E-ISSN: 2717-2937 | P-ISSN: 2783-1337

International Journal of Research in Industrial Engineering

www.riejournal.com

Int. J. Res. Ind. Eng. Vol. 12, No. 2 (2023) 197–204.

Paper Type: Research Paper

How Metaheuristic Algorithms Can Help in Feature Selection for Alzheimer's Diagnosis

Farzaneh Salami[1], Ali Bozorgi-Amiri[2,*], Reza Tavakkoli-Moghaddam[2]

[1] Department of Industrial Engineering, Alborz Campus, University of Tehran, Tehran, Iran; farzane.salami@ut.ac.ir.

[2] Department of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran; alibozorgi@ut.ac.ir; tavakoli@ut.ac.ir.

Citation:

Salami, F., Bozorgi-Amiri, A., & Tavakkoli-Moghaddam, R. (2023). How metaheuristic algorithms can help in feature selection for Alzheimer's diagnosis. *International journal of research in industrial engineering, 12*(2), 197-204.

Received: 19/06/2022 Reviewed: 21/07/2022 Revised: 08/09/2022 Accepted: 11/10/2022

Abstract

Feature selection is the process of picking the most effective feature among a considerable number of features in the dataset. However, choosing the best subset that gives a higher performance in classification is challenging. This study constructed and validated multiple metaheuristic algorithms to optimize Machine Learning (ML) models in diagnosing Alzheimer's. This study aims to classify Cognitively Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer's by selecting the best features. The features include Freesurfer features extracted from Magnetic Resonance Imaging (MRI) images and clinical data. We have used well-known ML algorithms for classifying, and after that, we used multiple metaheuristic methods for feature selection and optimizing the objective function of the classification. We considered the objective function a macro-average F1 score because of the imbalanced data. Our procedure not only reduces the irreverent features but also increases the classification performance. Results showed that metaheuristic algorithms could improve the performance of ML methods in diagnosing Alzheimer's by 20%. We found that classification performance can be significantly enhanced by using appropriate metaheuristic algorithms. Metaheuristic algorithms can help find the best features for medical classification problems, especially Alzheimer's.

Keywords: Metaheuristic algorithm, Alzheimer's disease, MRI, Machine learning, Feature selection, Data mining.

1 | Introduction

Nowadays, a considerable amount of data is generated daily, and data-related technologies are advancing rapidly to cope. Data Mining (DM) is one of the methods to conclude helpful information from this massive amount of data with the help of Machine Learning (ML), pattern recognition, algorithms, and statistics. ML algorithms employed in solving different problems [1] specially disease diagnosis. Alzheimer's Disease (AD) is one of the problems, with a large amount of data produced for diagnosing and treating it. The exact causes of AD are not yet clear. However, there are some risk factors with proven effects on the cognitive function of the brain such as White Matter (WM), Gray Matter (GM), and Cerebrospinal Fluid (CSF) CerebroSpinal Fluids (CSFs) [2]. These biomarkers can be found by examining different types of imaging modalities like Magnetic Resonance Imaging (MRI), Functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET), etc. There are also other neuropsychology tests used to classify Alzheimer's such as Clinical Dementia

