Konya Mühendislik Bilimleri Dergisi., c. 9, s. 1, 50-61, 2021
Konya Journal of Engineering Sciences v. 9, n. 1, 50-61, 2021
ISSN: 2667-8055 (Elektronik)
DOI: 10.36306/konjes.731624

EARLY DETECTION OF ALZHEIMER'S DISEASE USING DATA MINING: COMPARISON OF ENSEMBLE FEATURE SELECTION APPROACHES

Selim BUYRUKOĞLU

Çankırı Karatekin University, Computer Engineering Department, Çankırı, TURKEY
sbuyrukoglu@karatekin.edu.tr

(Geliş/Received: 04.05.2020; Kabul/Accepted in Revised Form: 29.09.2020)

ABSTRACT: Early Alzheimer's disease detection has become an important research area for many years. Various studies in the field of Alzheimer's disease detection have focused on applying individual feature selection methods. In addition to individual feature selection methods, the ensemble feature selection approach has become a creative field. It advocates the combination of the ranked features from various feature selection methods to obtain better results than the current approaches. Thus, this study aims to build a predictive model for early diagnosis of Alzheimer's disease using the ensemble feature selection approaches. Also, Alzheimer's disease dataset consists of three target classes: Normal (CN), Mild Cognitive Impairment (MCI) and Alzheimer's disease (AD). In this study, homogeneous and heterogeneous ensemble approaches have been applied in the feature selection process. Two feature subsets are created based on these ensemble feature selection approaches. A predictive model for early diagnosis of Alzheimer's disease has been build applying Random Forest, Artificial Neural Network, Logistic Regression, Support Vector Machine, and Naïve Bayes data mining algorithms. The predictive model uses the two feature subsets applying these algorithms separately. Then, the performance results are compared to determine which ensemble feature selection approach performs better than the other. This study revealed that better performance result is provided applying Random Forest algorithm with feature subset obtained using the heterogeneous ensemble feature selection approach (91%).

Key Words: *Alzheimer's disease prediction, Heterogeneous, Random forest, Data mining, Early diagnosis*

Veri Madenciliği Kullanılarak Alzheimer Hastalığının Erken Tespiti: Topluluk Özellik Seçim Yaklaşımlarının Karşılaştırılması

ÖZ: Erken Alzheimer hastalığı tespiti uzun yıllardır önemli bir araştırma alanı haline gelmiştir. Alzheimer hastalığı tespiti alanında yapılan çeşitli çalışmalar, bireysel özellik seçme yöntemlerini uygulamaya odaklanmıştır. Bireysel özellik seçme yöntemlerine ek olarak, topluluk özellik seçme yaklaşımı yaratıcı bir alan haline gelmiştir. Bu yaklaşım, mevcut yaklaşımlardan daha iyi sonuçlar elde etmek için çeşitli özellik seçim yöntemlerinden sıralanan özelliklerin kombinasyonunu savunur. Bu nedenle, bu çalışmanın amacı, topluluk özellik seçim yaklaşımlarını kullanarak Alzheimer hastalığının erken teşhisi için bir öngörücü model oluşturmaktır. Ayrıca, Alzheimer hastalığı veri seti üç hedef sınıftan oluşur: Normal (CN), Hafif Bilişsel Bozukluk (MCI) ve Alzheimer hastalığı (AD). Bu çalışmada, özellik seçim sürecinde homojen ve heterojen topluluk yaklaşımları uygulanmıştır. Bu topluluk özellik seçim yaklaşımlarına dayanarak iki özellik alt kümesi oluşturulmuştur. Rastgele Orman, Yapay Sinir Ağı, Lojistik Regresyon, Destek Vektör Makinesi ve Naïve Bayes veri madenciliği algoritmaları uygulanarak Alzheimer hastalığının erken teşhisi için bir tahmin modeli oluşturulmuştur. Bu tahmin modeli yukarıda bahsedilen algoritmaları her iki özellik alt kümesini de ayrı ayrı kullanarak bir tahminde bulunmuştur. Ardından, hangi topluluk özellik seçim yaklaşımının diğerinden daha iyi performans gösterdiğini belirlemek için performans sonuçları karşılaştırılmıştır. Bu çalışma, heterojen topluluk özellik seçim yaklaşımı kullanılarak elde edilen özellik altkümesi ile Rastgele Orman algoritması uygulanarak daha iyi performans sonucunun sağlandığını ortaya koymuştur (% 91).

Anahtar Kelimeler: *Alzheimer hastalığı tahmini, Heterojen, Rastgele orman, Veri madenciliği, Erken tanı*

1. INTRODUCTION

Alzheimer's disease (AD) is widespread worldwide and is usually seen in elderly people (Stamps *et al.,* 2010). It is possible to reduce the number of AD as a result of early and accurate detection (Lee *et al.,* 2019). Chaves *et al.,* (2013) points out that almost half of the early AD diagnoses are incorrect, and also the number of AD or a kind of dementia in the world is nearly forty-four million. Recently, researchers have investigated a variety of data mining methods to accomplish early and accurate AD detection (Zhang *et al.,* 2017; Supekar *et al.,* 2008; Farhan *et al.,* 2014; Dallora *et al.,* 2017). Classification can be defined as a function that is used to assign items to target classes. One could say that early detection is a classification problem. The overall structures of the prediction systems take the patient data as an input and then the systems provide output to determine whether the patient is Normal (CN)**,** Mild Cognitive Impairment (MCI) or Alzheimer's disease (AD). If the output for a patient is recognised as MCI, then this means that early detection of AD is achieved in this study. However, the prediction systems have produced mostly biased classification results which are higher for the majority of the class. (Khan and Usman, **2019**). Thus, the major objective of this study is to build an early and accurate AD detection model using the dataset of Alzheimer's disease. The dataset consists of three target classes: Normal (CN)**,** Mild Cognitive Impairment (MCI) and Alzheimer's disease (AD).

