

Improvement of Alzheimer disease diagnosis accuracy using ensemble methods

Mohammed Abdullah Al-Hagery¹, Ebtehal Ibrahim Al-Fairouz², Norah Ahmed Al-Humaidan³

^{1,2,3} Computer Science Department, College of Computer, Qassim University, Saudi Arabia

Article Info

Article history:

Received May 15, 2019

Revised Feb 18, 2020

Accepted Mar 13, 2020

Keywords:

Data Mining

Alzheimer Disease

Ensemble Methods

Classification Accuracy

Random Forest

ABSTRACT

Nowadays, there is a significant increase in the medical data that we should take advantage of it. The application of the machine learning via the data mining processes, such as data classification, depends on using a single classification algorithm or those combined such as ensemble models. The objective of this work is to improve the classification accuracy of previous results for Alzheimer disease diagnosing. The Decision Tree algorithm was combined with three types of ensemble methods, which are Boosting, Bagging and Stacking. The clinical dataset from the Open Access Series of Imaging Studies (OASIS) was used in the experiments. The experimental results of the proposed approach were better than the previous work results. Where the Random Forest (Bagging) achieved the highest accuracy among all algorithms with 96.66%, while the lowest result was Decision Tree with 73.33%, all these results generated in this paper are higher in accuracy than that done before.

Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Mohammed Abdullah Al-Hagery,

Department of Computer Science,

College of Computer, Qassim University, Saudi Arabia.

Email: hajry@qu.edu.sa, drmalhagery@gmail.com

1. INTRODUCTION

The importance of this paper comes from the accuracy of the results achieved in a very important subject related to human health, which is associated with the methods of diagnosing of one of the most serious diseases, Alzheimer's disease. From the beginning of the last two decades to present, Electroencephalography is using as a promising technique for early screening and to help to diagnose Alzheimer's disease. For this reason, several EEG-based classification algorithms have been developed by [1], [2], [3], [4]. Alzheimer is a disease discovered by specific symptoms, such as the decline in memory, weak focusing, and the patients cannot do daily activities [5]. This disease is known as the most common causes of dementia in elderly people [6]. About 75% of Dementia cases are Alzheimer's Disease patients [7]. Alzheimer's disease is a disorder that affects memory functions first, then slowly affect cognitive functions with behavioural deteriorations and cause death. It can be diagnosed by an accurate clinical examination, meeting the patient and relatives and doing a comprehensive interview with them, and neuropsychological assessment [8].

The research problem comes from the weakness found in the accuracy of current diagnosis methods, which are used for Alzheimer's disease classification and diagnosis. Where, the development of tools or techniques is very necessary because the medical field is data-rich but knowledge weak, because the analysis tools in this area, that used to recognize trends and relationships, are not powerful enough. Although there is a wealth of data possible inside the medical systems [9].

The objective of this paper is to improve and increase the accuracy of Alzheimer disease diagnosis using the proposed method to overcome the weakness of that proposed by [10]. Where instead of using a decision tree, logistic regression and discriminant analysis algorithms. We used a combination of algorithms as a multilevel of processing, which includes the Decision Tree algorithm with three types of ensemble methods, include; boosting, bagging and Stacking [11]. Although, there are various research works [1], [2], [3], [4] are focusing on Alzheimer disease diagnosis field, but they are using different datasets.

Currently, a number of methods have been developed for pattern extraction from small and big datasets, to benefit from these data. A major area of development is called knowledge discovery in databases. It includes a variety of statistical analysis, machine learning techniques and pattern recognition. In principle, knowledge discovery is a process by which the steps of understanding the domain, data cleaning, integration, extraction of knowledge from patterns, and knowledge post-processing are utilizing to exploit the knowledge in decision-making. The step of using pattern extraction by various methods from data is commonly referred to as data mining or knowledge discovery in many areas, such as the medical field, including diagnosis, predictions and treatment [12], [13]. Data mining involves many methods used for extracting hidden patterns and relationships from huge datasets. These methods include machine learning, statistical analysis and database technology. Data mining works on two types, which are supervised and unsupervised learning algorithms. In supervised learning, a known data set is used to train a model, but unsupervised learning does not use a training set. There are two common modelling purposes of supervised data mining techniques. The first includes classification models, which predict discrete labels, the second is covering prediction models, that predict continuous-valued functions [14]. Advantages of data mining appear in processing big data, finding information and discovering association rules [15]. In this paper, we concentrated on the improvement of the diagnostic accuracy of Alzheimer disease using ensemble methods, which combined various classifiers to satisfy a higher accuracy of results than what was achieved previously.

