Entorhinal, MidTemp, Fusiform, APOE4, Ventricles, WholeBrain, AGE, RAVLT_forgetting, PTEDUCAT, ICV (see Table 6). It is observed that all top-ranked features are not maximizing the performance of SVM and KNN. However, a combination of low ranked feature like AV45 or FDG with top-ranked features (cognitive tests) can increase the performance (see Table 7). For the selected cohort of 683 patient's, the features that are of high importance as found out by the proposed fusion of FS and greedy searching feature selection algorithm (NC v/s MCI v/s AD) on ADNI-TADPOLE dataset are: CDRSB, MMSE, ADAS-13, and AV45.

The performance metrics evaluated on ADNI-TADPOLE using the proposed features for SVM and KNN classifiers with LOOCV strategy are shown in Table 7. The classification of NC, MCI, and AD using CDRSB, ADAS-13, MMSE, and AV45 has achieved an overall BCA of 90% with SVM and 91% with KNN using LOOCV. Moreover, the proposed approach has reported a MAUC of 0.97 with SVM and 0.98 with KNN using LOOCV (see Table 7). The misclassification is comparatively higher while predicting NC as MCI using both SVM and KNN with LOOCV (see Tables 9 and 10). The confusion matrix for SVM, KNN on ADNI-TADPOLE using LOOCV are given in Tables 9 and 10. Table 8 contains the performance results using CDRSB, ADAS-13, MMSE, and AV45 with stratified 10-fold cross-validation. KNN has achieved the highest BCA and MAUC of 90% and 0.96 respectively using stratified 10-fold cross-validation.

Tables 11 and 12 contains the performance of the proposed features (returned by the proposed feature selection algorithm) on ADNI-TADPOLE for binary classifications using LOOCV and stratified 10-fold cross-validations respectively. For binary classification using SVM, NC v/s AD has achieved a BCA of 98% using cognitive tests such as MMSE and CDRSB on ADNI-TADPOLE and BCA of 100% for MCI v/s AD and MCI v/s NC classifications using MMSE, CDRSB, FDG, and AV45. However, KNN has reported tremendous performance accuracy for all binary classifications. It is also noteworthy that while classifying MCI and NC using KNN, MMSE, CDRSB, FDG, AV45 has reported with BCA of 100% (see Tables 11 and 12).

4.2. Results on AIBL

The FS ranking of features on AIBL dataset in decreasing order is as follows: LDELTOTAL, CDGLOBAL, LIMMTOTAL, MMSCORE, HMT100, RCT6, HMT102, HMT40, HMT3, RCT20, RCT392, AXT117, RCT11, HMT13, HMT7, BAT126 (see Table 13). However, similar to ADNI-TADPOLE, all the top ranked features are not maximizing the performance of SVM and KNN. A combination of low ranked feature with high ranked features can increase the performance of SVM which is a pattern seen in ADNI-TADPOLE as well

Table 6
FS Ranking of features on ADNI-TADPOLE dataset.

Rank number	Feature	FS value
1	CDRSB	3.26456
2	MMSE	2.825267
3	ADAS-13	2.3509701
4	ADAS-11	1.876137
5	FAQ	1.6059582
6	RAVLT_Immediate	1.104149
7	FDG	0.90156
8	RAVLT_Perc _forgetting	0.698552
9	RAVLT_Learning	0.599177
10	Hippocampus	0.50647
11	AV45	0.432278
12	Entorhinal	0.410530
13	MidTemp	0.277111
14	Fusiform	0.194059
15	APOE4	0.165384
16	Ventricles	0.119923
17	WholeBrain	0.04857
18	AGE	0.0197
19	RAVLT_forgetting	0.0134366
20	PTEDUCAT	0.0116322
21	ICV	0.00840

Table 7
Summary of performance metrics using SVM and KNN on ADNI-TADPOLE (3-way classification) with LOOCV.

Classifier	Total features	Features chosen	Features	BCA	Precision	Sensitivity	Specificity	MAUC
SVM	21	4	CDRSB ADAS-13 MMSE AV45	90.10%	93.63%	90.40%	95%	0.97
KNN	21	4	CDRSB ADAS-13 MMSE AV45	91%	90.61%	90.66%	94.66%	0.98

Table 8
Summary of performance metrics using SVM and KNN on ADNI-TADPOLE (3-way classification) with stratified 10-fold cross-validation.

Classifier	Total feature length	Proposed feature length	Features	BCA	Precision	Sensitivity	Specificity	MAUC
SVM	21	4	CDRSB ADAS-13 MMSE AV45	88%	89%	88%	91%	0.95
KNN	21	4	CDRSB ADAS-13 MMSE AV45	90%	90%	90%	90%	0.96

Table 9
Confusion matrix of SVM on ADNI-TADPOLE using LOOCV (3-way classifications).

	NC	AD	MCI
NC	209	0	22
AD	0	78	16
MCI	0	8	350

Table 10
Confusion matrix of KNN on ADNI-TADPOLE using LOOCV (3-way classifications).

	NC	AD	MCI
NC	214	0	17
AD	0	83	11
MCI	14	13	331

(see Table 14). For the selected cohort of patient's, the features that are of high importance as found out by the proposed fusion of FS and greedy searching feature selection algorithm (NC v/s MCI v/s AD) on AIBL dataset are: LDELTOTAL, CDGLOBAL, and MMSCORE.