Numerous studies have attempted to detect AD applying data mining methods (Williams *et al.,* 2013; Bhagyashree *et al.,* 2018; Shankle *et al.,* 1997; Chen and Herskovits, 2010; Klöppel *et al.,* 2008; Zhang and Shen, 2011; Khazaee *et al.,* 2016; Wee *et al.,* 2013). Williams *et al.,* (2013) implemented four models to predict clinical diagnosis. These models are SVM, DT, ANN, and NB. Missing values have been replaced with mean values, and, Williams *et al.,* (2013) found that NB shows the most accurate result on the prediction of clinical diagnosis. Bhagyashree *et al.,* (2018) has compared the NB, J48, and RF to detect dementia. Bhagyashree's *et al.,* (2018) comparative study has revealed that J48 provides the least accuracy in detecting. A study carried out by Maroco *et al.,* (2011) examined the performance of different models (SVM, RF, RBF neural networks and MLP neural network etc.) in the prediction of AD. The study demonstrated that RF performs strong accuracy and sensitivity. DT and NB have been identified as the two most effective methods for AD detection by Shankle *et al.,* (1997). Elsewhere, a different research has developed a model for predicting AD using 24 various neuropsychology attributes is a comparative study (Bhagyashree and Sheshadri, 2014). Four different methods have been compared using these attributes, and the research highlighted that NB is better in all. Machine learning methods (SVM, NB, DT, and MLP) and statistical methods (logistic regression and discriminant analysis) have been compared in a different study on MRI images (Chen and Herskovits, 2010). Similarly, Klöppel *et al.,* (2008) used MRI images for early AD detection with SVM in their study. Different studies also examined the performance of different models, and each of them obtained the highest accuracy results on AD detection with SVM (Zhang and Shen, 2011; Khazaee *et al.,* 2016; Wee *et al.,* 2013).

In different studies, Naïve Bayes (Nunes *et al.,* 2013), Logit Boost (Munteanu *et al.,* 2015), Support Vector Machines (Liu *et al.,* 2014; Moradi *et al.,* 2015; Maroco *et al.,* 2011; Zhao and He, 2014; Westman *et al.,* 2012), and Deep learning (Moradi *et al.,* 2015; Jo *et al.,* 2019) methods have also been used in early detection of AD. However, several important limitations need to be considered and four of them will be presented in this study. First of all and the most important limitation lies in the fact that these studies used feature selection methods neglecting rich attributes (Wordoffa and Wangoria, 2012; Escudero *et al.,* 2013). The second limitation is that datasets consist of missing values (Campos *et al.,* 2015). Thirdly, small sample sizes have been a serious limitation for many studies (Supekar *et al.,* 2008; Bookheimer *et al.,* 2000). Finally, much of the current literature on AD detection pays particular attention to evaluate the measurement of the accuracy of the classifiers while neglecting sensitivity and Area Under the Curve AUC (Zhang *et al.,* 2017). A prediction model is considered efficient if it provides satisfactory sensitivity and AUC rates (Huang *et al.*, 2017).

In addition to data mining methods, feature selection is also an important process in the prediction systems and plays a key role in the evaluation measure of the systems, such as accuracy, time and cost etc. (Seijo-Pardo *et al.*, 2017). The feature selection methods use variable ranking techniques to select the highly ranked features which make an important contribution to the performance of the prediction systems (Sana *et al.*, 2019). Feature selection methods can be divided into three main categories: filter, wrapper, and embedded methods (Seijo-Pardo *et al.*, 2017). Note that fast evaluation functions are used in filter-based methods. Additionally, this approach requires fewer computational resources than wrapper based methods (Seijo-Pardo *et al.*, 2017). On the other hand, the embedded methods can be seen as the intermediate of the others. The filter-based methods are chosen for this study because there are important advantages, such as the computational cost, fast evaluation functions, and independence of specific learners (Trambaiolli *et al.*, 2019).

It has been specified in the previous paragraphs that the performance of a model is dependent on the dataset and features used in the model. Although various feature selection methods are proposed in the literature, there appears to be some agreement by researchers that there are no ideal feature selection methods in the literature. The ensemble approaches have been applied to this situation, and therefore this (current) study concentrates on the two kinds of ensemble approaches, namely homogeneous and heterogeneous (Hand, 2007). The application of these ensemble approaches aims to determine highly ranked features to early AD detection and to compare these approaches in terms of performance evaluation. In a homogeneous approach, the same feature selection method is applied for generated models, but different training data need to be used for each model. That is, the number of models depends on the used different training data. The same feature selection method is executed to obtain highly ranked features and then these features are combined through a combination method. In contrast to the homogeneous approach, models are formed using more than one (various) feature selection method with the same training data in the heterogeneous approach. The process of the combination of the obtained features is the same as a homogenous approach.

To overcome the aforementioned limitations, an early AD detection model using data mining based on ensemble approaches has been proposed. The objectives of this study based on the factor are listed below.

a. To determine the significant feature selection methods based on the filter approach.
b. To create different feature sets applying the determined feature selection methods.
c. To create models using the feature sets based on the homogeneous and heterogonous ensemble. In this sense, two different feature sets are created (one for each approach).
d. To compare the models' performances in terms of accuracy, sensitivity, AUC.