The rest of the paper divided into four sections: Section 2 discusses related work. Section 3 describes the methodology and section 4 demonstrates the experimental results. Section 5 explains results comparisons and discussion, and Section 6 contains the conclusion, finally, section 7 provides the future work.

2. RELATED WORK

Data mining has been used in many areas of medicine, such as diagnosis, prognosis, and treatment. There are many research works concentrating on this part, where one of the important researches contributions in the field of Alzheimer's disease diagnosis was carried out to obtain a high-quality of results in Morocco, Benyoussef et al. [10], used clinical data instead of Magnetic Resonance Imaging (MRI) since it is rare in Morocco because of the high price. They offered a new approach for classifying Alzheimer's disease using three models, but all these three models gave low accuracy results. In addition, Moreira and Namen [16] presented an approach that diagnoses people who are clinically suspected of having dementia. They used a set of 19 attributes but also created a new one by using text mining on the patient's history information. Then by applying different classification algorithms, they got a model of Alzheimer's disease also a model of mild cognitive impairment predictive. Finally, they applied ensemble methods on two datasets; first on the original dataset, the other is the dataset with the new attribute (hybrid model). The results showed that the hybrid model got a high accuracy as compared to the other models. On the other hand, Ramírez et al.[8], proposed an automatic Computer-aided Diagnosis System (CADS) to improve early diagnosis of the Alzheimer Disease. Their suggested approach was built on the SVM classification and image parameter selection. They found the sagittal correlation parameters and coronal standard deviation are the most effective ways of improving the accuracy of diagnosis and reduction of the dimensionality of the input size. In addition, they found the CADS give the accuracy of 90.38% for the early diagnosis of the Alzheimer Disease.

Furthermore, Chen et al. [5], used logistic regression to clarify the relationship between dementia and other diseases through medications that have been consistently prescribed to treat patients with dementia. Their findings included many diseases associated with dementia, such as indigestion in females and other stomach disorders. The association between dementia and pneumonia was mainly attributable to patients 65 years or older. Ateeq et al. [17] offered an efficient method for detected of Cerebral Microbleeds (CMB) in Susceptibility Weighted Imaging (SWI) scans. Their suggested technique included using ensemble methods, SVM and Quadratic Discriminant Analysis (QDA). Their results showed that QDA obtained better results of the sensitivity of 93.7% with 5.3 false positives (FP) per CMB and 56 FP per patient.

Daghistani and Alshammari [18] used three algorithms; Self-Organizing Map, Decision tree (C4.5), and Random Forest on real health care datasets to predict diabetic patients. They used a dataset of the adult population from the Ministry of National Guard Health Affairs (MNGHA), Saudi Arabia with 18 attributes to predict diabetic patients. Compared to other algorithms, the Random Forest gave the best results. On the other hand, Perveen et al. [19] worked on classify diabetic patients by constructed good models with higher

performance, across three groups of ages in the Canadian population. They used the AdaBoost, Decision tree (J48) algorithm and bagging. Their results showed AdaBoost ensemble method outperforms bagging and J48. Aljumah et al. [20] presented an approach to predicate diabetic treatment. They used a regression-based data mining algorithm which is the SVM algorithm and divided data into two groups: young and old. The young age group is predicted to have a treatment in this order: controlling diet, controlling weight, drug, exercise, stop smoking, and finally, insulin. The objective of Dangare and Apte's study [21] is to increase the accuracy of heart disease prediction; therefore, they added two additional attributes, i.e. obesity and smoking, to get more accurate results. Neural Networks, Naive Bayes, and Decision Trees are applied to heart disease data set, and they compared the accuracy performance of these algorithms. Their results showed that Neural Networks got more accurate results than others did.

