Improvement of Machine Learning Models' Performances based on Ensemble Learning for the detection of Alzheimer Disease

Selim Buyrukoğlu
Dept. of Computer Engineering
Çankırı Karatekin University
Çankırı, Turkey
s.buyrukoglu@karatekin.edu.tr

2021 6th International Conference on Computer Science and Engineering (UBMK) | 978-1-6654-2908-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/UBMK52708.2021.9558994

***Abstract*— Failure to early detection of Alzheimer's disease (AD) can lead memory deterioration. Therefore, early detection of AD is essential affecting the points of the brain that control vital functions. Various early AD detection approaches have been employed using machine learning. In literature, most of the early detection of AD approaches has been developed using single machine learning methods. Due to the importance of early detection of AD, the goal of this study is to improve the classification performance of the previous studies for early detection of AD applying ensemble learning methods including bagging, boosting and stacking. ADNI clinical dataset was used in this study with three target classes: Normal (CN), Mild Cognitive Impairment (MCI) and Alzheimer's disease (AD). The proposed ensemble learning methods provided better classification performance compared to single machine learning methods. Besides, the best classification performance from the ensemble methods is obtained through the boosting (AdaBoost) ensemble (92.7%). This study revealed that the classification rate increased up to between 3.2% and 7.2% compared to single based machine learning approaches through the AdaBoost ensemble method.**

Keywords—machine learning, ensemble methods, AdaBoost, Alzheimer Disease

I. Introduction

Alzheimer's disease (AD) is a type of dementia and it is seen commonly in elderly people **[1]**. It is the seventh leading cause of death among diseases worldwide **[2]**. AD can be considered as a neurodegenerative disorder and age is one of the most important factors for AD **[3]**. Also, early detection of AD is considered important in order to avoid making mistakes in the diagnosis of other diseases of the patient and to eliminate the problems related to hospitalisation. AD stages are cognitively normal (CN), mild cognitive impairment (MCI) and Alzheimer's disease (AD). Early detection is a classification problem, and also if early and accurate detection of AD is achieved, the number of AD can be reduced as much as possible **[4]**. It should be note that patients diagnosed with MCI have an increased possibility of Alzheimer's disease than those without MCI. When the patient is classified as MCI, it is predicted that the patient will most likely develop Alzheimer's **[4]**. Therefore, early detection of AD is vital to plan the appropriate health interventions. For the early detection of AD, various studies have been proposed to achieve the early and accurate detection of AD based on artificial intelligence (AI), machine learning (ML) algorithms **[5-10]**.

In a study **[5]**, radial basis function (RBF) was implemented to classify the AD patients into CN and AD classes applying ten-fold cross validation. The RBF classification model achieved to provide 82% accuracy. In a different study **[6]**, a model was proposed for the early detection of AD disease applying ML algorithms including random forest (RF), logistic regression (LR), naïve bayes (NB), artificial neural network (ANN), support vector machine (SVM). In that study, feature extraction is done using these ML algorithms based on homogeneous and heterogeneous ensemble approaches. At the end, the RF algorithm with the heterogeneous ensemble feature extraction provided the best performance (91%) compared to the others. Its note that ten-fold cross-validation was applied to the algorithms, and the target classes were CN, MCI and AD. In **[7]**, SVM model was proposed for the prediction of diagnosis of AD using discriminant features, and the target classes were CN and AD. This SVM model performed 90.38% accuracy. Cui et al., **[8]** performed a hybrid model for the early detection of AD. ANN model was combined with recurrent neural network (RNN) for the early detection of AD. Structural MR images were used as dataset and the target classes were CN and AD. The proposed method performed a convincing result (89.7% accuracy). In a different study, multi-view multimodal spectral embedding model was proposed by Liu et al., **[9]**. This model performed 75% accuracy with five-fold cross validation. The target classes were cognitively normal (CN) and Alzheimer's disease (AD). In **[10]**, ensemble learning models (bagging, boosting and stacking) were compared to improve the classification accuracy of previous studies for AD diagnosing using Open Access Series of Imaging Studies (OASIS) dataset. 83.33% accuracy rate was obtained as a result of the stacking ensemble (DT and kNN with logistic regression) model in the study.

