ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 11(22), DOI: 10.17485/ijst/2018/v11i22/123310, June 2018*

Hybrid Feature Selection Framework for Identification of Alzheimer's Biomarkers

V. Thavavel[1]* and M. Karthiyayini[2]

[1]College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al Kharj, Saudi Arabia; thavavelmurugesanv@gmail.com
[2]Department of Computer Sciences and Engineering, Sethu Institute of Technology, Kariapatti, Virudhunagar – 626115, Tamil Nadu, India; Karthithava49@gmail.com

Abstract

Objectives: Alzheimer's Disease (AD) is a chronic disease that eventually leads to death. Early diagnosis is expected to improve the patient survival. Evidence from literature indicates that best combination of biomarkers can help in early diagnosis. Feature selection techniques are proven as a workhorse in biomarker selection. This work aims to contribute a new hybrid feature selection framework based on ensemble learning for biomarker identification with an objective to improve the robustness of the final set of selected biomarkers. Methods and Analysis: The proposed framework employs Significance Analysis of Microarray (SAM) filter to initially select most relevant biomarkers. Then, the selected biomarkers are reselected in wrapper phase using heterogeneous voting based ensemble of classifiers to enhance the robustness of the final set of selected biomarkers. An extensive experiment was conducted to demonstrate the effectiveness of the proposed framework against the individual base learner in term of sensitivity and specificity. Finally, the statistical significance of the selected biomarkers was also verified by ANOVA test. Findings: The reported experimental results demonstrated improvement in AD diagnosis accuracy and proved the potential of the proposed framework in selecting the stable set of biomarkers for early AD diagnosis alternative to its base learners. The diagnosis accuracy level obtained by the identified set of biomarkers is over 87%. The higher area under ROC curve revealed the advantage of the identified biomarkers in discriminating AD patients from healthy control diagnosis with sensitivity and specificity of 93%. The p-value of ANOVA test results confirmed the significance of the identified biomarkers in early AD diagnosis. Overall, the identified robust set of biomarkers is expected to open a new pathway for early AD diagnosis. Improvements: Although clinical studies are needed to conform the findings, in the light of the results reported in this paper, it is evident that the proposed framework stands out to find the significant biomarkers for early AD diagnosis.

Keywords: Alzheimer's Disease Early Diagnosis, Biomarker Identification, Filter-Wrapper Model, Greedy Search and Ensemble Classifiers, Hybrid Feature Selection, SAM Filter

1. Introduction

Alzheimer's Disease (AD) is a chronic neurodegenerative disorder and the leading cause of dementia worldwide amongst elderly population[1]. It is characterized by progressive and irreversible loss of memory damage, resulting in cognitive impairment and eventually to death. In 2016, it is reported that over 47 million adults are diagnosed with dementia worldwide and is likely to double every 20 years and reach 131 million by 2050[2]. The rising incidence of AD is mainly because of no approved pharmacological treatment and diagnosis in the advanced stages to derive

*Author for correspondence

clinical benefit from treatment. Growing evidence has illustrated that early diagnosis of AD can maximize the effectiveness of the treatment benefits. Therefore, there is critical need on detecting AD as early as possible potentially in its most reversible stage.

Recent literature studies have confirmed that biomarker assessment enhances the accuracy of early AD diagnosis than traditional clinical assessment[3–5]. On this basis, many efforts are devoted worldwide to investigate effective biomarkers in biological fluids that can improve the diagnostic confidence of early diagnosis of AD and can be implemented widely. Despite of best efforts and increasing rates of publications on biomarkers, the identification of reliable biomarkers for Alzheimer's Disease (AD) still remains challenging task and needs further investigation. Fortunately, it has been observed from the literature of last decade that feature selection techniques are used as workhorse in biomarker discovery with varying degrees of success[6,7]. In line to this, scientific research community is very active in responding to the above-mentioned challenge by designing feature selection methods for identifying early predictive biomarkers and creates great impact in improving treatment.