Nadia et al. [1], proposed a machine deep learning approach as a data-driven for differentiating subjects with Alzheimer's Disease and Mild Cognitive Impairment and Healthy Control, by analyzing noninvasive scalp EEG recordings. Similarly, but for another disease where Gomathy and Banu [22], proposed an approach to predict heart diseases. They apply three algorithms: K-mean Maximal Frequent Itemset Algorithm (MAFIA) alone, K-mean based MAFIA with ID3 and K-mean based MAFIA with ID3 and C4.5. The last one was the best algorithm with 94% accuracy. Palaniappan and Awang [23] presented an approach called Intelligent Heart Disease Prediction System (IHDPS) and used Naïve Bayes, Neural Network and Decision Trees. Answer complex "what if" queries cannot be done in traditional Decision support systems, but IHDPS can. It can predict the likelihood of patients who have heart disease by using previous information from medical profiles. It also enables establishing significant knowledge, e.g., patterns, relationships between medical factors related to heart disease. It is Web-based, scalable, user-friendly, expandable and reliable. Kim et al. [24], proposed a model to predict cardiovascular disease diagnosis, they process images and measure the thickness of carotid intima-media to extract a vector feature then invent the multiple feature vectors. They applied several machine learning algorithms, and the best results were 89.51% and 89.46% by SVM and classification based on Multiple Association Rules (CMAR) respectively.

Chauraisa and Pal [25], used three algorithms; Iterative Dichotomized (ID3), Classification and Regression Tree (CART) and Decision Table (DT) to develop heart disease, prediction models. Furthermore, they also used 10-fold cross-validation methods. When comparing the results of all the algorithms, it showed that CART had the best performance. In addition, Abdullah et al. [26] identified the best variants among Decision tree algorithm, e.g. (C4.5, C5.0) and Weighted Decision Tree (WDT). Their results showed that the C4.5 algorithm had the best accuracy compared to C5.0 and WDT.

Delen et al. [12] used two data mining algorithms to predict breast cancer; artificial neural networks and Decision trees. They used logistic regression with a database contain 200,000 cases, for performance comparison purposes. They used 10-fold cross-validation methods. Their results pointed out that the Decision tree (C5.0) is the best predictor, then the artificial neural networks are the second, and the worst of the three was logistic regression. Additionally, Yang and Chen [27] proposed a data mining method to diagnose lung cancer pathologic staging. They used the Decision tree method to obtain and classify clinical data attributes. To select the suitable rules for evaluation, they used two approaches, support-then-confidence-then-lift (SCL) and confidence-then-lift then-support (CLS). Both approaches provide satisfying results. Likewise, Li et al. [28], proposed a model for pancreas cancer diagnosis using Positron Emission Tomography/Computed Tomography (PET/CT) images. The model includes three parts; pancreas segmentation, extracting & selecting features and designing a classifier. They designed a model contains hybrid feedback, SVM and Random Forest. The model applied to a data set of 80 instances and got an accuracy of 96.47%. As well, Xiao et al. [3], presented a cancer prediction approach by using deep learning method with an ensemble of different machine learning algorithms, they applied it to three data sets of different types of cancer. The results show an increase of accuracy as compared to one algorithm or algorithm based on majority voting. Besides, Kiranmayee et al. [29], applied Hybrid data mining methods, to predict Brain Tumours, which contains clustering, classification and association techniques and they analyzed the results using some statistical techniques. The results showed that the performance of classifiers Nearest Neighbor with Generalization (NNge) and Random Forest performed better in classification and followed by the Logical Analysis of Data Tree when compared to other classifiers. In addition, the results of J48 classifier were weak when compared to the other classifiers. They improved these results on the Hybridization through clustering and associations. They concluded from the results that the rate of Males Mortality is more when compared to females' mortality rate.

In addition, Chen et al. [30], proposed a data mining-based method called MyPHI for personal health index (PHI) prediction. The MyPHI performed better than other classifiers, such as logistic regression, linear Support Vector Machine (SVM) and their class-weighted versions. The result showed that the MyPHI achieved 89.95%, while Zhang et al. [15], were used the Apriori algorithm. By improving this algorithm, they found it to be a great benefit to research and drill large databases associated with the diagnosis in the medical area. Based on this improvement, they found that the results could help others who are using the data mining

techniques in extracting, analyzing and processing data to prevent disease and choose best medical treatment in the area of medicine.

In addition, there are three studies implemented ensemble methods on other fields, for instance, predicting chronic kidney disease, using the ensemble [31], Graczyk et al.[32], compared three ensemble methods with six machine learning algorithms. The results showed that ensemble methods increase the quality of performance, but no particular algorithm that makes ensemble methods perform the best. On the other hand, Dietterich [33], compared the effect of three ensemble methods: bagging, boosting and Randomization on improving the C4.5 algorithm performance. Boosting had the best results when there is no noise or little noise in data while randomization was quite better than bagging. When added some noise, bagging was definitely the best. However, this paper is focusing on the accuracy improvement of Alzheimer disease diagnoses results discussed above in [10], based on ensemble models as a combination of various algorithms.