Even if the aforementioned studies achieved to provide promising results, these studies have not applied the ensemble learning models to improve and compare the performance of ML algorithms for Alzheimer's Disease Neuroimaging Initiative (ADNI) (see Section 2 for detailed information) dataset. Also, boosting (especially AdaBoost algorithm) algorithms have not been used in a stacking ensemble for the early detection of AD. The used ADNI dataset consists of three target classes: Cognitively Normal (CN), Mild Cognitive Impairment (MCI) and Alzheimer's disease (AD). Ensemble learning methods are classified as bagging, boosting and staking ensemble **[10]**. In bagging, same ML algorithm is trained with different subsets sampled from training set in

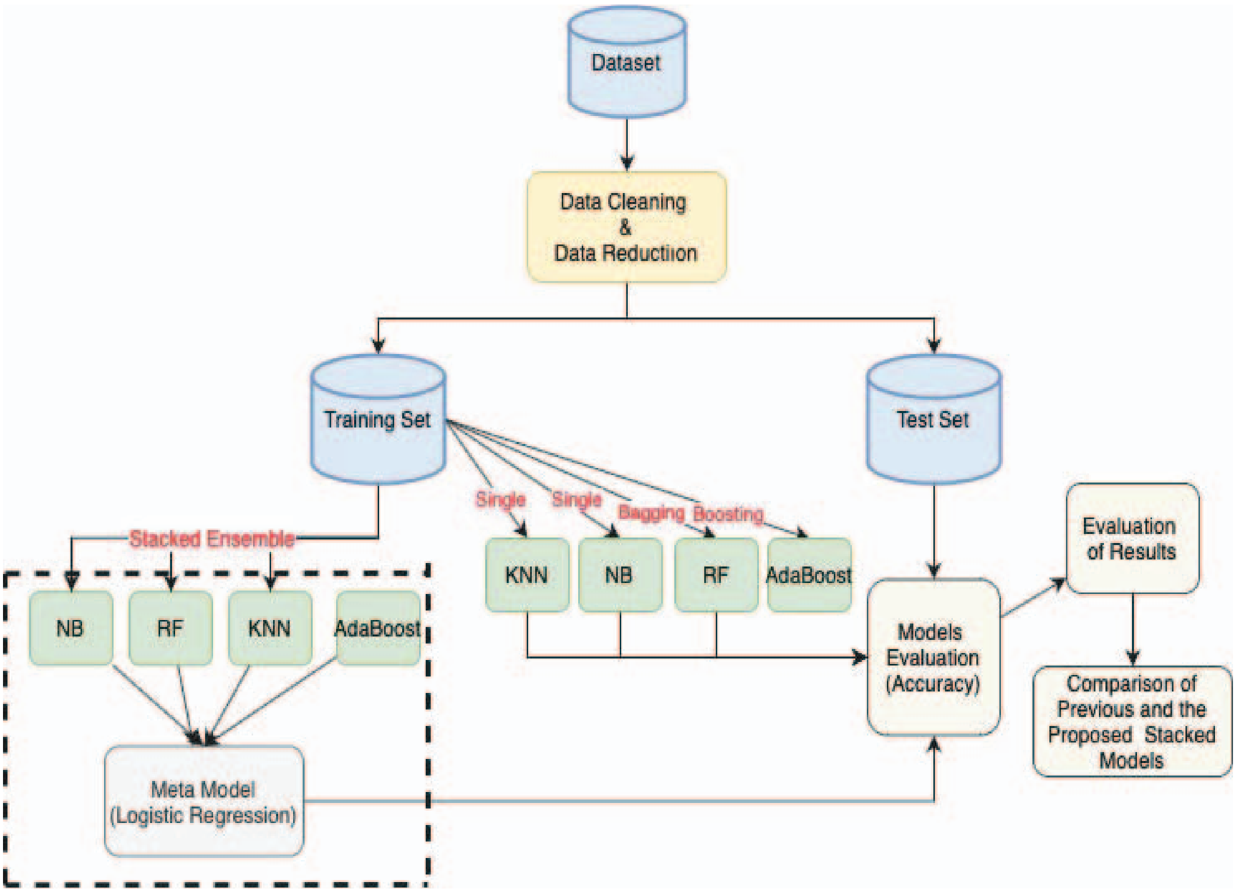

Fig. 1 Proposed early AD detection approach

order to improve the performance of these algorithms **[11]**. RF and DT are some of the bagging algorithms. The boosting method has an iterative process, and each individual model is created sequentially, iterating over the previous one. In this case, especially, the wrongly classified any data points by the previous model is highlighted in the next model **[12]**. Thus, this iterative process purposes to improve the performance of the model. AdaBoost, Gradient Boosting and XGBoost are some of boosting algorithms. In contrast to bagging and boosting ensembles, results of different ML algorithms are combined for the purpose of creation of a more strong and effective learning model in stacking ensemble **[12]**. Voting, stacking, blending and super learner are some of the stacking ensembles. On the other hand, boosting types of algorithms (AdaBoost, Gradient Boosting and XGBoost) have never been used in a stacking ensemble model to create more strong and effective learning model using ADNI dataset. Thus, the major objective of this study is;

- to improve the performance of ML algorithms based on ensemble learning models. Especially using one of the boosting algorithms in stacking ensemble (such as AdaBoost etc.).
- to compare the performance of previous learning models and the proposed ensemble learning models in this study for Alzheimer's Datasets (ADNI1, OASIS etc.).

The rest of this paper is structured as follows: Section II presents the proposed early AD detection approach including data pre-processing, single and ensemble methods (including bagging, boosting and stacking), and model evaluation. The next section includes analysis of the proposed approaches, discusses of the results, and compares the previous and the proposed ensemble studies in the early detection of AD. The final section is about the conclusion and further direction.

II. Proposed Approach

This section describes the proposed approach for the detection of Alzheimer disease aiming to improve machine learning algorithms applying ensemble learning methods. Fig.1 illustrates the proposed approach.

A. Data Collection, Cleaning and Reduction

The used dataset was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). Features have more than 50% missing values are removed from the dataset. Then, Gain Ratio (GR) method is applied to the cleaned data as a feature selection approach. The used dataset consists of 389 instances (111 NC, 193MCI, 85AD) with 9 features showing in Table I. GR is a revision of information gain which was developed to reduce the bias of it **[13]**. Also, Fig. 2 presents the sample brain MRIs of CN, MCI and AD patients.