This research aims to respond the above challenge contributing a new feature selection framework for identifying stable biomarkers and optimizing the early diagnosis of AD. The approach is based on hybrid filter-wrapper approach. The filter phase employs SAM filter to reduce the search space and time complexity of the subsequent phase. In wrapper phase, ensembles of classifiers are utilized to increase the selected biomarker reproducibility and thereby enhance the accuracy of early AD diagnosis. The final set of biomarkers selected by the proposed framework is evaluated using the state of art classifiers such as SVM, K-Nearest Neighbhor (KNN) and Logistic regression classifier in term of four popular evaluation metrics namely, accuracy, and sensitivity, specificity and Receiver Operator Characteristic (ROC) value.

The main contribution of this study can be summarized as follows:

- For first time, the proposed feature selection method employs ensemble of classifiers in hybrid framework for AD biomarker identification to find more stable set of AD biomarkers.

- The proposed framework is expected to identify the more stable set of biomarkers for discriminating AD subjects in its early stage and enhance the treatment benefits.

2. Hybrid Feature Selection Framework

The working procedure of general hybrid feature selection framework is summarized in algorithm-1. The framework comprises two stages working sequentially. In the initial stage, filter is used to select the most relevant features based on its ability to discriminate between different target classes and reduce the search space for subsequent tasks. The next stage, called wrapper is based on two components namely, search algorithm and evaluation algorithm (classifier). The search algorithm is responsible for finding the most promising subset of features from feature space for evaluation. The evaluation algorithm assesses the goodness of feature subset identified by the search algorithm so as to discover the optimum subset that gives the best accuracy[8]. Therein this stage is also known as fine-tuning stage.

Algorithm 1: Hybrid Feature Selection

Input **:** Training dataset
Output **:** Subset Features that

Filter phase

a) Calculate score for all attributes applying ranking method.
b) Rank the features based on calculated score from the highest to the lowest.
c) Select top n ranking features from the list.

Wrapper Phase

a) Initialize this phase with above selected n features
b) Repeat
   i. Apply search algorithm for feature subset selection
   ii. Evaluate the selected feature subset for its classification accuracy
c) Until termination criteria.

3. Proposed Hybrid Feature Selection Framework

The proposed hybrid feature selection framework is based on ensemble paradigm to identify the most stable predictive biomarkers for early AD diagnosis. It comprises two essential parts and this is illustrated in algorithm-2. Each these parts are briefed in the following subsection,

3.1 Filter Phase

The objective of this phase is to remove the most redundant and irrelevant features and identify a subset of features that best characterize the statistical significance of the target class for the given sample during classification task[8]. In this context, SAM is used in the present work based on its performance reported in the previous literature[9]. To our knowledge, it is the first time that the SAM is used for ranking AD biomarker and filter the most significant biomarker for AD classification. Authors in[10] proposed SAM in 2001 to deal with problem of small variances in small sample sizes by adding a small "fudge factor" to the denominator of the test statistic[10]. This fudge factor is calculated as given below from the distribution of feature-specific standard errors.

$$SAM(i) = \frac{\mu_{i1} - \mu_{i2}}{\sigma_i + \sigma_0} \quad (1)$$

$$\sigma_i = \sqrt{\left(\frac{1}{n_{i1}} + \frac{1}{n_{i2}}\right)\frac{\sigma_{i1} + \sigma_{i2}}{n_{i1} + n_{i2} - 2}} \quad (2)$$

Algorithm 2: Hybrid Feature Selection

Input **:** AD Training dataset.
Output **:** Best combination of AD Biomarkers (AD Biomarker subset).

Filter phase

a) Calculate statistical score for all Biomarkers applying SAM given in equation ().
b) Rank the Biomarker based on calculated score from the highest to the lowest.
c) Select top 25 ranking Biomarker from the list.