The performance metrics evaluated on AIBL dataset using SVM and KNN with LOOCV are shown in Table 14. The proposed feature selection algorithm has reported BCA of 77%, 76% with SVM, KNN respectively using LDELTOTAL, MMSCORE, and CDGLOBAL with LOOCV. Table 15 contains the performace on AIBL dataset using SVM and KNN with stratified 10-fold cross-validation. It is worth mentioning that SVM has achieved much better BCA, precision, sensitivity, and specificity of 92.5%, 87%, 92%, and 92% respectively using stratified 10-fold cross-validation (see Table 15). The misclassification rate is comparatively higher in AIBL while predicting AD as MCI using SVM (see Table 16) and MCI as AD using KNN with LOOCV (see Table 17). The confusion matrix for SVM on AIBL using LOOCV is given in Table 16. The confusion matrix for KNN on AIBL using LOOCV is given in Table 17.

Tables 18 and 19 contains the performance of the proposed features (returned by the proposed feature selection algorithm) on AIBL for binary classifications using LOOCV and stratified 10-fold cross-validations. Binary classifications on AIBL using both SVM and KNN has reported comparatively lower results for MCI v/s AD classifications with BCA of 84% and 82% respectively using CDGLOBAL and MMSCORE with LOOCV. However, our proposed methodology has reported a better both sensitivity of 84% with SVM and specificity of 82.5% with KNN in classifying MCI and AD also highlights the importance of feature selection algorithm using LOOCV. Cognitive tests like CDGLOBAL, LIMMTOTAL, LDELTOTAL, and MMSCORE are promising features in classifying MCI v/s AD (see Tables 18 and 19) (see Table 20).

4.3. Missing data handling

It is to be noted that unlike other compared studies (Nguyen et al., 2018; Iddi et al., 2019), the missing data handling of our methodology involves the deletion of all patients with missing feature data. However, studies (Ghazi et al., 2019; Iddi et al., 2019) performed data imputation using RNN and R tool Nguyen et al., 2018 has achieved lower BCA, MAUC than our approach on ADNI-TADPOLE dataset (see Table 21). The study conducted by

Table 11
Summary of performance metrics using SVM and KNN on ADNI-TADPOLE for binary classifications with LOOCV.

Classifier	Classification type	Proposed features	BCA	Precision	Sensitivity	Specificity
SVM	NC v/s AD	MMSE, CDRSB	98%	96%	99%	98%
	MCI v/s AD	MMSE, CDRSB	100%	100%	100%	100%
	MCI v/s NC	MMSE, CDRSB FDG, AV45	100%	100%	100%	100%
KNN	NC v/s AD	MMSE, CDRSB	100%	100%	100%	100%
	MCI v/s AD	ADAS11,Wholebrain	100%	100%	100%	100%
	MCI v/s NC	MMSE, CDRSB	100%	100%	100%	100%

Table 12
Summary of performance metrics using SVM and KNN on ADNI-TADPOLE for binary classifications with stratified 10-fold cross-validation.

Classifier	Classification type	Proposed features	BCA	Precision	Sensitivity	Specificity
SVM	NC v/s AD	MMSE, CDRSB	97%	96%	99%	98%
	MCI v/s AD	MMSE, CDRSB	100%	100%	100%	100%
	MCI v/s NC	MMSE, CDRSB FDG, AV45	100%	100%	100%	100%
KNN	NC v/s AD	MMSE, CDRSB	100%	100%	100%	100%
	MCI v/s AD	ADAS11,Wholebrain	100%	100%	100%	100%
	MCI v/s NC	MMSE, CDRSB	100%	100%	100%	100%

Table 13
FS Ranking of features on AIBL dataset.

Ranknumber	Feature	FS value
1	LDELTOTAL	2.446481
2	CDGLOBAL	2.008132
3	LIMMTOTAL	1.8561295
4	MMSCORE	1.534779
5	HMT100	0.0195
6	RCT6	0.01903488
7	HMT102	0.017035
8	HMT40	0.015722
9	HMT3	0.013544
10	RCT20	0.0131114
11	RCT392	0.011801822
12	AXT117	0.010621
13	RCT11	0.0072096
14	HMT13	0.00716506
15	HMT7	0.00539985
16	BAT126	0.001149

Table 16
Confusion matrix of SVM on AIBL dataset using LOOCV (3-way classifications).

	NC	AD	MCI
NC	587	21	1
AD	43	81	19
MCI	6	14	85

Table 17
Confusion matrix of KNN on AIBL dataset using LOOCV (3-way classifications).

	NC	AD	MCI
NC	566	38	5
AD	34	87	22
MCI	4	23	781

Albright (2019) deleted the participants with missing data is the same as our approach to missing data handling also achieved a lower MAUC of 0.866. Besides, the study conducted by researchers in Moore et al. (2019) has ignored the time point missing data (without deleting the patients data) also resulted in lower BCA and MAUC of 73% and 0.82 respectively. But, the FS based feature selection is not implemented in Albright (2019) making our approach better. Moreover, the researchers find it difficult to choose the imputation strategy for missing health care data due to its inconsistency and unpredictability. That is why using only the complete data patients for our algorithm is a much better approach ensuring the study to be conducted purely on data from health sources without any external interventions like imputation. The significant count of samples even after deleting missing data is also a good indicator for conducting the study. Therefore, deleting the incomplete participant's data instead of using imputed values for missing data can improve the performance to a greater extent.

4.4. Proposed methodology without feature selection

The feature selection algorithm has significantly improved the performance in the classification of AD, NC, and MCI in both datasets. Fig. 2 contains the performance of the methodology without

Table 14
Summary of performance metrics using SVM and KNN on AIBL (3-way classification) with LOOCV.

Classifier	Total feature length	Proposed feature length	Proposed features	BCA	Precision	Sensitivity	Specificity
SVM	18	3	LDELTOTAL CDGLOBAL MMSCORE	77%	87.5%	79%	95%
KNN	18	3	LDELTOTAL CDGLOBAL HMT3	76%	82%	82.66%	95%

Table 15
Summary of performance metrics using SVM and KNN on AIBL (3-way classification) with stratified 10-fold cross-validation.