Fig. 2. The sample brain MRIs of CN, MCI and AD patients.

TABLE I. SELECTED FEATURES BASED ON GAIN RATIO.

No	Feature Name
1	Age
2	Gender
3	Education
4	Ethnic Category
5	Racial Category
6	PT_marry
7	Site
8	APOE4
9	ADAS13

B. Single and ensemble learning based methods

This section provides information about the used single machine learning methods (NB, KNN), bagging (RF), boosting (AdaBoost) and stacking ensemble learning methods to early detect of AD in this study, respectively.

- Naïve Bayes: Naïve Bayes is an efficient classifier and provides well perform on unstable data. **[14]**. Probability value for an element is used in the classification process of this algorithm. In addition to this, this method produces convincing performance if the parameters are independent **[15]**.
- k-Nearest Neighbors (KNN): It is a supervised learning algorithm. This algorithm is effectively used in classification and regression problems. Initially, T. M. Cover and P. E. Hart proposed this algorithm in 1967. The goal of this algorithm is to find the samples nearest to the sample to be classified. Thus, there is a various distance functions are used for the purpose of distance calculation including Minkowski, Euclidean and Manhattan **[16]**.
- Bagging Ensemble (Random Forest-RF): The Random Forest (RF) was initially proposed by Tin Kam Ho in 1995. The aim of the RF is to derive more than one decision trees in terms of achieving better performance compared to a single based decision tree. That is, created single based decision trees form the RF **[17]**. As highlighted in Section 1, early detection of AD is a classification problem, and the classification is made based on the superiority of the predictions obtained from trees. In our study, different number of trees have been used in the creation of the RF such as 42, 50, 100 and 200. At the end, the best performance was obtained by using 42 trees.
- Boosting Ensemble (AdaBoost) : AdaBoost algorithm is aimed to increase the performance of the classification model by focusing more on samples that are difficult to classify **[18]**. In each iteration of the method, while the weight values for the incorrectly classified samples are increased, the weight values for the correctly classified samples are decreased, allowing the basic learning algorithms to allocate more iterations to the data samples that are difficult to classify in the training set. In addition, by assigning weight values to classification algorithms in the AdaBoost algorithm, classifiers with higher correct classification performance are represented with higher weight values **[19]**. In this study, the best performance of AdaBoost algorithm is obtained with 10 estimators.
- Stacked Ensemble Method: The stacking method consists of two stages. Initially, ensemble of various algorithms applies on training set. In the second stage, meta classifier gets the results of these algorithms, and then unseen data is used in the evaluation part of the stacking method **[20]**. In our study, the Naïve Bayes, K-Nearest Neighbor, Random Forest and AdaBoost are used with Logistic Regression as the Meta Classifier. Logistic regression is a statistical method used to analyse a dataset with one or more independent variables that determine an outcome. The most distinctive feature that distinguishes logistic regression from linear regression is that the outcome variable in logistic regression is binary or multiple **[21]**.