Corresponding Author: alibozorgi@ut.ac.ir

https://doi.org/10.22105/riej.2023.347524.1321

Rating (CDR), Mini-Mental State Examinations (MMSE), and AD Assessment Scale Cognitive (ADAS-Cog) [3]. However, physicians still have difficulty predicting Alzheimer's because these biomarker variations are difficult to generalize. This is where DM and ML models, that are trained on multiple factors, can help in diagnosing Alzheimer's [4], [5]. Raw data from MRI scans need to be adjusted and preprossedsed to be used by before-mentioned models. Preprossessing data also helps with models accuracy [6]. One commonly used methods in data preprocessing is Feature Selection (FS) [6], which tries to select the smallest and most valuable features from data, leading to better classification performance. As well as performance, FS can speed up the decision-making process. The FS process is considered an NP-hard searching problem [7]. When there are n features, it will be a 2n subset of features that can be used, and when n is large, the computation cost increases. Therefore, for high-dimensional data, exhaustively generating all possible subsets becomes impractical and computationally intensive. Forward sequential, backward sequential, random, and heuristic searches are all examples of search strategies that can be used to improve the efficiency of the FS method. Recently, metaheuristic algorithms have been successfully applied to many FS methods [8]. As there is still a lot of work to do to find the main features of AD in this paper, we proposed a novel method for selecting the best features among many extracted from MRI images in combination with other health factors. Our proposed approach focuses on the choice of feature selection algorithm to overcome computation cost problems employimg metaheuristic methods. Also, classified AD into three different groups. The article is planned as follows. The next section discusses the literature review of other authors who have used DM and ML algorithms to analyze Alzheimer's. Section 3 describes the proposed technique used for feature extraction from MRI images and optimizing the result by metaheuristics. Section 4 discusses the experiments and evaluation results. Finally, Section 5 presents the paper summary and conclusions.

Table 1. Demographic details of OASIS-3 subjects.

	Male	Female
Number of subjects	487	611
Age	70.17 (42.5-91.7)	67.78 (43.2-95.6)
Right-handed	433	546

2 | Literature Review

Healthcare costs today are much higher due to technological advances and demographic changes. Planning and managing treatment resources and facilities is critical to controlling and reducing these costs and providing the desired services to patients [9]. Multiple criteria are needed to be managed in health care like infection control surveillance, diagnosis and treatment of various diseases, healthcare resource management, customer relationship management, healthcare administration, hospital management and public health.

Artificial Intelligence (AI) models played a successful role in health care issues, especially managing detection, determination, and disease prediction costs [10]. Hariri et al. [11] developed a model for diagnosis of hear failure using AI models to manage treatment costs. In all the AI models developed for disease classification, feature selection was the most challenging task. It involves selecting the most distinct and relevant subset of features from a large set of features to represent the dataset. In recent years we have seen studies on feature selection in medical imaging. For example, an ML model developed by Zhang et al. [5] combines AD baseline features from MRI images, hypometabolism, and CSF to classify AD vs CN. In the proposed model, they used a multiple-kernel support vector machine. Non-imaging biomarkers with CSF and clinical data were also used to develop and compare ML models to classify CN versus Mild Cognitive Impairment (MCI) [12]. More recently, Stamate et al. [13] compared three ML models to classify CN versus AD, which XGBoost claimed to be the best one that used blood metabolite data with clinical and cognitive data. Demir et al. [14] implemented a convolutional neural network to extract features from MRI images. Multiple methods have been proposed to solve the feature selection problem, such as random and greedy search, exhaustive search, etc. Most of these methods suffer significant complications and high computational costs.

Salami et al. | Int. J. Res. Ind. Eng. 12(2) (2023) 197-204

Metaheuristics provide practical solutions in a considerable amount of time, optimizing classification performance and overcoming the curse of dimensionality by alleviating the consumption of large functions such as computational resources, storage, etc. Bahmani et al. [15] employed improved Whale Optimization Algorithm (WOA) and for flow shop scheduling and vehicle routing and showed efficiency of improved WOA in converging to optimal solution and achieve better solution in comparison to the Genetic Algorithm (GA). Dirik [16] applied fuzzy GA to detection of counterfeit banknotes. Zhang et al. [17] fit GA to solve the appointment scheduling problem.

Therefore, metaheuristic algorithms as a solution received extra attention in classification. For example, Negahbani et al. [18] examined coronary artery disease and proposed an evolution based search algorithm with fuzzy c-means classifier. Al-Tashi et al. [19] used Grey Wolf Optimizer (GWO) to select the best features and used SVM as a classifier for diagnosing cardiovascular disease. Mafarja and Mirjalili [20] applied Simulated Annealing (SA) algorithm as a local search method to enhance the WOA to choose the optimal subset of features to develop classification accuracy on datasets from the UCI data repository [21].