Although all the previously mentioned studies achieve to predict AD so far, they do suffer from several drawbacks. A serious drawback with some of these studies is that the dataset did not pre-processed. In this case, models' performances are negatively affected since they take unnecessary data, garbage, and noise values into account. On the other hand, if the dataset has been pre-processed, complexity would be reduced while the good results would be obtained with higher accuracy. An arguable weakness is that to the best of our knowledge, these two homogeneous and heterogonous ensemble feature selection approaches have never been compared in terms of evaluation measure in a single study for the early AD detection. Thus, this study aims to build a predictive model for early diagnosis of Alzheimer's disease comparing the ensemble feature selection approaches. These studies would have been more useful if researchers had focused on both approaches. The reason for this is that different feature sets would be obtained applying both homogeneous and heterogeneous approaches, and then different models could be built based on the feature sets. At the end of this argument, model performance could be compared to decide which approach is better. This study has proposed a predictive model of AD based on the drawbacks presented in the light of the above.

The structure of the paper is as follows: the next section introduces material and method which can be listed under two headings: dataset and system framework. Section 3 presents results and discussion and the final section provides conclusions and outlines the potential for future work in this area.

2. MATERIAL AND METHOD

This section may be divided into six subsections: the dataset, data cleaning and transformation, ensemble feature selection, sizes of training and testing sets, model selection and application, and also performance evaluation. Figure 1 presents the framework of early Alzheimer's Disease prediction.

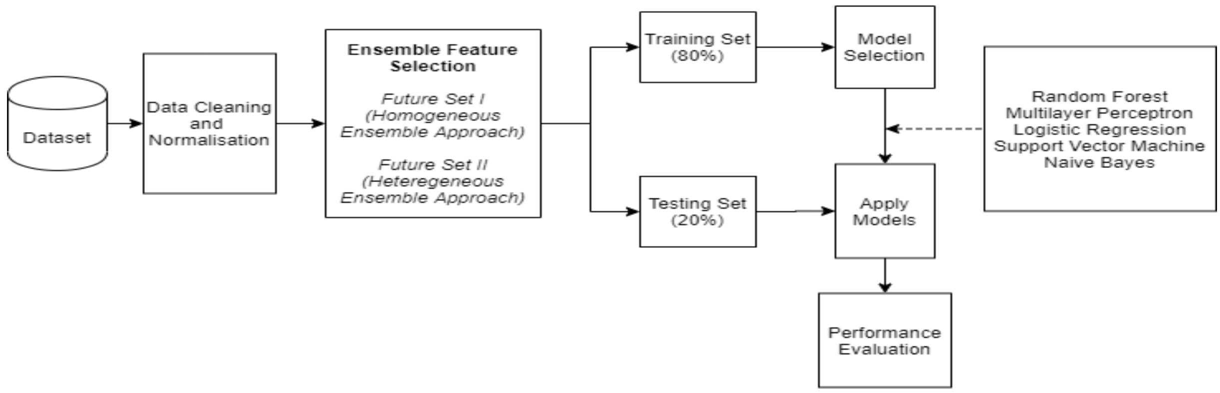

Figure 1. Framework of AD prediction

2.1. Dataset

Dataset was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). It was established in 2003 and its main goal was to measure the proceeding of early Alzheimer's disease (AD) and Mild Cognitive Impairment (MCI) of patients. Table 1 illustrates the details of the dataset used in this study. The baseline combined dataset consisted of 819 instances, 229 is CN, 402 MCI, and 188 AD. The sample size used in this study (819) is adequate to obtain reliable results. The third limitation mentioned in Section 1 is overcome using a large sample size.

Table 1. Dataset used in this study

Data Set (Baseline combined data)			Number of Instances	Number of Attributes		Number of Classes
CN	MCI	AD		Raw Dataset	After Cleaning Process	
229	402	188	819	113	35	3

The raw dataset consists of 113 attributes which include both 57 attributes are double recorded and 21 attributes include more than 50% missing value. The 78 (57+21) attributes have been considered as unnecessary data in the current study. On the other hand, the rest of the 35 attributes have been used for the objectives of this study. More information about whether the attributes are continuous or discrete is presented in Table 3.

2.2. Data cleaning and normalization

During the process of data cleaning, 57 unnecessary data, garbage, and noise values has been removed from the dataset as highlighted in Section 3.1. Additionally, 21 attributes including more than 50% missing value were removed from the dataset. Thus, the second limitation mentioned in Section 1 is eliminated without using redundant missing values. The main motivation for removing these attributes was to obtain correct classification and unbiased results. After applying this process, the rest of the data (35 attributes) consists of enough quality content to meet the demands of this study. In other words, rich attributes have been used in this study, and so the first limitation mentioned in Section 1 section is eliminated. On the other hand, if the attributes include more than 50% missing value, mean imputation method, where the missing observations are replaced by the mean of the available cases, is applied. The data were normalized using z-score normalization which means that values of attributes ranged from 0 – 1 (Little and Rubin, 2019).

2.3. Ensemble feature selection

Ensemble feature selection approach advocates the output combination of various feature selection methods to obtain the most appropriate feature set; it enables researchers to get better results than the output which has been generated from any single model. As explained in the introduction, this study uses the homogeneous and heterogeneous approaches applying the filter-based methods. The methods order all features using feature selection methods and then rank them based on their importance. However, threshold values should be particularly used before the ranked features are combined. Note that highly ranked features are used in models enabling significant results.