C. Evaluation Metrics

Accuracy, precision and sensitivity (recall) are used as evaluation metrics in this study **[22]**. Ten-fold cross-validation is used in the training process of the models. In the evaluation process of the models, 20% of data set was used as test set (unseen data). The accuracy, precision and sensitivity (recall) formula is presented as follows.

- Accuracy = (TN + TP)/(TN+TP+FN+FP),
- Precision = TN/(TN + FP),
- Sensitivity (Recall) = TP/(TP + FN).

The definition of TP, FP, TN, and FN presented in Table II.

TABLE II. DEFINITION OF TP, FP, TN, AND FN

		Actual Class	
		Positive	Negative
Predicted Class	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

III. ANALYSIS AND DISCUSSION

This section discusses the results and performances of the proposed methods including single, bagging, boosting and stacking models. Besides, the achievements of the implemented methods are compared with the proposed approaches developing for the early detection of AD. Finally, a general inference is made about the ensemble methods.

A. Performance Results

As it can be seen from Fig. 3, the proposed boosting (AdaBoost) ensemble model has better performance (92.7%) compared to the others. In contrast to the proposed boosting ensemble model, stacking (NB, KNN, RF, AdaBoost) and Bagging (RF) ensemble models also provide same and convincing result (91.2%) closing to the result of the boosting ensemble method. It can be easily inferred that the boosting ensemble method may play key role in the early detection of AD. On the other hand, worst performance is obtained from the NB algorithm (84%).

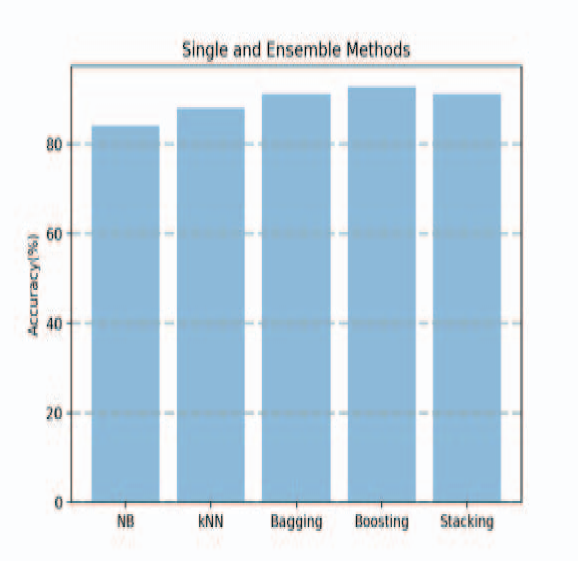

Fig. 3. Performances of the proposed single and ensemble methods for AD

In addition to this, Table III depicts information about the precision and sensitivity (recall) rates of the proposed methods. Boosting ensemble learning method achieved the best precision (93%) and sensitivity rates (93%). In contrast to boosting ensemble, NB algorithm has the lowest accuracy rate (84%) compared to the others.