Overall, studies are conducted extensively on various modalities for AD diagnosis, but still, researchers are facing difficulties with finding the best features. Therefore, an active feature selection algorithm is essential in identifying key features. Also, most of the research concentrated on 2 class classifications and did not consider the MCI phase, which is a more critical phase in AD diagnosis.

3 | Research Methodology

In this section, we will discuss the preprocessing methods and steps we did for the classification and optimization of the classification method by feature selection. In the first step, we compared well-known algorithms of ML; Adaboost, GradientBoost, XGBoost, Light GBM, and Gaussian Naive Bayes for our data classification with 192 features extracted from brain images. In the next step, we optimize the ML algorithm using metaheuristic algorithms for feature selection.

The data we have used is from the OASIS-3 dataset, a newly introduced dataset for AD. It was prepared across studies in Washington University Knight Alzheimer disease research center over 15 years. Demographic information about it is illustrated in *Table 1*. Participants in this study ranging from 42 to 95 years old including Cognitively Normal (CN) and different stages of cognitive decline adults. OASIS-3 dataset contains over 2000 MR sessions, including structural and functional sequences. Imaging data is accompanied by dementia, and APOE status, which is the strongest genetic risk factor for AD diagnosis suggested in multiple papers [22]–[24], and longitudinal clinical and cognitive outcomes.

Table 2. Features description.

Variable	Description
SubCortGrayVol	Subcortical GM volume
CortexVol	Total cortical GM volume
IntraCranialVol	ICV, intracranial volume
CorticalWhiteMatterVol	Total cortical WM volume
TotalGrayVol	Total GM volume
SupraTentorialVol	Supratentorial volume
lhCortexVol	Left hemisphere cortical GM volume
rhCortexVol	Right hemisphere cortical GM volume
lhCortical WM Vol	Left hemisphere cortical WM volume
rhCortical WM Vol	Right hemisphere cortical WM volume
APOE	Apolipoprotein E
MMSE	Mini-mental state examination

Salami et al. | Int. J. Res. Ind. Eng. 12(2) (2023) 197-204

Table 3. CDR details of OASIS-3 subjects.

Categories	Max CDR=0	Max CDR=0.5	Max CDR=1	Max CDR>2
Min CDR=0	605	192	39	14
Min CDR=0.5	-	66	61	50
Min CDR>1	-	-	31	36
Total	605	258	131	100

3.1 | Preprocessing

In this study, we used features extracted from T1w MRI images. These images were processed through Freesurfer. FreeSurfer is an imaging tool and the most widely used software for analysing brain images. It is used for the depiction of the cortical surface between white and GM, segmentation of WM from the rest of the brain, skull stripping, and many other purposes. We have used the statistical output from the subcortical segmentation. We also considered APOE and MMSE values because they are important risk factors for AD diagnosis. Some of the features we have used are described in *Table 3*. Overall, there were 2047 records, and from that, we eliminated records with missing APOE or MMSE.

We have created three classes based on the provided CDR scores in which a CDR value equaling zero represents the non-existence of Alzheimer's, a CDR equaling 0.5 represents MCI subjects, and a CDR greater than 0.5 represents the presence of AD. Detailed information about CDR values is shown in *Table 2*. Some patients have multiple records because they had multiple visits; we didn't want to have any overlap between train and test dataset. To overcome this problem for each subject, we kept the records with the biggest CDR value indicating Alzheimer's or MCI. After these steps, the number of data became 1006 records and included 659 CN, 265 MCI, and 82 severe Alzheimer's (AD). It can be seen that data distribution is imbalanced. We split the data as 70% for training and 30% for testing.

The impact of class imbalance on classification performance is a major issue [25], [26]. In these studies, authors discussed the impact of im-balanced classification on accuracy using various examples and showed reporting classification accuracy for a severely imbalanced classification problem could be dangerously misleading. Ac-curacy is not considered a good evaluation measure as the majority class overwhelms the errors of the minority class. It is appropriate to use macro-averaging metrics over micro-averaging to avoid a dominance of majority classes [27]. In this work, we used macro average F1 score as the performance metric and the objective function for optimizing. F1 score is the variant most often used when learning from imbalanced data, which weights precision and recall equally [28]. After preprocessing the data for discrimination of CN, MCI and AD we used Adaboost [29], Gradient Boost [30], XGBoost [31], Light GBM, and Gaussian Naive Bayes [32] classifiers with all the extracted features.