Two different models have been created based on the homogeneous and heterogeneous ensemble approaches. Most popular feature selection methods were chosen in filter-based methods. In this study, the total number of used feature selection methods for both the models are four, namely Gain Ratio, ReliefF, Chi-Square, and Fast Correlation Based Filter (FCBC). In Gain Ratio, a feature value is evaluated measuring the gain ratio (Kumar and Vanaja, 2014). Also, it is a modified version of the information gain which aims to reduce its bias. A feature score is calculated for each feature to rank and select the highest features by ReliefF. Additionally, an instance from the dataset is randomly sampled and then its nearest neigbour is determined from the same and different class (Tang *et al.*, **2014**; Kumar and Vanaja, 2014). Chi-square is a key method for categorical features, and also Chi-square is calculated between each feature and the target to select the top features (Kumar and Vanaja, 2014). FCBC is used to find a set of principles features for the class conception (Aldehim, 2015).

Moreover, a small case study has been carried out on choosing feature selection methods. The detail of this case study is specified below.

- Model 1 (Homogeneous Ensemble): Training data is divided into four main parts based on the homogeneous ensemble approach, each of which has the same size. In a case study, various feature selection methods (Gain Ratio, ReliefF, Chi-Square, and FCBC) were applied to the dataset to obtain an optimal method. Chi-Square feature selection method was chosen for this model (Model 1) based on the case study result. That is, this model provided the best accuracy result using the attributes which are ranked from the Chi-Square. Then, from each part, the top 25% of the features were selected and combined. Note that different threshold values (10%, 25% and 50%) were applied to obtain the best feature set. The best result was achieved from the feature set created with the 25% threshold value.
- Model 2 (Heterogamous Ensemble): These four feature selection methods are used in this model. In the same vein, the reason for choosing the methods is that this model provided the best accuracy result using the attributes which are ranked from the Gain Ratio, ReliefF, Chi-Square, and FCBC, respectively. As in Model 1, the top 25% of the features were selected providing from each feature selection method and then they were combined. Then, Feature Set II was formed which is presented in Table 3.

To conclude, models 1 and 2 have been built based on homogeneous and heterogeneous approaches. The threshold value is set as 25% in these models which means the top 25% of the most applicable features are selected. The selected features are shown in Table 3.

2.4. Determining the size of training and testing set

Dataset is divided depending on the holdout method into a training and testing set (Bookheimer *et al.*, 2000). In the literature, many researchers agreed that 20% of the dataset (as a testing set) is enough to reach reliable results (Huang *et al.*, 2017; Seijo-Pardo *et al.*, 2017). Additionally, a small case study was carried out by taking different sizes data set such as 60%/40%, 70%/30%, and 80%/20% applying RF and SVM methods. As a result, the size of training and testing set to obtain the best result was to 80%/20%. Therefore, the prediction model has been built by taking these sizes.

2.5. Model selection and performance evaluation

To date, various data mining methods have been used to measure early AD detection. In this study, a comparative analysis has been done with well-known methods using Orange Data Mining platform (https://orange.biolab.si/). Naïve Bayes, SVM, Multilayer Perceptron (MLP), Random Forest, and Logistic Regression are one of the most widely used methods on early AD detection (Khan and Usman, 2019; Seijo-Pardo *et al.*, 2017; Bansal *et al.*, 2018; Farid *et al.*, 2020). The rest of this section provides brief information on the chosen methods (Vapnik, 1995; Patel *et al.*, 2015; Balakrishnan and Puthusserypady, 2005; Patil and Shimpi, 2011).

- Naïve Bayesian classifier: This classifier is used for probabilistic learning in machine learning. It provides high accuracy results when features are independent (Patel et al., 2015).
- Support Vector Machines (SVM): Its main purpose is to simplify the data and provide more understandable information to the user. Support Vector Machines are used to separate the data of the two classes in the most appropriate way (Patel *et al.*, 2015).
- Random Forest (RF): The main process with this algorithm is that it creates a great number of decision trees based on the random selection of data and variables. Individually created decision trees compose decision forest. Results are obtained during the creation of decision forests that are combined for the latest estimates (Patel *et al.*, 2015).
- Logistic Regression: It has been used to classify data based on historical data. This regression aims to determine a relationship between historical and output data and finally makes a prediction of the probabilities of events (Muralidharan *et al.*, 2018).
- Multilayer Perceptron: MLP consists of one input layer, one or more hidden layers, and one output layer. The information in the input layer is transmitted to the output of the network by processing each cell individually (Patel *et al.*, 2015; Muralidharan *et al.*, 2018).

The results are evaluated with unseen data (20% - testing set). The evaluation metrics in this study are Accuracy, Sensitivity, Specificity, and Area under the curve (AUC). Moreover, this study evaluates the threshold factor for the prediction of AD on the models. This study also compares the current models and the proposed model in terms of Accuracy, Sensitivity, and Specificity. Table 2 provides information on the terms used to define sensitivity, specificity, and accuracy. Accuracy, sensitivity, and specificity are described in terms of TP, TN, FN, and FP as follows.