Wrapper Phase

a) Initialize this phase with above selected 25 Biomarkers.
b) Ensemble Model is created using KNN, SVM and Logistic regression classifiers.
c) Repeat.
   i. Apply greedy search algorithm for Biomarker subset selection.
   ii. Train the above created ensemble classifier with the selected Biomarker subset.
   iii. Biomarker subset evaluation is performed based on classification accuracy using the created ensemble classifier. In this step, 10-fold cross validation and classification accuracy.
d) Until termination criteria.

3.2 Wrapper Phase

As pointed out in literature, the reproducibility of biomarker is key requirement for its application in clinical practice[6,11]. The reproducibility of biomarker is interpreted as its stability to demonstrate good performance across diversified situations/subjects. Numerous studies have confirmed that ensemble approaches can help increase the accuracy and stability of the selection results[11,12]. In this context, this phase employs ensemble of heterogeneous classifiers within wrapper to exploit the strengths

of each classifier and obtain not only enhanced performance by their combination in classification accuracy but also in robustness of the final set of selected features.

The important success factor in the design of an ensemble relies on its member component diversity at the same time this diversity should not diminish its accuracy. Bearing this in mind, this work introduces as base learners' three classical classifiers namely SVM, Logistic Regression (LR) and K-Nearest Neighbor (KNN) for ensemble learning in wrappers considering their superior performance in previous literatures in achieving a good balance between accuracy and efficiency for feature selection[13–15]. The diversity here comes from their ability to learn from given training set as well their ability to discriminate the unknown new sample.

Once the classifier ensemble is constructed, a suitable aggregation function is required to combine the results of base classifiers and form the final ensemble decision. In this work, Majority voting is applied. This strategy involves finding the votes of each classifier for a class and then the class with the maximum vote is considered as the final result. The chance for majority voting to give wrong results is possible only if more than half of the base classifiers are with wrong results[16]. Encouraged by this capability of majority voting as well its performance in previous experimental studies, it was considered in the present work.

4. Experimental Setup

This section describes experimental setup used to evaluate the performance of the proposed framework. First subsection deliberates the AD datasets used in all of our experiments. Subsequently, the parameter settings for the base classifiers used in our framework is presented.

4.1 Data Source

AD dataset chosen to evaluate our framework was obtained from Kaggle at https://www.kaggle.com. The clinical dataset used in the experiments include 333 subjects. Out of which 91 subjects are mild cognitive impaired (AD subjects) and 242 are healthy control. Data collected from each subject consisted 128 CSF biomarkers measuring variety of cytokines, chemokines, metabolic markers, growth factors and other markers alongside demographic parameters namely gender and age[17].

4.2 Data Cross Validation

To avoid over-fitting as well to improve the robustness of the results obtained, this work applies 5-fold Cross Validation (CV) approach to partition the original dataset into 5 folds evenly of equal size[18]. Of 5-folds, 4 folds are used for training our framework and one fold is held for testing.

4.3 Parameter Settings for Classifiers

As demonstrated in several literature results, the accuracy of a classifier solely relies on parameter settings for a given dataset. Therein, the main objective of our experimental study was to build a unique model in which parameters are fine tuned for all classifiers to achieve the best results across all datasets. In achieving this objective, the parameter setting for the base classifiers were found on trial and

Table 1. Parameter settings for the ensemble classifiers in our proposed framework

Base Classifier	Parameter Setting
K-NN	Euclidean distance with linear search was employed with K = 7
Logistic	Conjugate gradient descent is applied with ridge = 1.0 E-7
SVM	C-SVC was employed with linear Kernel, gamma = 5 and cost error = 3

error basis with 10-fold cross validation and is listed in Table 1.