Table 4. Metaheuristic algorithms performance.

Classifier	Default F1-MAVG	NSF	PSO F1-MAVG	NSF	GWO F1-MAVG	NSF	DFO F1-MAVG	NSF	GA F1-MAVG	NSF
AdaBoost	62.55%	198	78.04%	90	76.56%	181	77.00%	94	76.77%	198
Gradient boosting	78.09%	198	83.81%	96	82.40%	192	82.90%	112	83.13%	198
XGBoost	75.65%	198	85.13%	105	84.32%	189	83.60%	91	84.08%	198
LightGBM	75.14%	198	82.97%	89	82.80%	183	82.00%	102	82.37%	198
Gaussian naive bayes	52.92%	198	72.30%	87	67.72%	148	69.20%	87	66.29%	198

*Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO), Dragonfly Optimization (DFO), Genetic Algorithm (GA), F1 score Macro Average (F1 MAVG), Number of Selected Features (NSF).

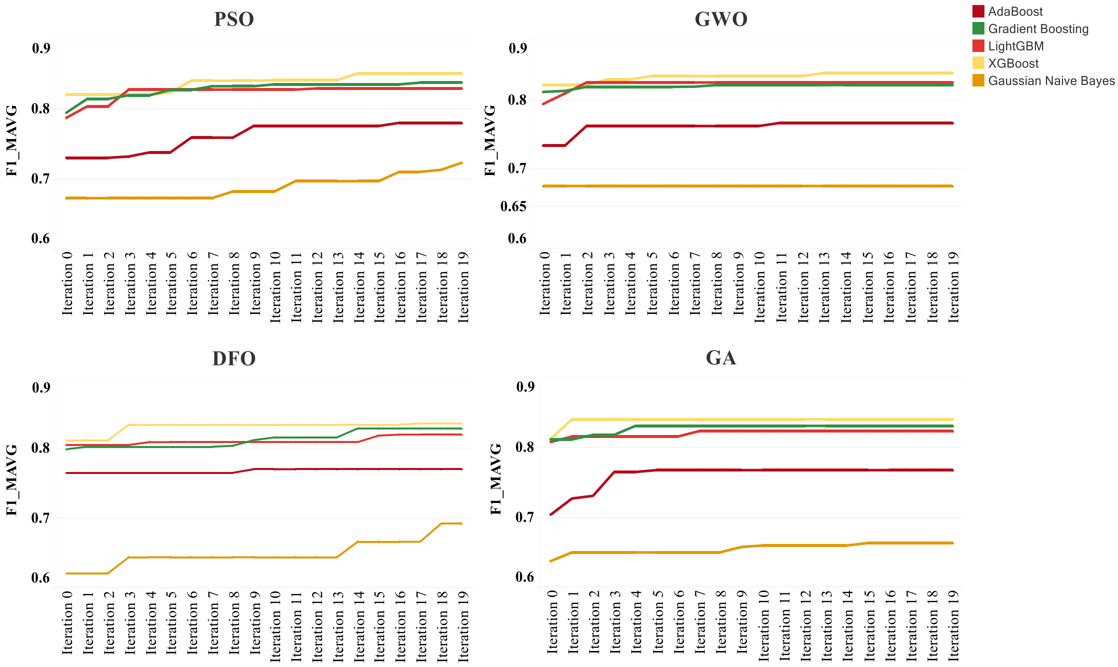

Fig. 1. Different meta-heuristic algorithms' performance on feature selection and different classifiers.