3. METHOD

The methodology developed to improve the previous results [10], where they applied the decision tree, logistic regression and discriminant analysis algorithms. In this approach, the Decision Tree algorithm with three types of ensemble methods was combined. These algorithms include; boosting, bagging and Stacking to improve results for Alzheimer's disease diagnosis results [10], and using the same dataset. Figure 1 shows the method chart. The proposed methodology includes the following steps:

1. Dataset: the dataset was collected from OASIS.
2. Pre-processing: it includes clean missing or inconsistent data, removes attributes that irrelevant to our study.
3. Classification: various algorithms were used to diagnose Alzheimer's disease, includes; Decision tree, Boosting, Stacking and Bagging. These algorithms will be explained with more details in section 3.3.

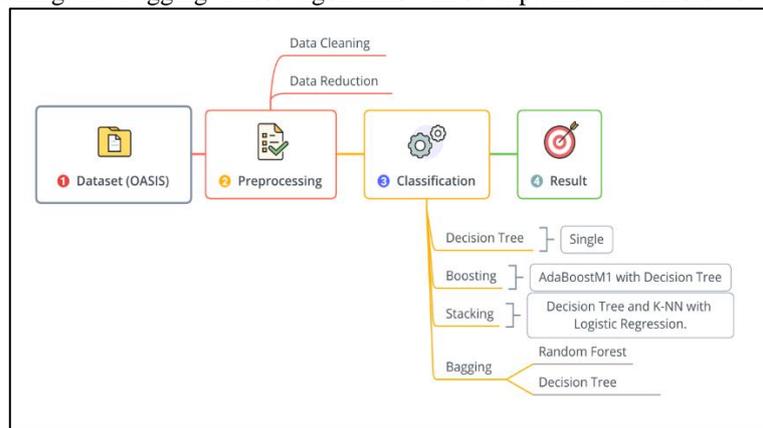


Figure 1. Proposed Methodology

3.1 Data Source

The clinical data sets of the OASIS used in this study, it contains cross-sectional data from 416 people at age 18 to 96 years, it has 100 people between very mild to moderate Alzheimer's disease who were diagnosed clinically and characterized by the Clinical Dementia Rating (CDR) scale [34]. Attributes and their measures are shown in Table 1.

Table 1. Measures of Attributes [35]

Attribute name	Measure
Age	Age at time of image acquisition (years)
Sex	Sex (male or female)
Education	Years of education
Socioeconomic status	Assessed by the Hollingshead Index of Social Position and classified into categories from 1 (highest status) to 5 (lowest status)
MMSE	score Ranges from 0 (worst) to 30 (best)
CDR	scale 0 = no dementia, 0.5 = very mild
AD	1 = mild AD, 2 = moderate AD
Atlas Scaling Factor (ASF)	A computed scaling factor that transforms the native-space brain and skull to the atlas target (i.e., the determinant of the transform matrix)
eTIV	Estimated total intracranial volume
nWBV	Expressed as the percentage of all voxels in the atlas-masked image that is labelled as grey or white matter by the automated tissue segmentation process

3.2 Data pre-processing

The data set includes 416 records, after eliminating the Null values it becomes 216, then two outliers removed, and the rest of the data set becomes 214 records. While selecting the interesting attributes, the value of ASF [10] eliminated and represented as in equation (1). Table 2 shows a sample of the dataset.

$$ASF = \frac{1755}{eTIV} \quad (1)$$

Table 2. Sample of the Dataset from OASIS [35]

ID	AGE	EDUC	SES	MMSE	CDR	ETIV	NWBV
OAS1_0001_MR1	74	2	3	29	0	1344	0.743
OAS1_0002_MR1	55	4	1	29	0	1147	0.81
OAS1_0003_MR1	73	4	3	27	0.5	1454	0.708
OAS1_0018_MR1	39	3	4	28	0	1636	0.813
OAS1_0019_MR1	89	5	1	30	0	1536	0.715
OAS1_0023_MR1	82	2	3	27	0.5	1420	0.71

3.3 Ensemble Models

Ensemble methods have multiple algorithms trained together for solving the same problem and to improve the results [36]. The objective of ensemble methods is to combine a group of classifiers that are varied and yet accurate, to achieve high accuracy of classification [37]. In this paper, we used three methods applied in [11], then we combined it with the Random Forest algorithm as a new approach.