4.4 Performance Metrics

In all the above designed experiments, the final set of biomarker identified by the proposed framework was assessed for its quality using individual base classifiers and designed ensemble classifier against three performance metrics which are computed as below[19,20].

$$Accuracy = \frac{TP_i + TN_i}{Total\ number\ of\ Sample} \times 100 \quad (3)$$

$$Sensitive = \frac{TP_i}{TP_i + FP_i} \times 100 \quad (4)$$

$$Specificity = \frac{TN_i}{TN_i + FN_i} \times 100 \quad (5)$$

5. Results and Discussion

The experiments carried out to demonstrate the efficiency of our framework and the results obtained are summarized in this section.

5.1 Performance Analysis with Different Ranking Filters

In this experimental study, the impact of different filtering methods in selecting the most promising biomarkers that can best discriminate the target classes eliminating the irrelevant one is analyzed. For this preliminary screening, the number of candidate biomarkers was set to 25. These

Table 2. Biomarker subset selected by different ranking filter

Ranking Methods	Feature Set
Chi-Square	Tau, Ab_42, GRO_alpha, MMP10, B_Lymphocyte_Chemoattractant_BL, Eotaxin_3, p_tau, PAI_1, FAS, TRAIL_R3, Pancreatic_polypeptide, NT_proBNP, MMP7, Genotype, age, IGF_BP_2, MIF, MIP_1alpha, Gamma_Interferon_induced_Monokin, Creatine_Kinase_MB
Information Gain	tau, Ab_42, B_Lymphocyte_Chemoattractant_BL, GRO_alpha, p_tau MMP10, Eotaxin_3, PAI_1, MMP7, TRAIL_R3, Pancreatic_polypeptide, Creatine_Kinase_MB, Gamma_Interferon_induced_Monokin, age, Genotype, MIF, FAS, NT_proBNP
SAM	Apolipoprotein_D, IL_7, IL_8, Ab_42, tau, GRO_alpha, FAS, PAI_1, Eotaxin_3, B_Lymphocyte_Chemoattractant_BL, MMP10, p_tau, TRAIL_R3, NT_proBNP, Pancreatic_polypeptide, TNF_RII, MMP7, Genotype, IGF_BP_2, Fatty_Acid_Binding_Protein
Correlation	tau, Ab_42, p_tau, GRO_alpha, MMP10, TRAIL_R3, PAI_1, Pancreatic_polypeptide, NT_proBNP, FAS, MIF, Fibrinogen, MMP7, Gamma_Interferon_induced_Monokin, Eotaxin_3, IGF_BP_2, age, Thymus_Expressed_Chemokine_TECK, Resistin, TNF_RII

Table 3. Performance analysis of our proposed framework with different ranking filters

Classifiers	Without filters	Chi-square	IG	SAM	Correlation
SVM	82.5%	82%	81.6%	82%	83%
Logistic	80%	82.2%	81%	83.4%	81%
K-NN	77%	78%	80%	82.5%	80
Designed Ensemble Classifier	84%	82.5%	82%	84%	83

25 biomarkers were then fed into the wrapper phase to select the optimal set of biomarkers while maintaining the highest accuracy. The top 25 features selected by different ranking methods such as chi-square, information gain, SAM and correlation are tabulated in Table 2.

For the sake of comparison, Table 3 depicts the classification accuracy of the designed ensemble classifiers and the individual base classifiers SVM, Logistic, KNN and K-Nearest Neighbor without feature selection in column 2 and with features selection by chi-square, Information Gain (IG), SAM and correlation in column 3-5 respectively. Comparing these results, it is apparent that the SAM filter outperforms its counterparts and proves to be the best filter method for AD dataset.

5.2 Performance Analysis with Different Search Methods

In this experimental study, the impact of different search algorithms along with the designed ensemble of classifier within wrapper phase is analyzed for finding the most promising subset of biomarkers. For this feature reselection process, the top 25 biomarkers selected by SAM filter were fed into the wrapper phase and the results of feature reselection are tabulated in column-1 of Table 4.