3.2 | Feature Selection

Feature selection has been considered to be an NP-hard problem [33] as it is a challenging problem to find the best subset of features. Therefore, more like other NP-hard problems, a high-performance metaheuristic method is required to reduce processing time. The proposed method in this study aims to compare and propose the best metaheuristic method to find the optimal combination of features. We used Particle Swarm Optimization (PSO), GWO, Dragonfly Optimization (DFO) Algorithm and GA as our metaheuristics algorithms because of their exploration and exploitation ability. PSO is a computational method that optimizes an objective function by iteratively improving candidate solutions with respect to a given quality measure. It takes a population of candidate solutions and solves the problem by moving these particles in the search space according to a simple formula via their positions and velocities. Each particle's motion is influenced by its best-known position locally. It also leads to the best-known position in the search space that is updated as better positions found by other particles. GWO mimics the natural gray wolf leadership hierarchy and hunting mechanisms. GWO implements and optimizes multiple steps: hunt, loot search, loot siege and loot attack. The DFO algorithm is derived from static and dynamic swarm behavior. These two herd behaviors are very similar to exploration and exploitation in the metaheuristic phase. Dragonflies form subflocks and fly in static flocks over different regions, which is the main purpose of the exploration phase. In static flocks, on the other hand, dragonflies fly in one direction in larger flocks. This is an advantage during the recovery phase. GA was inspired by the process of natural selection, which belongs to the larger class of Evolutionary Algorithms (EAs). GAs, which rely on biologically-inspired operators such as mutation, crossover, and selection, are widely used to generate high-quality solutions to optimization and search problems. We considered the objective function as macro average F1 score. For each algorithm, we have trained for 20 iterations to reach a maximum macro average F1 score in Alzheimer's classification.

OASIS-3 has been used in experiments that demonstrated the proposed approach produces a statistically significant compact set of features and improves the F1 score in classification. Results showed that PSO outperformed other methods in selecting several features and leads to improving classification metrics.

4 | Results and Discussion

In this research, multiple experiments are set up to analyze the performance of metaheuristic feature selection methods among different classifiers for diagnosing Alzheimer's. Five benchmark classifiers AdaBoost, Gradient Boosting, XGBoost, LightGBM, and Gaussian Naive bayes were applied in these experiments to evaluate the performance of feature selection methods. This section discusses the experimental results of different classifiers with PSO, GWO, DFO, and GA feature selection methods on the OASIS-3 dataset.

Table 4 outlines the Macro average F1 score obtained over different algorithms. The training process is also shown in *Fig. 1*. We performed multiple experiments comparing the effect of metaheuristic methods on optimizing the classifiers. Results show metaheuristics increase the macro average F1 score between 4.31% to 19.38%. We can remark that based on the selected features by different optimizers, the performance of the PSO with the XGBoost classifier outperforms other optimizers and classifiers because it reaches 85.13% macro average F1 score which is about 9.48% better than using XGBoost alone. This proves PSO optimized XGBoost classifier's future performance on the unseen data, and hence it can be used as a candidate method for Alzheimer's diagnosis.

5 | Conclusion

This paper tackles the optimization of classification methods for AD diagnosing problems. The ML classification methods are modified, and feature selection is applied in hybrid mode. Four metaheuristic methods are proposed besides ML methods and compared against the previous ones. To assess our proposed method, OASIS3 dataset is used. This dataset contains MRI images for people aged 42 to 95 years old. The dataset is highly imbalanced and finding a proper solution for AD classification in this dataset is difficult.

Evaluation is performed using appropriate measures (Macro-average F1 score) for imbalanced data and results compared. We found that the performance of classifiers can be improved significantly by using appropriate features, and metaheuristic algorithms can help to find the best features for medical classification problems. The proposed method can be improved by using other metaheuristics. Furthermore, other ML or deep learning algorithms can be applied to investigate the solutions. Finally, another way of future research could be to study the effect of hyperparameters values on performance.

Acknowledgments

OASIS-3: principal investigators: T. Benzinger, D. Marcus, J. Morris; NIH P50 AG00561, P30 NS09857781, P01 AG026276, P01 AG003991, R01 AG043434, UL1 TR000448, R01 EB009352. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly-owned subsidiary of Eli Lilly.

Conflicts of Interest

All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report. We certify that the submission is original work and is not under review at any other publication.