- Accuracy = (TN + TP)/(TN+TP+FN+FP)
- Sensitivity = TP/(TP + FN)
- Specificity = TN/(TN + FP)

Table 2. Terms used to define sensitivity, specificity, and accuracy

Outcome of the diagnostic test	Positive	Negative	Row Total
Positive	TP	FP	TP+FP (Total number of subjects with a positive test)
Negative	FN	TN	FN + TN (Total number of subjects with negative test)
Column total	TP+FN (Total number of subjects with the given condition)	FP+TN (Total number of subjects without given condition)	N = TP+TN+FP+FN (Total number of subjects in study)

3. RESULTS AND DISCUSSION

3.1. Obtaining of feature sets

Table 3 presents the results obtained from the homogeneous and heterogeneous ensemble feature selection approaches. As was mentioned in Section 3.3, Gain Ratio, ReliefF, Chi-Square, and FCBC are feature selection methods in the ranking of features with a 25% threshold value. While Chi-Square has been used on both homogeneous and heterogeneous approaches, ReliefF, Gain Ratio, and FCBC have been used on only a heterogeneous approach. In Table 3, continuous and discrete attributes are represented by 'C' and 'D' signs respectively.

Table 3. Feature subsets based on the ensemble approaches

Ensemble Approach	Ranker Feature (Selection Method)	Threshold Value	Ranking Combination (Ranking of Features)
Homogenous	Chi-Square	25%	**CDRSB (D), ADAS11 (C), ADAS13 (C), MMSE (D)**, **RAVLT perc. Forgetting (C)**, FAQ (C), AGE (C), SITE (D), PTGENDER (D)
Heterogeneous	ReliffF Gain Ratio Chi-Square FCBF	25%	**CDRSB (D), MMSE (D)**, APOE4 (D), **ADAS13 (C),** RAVLT forgetting (D), **ADAS11 (C)**, RAVLT immediate (C), EXAM DATE (C), PTMARRY (D), **RAVLT perc. Forgetting (C)**

Abbreviations: CDRSB, Clinical Dementia Rating Sum of Boxes; ADAS, Alzheimer's disease assessment scale; MMSE, Mini-Mental Scale Examination; RAVLT, Rey Auditory Verbal Learning Test; FAQ, Functional Assessment Questionnaire; PTGENDER, participant's gender; APOE4, APOE e4 allele; PTMARRY, participant's marital status.

Before applying these ensemble feature selection approaches to the ADNI dataset (see Table 1 for detail), there were 35 features are available after the data cleaning process. Nine and ten features were selected after applying homogeneous and heterogeneous approaches respectively. Also, note that five features were mutual for both approaches which are signalled by the words in bold in Table 3.

3.2. Comparison of classification accuracy

For this study, the RF, ANN, LR, SVM, and NB algorithms were used to classify ADNI dataset.

Table 4 provides information on the measurement of these algorithms. The used feature selection methods have been highlighted in the previous section. Firstly, classification accuracy has been measured without feature selection. Then, the accuracy has been measured again with feature subsets I and II, respectively. In other words, the classification algorithms are compared by non-applied and applied (homogeneous and heterogeneous ensemble) feature selection approaches. The classification accuracy was performed according to the sizes of training (80%) and testing (20%) set. From Table 4 we can see that the highest classification accuracy result is obtained applying the Random Forest algorithm with feature set II (heterogeneous ensemble feature selection approach). In addition to this, Table 5 depicts the evaluation measure for this study in terms of Precision, Sensitivity, Specificity, and Area under the curve (AUC). This study provides information on AUC values (see Table 5) of algorithms in opposition to current studies presented in Section 1. Thus, the fourth limitation mentioned in Section 1 is overcome.

Table 4. Comparison of classification accuracy based on ensemble approaches

Classification Algorithm	Classification Accuracy (%)	Accuracy after Feature Selection (%)	
		Feature Set I (Chi-Square)	Feature Set II (RelifF, Gain Ratio, Chi-Square)
Random Forest	0.83	**0.88**	**0.91**
Multilayer Perceptron	0.71	0.82	0.87
Logistic Regression	0.61	0.78	0.81
SVM	0.74	0.85	0.86
Naïve Bayes	0.70	0.82	0.83

Table 5. Evaluation of applied ensemble approaches for Alzheimer dataset

Ensemble Approach	Classification Algorithm	Precision (%)	Sensitivity (%)	Specificity (%)	AUC (%)
Homogenous (Feature Set I)	RF	**0.88**	**0.88**	**0.90**	**0.91**
	MLP	0.83	0.82	0.82	0.83
	LR	0.78	0.79	0.79	0.81
	SVM	0.85	0.85	0.88	0.90
	NB	0.82	0.82	0.86	0.89
Heterogeneous (Feature Set II)	RF	**0.91**	**0.91**	**0.92**	**0.94**
	MLP	0.87	0.87	0.90	0.92
	LR	0.81	0.82	0.84	0.88
	SVM	0.87	0.87	0.90	0.92
	NB	0.83	0.84	0.89	0.91

From the data in Table 5, it is apparent that the Radom Forest algorithm (heterogeneous ensemble feature selection approach) performs better than the other algorithms on early AD prediction. These findings help us to highlight that a heterogeneous ensemble feature selection approach provides significantly better results than the homogeneous ensemble feature selection approach. This approach can be considered helpful for the people studying Alzheimer's Data in the early AD prediction.