Table 4. Performance analysis of our proposed framework with different search methods

Search Method	# Feature selected	SVM	Logistic	K-NN	ensemble	Run Time (s)
Greedy stepwise	6 Features { 3,6,8,15,21,23}	85.8%	84.6%	82%	86.7	1.2
PSO	11 Features {1,3,4,6,7,11,15,17, 21,22,23 }	84.6	85.8	84	86	2.6
Incremental	8 Features { 23,21,16,10,1,15,3,6 }	86	85	82.5	86.2	1.5
Best-First	13 Features {1,3,6,7,8,10,13,15,16,18, 19,21,23}	87	87	85	87.6	3.3

For sake of comparison, the obtained final set of features (column 2) is analyzed for its classification accuracy with the designed ensemble classifier (column 6) as well with the individual base classifiers (column 3-5). Also run time on i5 processor is noted and included in column 7.

Comparing the results in the above table, it can be inferred that greedy search algorithm outperforms other algorithms not only in execution time but also in defining small set of features while maintaining good classification accuracy with both ensemble classifier and its base classifiers.

5.3 Performance Analysis of the Biomarkers Selected by the Proposed Framework

These subsection summaries the effectiveness of the biomarkers selected using ROC and statistical analysis. .

5.4 ROC Analysis

ROC is 2D graphs that depict the trade-off between sensitivity and specificity. This performance criterion measures the area under ROC curve[21]. It enables to illustrate the behaviour of a classifier irrespective of class distribution and misclassification cost. The ROC curve of the designed ensemble classifiers and its baseline classifier is depicted in Figure 1 for set of biomarkers identified by the proposed framework. Also the confusion matrix of the designed ensemble classifier for the set of biomarkers identified is given in Table 5. Both ROC curve value and the values in the table reveal the significant relationship between the identified set of biomarkers and its predictive power in AD with 93% ***sensitivity*** and specificity.

5.5 Statistical Analysis

In addition to the above assessment of the identified set of biomarkers, bivariate analysis between each pair of the identified biomarkers and the cognitive impairment level was performed to confirm its significant performance and the results are depicted in Figure 2(a). Furthermore, the association levels of the identified biomarkers with the presence of AD were analyzed and are presented in Figure 2(b). Finally, the statistical significance of the identified Biomarkers was analyzed using ANOVA test. Here the P value of less than .05 was considered to indicate statistical significance.

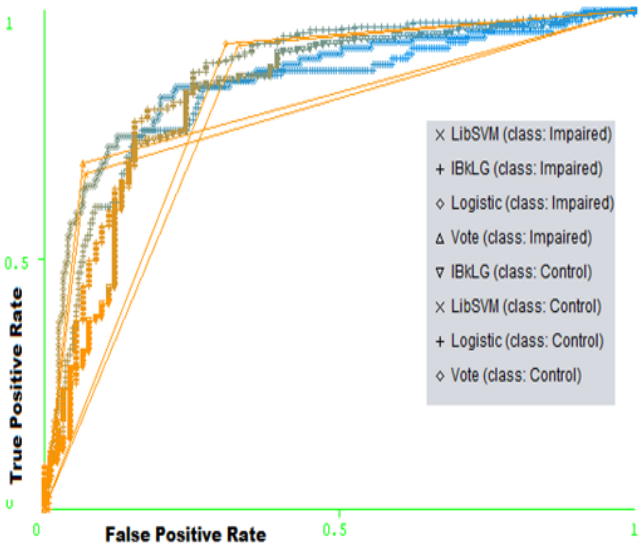
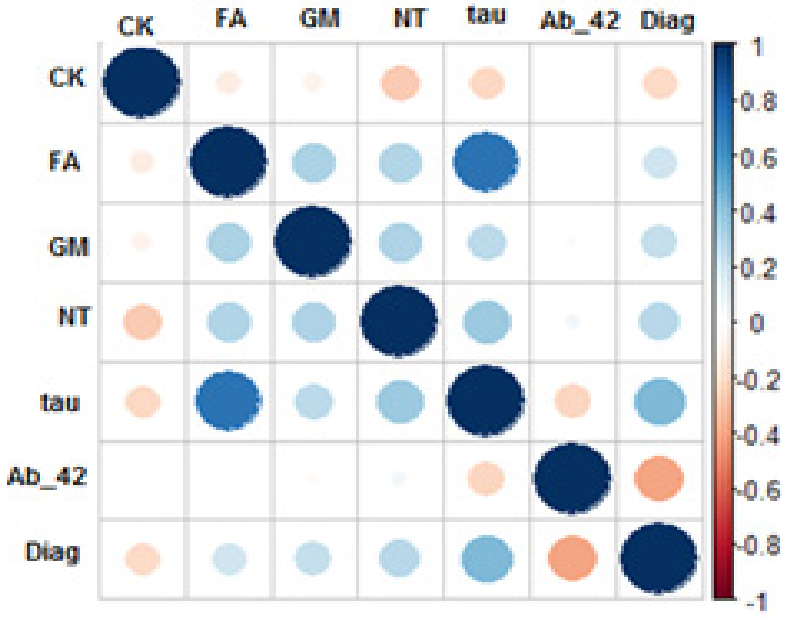


