IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 3, June 2025, pp. 1772~1780
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i3.pp1772-1780

ApDeC: A rule generator for Alzheimer's disease prediction

Sonam Vayaliparambil Maju[1,2], Gnana Prakasi Oliver Siryapushpam[1]
[1]Department of Computer Science and Engineering, Christ University, Bangalore, India
[2]School of Computer Science and Engineering, RV University, Bangalore, India

Article Info

Article history:

Received Mar 28, 2024
Revised Feb 23, 2025
Accepted Mar 15, 2025

Keywords:

Alzheimer's disease
Apriori algorithm
Artificial intelligence
Association rules
Machine learning

ABSTRACT

Artificial intelligence (AI) paved the way and helping hand for the medical practitioners in various aspects and early disease prediction is one among many. Interdisciplinary research studies on the early prediction of diseases are often analyzed based on the accuracy of the prediction model. But how early these diseases can be predicted will not be answered in many of the research studies unless they have a time series data. This work proposes a machine learning model, ApDeC which solves the above-mentioned problem by generating association rules for the early disease prediction of Alzheimer patients. The ApDeC model calculates the probability of occurrence of eleven Alzheimer disease prediction risk factors and identifies the combination of diseases that can lead to Alzheimer disease. The association rules will be generated by considering the observed combination of risk factors. The research introduces an innovative approach that helps in the early prediction of Alzheimer disease from the risk factors/symptoms. The results show the strong correlation of diabetes and blood pressure with Alzheimer disease.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Sonam Vayaliparambil Maju
Department of Computer Science and Engineering, Christ University
Bangalore, Karnataka, India
Email: sonam.maju@res.christuniversity.in

1. INTRODUCTION

In the era of data analytics and artificial intelligence (AI), the prevalence of conditions such as Alzheimer's disease has emerged as a significant 21st-century public health challenge. Despite Alzheimer's being incurable, leveraging data analytics on electronic health records (EHR) have become instrumental in predicting the disease at an early stage [1]. AI algorithms help in understanding the correlation of diseases, long distance monitoring of patients, analyzing the medical images for identifying the foreign bodies and what not [2]. Early disease prediction is one segment that hails the contributions of AI [3], but most of the research work halts with the prediction accuracy but how early a disease can be predicted or the tentative time a disease will take to evolve in the human body is not explored much. The disease's evolution time in a human body change from person to person depending on his/her lifestyle, stress levels, eating habits, and exercise. So, generalizing the time duration or age is practically impossible. However, technically it is possible to prove that a set of symptoms can lead to a particular disease in future and thereby undergoing consistent monitoring and preventive treatments to push the disease for a certain time.

Alzheimer being a neuro degenerative disease is caused by the formation of abnormal deposits of protein in the brain. The brain of a person with Alzheimer's also has lower levels of neurotransmitter, which may cause the remaining cells to communicate with each other less effectively and creates problems with memory and thinking [4]. Alzheimer disease can be predicted early by observing the symptoms and diseases which can lead to the condition of Alzheimer. Studies show diverse diseases like heart disease, diabetes,

blood pressure, and cholesterol as factors that tend to cause Alzheimer disease [5]. It is possible to develop algorithms to select and rank the factors which cause diseases, to analyze the correlation of diseases and to identify the combination of diseases which trigger the condition of other diseases.

Numerous algorithms are available to uncover the associations between the symptoms or the risk factors and to predict the association between diseases [6], [7]. EHR data is being utilized for various clinical and research applications, including disease prediction and association rule mining. Supervised machine learning algorithms are often used in disease prediction research. However, very few articles explore the possibilities of association rules and ensemble models of apriori and other algorithms. Association rules are used to extract matched features from medical records with the application of keyword-based clustering and multiple-criteria decision analysis to diagnose the exact disease of the patient [8], [9]. According to Inamdar *et al.* [10], heart attack risk factors were analyzed using association rules, and authors states the advantages of unveiling the association between risk factors for disease prediction [11] uses frequent pattern growth (FP-Growth) to mine patterns from the consortium to establish a registry for Alzheimer's disease-neuropsychological battery (CERAD-NB) database. These algorithms offer supplementary data analysis, predictive item response capabilities, and aid in clinical decision-making. Finally, Liu *et al.* [12] addresses a few drawbacks of apriori algorithm and proposes a data mining algorithm of association rules combining clustering matrix and pruning strategy.