TABLE III. EVALUATION OF THE PROPOSED SINGLE AND ENSEMBLE METHODS IN TERMS OF PRECISION AND SENSITIVITY

Model Name	Precision (%)	Sensitivity (Recall) (%)
NB	83	84
kNN	87	87
Bagging (RF)	91	91
Boosting (AdaBoost with ten estimators)	93	93
Stacking (NB, kNN, RF, Adaboost)	91	91

B. *Comparison of the results of the proposed and the previous approaches*

Table IV presents the performance results of previous ensemble studies and the proposed ensemble model in terms of early detection of AD. Due to the using of clinical data in our study, the accuracy results of the previous approaches using the clinical data are presented in Table IV. In **[10]**, a stacking ensemble model achieved to provide 83.33% accuracy using DT-kNN with logistic regression. Also, a boosting ensemble model provides 90.1% accuracy based on AdaBoost with J48 in **[23]**. On the other hand, a stacking (heterogeneous) ensemble feature selection approach proposed by **[6]**. ReliefF, Gain Ratio, Chi-Square and FCBF feature selection approaches have been used to obtain an optimal future subset. Then, bagging ensemble (random forest) approach is used based on the optimal subset. 91% accuracy has been obtained through the proposed bagging ensemble. As it can be seen from Table IV, the proposed boosting ensemble approach provided 92.7% accuracy, and we believe that the AdaBoost ensemble method plays an important role in early detection of AD.

TABLE IV. PERFORMANCE RESULTS OF PREVIOUS ENSEMBLE STUDIES AND THE PROPOSED ENSEMBLE MODEL IN TERMS OF EARLY DETECTION OF AD.

Study	Sample Size	Model	Accuracy (%)
[6]	819	Hybrid heterogeneous ensemble model (ReliefF Gain Ratio Chi-Square FCBF with RF)	91%
[10]	416	Stacking (DT-kNN with logistic regression)	83.33
[23]	605	AdaBoost	90.1%
Proposed Boosting (AdaBoost)	389	AdaBoost (with ten estimators)	92.7%

C. *General interpretation of the ensemble models*

Ideally, the stacking ensemble model should provide better performance compared to the others. The reason behind this is that stacking ensemble combined four different models with the logistic regression as a meta classifier. However, boosting ensemble model provided the best performance in our study. Thus, it can be inferred from this result is that stacking ensemble model does not always provide better performance compared to the bagging and boosting ensemble. Also, it should be noted that different stacking algorithms can be used to improve the performance of it in future (i.e. super learner etc.).

Even if the AdaBoost ensemble algorithm has the best performance compared to the bagging and boosting ensemble models, there are a few differences between them in terms of performances as it is seen in Fig. 3. In this case, these ensemble models can be used in different fields behind the health science including sport science to classify football players' performances, agricultural science to detection of pathogen (*salmonella* or *E. coli*) in agricultural surface waters, and education science to detect and classify students etc.

IV. CONCLUSION AND FURTHER DIRECTION

This study proposed ensemble learning approaches to early prediction of AD. Alzheimer's Disease Neuroimaging Initiative (ADNI) clinical data was used in our study. Bagging, boosting and stacking ensemble learning methods were compared in terms of accuracy performance. Boosting ensemble model (AdaBoost) achieved providing the best accuracy (92.7%) performance compared to the others. Also, the boosting ensemble model provided the best precision (93%) and sensitivity (recall) (93%) rates compared to the others. The findings reflect the importance of ensemble methods compared to single based machine learning models in the early prediction of AD. Overall, the proposed boosting ensemble method enable to obtain significant performance for the early detection of AD.

As a further study, different boosting algorithms (such as Gradient Boosting and XGBoost) can be used in the stacked ensemble to improve and compare performances of the models in terms of classification. Also, ensemble models can be combined with deep learning algorithms (such as deep belief networks – DBNs) to create a hybrid model for the purpose of improving models' performances in the early AD detection. Besides, stacking restricted Boltzman machine (RBMs) can be used to create a DBN in the early detection of AD.

REFERENCES

[1] O. Zanetti, S. B. Solerte, and F. Cantoni, “Life expectancy in Alzheimer's disease (AD)”, Archives of gerontology and geriatrics, *49*, 237-243, 2009.

[2] C. Patterson, “The state of the art of dementia research: New frontiers”, World Alzheimer Report, 2018.