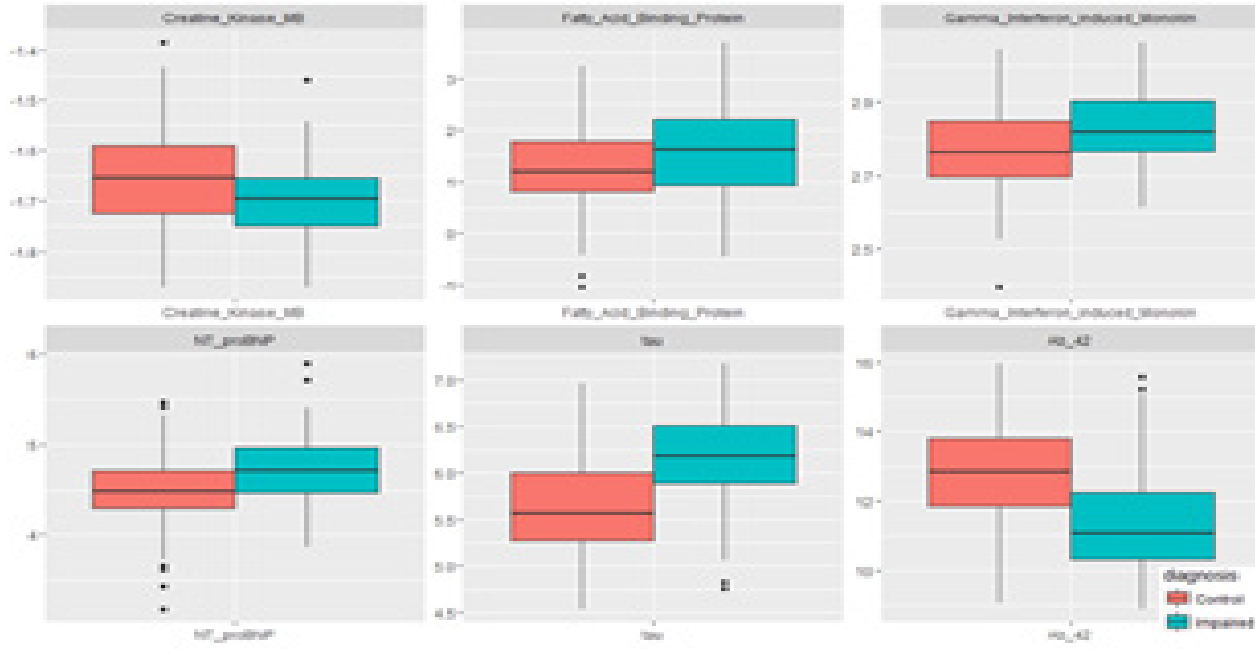
Figure 1. ROC measures of base classifiers and proposed ensemble in detecting Alzheimer impaired samples from healthy control samples.