1. *Decision Tree (Random forest)*: which is a Decision Tree collection, is a popular machine-learning algorithm. It is one of the bagging techniques. It can trace nodes values and therefore, we can understand what is happening inside the algorithm [38].
2. *Boosting*: is an ensemble method. In the beginning, a classifier applied to a training set. Then the second classifier applied to focus on the wrong instances that the first algorithm got. Then it adds more classifiers until it reaches the maximum number of models or accuracy. For this study, we used AdaBoostM1 as the first classifier; then the second classifier was the Decision Tree.
3. *Bagging*: is an ensemble method that split the training set into groups and creates a classifier for each group, each group of the training dataset is different. It takes the average or majority voting to combine the multiple classifier results. For this study, two bagging methods were used, which are the Decision Tree and Random Forest algorithm.
4. *Stacking*: is an ensemble of different algorithms are applied on the training set and one the Meta classifier going to take the results of all classifiers and make accurate results on invisible data. For this study, the Decision Tree, Random Forest and K-Nearest Neighbor applied on the training set with Logistic Regression as the Meta classifier.

4. EXPERIMENTAL RESULTS

The previously mentioned algorithms applied to the data set using Python Programming Language. The data set was divided into 86% training set, and 14% testing set same as [10]. Then the parameters were investigated to get the best values, and the experiments' results presented then compared with previous work results.

The best result we got was boosting algorithm with 90% at maximum depth 5 and random state 4. Then Stacking becomes second with 83.33% with random state 200. After that, bagging with Decision Tree result was 80% at maximum depth 3 and random state 200 while the lowest result was Decision Tree with 73.33% accuracy.

After getting the previous results, we tried to improve the accuracy more. Random Forest had chosen because it is a popular ensemble method.

The experiment result of the Random Forest was excellent, and it got 96.66% accuracy, with maximum depth 6, random state 63 and number of estimators 63 makes it the best among all the experiments. The results are shown in Table 3.

Table 3. Results of Ensemble Methods

Method Name	Description Of Algorithms	Accuracy
Decision Tree	Single	73.33%
Boosting	AdaBoostM1 with Decision Tree	90%
Stacking	Decision Tree and K-Nearest Neighbour with Logistic Regression.	83.33%
Bagging	Decision Trees	80%
Bagging	Random Forest	96.66%

5. RESULTS COMPARISONS AND DISCUSSION

Ensemble methods proved to be effective in improving the accuracy, and this improvement is shown in two algorithms, which are Boosting and Random Forest, they showed an increase in the accuracy as compared to Decision Tree alone.

Benyoussef et al. [10] used a decision tree, logistic regression and discriminant analysis algorithms and the accuracy of the test set were 60%, 59%, and 66% respectively.

Table 4 Results Comparison with the Previous Work

Results	Algorithm	Accuracy
Our results	Decision tree	73.33%
	Boosting	90%
	Bagging (Decision Trees)	80%
	Bagging (Random Forest)	96.66%
	Stacking	83.33%
Previous results [10]	Decision tree	60%
	Logistic regression	59%
	Discriminant analysis	66%

However, they did not mention about the parameters they used in their work. The experiments of our methods showed an improvement in the accuracy of the diagnosis, using the same dataset and the same splitting percentage for better comparison. The results are shown in Table 4. As well, Figure 2 demonstrates the previous results and Figure 3 shows our results.