In this research work, the concepts of supervised and unsupervised machine learning algorithms are used to design an ensemble ApDeC algorithm. Which evaluates the combinations of risk factors and generating rules from the risk factors. This helps in the early prediction Alzheimer disease and assists the health practitioners for further monitoring and treatments.

2. METHOD

Generally, all the Alzheimer prediction algorithms are successful in detecting the disease from the given risk factors, but not directing to the early prediction of the disease. Alzheimer being the non-curable disease only the early prediction can help the medical practitioners to treat from the early stage. Hence, this research analyzes the correlation of diseases, unveiling the association of risk factors and consistent monitoring of diseases. To attain this, an ensembled algorithm named ApDeC algorithm was proposed which entails by creating permutations and combinations of risk factors for Alzheimer disease prediction, employing decision trees to extract association rules. Subsequently, the association rules are tested by considering different performance metrics of apriori algorithm, such as support, lift and confidence which helps to identify the antecedent and consequent risk factors of the early Alzheimer's disease prediction. The flow of the proposed model is shown in Figure 1.

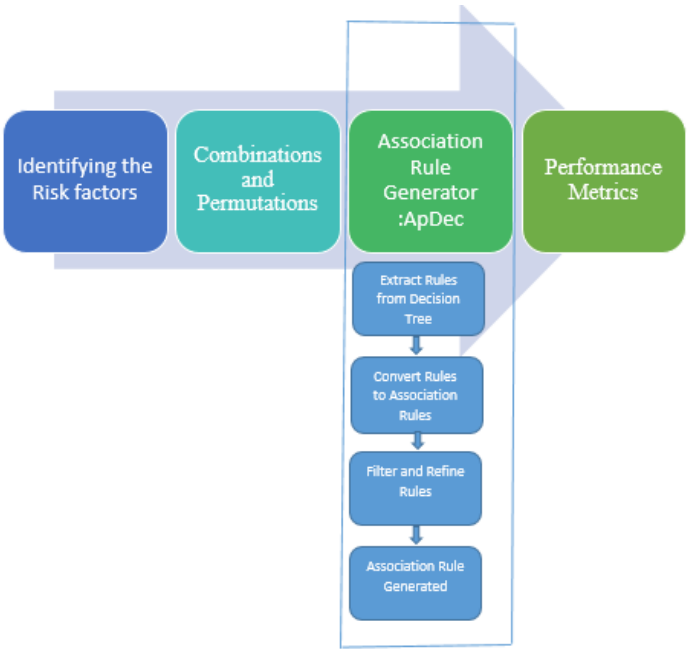

Figure 1. Workflow of ApDec algorithm

The proposed model starts by identifying the Alzheimer triggering risk factors and is obtained from existing literature reviews. Maju and Pushpam [13] shortlisted eleven Alzheimer triggering risk factors and is being considered in this research work. From the obtained risk factors, the combination of risk factors is generated, and the probability of each combination is calculated. Later, the ApDeC algorithm creates the antecedents and consequents from the combinations of risk factors through association rules. Initially, a decision tree is trained on a dataset using an iterative dichotomiser 3 (ID3) algorithms. The decision tree learns to make decisions based on the features of the data. Once the decision tree is trained, rules are extracted from it. The rules extracted from the decision tree can be converted into association rules. Association rules typically have the form "X=>Y," where X is the antecedent (the risk factors) and Y is the consequent (the Alzheimer disease). After filtering, the refined rules can be considered as association rules that provide insights into relationships between risk factors in the dataset and the performance is evaluated by calculating the lift, support, and confidence.