Classifier	Total feature length	Proposed feature length	Proposed features	BCA	Precision	Sensitivity	Specificity
SVM	18	3	LDELTOTAL CDGLOBAL MMSCORE	92.5%	87%	92%	92%
KNN	18	3	LDELTOTAL CDGLOBAL HMT3	82.5%	84%	81%	81%

Table 18
Summary of performance metrics using SVM and KNN on AIBL for binary classifications using LOOCV.

Classifier	Classification type	Proposed features	BCA	Precision	Sensitivity	Specificity
SVM	NC v/s AD	CDGLOBAL, MMSCORE LIMMTOTAL,LDELTOTAL	92%	97%	82%	91%
	MCI v/s AD	CDGLOBAL, MMSCORE	84%	85%	84%	84%
	MCI v/s NC	LIMMTOTAL, CDGLOBAL	93%	87.5%	93%	93%
KNN	NC v/s AD	CDGLOBAL, LDELTOTAL MMSSCORE	96%	95.5%	95.5%	95.5%
	MCI v/s AD	CDGLOBAL,LIMMTOTAL,LDELTOTAL	82%	83%	82.5%	82.5%
	MCI v/s NC	LDELTOTAL, CDGLOBAL	89%	86.5%	88.5%	88.5%

Table 19
Summary of performance metrics using SVM and KNN on AIBL for binary classifications using stratified 10-fold cross-validation.

Classifier	Classification type	Proposed features	BCA	Precision	Sensitivity	Specificity
SVM	NC v/s AD	CDGLOBAL, MMSCORE LIMMTOTAL,LDELTOTAL	92.5%	87%	92%	92%
	MCI v/s AD	CDGLOBAL, MMSCORE	81.5%	82%	81.5%	81%
	MCI v/s NC	LIMMTOTAL, CDGLOBAL	92.1%	97%	82%	92%
KNN	NC v/s AD	CDGLOBAL, LDELTOTAL MMSSCORE	82.5%	84%	81%	81%
	MCI v/s AD	CDGLOBAL,LIMMTOTAL,LDELTOTAL	81.5%	82%	81.5%	81%
	MCI v/s NC	LDELTOTAL, CDGLOBAL	96%	95.5%	95.5%	95.5%

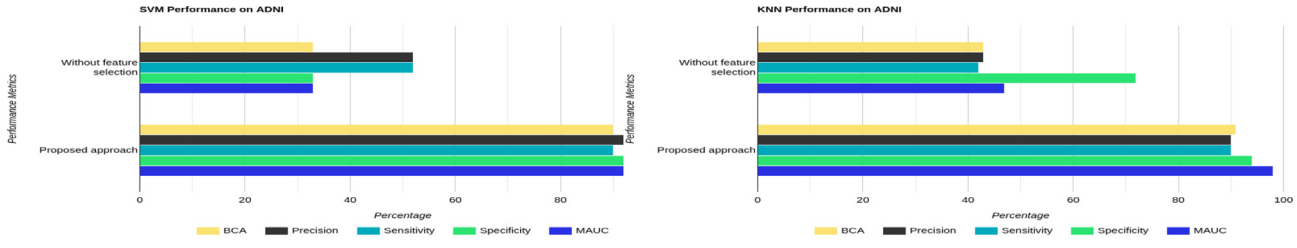
Table 20
Handling of missing data (ADNI-TADPOLE) in the compared papers.

Reference	Technique
Ghazi et al. (2019)	Imputation using RNN
Moore et al. (2019)	Ignored the missing time points of longitudinal data
Nguyen et al. (2018)	Imputation using RNN
Iddi et al. (2019)	Imputation using R tool
Albright (2019)	Deleted participants with missing data
Proposed approach	Deleted participants with missing data

using the proposed feature selection algorithm using SVM and KNN respectively evaluated using LOOCV for ADNI-TADPOLE. Fig. 3 contains the performance of the methodology without the proposed feature selection algorithm using SVM and KNN respectively evaluated with stratified 10-fold cross-validations for ADNI-TADPOLE. Fig. 4 contains the performance of the methodology without using the proposed feature selection algorithm for SVM and KNN respectively evaluated using LOOCV for AIBL. Fig. 5 contains the performance of the methodology without using the proposed feature

Table 21
Comparison of the proposed methodology for NC v/s MCI v/s AD classifications with the existing studies using ADNI-TADPOLE dataset. LDA-Linear Discriminant Analysis, NR-Not Reported.

Reference	Feature length	Modalities	Longitudinal/Cross-sectional	Classifier	BCA	MAUC
Ghazi et al. (2019)	6	MRI	Longitudinal	LDA	NR	0.75
Moore et al. (2019)	13	MRI Cognitive tests, Age, Gender	Longitudinal	RF	73%	0.82
Nguyen et al. (2018)	6	MRI, Cognitive, CSF	Longitudinal	SVM	86%	0.866
Iddi et al. (2019)	12	MRI, Cognitive tests	Longitudinal	RF	86%	NR
Albright (2019)	13	PET, Genetic risks Genetic, MRI Cognitive tests, Race, Age	Longitudinal	RNN	NR	0.866
Proposed approach	4	Cognitive tests, PET	Cross-sectional	SVM	90%	0.97
(LOOCV) Proposed approach	4	Cognitive tests, PET	Cross-sectional	KNN	91%	0.98
(LOOCV) Proposed approach	4	Cognitive tests, PET	Cross-sectional	SVM	88%	0.95
(Stratified 10-fold cross-validation) Proposed approach (Stratified 10-fold cross-validation)	4	Cognitive tests, PET	Cross-sectional	KNN	90%	0.96



(a) SVM (b) KNN

Fig. 2. Proposed methodology comparison after excluding proposed feature selection (on the proposed methodology) using SVM and KNN for NC v/s MCI v/s AD classification on ADNI-TADPOLE with LOOCV.