Salami et al. | Int. J. Res. Ind. Eng. 12(2) (2023) 197-204

References

[1] Mohammadi, S., Hemati, N., & Mohammadi, N. (2022). Speech recognition system based on machine learning in persian language. *Computational algorithms and numerical dimensions*, *1*(2), 72–83.

[2] Zimmermann, J., Perry, A., Breakspear, M., Schirner, M., Sachdev, P., Wen, W., … & Solodkin, A. (2018). Differentiation of Alzheimer's disease based on local and global parameters in personalized virtual brain models. *NeuroImage: clinical*, *19*, 240–251.

[3] Salmon, D. P., & Bondi, M. W. (2009). Neuropsychological assessment of dementia. *Annual review of psychology*, *60*, 257–282. https://doi.org/10.1146/annurev.psych.57.102904.190024

[4] Zhang, D., & Shen, D. (2012). Alzheimer's disease neuroimaging initiative. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *NeuroImage*, *59*(2), 895–907. https://doi.org/10.1016/j.neuroimage.2011.09.069

[5] Zhang, D., Wang, Y., Zhou, L., Yuan, H., & Shen, D. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage*, *55*(3), 856–867.

[6] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, *3*(Mar), 1157–1182. https://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf?ref=driverlayer.com/web

[7] Faris, H., Al-Zoubi, A. M., Heidari, A. A., Aljarah, I., Mafarja, M., Hassonah, M. A., & Fujita, H. (2019). An intelligent system for spam detection and identification of the most relevant features based on evolutionary random weight networks. *Information fusion*, *48*, 67–83. DOI:https://doi.org/10.1016/j.inffus.2018.08.002

[8] Zelinka, I. (2015). A survey on evolutionary algorithms dynamics and its complexity-mutual relations, past, present and future. *Swarm and evolutionary computation*, *25*, 2–14. https://www.sciencedirect.com/science/article/pii/S2210650215000437

[9] Ala, A., & Chen, F. (2020). Alternative mathematical formulation and hybrid meta-heuristics for patient scheduling problem in health care clinics. *Neural computing and applications*, *32*, 8993–9008.

[10] Aminoshariae, A., Kulild, J., & Nagendrababu, V. (2021). Artificial intelligence in endodontics: current applications and future directions. *Journal of endodontics*, *47*(9), 1352–1357.

[11] Hariri, A. H., Bagheri, E., & Davoodi, S. M. R. (2022). Presenting a model for the diagnosis of heart failure using cumulative and deep learning algorithms: a case study of tehran heart center. *Big data and computing visions*, *2*(1), 18–30. https://www.bidacv.com/article_145116.html

[12] Hassan, S. A., & Khan, T. (2017). A machine learning model to predict the onset of alzheimer disease using potential cerebrospinal fluid (CSF) biomarkers. *International journal of advanced computer science and applications*, *8*(12). https://pdfs.semanticscholar.org/f8f0/87dde0db02dd5bdf1ddfef61ee39b793d8ce.pdf

[13] Stamate, D., Kim, M., Proitsi, P., Westwood, S., Baird, A., Nevado-Holgado, A., … & Legido-Quigley, C. (2019). A metabolite-based machine learning approach to diagnose Alzheimer-type dementia in blood: results from the european medical information framework for alzheimer disease biomarker discovery cohort. *Alzheimer's & dementia: translational research & clinical interventions*, *5*(1), 933–938.

[14] Gu, F., Ma, S., Wang, X., Zhao, J., Yu, Y., & Song, X. (2022). Evaluation of feature selection for Alzheimer's disease diagnosis. *Frontiers in aging neuroscience*, *14*, 924113. https://pubmed.ncbi.nlm.nih.gov/35813964/

[15] Bahmani, V., Adibi, M. A., & Mehdizadeh, E. (2023). Integration of two-stage assembly flow shop scheduling and vehicle routing using improved whale optimization algorithm. *Journal of applied research on industrial engineering*, *10*(1). https://doi.org/10.22105/jarie.2022.329251.1450