[3] B. Lam, M. Masellis, M. Freedman, D. T. Stuss, and S. E. Black, “Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome”, Alzheimer's research & therapy, *5*(1), 1-14, 2013.

[4] B. Dubois, G. Picard, and M. Sarazin, “Early detection of Alzheimer's disease: new diagnostic criteria”, Dialogues in clinical neuroscience, 11(2), 135, 2009.

[5] N. Zeng, H. Qiu, Z. Wang, W. Liu, H. Zhang, and Y. Li, “A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer’s disease”, Neurocomputing, 320, 195-202, 2018.

[6] S. Buyrukoğlu, “Early Detection of Alzheimer’s Disease Using Data Mining: Comparison of Ensemble Feature Selection Approaches”, Konya Mühendislik Bilimleri Dergisi, 9(1), 50-61, 2021.

[7] J. Ramírez, J. M. Górriz, D. Salas-Gonzalez, A. Romero, M. López, I. Álvarez, and M. Gómez-Río, “Computer-aided diagnosis of Alzheimer’s type dementia combining support vector machines and discriminant set of features”, Information Sciences, 237, 59-72, 2013.

[8] R. Cui, M. Liu, and G. Li, “Longitudinal analysis for Alzheimer's disease diagnosis using RNN”, In 2018 IEEE 15th International Symposium on Biomedical Imaging, pp. 1398-1401, 2018, IEEE.

[9] F. Liu, C. Y. Wee, H. Chen, and D. Shen, “Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification”, NeuroImage, 84, 466-475, 2014.

[10] M. A. Al-Hagery, E. I. Al-Fairouz, and N. A. Al-Humaidan, “Improvement of Alzheimer disease diagnosis accuracy using ensemble methods”, Indonesian Journal of Electrical Engineering and Informatics (IJEEI), 8(1), 132-139, 2020.

[11] M. H. D. Ribeiro, and L. dos Santos Coelho, “Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series”, Applied Soft Computing, 86, 105837, 2020.

[12] P. Bühlmann, “Bagging, boosting and ensemble methods”, In Handbook of computational statistics, pp. 985-1022, Springer, Berlin, Heidelberg, 2012.

[13] A. G. Karegowda, A. S. Manjunath, and M. A. Jayaram, “Comparative study of attribute selection using gain ratio and correlation based feature selection”, International Journal of Information Technology and Knowledge Management, 2(2), 271-277, 2010.

[14] R. F. Fouladi, C. E. Kayatas, and E. Anarim, “Frequency based DDoS attack detection approach using naive Bayes classification”, In 2016 39th International Conference on Telecommunications and Signal Processing (TSP), pp. 104-107, 2016, IEEE.

[15] B. Zhang, T. Zhang, and Z. Yu, “DDoS detection and prevention based on artificial intelligence techniques”, In 2017 3rd IEEE International Conference on Computer and Communications (ICCC), pp. 1276-1280, 2017, IEEE.

[16] Y. Miah, C. N. E. Prima, S. J. Seema, M. Mahmud, and M. S. Kaiser, “Performance comparison of machine learning techniques in identifying dementia from open access clinical datasets”, In Advances on Smart and Soft Computing, pp. 79-89, 2021. Springer, Singapore.

[17] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, “Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques”, Expert systems with applications, 42(1), 259-268, 2015.

[18] Y. Freund, and R. E. Schapire, R. E, “Experiments with a new boosting algorithm”, In *icml*, vol. 96, pp. 148-156, 1996.

[19] A. Onan, “Classifier and feature set ensembles for web page classification”, *Journal of Information Science*, *42*(2), 150-165, 2016.

[20] A. Géron, “ Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems”, O'Reilly Media, 2019.

[21] G. Tripepi, K. J. Jager, F. W. Dekker, and C. Zoccali, “Linear and logistic regression analysis”, Kidney international, 73(7), 806-810, 2008.

[22] A. Baratloo, M. Hosseini, A. Negida, and G. El Ashal, “Part 1: simple definition and calculation of accuracy, sensitivity and specificity”, 2015.

[23] L. B. Moreira, and A. A. Namen, “A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia”, Computer methods and programs in biomedicine, 165, 139-149, 2018.