selection algorithm for SVM and KNN respectively evaluated using stratified 10-fold cross-validations for AIBL. The proposed methodology consisting of feature selection is better than using all the features (without feature selection) (see Figs. 2–5). The BCA without feature selection in the proposed methodology reported a BCA of 33.3% and 43.3% using SVM and KNN respectively with LOOCV strategy for ADNI-TADPOLE dataset (see Fig. 2). The BCA without feature selection in the proposed methodology reported a BCA of 33.3% and 40% using SVM and KNN respectively with stratified 10-fold cross-validation strategy for ADNI-TADPOLE dataset (see Fig. 3). The observed BCA for AIBL without feature selection is also less with 36.4% and 43.2% using SVM and KNN respectively using LOOCV (see Fig. 4). Also, the BCA for AIBL without feature selection is less with 35% and 44% respectively for SVM and KNN using stratified 10-fold cross-validations (see Fig. 5). Our algorithm with feature selection on the same datasets has a significant improvement of BCA more than 90% using SVM and KNN with both LOOCV and stratified 10-fold-cross-validations. The fusion algorithm also tripled the MAUC from 0.33 to 0.97 using SVM and doubled from 0.47 to 0.98 using KNN on ADNI-TADPOLE when compared to the methodology excluding the proposed feature selection with LOOCV. The sensitivity also increased significantly from 36% to 79% using SVM and 23% to 95% using KNN with our novel fusion algorithm (see Figs. 2 and 4) on AIBL with LOOCV. It is also worth mentioning that the performance is increased by using a minimal feature set consisting of only 4 features on ADNI-TADPOLE and 3 features on AIBL.

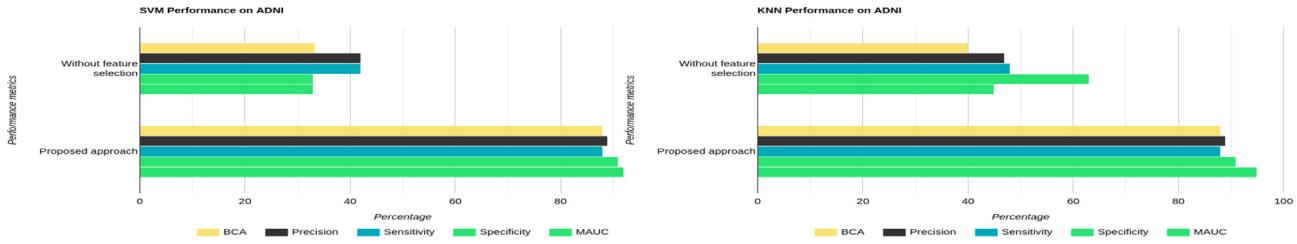
4.5. Comparison with other studies

The performance of the proposed fusion algorithm is compared with those studies that have predicted the diagnosis of AD, MCI, and NC on ADNI-TADPOLE and AIBL datasets. As the main focus is on the datasets rather than the algorithm, the studies that implemented FS feature selection on other datasets are not taken into consideration for comparisons. The studies in Ghazi et al. (2019), Nguyen et al. (2018), Iddi et al. (2019), Albright (2019) and Moore et al. (2019) also implemented the methodologies on ADNI-TADPOLE dataset in the same experimental setting as ours has paved the attention to choose them for comparison. Our study has used 22 features from the ADNI-TADPOLE dataset which is higher than the other five studies. But, our proposed feature selection algorithm finds an optimal set of 4 features from the 22 features and reported with a BCA of 90% and 91% and MAUC of 0.97 and 0.98 using SVM and KNN classifiers which is better than the other compared studies (Ghazi et al., 2019; Nguyen et al., 2018; Iddi et al., 2019; Albright, 2019; Moore et al., 2019). The proposed approach is designed to find out the sub-optimal set of features using FS ranking and a greedy selection based on BCA performance as the selection criteria will eventually finish with a set of features that aims to increase the performance of dimension based classifiers like SVM and KNN. Thus, the feature selection algorithm is more focused on the improvement of classifier performance. But, in the other studies (Ghazi et al., 2019; Nguyen et al., 2018; Iddi et al., 2019; Albright, 2019; Moore et al., 2019), the classifier is designed for the features.

In our study, it is also noteworthy that the features are added and removed to the optimal set using BCA as the evaluation function as BCA is an appropriate metric for class imbalanced datasets. The proposed algorithm has also achieved much better BCA, sensitivity and specificity of 91%, 92% and 87% respectively for NC v/s AD classifications on AIBL compared to the study in Doecke et al. (2012). Also, MCI v/s AD classification on AIBL also reported with better sensitivity and specificity of 84% respectively as compared to the study in Ashton et al. (2019). The feature selection technique on cross-sectional data is also used in both the studies (Doecke et al., 2012; Ashton et al., 2019). But, the feature selection technique of our proposed fusion algorithm is focused on the classifier performance on both datasets. The comparison of our study with similar studies in ADNI-TADPOLE and AIBL are given in Table 21.