3.3. Comparison of previous studies and the proposed model

Table 6 compares the results obtained from the previous studies and the proposed model for early detection of Alzheimer's disease, MCI diagnosis, and CN. The purpose of the studies is the same as our study. Zhang and Shen, (2011) applied the Support Vector Regression (SVR) algorithm to early AD detection and achieved 85% accuracy. However, they did not provide any information for sensitivity and specificity. On the other hand, the study of Khazaee *et al.* (2016) provided a higher accuracy value (87%), but the used data set was small (168 instances). Additionally, the used feature selection methods are presented in Table 6. The study of Wee *et al.,* (2013) presented the best result applying the SVM algorithm with Support Vector Machine Recursive Feature Elimination (SVM-RFE) on the classification accuracy (92%).

Table 6. Comparisons of the proposed model and previous studies

REF (Authors)	Target	Best Classifier	Data set Sample	ACC (%)	Sen (%)	Spe (%)	FSM
Zhang and Shen, 2011	CN, MCI, AD	SVR	42 CN, 99 MCI, 51 AD	85	-	-	MTFS
Ahmed *et al.,* 2015	CN, MCI, AD	SVM	162 CN, 210 MCI, 137 AD	84	79	88	-
Zhang *et al.,* 2015	CN, MCI, AD	SVM	97 CN, 57 MCI, 24 AD	81	-	-	PCA
Quintana *et al.,* 2012	CN, MCI, AD	ANN	346 CN, 79 MCI, 97 AD	66	-	-	-
Khazaee *et al.,* 2016	CN, MCI, AD	SVM	45 CN, 89 MCI, 34 AD	87	-	-	SFS
Suk *et al.,* **2015**	CN, MCI, AD	SVM	52 CN, 99MCI, 51 AD	55	-	-	LBFS
Lama *et al.,* 2017	CN, MCI, AD	SVM	70 CN, 74 MCI, 70 AD	77	62	79	PCA
Tong *et al.,* 2017	CN, MCI, AD	PCA+RELM	35 CN, 37 AD, 75 MCI	60	-	-	-
Cuingnet *et al.,* 2011	CN, MCI, AD	SVM	81 CN, 68 AD, 104 MCI	87	91	95	V_Std & V_Com
Teipel *et al.,* 2007	CN, MCI, AD	Logistic regression	18 CN, 32 AD, 24 MCI	83	88	78	PCA
Wee *et al.,* 2013	CN, MCI, AD	SVM	200 CN, 198 AD, 200 MCI	92	90	94	SVM-RFE
Proposed Model (Heterogeneous Feature Set II)	CN, MCI, AD	RF	229 CN, 402 MCI, 188 AD	91	91	92	HEFS

**ACC: Accuracy, Sen: Sensitivity, Spe: Specificity, FSM: Feature Selection Method, MTFS: Multi-Task Feature Selection, PCA: Principal Component Analysis, SFS: Sequential Forward Selection, Lasso-Based Feature Selection, V_Std: Voxel-STAND, V_Com: Voxel-COMPARE, SVM-RFE: Support Vector Machine Recursive Feature Elimination, HEFS: Heterogeneous Ensemble Feature Selection*

As shown in Table 6, the authors obtained the best results in their studies mostly using the SVM algorithm. In the current study, the proposed early AD detection model obtained better results applying the RF algorithm with a heterogeneous ensemble feature selection approach. We believe that a heterogeneous ensemble feature selection approach plays an important role in obtaining the highest measurement results.

4. CONCLUSIONS

The use of data mining methods has increased to provide prediction models in recent years in various areas such as health, education, and real estate, etc. Datasets may consist of many features, but it may also contain unnecessary data, garbage, and noise values. Features can be chosen using significant feature selection methods to build a well predictive model. The reason for this is that all feature selection methods cannot take into account important features to enhance the predicting process. Note that significant features are used to build well predictive models that help healthcare professionals to treat patients.

A reasonable approach to tackle this issue could be to describe different feature selection approaches. In this case, a predictive model has been built based on two different ensemble feature selection approaches, namely homogeneous and heterogeneous. Features were ranked using various feature ranking methods and then the ranked different features were combined to obtain different feature subsets applying a threshold value (25%). In this case, two different feature subsets were obtained based on the homogeneous and heterogeneous approaches. Five different data mining methods applied to the feature subsets and then their performances were compared to reveal which approach is better on the early AD detection. This study revealed that the best classification accuracy is obtained by applying the Random Forest algorithm with feature set generating from the heterogeneous approach. Overall, these results indicate that ensemble feature selection approaches enable them to obtain significant performance comparison.

Our future research will be about the combination of different data mining methods to improve classification accuracy applying the heterogeneous approaches.

REFERENCES

Ahmed O.B., Mizotin, M., Benois-Pineau, J., Allard, M., Catheline. G., Amar, C. B., Alzheimer's Disease Neuroimaging Initiative., 2015, "Alzheimer's disease diagnosis on structural MR images using circular harmonic functions descriptors on the hippocampus and posterior cingulate cortex", *Computerized Medical Imaging and Graphics*, Vol. 44, pp. 13-25.

Aldehim, G., 2015, *Heuristic ensembles of filters for accurate and reliable feature selection*, Doctoral dissertation, University of East Anglia.

Balakrishnan, D., Puthusserypady, S., “Multilayer perceptrons for the classification of brain-computer interface data”, *In Proceedings of the IEEE 31st Annual Northeast Bioengineering Conference*, 118-119, 2005.

Bansal, D., Chhikara, R., Khanna, K., Gupta, P., 2018, “Comparative analysis of various machine learning algorithms for detecting dementia”, *Procedia computer science*, 132, 1497-1502.