[16] Dirik, M. (2022). Detection of counterfeit banknotes using genetic fuzzy system. *Journal of fuzzy extension and applications*, *3*(4), 302–312. https://www.journal-fea.com/article_154622.html

[17] Zhang, J., Chung, H. S. H., & Lo, W. L. (2007). Clustering-based adaptive crossover and mutation probabilities for genetic algorithms. *IEEE transactions on evolutionary computation*, *11*(3), 326–335. DOI:10.1109/TEVC.2006.880727

[18] Negahbani, M., Joulazadeh, S., Marateb, H. R., & Mansourian, M. (2015). Coronary artery disease diagnosis using supervised fuzzy c-means with differential search algorithm-based generalized Minkowski metrics. *Archive of biomedical science and engineering*, *1*(1), 006-0014. https://www.peertechzpublications.org/articles/ABSE-1-102.php

[19] Al-Tashi, Q., Rais, H., & Jadid, S. (2019). Feature selection method based on grey wolf optimization for coronary artery disease classification. *Recent trends in data science and soft computing: proceedings of the 3rd international conference of reliable information and communication technology (IRICT 2018)* (pp. 257-266). Springer International Publishing. https://doi.org/10.1007/978-3-319-99007-1_25

[20] Mafarja, M. M., & Mirjalili, S. (2017). Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing*, *260*, 302–312.

[21] UCI. (2015). *UCI repository of machine learning databases*. https://archive.ics.uci.edu/

[22] Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., … & Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia*, *7*(3), 280–292.

[23] Crous-Bou, M., Minguillón, C., Gramunt, N., & Molinuevo, J. L. (2017). Alzheimer's disease prevention: from risk factors to early intervention. *Alzheimer's research & therapy*, *9*(1), 71. https://doi.org/10.1186/s13195-017-0297-z

[24] Van Cauwenberghe, C., Van Broeckhoven, C., & Sleegers, K. (2016). The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genetics in medicine*, *18*(5), 421–430. https://doi.org/10.1038/gim.2015.117

[25] Chawla, N. V. (2010). Data mining for imbalanced datasets: an overview. In *Data mining and knowledge discovery handbook* (pp. 875-886). Springer, Boston, MA. https://doi.org/10.1007/978-0-387-09823-4_45

[26] Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, *5*(2), 1-11. https://www.academia.edu/download/37219940/5215ijdkp01.pdf

[27] Choeikiwong, T., & Vateekul, P. (2016). *Improve accuracy of defect severity categorization using semi-supervised approach on imbalanced data sets* [presentation]. Proceedings of the international multi conference of engineers and computer scientists (Vol. 1). https://www.iaeng.org/publication/IMECS2016/IMECS2016_pp434-438.pdf

[28] He, H., & Ma, Y. (2013). *Imbalanced learning: foundations, algorithms, and applications*. Wiley-IEEE Press. http://dx.doi.org/10.1002/9781118646106

[29] Freund, Y., & Schapire, R. E. (1996). *Experiments with a new boosting algorithm* [presentation]. Proceedings of the thirteenth international conference on international conference on machine learning (Vol. 96, pp. 148–156). https://dl.acm.org/doi/10.5555/3091696.3091715

[30] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *The annals of statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

[31] Chen, T., & Guestrin, C. (2016). *Xgboost: a scalable tree boosting system* [presentation]. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785–794). https://dl.acm.org/doi/abs/10.1145/2939672.2939785

[32] Rish, I. (2001). *An empirical study of the naive bayes classifier* [presentation]. IJCAI 2001 workshop on empirical methods in artificial intelligence (Vol. 3, No. 22, pp. 41-46). https://www.researchgate.net/publication/228845263_An_Empirical_Study_of_the_Naive_Bayes_Classifier

[33] Pal, S. K., & Skowron, A. (1999). *Rough-fuzzy hybridization: a new trend in decision making*. Springer-Verlag. https://dl.acm.org/doi/abs/10.5555/553340

Salami et al. | Int. J. Res. Ind. Eng. 12(2) (2023) 197-204