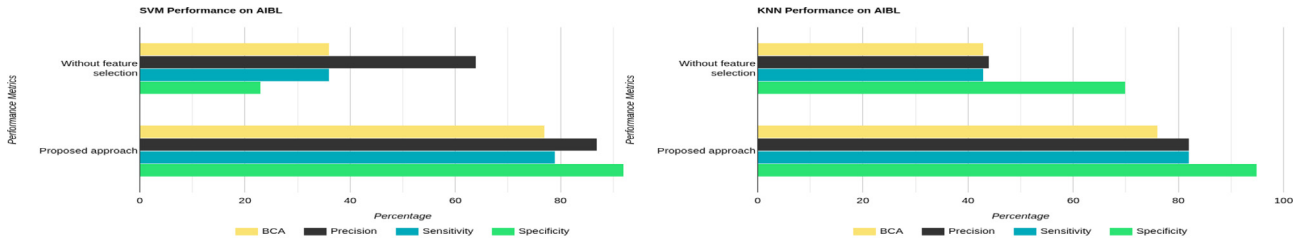
The main benefits of the proposed methodology are:

1. A very good performance with high BCA and MAUC and minimal feature set for SVM and KNN.



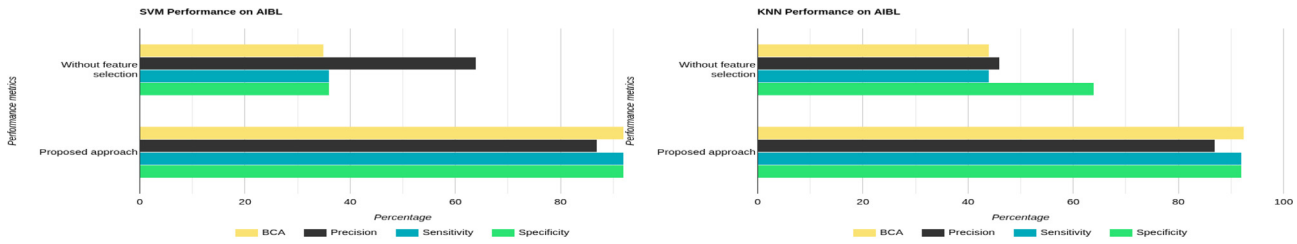
(a) SVM (b) KNN

Fig. 3. Proposed methodology comparison after excluding proposed feature selection (on the proposed methodology) using SVM and KNN for NC v/s MCI v/s AD classification on ADNI-TADPOLE with stratified 10-fold cross-validation.



(a) SVM (b) KNN

Fig. 4. Proposed methodology comparison after excluding proposed feature selection (on the proposed methodology) using SVM and KNN for NC v/s MCI v/s AD classification on AIBL with LOOCV.



(a) SVM (b) KNN

Fig. 5. Proposed methodology comparison after excluding proposed feature selection (on the proposed methodology) using SVM and KNN for NC v/s MCI v/s AD classification on AIBL with stratified 10-fold cross-validation.

Table 22
Comparison of the proposed methodology with the existing studies using AIBL dataset. GLM-Generalized Linear Models, NR-Not Reported.

Reference	Feature length	Modalities	Longitudinal/Cross-sectional	Classifier	BCA	Sensitivity	Specificity
NC v/s AD:							
Doecke et al. (2012)	18	Blood biomarker	NR	RF	NR	83%	83%
Proposed Approach (LOOCV)	3	Cognitive tests Lab tests	Cross-sectional	SVM	91%	92%	87%
Proposed Approach (Stratified 10-fold cross-validation)	3	Cognitive tests Lab tests	Cross-sectional	SVM	92.5%	92%	92%
MCI v/s AD:							
Ashton et al. (2019)	12	Blood biomarker	Cross-sectional	GLM	NR	78%	77%
Proposed approach (LOOCV)	3	Cognitive tests Lab tests	Cross-sectional	SVM	84%	84%	84%
Proposed approach (LOOCV)	3	Cognitive tests Lab tests	Cross-sectional	KNN	76%	82.5%	82.5%
NC v/s MCI v/s AD:							
Proposed approach (Stratified 10-fold cross-validation)	3	Cognitive tests Lab tests	Cross-sectional	SVM	92.5%	92%	92%
Proposed approach (Stratified 10-fold cross-validation)	3	Cognitive tests Lab tests	Cross-sectional	KNN	82.5%	81%	81%

2. The core element of our methodology is the feature selection. The main drawbacks of the FS based algorithms are setting the threshold for choosing the top-ranked features and finding the best possible combination of features within that threshold. However, the proposed feature selection algorithm does not set a threshold. But, it considers all the features in the FS rank wise order and finds the features that maximize each class's accuracy. The main benefit of the approach is that it finds out a better minimal set of features that can improve the performance of SVM and KNN. This is because of the greedy strategy where the searching initially starts from the top-ranked FS feature and then adds or removes the next features in the rank wise order is guaranteed to give a suitable minimal feature set. It is also to be noted that not all features in the top FS ranking will not maximize the performance. Hence, the inability of FS in dealing with a combination of features is dealt with by fusing FS with a greedy search strategy.
The proposed approach will find a minimal combination of features (that maximizes the BCA) by comparing the performance of each feature only once. Hence, the set of features can find out using less time. Here, the algorithm will take O(n) time as each 'n' feature will be searched and compared only once, but it still finds a locally better solution using the greedy searching heuristic based on the performance which makes it time-efficient. Besides, our study finds a minimal feature set with less time complexity as compared to other studies (Beheshti et al., 2016; Zhou, 2016; Song et al., 2017) where the hybrid approaches have many optimizations and searching resulting in high time complexity. Hence, the sub optimal minimal feature set for classification of AD is found out without any exhaustive searching. Thus, the proposed fusion algorithm is a novel approach for AD classification.

4.6. Baseline visit data

It is also worth mentioning that the proposed algorithm has achieved better performance by using only the baseline data of the patients when compared to other studies (Nguyen et al., 2018; Iddi et al., 2019; Moore et al., 2019) that used consequent visit (longitudinal) data. This kind of approach has two benefits mainly, 1. Our approach has used only fewer amount of data as compared to other studies (Nguyen et al., 2018; Iddi et al., 2019; Moore et al., 2019), 2. It is possible to make predictions with cross-sectional data alone. The diagnosis prediction for new testing data requires only baseline data. There is no requirement of the consequent visit longitudinal data of the patients (both training and testing) for making predictions (see Tables 21 and 22).