Bhagyashree, S.R., Nagaraj, K., Prince, M., Fall, C.H., Krishna, M., 2018, “Diagnosis of Dementia by Machine learning methods in Epidemiological studies: a pilot exploratory study from south India”, *Social Psychiatry and Psychiatric Epidemiology*, Vol. 53, No. 1, pp.77-86.

Bhagyashree S.R., Sheshadri H.S., "An initial investigation in the diagnosis of Alzheimer’s disease using various classification techniques", *IEEE International Conference on Computational Intelligence and Computing Research*, 2014.

Bookheimer, S.Y., Strojwas, M.H., Cohen, M.S., Saunders, A.M., Pericak-Vance, M.A., Mazziotta, J.C., Small, G. W., 2000, "Patterns of brain activation in people at risk of Alzheimer's disease", *New England journal of medicine,* Vol. 343, No. 7, pp.450-456.

Campos, S., Pizarro, L., Valle, C., Gray, K.R., Rueckert, D., Allende, H., “Evaluating Imputation Techniques for Missing Data in ADNI: A Patient Classification Study”, *In Iberoamerican Congress on Pattern Recognition*, 3-10, November 2015.

Chaves, R., Ramírez, J., Gorriz, J. M., 2013, "Integrating discretization and association rule-based classification for Alzheimer’s disease diagnosis", *Expert Systems with Applications*, Vol. 40, No. 5, pp.1571-1578.

Chen, R., Herskovits, E.H, 2010, “Machine-learning techniques for building a diagnostic model for very mild dementia”, *Neuroimage*, Vol. 52, No. 1, pp.234-244.

Cuingnet, R., Gerardin, E., Tessieras J., Auzias, G., Lehéricy, S., Habert, M.O., Alzheimer's Disease Neuroimaging Initiative, 2011, “Automatic classification of patients with Alzheimer’s disease from structural MRI: a comparison of ten methods using the ADNI database,” *NeuroImage*, Vol. 56, No. 2, pp.766-781.

Dallora, A.L., Eivazzadeh, S., Mendes, E., Berglund, J., Anderberg, P., 2017, “Machine learning and microsimulation techniques on the prognosis of dementia: A systematic literature review”, *PloS* *one*, Vol. 12, No. 6.

Escudero, J., Ifeachor, E., Zajicek, J.P., Green, C., Shearer, J., Pearson, S, 2013, "Machine learning-based method for personalized and cost-effective detection of Alzheimer's disease", *IEEE transactions on biomedical engineering*, Vol. 60, No. 1, pp.164-168.

Farhan, S., Fahiem, M.A., Tauseef, H., 2014, "An Ensemble-of-Classifiers Based Approach for Early Diagnosis of Alzheimer's Disease: Classification Using Structural Features of Brain Images", *Computational and mathematical methods in medicine*.

Farid, A.A., Selim, G., Khater, H., 2020, "Applying Artificial Intelligence Techniques to Improve Clinical Diagnosis of Alzheimer's disease".

Hand, D.J., 2007, "Principles of data mining", *Drug Safety*, 30(7), 621-622. Vol. 30, No. 7, pp.621-622.

Huang, M., Yang, W., Feng, Q., & Chen, W., 2017, "Longitudinal measurement and hierarchical classification framework for the prediction of Alzheimer's disease". *Scientific reports*, Vol. 7, No. 1, pp. 1-13.

Jo, T., Nho, K., Saykin, A.J., 2019, "Deep Learning in Alzheimer's disease: Diagnostic Classification and Prognostic Prediction using Neuroimaging Data", *Frontiers in aging neuroscience*, Vol. 11, pp.272.

Khan, A., Usman, M., 2019, "Alzheimer's Disease Prediction Model Using Demographics and Categorical Data", *International Journal of Online and Biomedical Engineering (iJOE),* Vol. 15, No. 15, pp.96-109.

Khazaee, A., Ebrahimzadeh, A., Babajani-Feremi, A., 2016, "Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease", *Brain imaging and behavior*, Vol. 10, No. 3, pp.799-817.

Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R. I., Rohrer, J.D., Frackowiak, R.S., 2008, "Automatic classification of MR scans in Alzheimer's disease", *Brain*, Vol. 131, No. 3, pp.681-689.

Kumar, K.R., Vanaja, S., 2014, "Analysis of feature selection algorithms on classification: a survey".

Lama, R.K., Gwak, J., Park, J.S., Lee, S.W., 2017, "Diagnosis of Alzheimer's disease based on structural MRI images using a regularized extreme learning machine and PCA features", *Journal of healthcare engineering*.

Lee, G., Nho, K., Kang, B., Sohn, K.A., Kim, D, 2019, "Predicting Alzheimer's disease progression using multi-modal deep learning approach", *Scientific reports*, Vol. 9, No. 1.

Little, R.J., Rubin, D.B., 2019, "Statistical analysis with missing data ", *John Wiley & Sons*, Vol. 793.

Liu, S., Liu, S., Cai, W., Pujol, S., Kikinis, R., Feng, D., "Early diagnosis of Alzheimer's disease with deep learning", *IEEE 11th international symposium on biomedical imaging (ISBI)*, 1015-1018, 2014.

Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., de Mendonça, A., 2011, "Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests", *BMC research notes*, Vol. 4, No. 1, pp.299-

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., 2015, "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects", *Neuroimage*, Vol. 104, pp. 398-412.

Munteanu, C.R., Fernandez-Lozano, C., Abad, V.M., Fernández, S.P., Álvarez-Linera, J., Hernández-Tamames, J.A., Pazos, A., 2015, "Classification of mild cognitive impairment and Alzheimer's Disease with machine-learning techniques using 1 H Magnetic Resonance Spectroscopy data", *Expert Systems with Applications*, Vol. 42, No. 15, pp.6205-6214.