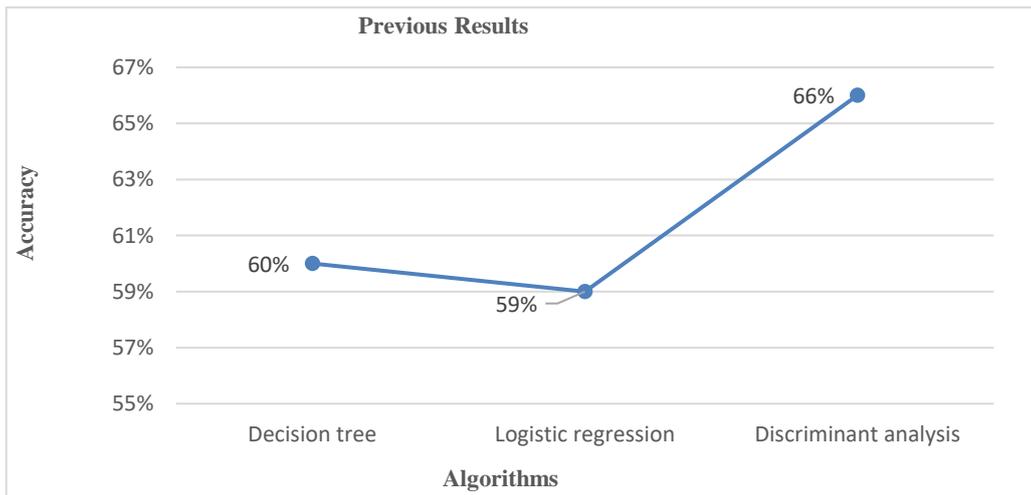


Figure 2. Previous Results Achieved by [10]

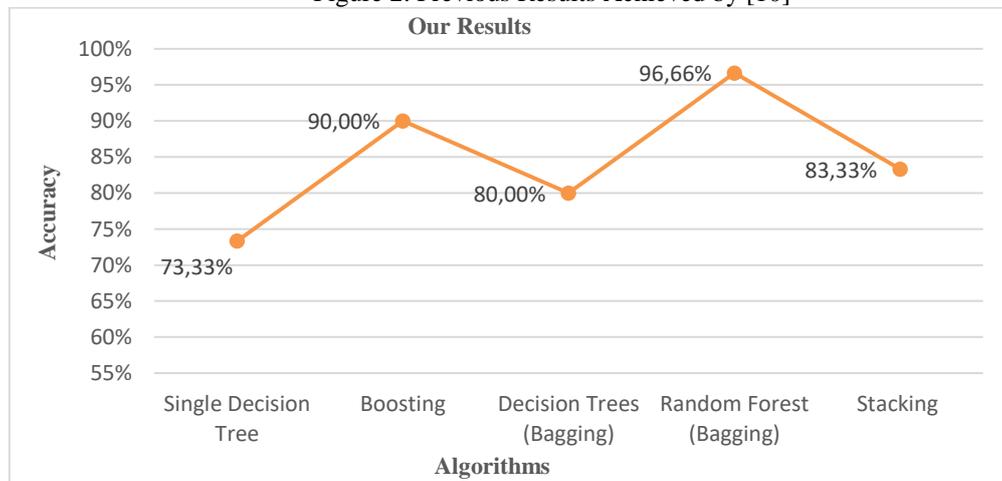


Figure 3. Results of Our Approach.

6. CONCLUSION

This paper used data mining techniques to diagnose Alzheimer disease. We used clinical datasets from the OASIS that include cross-sectional data. We combined the Decision Tree algorithm with three types of ensemble methods: boosting, bagging and Stacking. The algorithms implemented using Python, where parameters were investigated to get the best values, and the results were compared with previous work. The results of our experiments are better than the results in [10]. Boosting and bagging (Random Forest) had the best accuracy, with 90% and 96.66% respectively. These results proved the efficiency of ensemble methods.

The finding reflects the importance of this paper in regards to improving the classification accuracy of Alzheimer's disease diagnosis, this, in turn, will help doctors and patients to get highly accurate results, depends on the analysis of the accumulated data of patients according to the proposed approach.

7. FUTURE WORK

In future, a larger dataset size of this disease can be used. Furthermore, MRI dataset can be utilized to predict Alzheimer Disease since we used cross-sectional data in this study. The MRI is a powerful tool to diagnose Alzheimer Disease. In addition, other machine learning tools can be applied. The approach can be applied for the diagnosis of other diseases such as diabetes or heart disease or others. Based on the idea of this paper, researchers can create ensemble models for different purposes such as prediction or clustering processes. Moreover, this approach can be applied for classification to different types of datasets, such as the text data written in various languages, tweets contents, Emails messages, news, etc.