4.7. Common proposed features in the two datasets

The proposed feature selection algorithm found out MMSE as a common feature of high importance on the selected cohorts of both ADNI-TADPOLE and AIBL datasets in NC, MCI, and AD classifications using SVM. Also, the proposed feature selection algorithm found out CDGLOBAL as a feature of high importance on the selected cohorts of both ADNI-TADPOLE and AIBL datasets in NC, MCI, and AD classifications using KNN (see Tables 7, 8, 14 and 15). As far as the binary classifications are concerned, both MMSE

and CDRSB again are found to be the high important common features for classifying NC and AD using SVM on both the datasets. However, CDRSB alone is the common distinguishing feature for classifying MCI and NC patients using SVM and KNN on both the datasets (see Tables 11, 12, 18 and 19).

4.8. Limitations and future works

The limitations of the proposed method are:

- The initial selection of features from both ADNI-TADPOLE and AIBL is based on the features that are already used in the literature. However, the proposed methodology can be extended to more number of features from multiple modalities.
- The features sets that are selected need to be tested with more number of data. The proposed methodology is tested with less number of patients in both the datasets. However, we are planning to test our model with more number of test data in the future.
- The proposed feature selection algorithm will find a better minimal set of features but not the best set. Finding the most informative set of features for a classifier in a supervised wrapper feature selection model is also a challenging task.
- Needs to work out the proposed feature selection algorithm on the same datasets after performing better data imputation techniques. The influence of missing data can often leads to bad performance of a classifier. However, intelligent data imputation techniques can add more samples for training and improves the performance of a classifier. Moreover, the data imputation technique also needs to be meaningful. As an extension of this work, we are planning to design an intelligent data imputation technique on the same datasets.
- A novel feature selection method also needs to be applied to a local cohort population in a clinical setting. We are planning to validate such a model by a physician also. This is also considered as future work.

5. Conclusions

Detecting AD from other MCI at the baseline stage is a difficult task for a doctor. In this paper, a fusion of FS and greedy searching heuristic feature selection approach is used for classifying MCI, AD, and NC patients at baseline visit. Our main focus is on finding a sub-optimal minimal set of features that can predict the diagnosis of AD, MCI, and NC. The proposed approach is implemented on two standard datasets namely ADNI-TADPOLE and AIBL. Our approach has reported a BCA, MAUC of 90%, 0.97 using SVM and 91%, 0.98 using KNN respectively on ADNI-TADPOLE. Moreover, the proposed algorithm also resulted in better sensitivity and specificity of 84%, 82.5% using SVM and KNN respectively for binary classification of MCI and AD on AIBL dataset. Our findings indicate that the proposed fusion approach can out-perform other algorithms that predict the diagnosis using multi-modality data. Further, our study pinpoints the necessity for advanced feature selection algorithms depending on the classifier for predicting the diagnosis of AD.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors of this study wish to express their gratitude to the ADNI-TADPOLE for providing access to the database; to Department of Computer Science, Central University of Tamilnadu, Thiruvarur, India for supporting this study. This work was funded by the Central University of Tamilnadu, Thiruvarur, India.

References

Albright, Jack, 2019. Forecasting the progression of alzheimer's disease using neural networks and a novel pre-processing algorithm. arXiv preprint arXiv:1903.07510.

Alzheimer's Association, 2019. 2019 alzheimer's disease facts and figures. Alzheimer's & Dementia 15 (3), 321–387.

Ashton, Nicholas J., Nevado-Holgado, Alejo J., Barber, Imelda S., Lynham, Steven, Gupta, Veer, Chatterjee, Pratishtha, Goozee, Kathryn, Hone, Eugene, Pedrini, Steve, Blennow, Kaj, 2019. A plasma protein classifier for predicting amyloid burden for preclinical alzheimer's disease. Science Advances, 5(2), eaau7220.

Battineni, Gopi, Chintalapudi, Nalini, Amenta, Francesco, 2019. Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (svm). Informatics in Medicine Unlocked 16, 100200.

Beheshti, Iman, Demirel, Hasan, Alzheimer's Disease Neuroimaging Initiative, 2016. Feature-ranking-based alzheimer's disease classification from structural mri. Magnetic resonance imaging 34 (3), 252–263.

Doecke, James D., Laws, Simon M., Faux, Noel G., Wilson, William, Burnham, Samantha C., Lam, Chiou-Peng, Mondal, Alinda, Bedo, Justin, Bush, Ashley I., Brown, Belinda, 2012. Blood-based protein biomarkers for diagnosis of alzheimer disease. Archives of Neurology, 69 (10), 1318–1325.

Eldholm, Rannveig Sakshaug, Barca, Maria Lage, Persson, Karin, Knapskog, Anne-Brita, Kersten, Hege, Engedal, Knut, Selbaek, Geir, Braekhus, Anne, Skovlund, Eva, Saltvedt, Ingvild, 2018. Progression of alzheimer's disease: A longitudinal study in norwegian memory clinics. Journal of Alzheimer's Disease 61 (3), 1221–1232.

Ellis, Kathryn A., Bush, Ashley I., Darby, David, De Fazio, Daniela, Foster, Jonathan, Hudson, Peter, Lautenschlager, Nicola T., Lenzo, Nat, Martins, Ralph N., Maruff, Paul, Masters, Colin, Milner, Andrew, Pike, Kerryn, Rowe, Christopher, Savage, Greg, Szoeke, Cassandra, Taddei, Kevin, Villemagne, Victor, Woodward, Micheal, Ames, David, AIBL Research Group, 2009. The australian imaging, biomarkers and lifestyle (aibl) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of alzheimer's disease. International Psychogeriatrics, 21 (4), 672–687.

Farooq, Ammarah, Anwar, SyedMuhammad, Awais, Muhammad, Rehman, Sa.ad., 2017. A deep cnn based multi-class classification of alzheimer's disease using mri. In: 2017 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, pp. 1–6.