Muralidharan, S., Phiri, K., Sinha, S.K., Kim, B., 2018, "ANALYSIS AND PREDICTION OF REAL ESTATE PRICES: A CASE OF THE BOSTON HOUSING MARKET"i, *Issues in Information Systems*, Vol. 19, No. 2, pp. 109-118.

Nunes, C., Silva, D., Guerreiro, M., de Mendonça, A., Carvalho, A. M., Madeira, S. C., "Class Imbalance in the Prediction of Dementia from Neuropsychological Data", *In Portuguese Conference on Artificial Intelligence, Springer Berlin Heidelberg*., 138-151, 2013.

Patel, J., Shah, S., Thakkar, P., Kotecha, K., 2015, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques", *Expert Systems with Applications*, Vol. 42, No. 1, pp. 259–268.

Patil, V., Shimpi, S., 2011, "Handwritten English character recognition using neural network", *Elixir Comput Sci Eng*, Vol. 41, pp. 5587-5591.

Quintana, M., Guàrdia, J., Sánchez-Benavides, G., Aguilar, M., Molinuevo, J.L., Robles, A., Fernández, M., 2012, "Using artificial neural networks in clinical neuropsychology: High performance in mild cognitive impairment and Alzheimer's disease", *Journal of Clinical and Experimental Neuropsychology*, Vol. 34, No. 2, pp. 195–208.

Sana, B., Siddiqui, I.F., Arain, Q. A., 2019, "Analyzing Students' Academic Performance through Educational Data Mining", *3c Tecnología: glosas de innovación aplicadas a la pyme*, Vol. 8, No. 29, pp. 402-421.

Seijo-Pardo, B., Porto-Díaz, I., Bolón-Canedo, V., Alonso-Betanzos, A, 2017, "Ensemble feature selection: homogeneous and heterogeneous approaches", *Knowledge-Based Systems*, Vol. 118, pp. 124-139.

Shankle, W.R., Mani, S., Pazzani, M.J., Smyth, P., "Detecting very early stages of dementia from normal aging with machine learning methods", *In Conference on Artificial Intelligence in Medicine in Europe Springer, Berlin, Heidelberg*, 71-85, 1997.

Stamps, J.J., Bartoshuk, L.M., Heilman, K.M., 2013, "A brief olfactory test for Alzheimer's disease", *Journal of the neurological sciences,* Vol. 333, No. 1-2, pp. 19-24.

Suk, H.I., Lee, S.W., Shen, D, 2015, "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis.", *Brain Structure and Function*, Vol. 220, No. 2, pp. 841–859.

Supekar, K, Menon, V, Rubin, D, Musen, M, Greicius M.D, 2008, "Network analysis of intrinsic functional brain connectivity in Alzheimer's disease", *PLoSComputBiol*, Vol. 4, No. 6, pp. 1-11.

Tang, J., Alelyani, S., Liu, H., 2014, "Feature selection for classification: A review.", *Data classification: Algorithms and applications*, 37.

Teipel, S.J., Born, C., Ewers M. 2007, "Multivariate deformation based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment", *NeuroImage*, Vol. 38, No. 1, pp. 13–24.

Tong, T., Gray, K., Gao, Q., Chen, L., Rueckert, D., 2017, "Multi-modal classification of Alzheimer's disease using nonlinear graph fusion", *Pattern recognition*, Vol. 63, pp. 171-181.

Trambaiolli, L.R., Spolaôr, N., Lorena, A.C., Anghinah, R., Sato, J.R., 2019, "Feature selection before EEG classification supports the diagnosis of Alzheimer's disease", Clinical Neurophysiology, 128(10), 2058-2067. Vol. 128, No. 10, pp. 2058-2067.

Vapnik, V.N., 1995, "The Nature of Statistical Learning", *Theory*.

Wee, C.Y., Yap, P.T., Shen, D., 2013, "Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns", *Human Brain Mapping*, Vol. 34, No. 12, pp. 3411–3425.

Westman, E., Muehlboeck, J.S., Simmons, A, 2012, "Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion.", *Neuroimage*, Vol. 62, No. 1, pp. 229-238.

Williams, J.A., Alyssa W., Diane J.C, Maureen S., "Machine learning techniques for diagnostic differentiation of mild cognitive impairment and dementia.", *In Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence*, pp.71-76, 2013.

Wordoffa, H., Wangoria, E., 2012, "Alzheimer's Disease Stage Prediction using Machine Learning and Multi-Agent System".

Zhang, R., Simon, G., Yu, F, 2017, "Advancing Alzheimer's research: A review of big data promises", International Journal of medical informatics, Vol. 106, pp. 48-56.

Zhang D, Shen D, 2011, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease", *NeuroImage*, Vol. 59, No. 2, pp. 895–907.

Zhang, Y., Wang, S., Phillips, P., Dong, Z., Ji, G., Yang, J., 2015, "Detection of Alzheimer's disease and mild cognitive impairment based on structural volumetric MR images using 3D-DWT and WTA-KSVM trained by PSOTVAC", *Biomedical Signal Processing and Control*, Vol. 21, pp. 58–73.

Zhao, Y., He, L., "Deep learning in the EEG diagnosis of Alzheimer's disease. In Asian Conference on Computer Vision", Springer International Publishing, pp. 340-353, November 2014.