REFERENCES

- [1] C. Ieracitano, N. Mammone, A. Bramanti, A. Hussain, and F. C. Morabito, "A Convolutional Neural Network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings," *Neurocomputing*, vol. 323, pp. 96–107, 2019.
- [2] N. Zeng, Z. Wang, H. Zhang, and F. E. Alsaadi, "A Novel Switching Delayed PSO Algorithm for Estimating Unknown Parameters of Lateral Flow Immunoassay," *Cognit. Comput.*, vol. 8, no. 2, pp. 143–152, 2016.
- [3] Y. Xiao, J. Wu, Z. Lin, and X. Zhao, "A deep learning-based multi-model ensemble method for cancer prediction," *Comput. Methods Programs Biomed.*, vol. 153, pp. 1–9, 2018.
- [4] S. Gasparini *et al.*, "Information theoretic-based interpretation of a deep neural network approach in diagnosing psychogenic non-epileptic seizures," *Entropy*, vol. 20, no. 2, 2018.
- [5] P. H. Chen, F. Y. Yang, D. D. Lee, and M. H. Yang, "Data mining the comorbid associations between dementia and various kinds of illnesses using a medicine database," *Comput. Electr. Eng.*, vol. 70, no. September 2017, pp. 12–20, 2018.
- [6] *et. al.* Al-Naami, "Automated detection of Alzheimer disease using region growing technique and artificial neural network," *Waset.Org*, vol. 7, no. 5, pp. 204–208, 2013.
- [7] D. Bansal, R. Chhikara, K. Khanna, and P. Gupta, "Comparative Analysis of Various Machine Learning Algorithms for Detecting Dementia," *Procedia Comput. Sci.*, vol. 132, pp. 1497–1502, 2018.
- [8] J. Ramírez *et al.*, "Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features," *Inf. Sci. (Ny)*, vol. 237, pp. 59–72, 2013.
- [9] M. A. N. Banu and B. Gomathy, "Disease Predicting System Using Data Mining Techniques," *Int. J. Tech. Res. Appl.*, vol. 1, no. 5, pp. 41–45, 2013.
- [10] E. M. Benyoussef, A. Elbyed, and H. El Hadiri, "Data Mining Approaches for Alzheimer's Disease Diagnosis," vol. 10542, pp. 619–631, 2017.
- [11] A. Tiwari and A. Prakash, "Improving classification of J48 algorithm using bagging , boosting and blending ensemble methods on SONAR dataset using WEKA," *Int. J. Eng. Tech. Res.*, vol. 2, no. 9, pp. 207–209, 2014.
- [12] D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: A comparison of three data mining methods," *Artif. Intell. Med.*, vol. 34, no. 2, pp. 113–127, 2005.
- [13] T. R. S. Mary and S. Sebastian, "Predicting heart ailment in patients with varying number of features using data mining techniques," *Int. J. Informatics Commun. Technol.*, vol. 8, no. 1, p. 56, 2019.
- [14] V. Chaurasia, "Early Prediction of Heart Diseases Using Data Mining," *Caribb. J. Sci. Technol.*, vol. 1, no. December, pp. 208–217, 2013.
- [15] W. Zhang, D. Ma, and W. Yao, "Medical diagnosis data mining based on improved Apriori algorithm," *J. Networks*, vol. 9, no. 5, pp. 1339–1345, 2014.
- [16] L. B. Moreira and A. A. Namen, "A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia," *Comput. Methods Programs Biomed.*, vol. 165, pp. 139–149, 2018.
- [17] T. Ateeq *et al.*, "Ensemble-classifiers-assisted detection of cerebral microbleeds in brain MRI," *Comput. Electr. Eng.*, vol. 69, pp. 768–781, 2018.
- [18] T. Daghistani and R. Alshammari, "Diagnosis of Diabetes by Applying Data Mining Classification Techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 7, pp. 329–332, 2016.
- [19] S. Perveen, M. Shahbaz, A. Guergachi, and K. Keshavjee, "Performance Analysis of Data Mining Classification Techniques to Predict Diabetes," *Procedia Comput. Sci.*, vol. 82, no. March, pp. 115–121, 2016.
- [20] A. A. Aljumah, M. G. Ahamad, and M. K. Siddiqui, "Application of data mining: Diabetes health care in young