Ferrari, Camilla, Lombardi, Gemma, Polito, Cristina, Lucidi, Giulia, Bagnoli, Silvia, Piaceri, Irene, Nacmias, Benedetta, Berti, Valentina, Rizzuto, Debora, Fratiglioni, Laura, Sandra, S., 2018. Alzheimer's disease progression: factors influencing cognitive decline. Journal of Alzheimer's Disease 61 (2), 785–791.

Forouzannezhad, Parisa, Abbaspour, Alireza, Li, Chunfei, Fang, Chen, Williams, Ulyana, Cabrerizo, Mercedes, Barreto, Armando, Andrian, Jean, Rishe, Naphtali, Curiel, Rosie E., Loewenstein, David, Duara, Ranjan, Adjouadi, Malek, 2020. A gaussian-based model for early detection of mild cognitive impairment using multimodal neuroimaging. Journal of Neuroscience Methods, 333, 108544.

Ghazi, Mostafa Mehdipour, Nielsen, Mads, Pai, Akshay, Jorge Cardoso, M., Modat, Marc, Ourselin, Sébastien, Sørensen, Lauge, Alzheimer's Disease Neuroimaging Initiative, 2019. Training recurrent neural networks robust to incomplete data: Application to alzheimer's disease progression modeling. Medical Image Analysis, 53, 39–46.

Gu, Quanquan, Li, Zhenhui, Han, Jiawei, 2012. Generalized fisher score for feature selection. arXiv preprint arXiv:1202.3725.

Hand, David J., Till, Robert J., 2001. A simple generalisation of the area under the roc curve for multiple class classification problems. Machine Learning 45 (2), 171–186.

Hao, Xiaoke, Bao, Yongjin, Guo, Yingchun, Yu, Ming, Zhang, Daoqiang, Risacher, Shannon L., Saykin, Shannon L., Yao, Xiaohui, Shen, Li, Alzheimer's Disease Neuroimaging Initiative, 2020. Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of alzheimer's disease. Medical Image Analysis, 60, 101625.

Iddi, Samuel, Li, Dan, Aisen, Paul S., Rafii, Michael S., Thompson, Wesley K., Donohue, Michael C., Alzheimer's Disease Neuroimaging Initiative, 2019. Predicting the course of alzheimer's progression. Brain Informatics 6 (1), 6.

Kaucha, Deep Prakash, Prasad, P.W.C., Alsadoon, Abeer, Elchouemi, A., Sreedharan, Sasikumaran, 2017. Early detection of lung cancer using svm classifier in biomedical image processing. In: 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI). IEEE, pp. 3143–3148.

Khagi, Bijen, Kwon, Goo-Rak, Lama, Ramesh, 2019. Comparative analysis of alzheimer's disease classification by cdr level using cnn, feature selection, and machine-learning techniques. International Journal of Imaging Systems and Technology 29 (3), 297–310.

Khan, Afreen, Zubair, Swaleha, 2020. An improved multi-modal based machine learning approach for the prognosis of alzheimer's disease. Journal of King Saud University-Computer and Information Sciences.

Kira, Kenji, Rendell, Larry A., 1992. The feature selection problem: Traditional methods and a new algorithm. Aaai 2, 129–134.

Kurlowicz, Lenore, Wallace, Meredith, 1999. The mini-mental state examination (mmse). Journal of Gerontological Nursing 25 (5), 8–9.

Langa, Kenneth M., 2015. Is the risk of alzheimer's disease and dementia declining? Alzheimer's Research & Therapy 7 (1), 34.

Long, Xiaojing, Chen, Lifang, Jiang, Chunxiang, Zhang, Lijuan, Alzheimer's Disease Neuroimaging Initiative, 2017. Prediction and classification of alzheimer disease based on quantification of mri deformation. PloS One 12, (3) e0173372.

Lorenz, Klara, Freddolino, Paul P, Comas-Herrera, Adelina, Knapp, Martin, Damant, Jacqueline, 2019. Technology-based tools and services for people with dementia and carers: Mapping technology onto the dementia care pathway. Dementia 18 (2), 725–741.

Marinescu, Razvan V., Oxtoby, Neil P., Young, Alexandra L., Bron, Esther E., Toga, Arthur W., Weiner, Michael W., Barkhof, Frederik, Fox, Nick C., Klein, Stefan, Alexander, Daniel C., 2018. Tadpole challenge: Prediction of longitudinal evolution in alzheimer's disease. arXiv preprint arXiv:1805.03909.

Marques, Gonçalo, Agarwal, Deevyankar, de la Torre Díez, Isabel, 2020. Automated medical diagnosis of covid-19 through efficientnet convolutional neural network. Applied Soft Computing 96, 106691.

Matthews, Kevin A., Wei, Xu, Gaglioti, Anne H., Holt, James B., Croft, Janet B., Mack, Dominic, McGuire, Lisa C., 2019. Racial and ethnic estimates of alzheimer's disease and related dementias in the united states (2015–2060) in adults aged 65 years. Alzheimer's & Dementia 15 (1), 17–24.

McKhann, Guy M., Knopman, David S., Chertkow, Howard, Hyman, Bradley T., Jack, Clifford R., Jr, Kawas, Claudia H., Klunk, William E., Koroshetz, Walter J., Manly, Jennifer J., Mayeux, Richard, Mohs, Richard C., Morris, John C., Rossor, Martin N., Scheltens, Phillip, Carillo, Maria C., Thies, Bill, Weintraub, Sandra, Phelps, Creighton H., 2011. The diagnosis of dementia due to alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. Alzheimer's & Dementia, 7 (3), 263–269.

Molina, Luis Carlos, Belanche, Lluís, Nebot, Àngela, 2002. Feature selection algorithms: A survey and experimental evaluation. In: 2002 IEEE International Conference on Data Mining, 2002. Proceedings, IEEE, pp. 306–313.