Table 5. Confusion matrix for the biomarkers subset selected by our framework

	Control	Impaired	Sensitivity	Specificity	ROC Area
Control	225	17	0.930	0.286	0.822
Impaired	26	65	0.714	0.070	0.822



(a)



(b)

Figure 2. (a) Depicts correlation between the identified biomarkers and cognitive decline in control and AD patients. (b) Depicts level of the identified biomarker in control and AD patients.

From last column in Figure 2(a), it is observed that all the identified biomarker are significantly correlated with the cognitive decline in patients with AD. Also the Figure (b) confirms that the level of all identified biomarkers are either significantly increased (tau, NT, FA, GM) or decreased (Ab_42, CK) in AD patients compared to the

Table 6. Summary of ANOVA results for biomarker subset selected by our framework

Biomarker	Diff	Sum Square	Mean Square	F Value	Pr (>F)	Significant Value
Creatine_Kinase_MB (CK)	1	2.52	2.518	19.30	1.51e-05	0.001
Fatty_Acid_Binding_Protein (FA)	1	2.16	2.158	16.54	5.98e-05	0.001
Gamma_Interferon_induced_ Monokin (GM)	1	2.16	2.158	16.54	5.98e-05	0.001
NT_proBNP (NT)	1	2.04	2.040	15.63	9.44 e-05	0.001
Tau	1	9.83	9.829	75.32	<2 e-16	0.001
Ab_42	1	5.38	5.382	41.24	4.75e-10	0.001

control group. Further confirmation by the results from Analysis of Variance (ANOVA) in Table 6 reveals that all the identified biomarkers significantly contribute to the change in cognitive decline in AD patients.

In summary, it is clear from ROC analysis that the identified set of biomarkers can precisely and perfectly classify all AD patients from healthy control subjects with 93% sensitivity and with accuracy of 88%. Further, the significance of the identified biomarkers is revealed from our statistical analysis. Although clinical studies are needed to conform the findings, in the light of the results reported above, it is evident that the final set of biomarkers from our framework will stand out to be precise and robust biomarkers for AD diagnosis with sensitivity of 93% and specificity of 71%.

6. Conclusion

A hybrid filter-wrapper framework is presented for AD biomarker identification. Leveraging the best practices from previous literature, SAM was used initially to select the most promising 25 biomarkers and reduce the search space for subsequent tasks. In wrapper phase, a heterogeneous voting-based ensemble of classifiers was explored for its potential in selecting the best combination of AD biomarkers that can help to predict AD with high accuracy. The stability and predictive performance of the selected biomarkers was analyzed against designed ensemble of classifiers and its individual base classifier in terms of sensitivity, specificity and ROC value. Results encouraged performing statistical analysis and confirming the effectiveness of the identified biomarkers for detecting AD in its early stage.

7. References

1. Shui B, Tao D, Florea A, Cheng J, Zhao Q, Gu Y, Li W, Jaffrezic-Renault N, Mei Y, Guo Z. Biosensors for Alzheimer's Disease biomarker detection: A review. Biochimie. 2018; 147:13–24. PMid: 29307704. Crossref. Crossref.
2. Long J, Pan G, Ifeachor E, Belshaw R, Li X. Discovery of novel biomarkers for Alzheimer's Disease from blood. Disease Markers. 2016; 2016:1–9.
3. Blennow K, Dubois B, Fagan AM, Lewczuk P, de Leon MJ, Hampel H. Clinical utility of cerebrospinal fluid biomarkers in the diagnosis of early Alzheimer's Disease. Alzheimer's