- and old patients,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 25, no. 2, pp. 127–136, 2013.
- [21] C. S. Dangare and S. S. Apte, “Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques,” *Int. J. Comput. Appl.*, vol. 47, no. 10, pp. 44–48, 2012.
- [22] B. Gomathy and M. A. N. Banu, “Disease Predicting System Using Data Mining Techniques,” *Int. J. Tech. Res. Appl.*, vol. 1, no. 5, pp. 41–45, 2013.
- [23] S. Palaniappan and R. Awang, “Intelligent heart disease prediction system using data mining techniques,” in *AICCSA 08 - 6th IEEE/ACS International Conference on Computer Systems and Applications*, 2008, pp. 108–115.
- [24] H. Kim, M. I. M. Ishag, M. Piao, T. Kwon, and K. H. Ryu, “A data mining approach for cardiovascular disease diagnosis using heart rate variability and images of carotid arteries,” *Symmetry (Basel)*, vol. 8, no. 6, p. 47, 2016.
- [25] V. Chauraisa and S. Pal, “Early Prediction of Heart Diseases Using Data Mining Techniques,” no. December, 2013.
- [26] A. S. Abdullah, S. Selvakumar, P. Karthikeyan, and M. Venkatesh, “Comparing the Efficacy of Decision Tree and its Variants using Medical Data,” *Indian J. Sci. Technol.*, vol. 10, no. 18, pp. 1–8, 2017.
- [27] H. Yang and Y. P. P. Chen, “Data mining in lung cancer pathologic staging diagnosis: Correlation between clinical and pathology information,” *Expert Syst. Appl.*, vol. 42, no. 15–16, pp. 6168–6176, 2015.
- [28] S. Li, H. Jiang, Z. Wang, G. Zhang, and Y. dong Yao, “An effective computer aided diagnosis model for pancreas cancer on PET/CT images,” *Comput. Methods Programs Biomed.*, vol. 165, pp. 205–214, 2018.
- [29] B. . Kiranmayee, “Effective Analysis of Brain Tumor Using Hybrid Data Mining Techniques,” *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 7, pp. 286–293, 2017.
- [30] L. Chen *et al.*, “Personal health indexing based on medical examinations: A data mining approach,” *Decis. Support Syst.*, vol. 81, pp. 54–65, 2016.
- [31] K. K. N, R. L. Tulasi, and V. D, “An ensemble multi-model technique for predicting chronic kidney disease,” *Int. J. Electr. Comput. Eng.*, vol. 9, no. 2, p. 1321, 2019.
- [32] M. Graczyk, T. Lasota, B. Trawiński, and K. Trawiński, “Comparison of Bagging , Boosting and Stacking,” *Asian Conf. Intell. Inf. database Syst.*, pp. 340–350, 2010.
- [33] T. G. Dietterich, “Experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization,” *Mach. Learn.*, vol. 40, no. 2, pp. 139–157, 2000.
- [34] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, “Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults,” *J. Cogn. Neurosci.*, vol. 19, no. 9, pp. 1498–1507, 2007.
- [35] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, “Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults,” *J. Cogn. Neurosci.*, vol. 19, no. 9, pp. 1498–1507, 2007.
- [36] J. Kozak, “Ensemble methods,” *Stud. Comput. Intell.*, vol. 781, pp. 107–118, 2019.
- [37] T. G. Dietterich, “An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees,” *Mach. Learn.*, vol. 40, pp. 139–157, 2000.
- [38] N. Altman and M. Krzywinski, “Points of Significance: Ensemble methods: Bagging and random forests,” *Nat. Methods*, vol. 14, no. 10, pp. 933–934, 2017.

BIOGRAPHIES OF AUTHORS



Mohammed Abdullah Al-Hagery: received his B.Sc in Computer Science from the University of Technology in Baghdad Iraq-1994. He got his MSc. in Computer Science from the University of Science and Technology Yemen-1998. Al-Hagery finished his Ph.D. in Computer Science and Information Technology, (Software Engineering) from the Faculty of Computer Science and IT, University of Putra Malaysia (UPM), November 2004.

He was a head of the Computer Science Department at the College of Science and Engineering, USTY, Sana’a from 2004 to 2007. From 2007 to this date, he is a staff member at the College of Computer, Department of Computer Science, Qassim University, KSA. He published more than 20 papers in various international journals. Dr. Al-Hagery was appointed a head of the Research Centre at the Computer College, and a council member of the Scientific Research Deanship Qassim University, KSA from September 2012 to October 2018. Currently, he is teaching the master degree students and a supervisor of four master thesis. He is a jury member of a number of PhD and master thesis, as an internal and external examiner in his field of his specialist.

Ebtehal Ibrahim Al-Fairouz : Teaching assistant in Department of MIS and a Master student in Computer Science Department, College of Computer, Qassim University, KSA.

Norah Ahmed Al-Humaidan: Master student in Computer Science Department, College of Computer, Qassim University, KSA.