Moore, P.J., Lyons, T.J., Gallacher, John, Alzheimer's Disease Neuroimaging Initiative, 2019. Random forest prediction of alzheimer's disease using pairwise selection from time series data. PloS One 14, (2) e0211558.

Nguyen, Minh, Sun, Nanbo, Alexander, Daniel C., Feng, Jiashi, Thomas, B.T., 2018. Yeo: Modeling alzheimer's disease progression using deep recurrent neural networks. In: 2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI). IEEE, pp. 1–4.

Pan, Yijun, Nicolazzo, Joseph A., 2018. Impact of aging, alzheimer's disease and parkinson's disease on the blood-brain barrier transport of therapeutics. Advanced Drug Delivery Reviews 135, 62–74.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Petersen, Ronald C., Smith, Glenn E., Waring, Stephen C., Ivnik, Robert J., Tangalos, Eric G., Kokmen, Emre, 1999. Mild cognitive impairment: Clinical characterization and outcome. Archives of Neurology 56 (3), 303–308.

Salmon, David P., Bondi, Mark W., 2009. Neuropsychological assessment of dementia. Annual Review of Psychology 60, 257–282.

Sarawgi, Utkarsh, Zulfikar,Wazeer, Soliman, Nouran, Maes, Pattie, (2020). Multimodal inductive transfer learning for detection of alzheimer's dementia and its severity. arXiv preprint arXiv:2009.00700.

Song, QingJun, Jiang, HaiYan, Liu, Jing, 2017. Feature selection based on fda and f-score for multi-class classification. Expert Systems with Applications 81, 22–27.

Tsoy, Radmila T., Turuspekova, Saule T., Klipitskaya, Nina K., Mereke, Alibek, Cumming, Robert G., 2019. Prevalence of mild cognitive impairment among older people in kazakhstan and potential risk factors. Alzheimer Disease & Associated Disorders 33 (2), 136–141.

Valenchon, Juliette, Coates, Mark, 2019. Multiple-graph recurrent graph convolutional neural network architectures for predicting disease outcomes. In: ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 3157–3161.

Verma, Jyoti, Nath, Madhwendra, Tripathi, Priyanshu, Saini, K.K., 2017. Analysis and identification of kidney stone using k th nearest neighbour (knn) and support vector machine (svm) classification techniques. Pattern Recognition and Image Analysis 27 (3), 574–580.

Vivar, Gerome, Burwinkel, Hendrik, Kazi, Anees, Zwergal, Andreas, Navab, Nassir, Ahmadi, Seyed-Ahmad, 2019. Multi-modal graph fusion for inductive disease classification in incomplete datasets. arXiv preprint arXiv:1905.03053.

Winblad, Bengt, Amouyel, Philippe, Andrieu, Sandrine, Ballard, Clive, Brayne, Carol, Brodaty, Henry, Cedazo-Minguez, Angel, Dubois, Bruno, Edvardsson, David, Feldman, Howard, Fratiglioni, Laura, Frisonni, Giovanni B.. Gauthir, Serge, Georges, Jean, Graff, Caroline, Iqbal, Khalid, Jessen, Frank, Johansson, Gunilla, Jonsonn, Linus, Kivipelto, Milia, Martin, Knapp, Mangialasche, Francesca, Melis, Rena, Nordberg, Agneta, Olde Rikket, Marcel, Qiu, Chengxuan, Sakmar, Thomas P., Scheltens, Philips, Schneider, Lon S., Sperling, Reisa, Tjernberg, Lars O., Waldemar, Gunhild, Wimo, Anders, Henrik, Z., 2016. Defeating alzheimer's disease and other dementias: a priority for european science and society. The Lancet Neurology, 15 (5), 455–532.

Wortmann, Marc, 2012. Dementia: a global health priority-highlights from an adi and world health organization report. Alzheimer's Research & Therapy 4 (5), 40.

Xing, Wenchao, Bei, Yilin, 2019. Medical health big data classification based on knn classification algorithm. IEEE Access 8, 28808–28819.

Zhang, Daoqiang, Shen, Dinggang, 2011. Multi-modal multi-task learning for joint prediction of clinical scores in alzheimer's disease. In: International Workshop on Multimodal Brain Image Analysis. Springer, pp. 60–67.

Zhang, Daoqiang, Wang, Yaping, Zhou, Luping, Yuan, Hong, Shen, Dinggang, Alzheimer's Disease Neuroimaging Initiative, 2011. Multimodal classification of alzheimer's disease and mild cognitive impairment. Neuroimage 55 (3), 856–867.

Zhang, Daoqiang, Shen, Dinggang, Alzheimer's Disease Neuroimaging Initiative, 2012. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease. NeuroImage 59 (2), 895–907.

Zhang, Yu-Dong, Zhang, Yin, Lv, Yi-Ding, Hou, Xiao-Xia, Liu, Fang-Yuan, Jia, Wen-Juan, Yang, Meng-Meng, Phillips, Preetha, Wang, Shui-Hua, 2017. Alcoholism detection by medical robots based on hu moment invariants and predator–prey adaptive-inertia chaotic particle swarm optimization. Computers & Electrical Engineering 63, 126–138.

Zhou, M., 2016. Hybrid feature selection method based on fisher score and genetic algorithm. Journal of Mathematical Sciences: Advances and Applications 37, 51–78.

Zimmermann, Joelle, Perry, Alistair, Breakspear, Michael, Schirner, Michael, Sachdev, Perminder, Wen, Wei, Kochan, Nicole A., Mapstone, Michael, Ritter, Petra, McIntosh, Anthony R., 2018. Differentiation of alzheimer's disease based on local and global parameters in personalized virtual brain models. NeuroImage: Clinical, 19, 240–251.