and Dementia: The Journal of the Alzheimer's Association. 2015; 11(1):58–69.
4. Blennow K, Zetterberg H. The application of cerebrospinal fluid biomarkers in early diagnosis of Alzheimer Disease. Medical Clinics. 2013; 97(3):369–76. Crossref.
5. Frisoni GB, Boccardi M, Barkhof F, Blennow K, Cappa S, Chiotis K, Demonet JF, Garibotto V, Giannakopoulos P, Gietl A, Hansson O. Strategic roadmap for an early diagnosis of Alzheimer's Disease based on biomarkers. The Lancet Neurology. 2017; 16(8):661–76. Crossref.
6. He Z, Yu W. Stable feature selection for biomarker discovery. Computational Biology and Chemistry. 2010; 34(4):215–25. PMid: 20702140. Crossref.
7. Zhou W, Dickerson JA. A novel class dependent feature selection method for cancer biomarker discovery. Computers in Biology and Medicine. 2014; 47:66–75. PMid: 24561345. Crossref.
8. Hsu HH, Hsieh CW, Lu MD. Hybrid feature selection by combining filters and wrappers. Expert Systems with Applications. 2011; 38(7):8144–50. Crossref.
9. Liu HC, Peng PC, Hsieh TC, Yeh TC, Lin CJ, Chen CY, Hou JY, Shih LY, Liang DC. Comparison of feature selection methods for cross-laboratory microarray analysis. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2013; 10(3):593–604. PMid: 24091394. Crossref.
10. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences. 2001; 98(9):5116–21. PMid: 11309499 PMCid: PMC33173. Crossref.
11. Abeel T, Helleputte T, Van de Peer Y, Dupont P, Saeys Y. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. Bioinformatics. 2009 Nov 25; 26(3):392–8. PMid: 19942583. Crossref.
12. Pes B, Dessì N, Angioni M. Exploiting the ensemble paradigm for stable feature selection: A case study on high-dimensional genomic data. Information Fusion. 2017; 35:132–47. Crossref.
13. Pradhan B, Jebur MN. Spatial prediction of landslide-prone areas through K-Nearest Neighbor algorithm and logistic regression model using high resolution airborne laser scanning data. Laser Scanning Applications in Landslide Assessment; 2017. p. 151–65. Crossref.
14. Son H, Kim C, Hwang N, Kim C, Kang Y. Classification of major construction materials in construction environments using ensemble classifiers. Advanced Engineering Informatics. 2014; 28(1):1–10. Crossref.
15. Oreski G, Oreski S. An experimental comparison of classification algorithm performances for highly imbalanced datasets. 25th Central European Conference on Information and Intelligent Systems; Varazdin, Croatia. 2014. p. 1–4.
16. Albukhanajer WA, Jin Y, Briffa JA. Classifier ensembles for image identification using multi-objective Pareto features. Neurocomputing. 2017; 238:316–27. Crossref.
17. Craig-Schapiro R, Kuhn M, Xiong C, Pickering EH, Liu J, Misko TP, Perrin RJ, Bales KR, Soares H, Fagan AM, Holtzman DM. Multiplexed immunoassay panel identifies novel CSF biomarkers for Alzheimer's Disease diagnosis and prognosis. PloS one. 2011; 6(4):18850. PMid: 21526197 PMCid: PMC3079734. Crossref.
18. Fergus P, Selvaraj M, Chalmers C. Machine learning ensemble modeling to classify caesarean section and vaginal delivery types using Cardiotocography traces. Computers in Biology and Medicine. 2018; 93:7–16. PMid: 29248699. Crossref.
19. Singh BK, Verma K, Thoke AS, Suri JS. Risk stratification of 2D ultrasound-based breast lesions using hybrid feature selection in machine learning paradigm. Measurement. 2017; 105:146–57. Crossref.
20. Ko J, Baldassano SN, Loh PL, Kording K, Litt B, Issadore D. Machine learning to detect signatures of disease in liquid biopsies- a user's guide. Lab on a Chip. 2018; 18(3):395–405. PMid: 29192299. Crossref.
21. Baesens B, Van Gestel T, Viaene S, Stepanova M, Suykens J, Vanthienen J. Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the Operational Research Society. 2003; 54(6):627–35. Crossref.