2.1. Combinatorics and probability

The first step is to identify the association between possible risk factors of Alzheimer disease. The principles of combinatory are used to extract all possible combinations of these risk factors. Suppose if risk factor 1 and risk factor 2 is present, then the probability of having Alzheimer in the future will be high. Using the formula given in (1), all possible combinations have been obtained. Thus, a function systematically generates permutations of variables to explore different combinations and is shown in Table 1.

$$nC_r = \frac{n!}{r!(n-r)!} \quad (1)$$

To calculate the probability of each combination of risk factors by their occurrence in the EHR. For this, the calculation is done with each number of combinations selected against the total number of combinations created. To calculate the probability of each combination created. Finally, a mapping between combinations of risk factors and their corresponding probabilities was calculated for further evaluation of ApDeC algorithm. In a decision tree for classification, each leaf node represents a class or outcome. To calculate the probability associated with a specific combination of features that leads to a particular leaf node, distribution of instances in the training data that reached that leaf node during the construction of the tree will be observed.

For each leaf node in the decision tree, examine the training data instances that reached that leaf. Count how many instances belong to each class (e.g., Alzheimer's disease or no Alzheimer's disease). Calculate the probability of each class by dividing the count of instances for that class by the total number of instances at the leaf node. Example: suppose one is interested in the leaf node where the prediction is "predict Alzheimer's disease." If there are 20 instances that reached this leaf, and 15 of them are labelled as having Alzheimer's disease, the probability of Alzheimer's disease would be 15/20=0.75. Similarly, the probability of no Alzheimer's disease would be 5/20=0.25. This probability is specific to the training data and the combinations of features that lead to that leaf node. In a decision tree, predictions are often made by assigning the class with the highest probability.

Table 1. Combinations of risk factors

Risk factors	Set of 1	Set of 2	Set of 3	Set of 4	Set of 5
RF1	Diabetes	Diabetes, Age	Diabetes, Age, Cholesterol	Diabetes, Age, Cholesterol, Obesity	Diabetes, Age, Cholesterol, Obesity, SES
RF2	Blood Pressure	Diabetes, Cholesterol	Diabetes, Age, Obesity	Diabetes, Age, Cholesterol, Blood Pressure	Diabetes, Age, Cholesterol, Obesity, Blood Pressure
RF3	Cholesterol	Diabetes, Obesity	Diabetes, Age, SES	Diabetes, Age, Obesity, SES	Diabetes, Age, Cholesterol, SES, Blood Pressure
RF4	Obesity	Diabetes, SES	Diabetes, Cholesterol, Obesity	Diabetes, Age, Obesity, Blood Pressure	Diabetes, Age, Obesity, SES, Blood Pressure
RF5	EDUC	Diabetes, Blood Pressure	Diabetes, Cholesterol, SES	Diabetes, Age, SES, Blood Pressure	Diabetes, Cholesterol, Obesity, SES, Blood Pressure
RF6	SES	Age, Cholesterol	Diabetes, Cholesterol, Blood Pressure	Diabetes, Cholesterol, Obesity, SES	Age, Cholesterol, Obesity, SES, Blood Pressure
RF7	Age	Age, Obesity	Diabetes, Obesity, SES	Diabetes, Cholesterol, Obesity, Blood Pressure	

2.2. ApDec: association rule generation

To extract association rules, the algorithm iterates through the leaf nodes. The first step is to constructs a lineage of the node's path from leaf to root, incorporating whether it's a left or right child, the threshold, and the feature used for the split. It identifies the risk factor and threshold that led to a right split and constructs association rules accordingly. A comprehensive set of association rules were generated, which were further used to find the association between the diseases. Apriori algorithm is used identify the best combination of risk factors that can trigger the condition of Alzheimer disease.

2.2.1. Association rule generator through trees

In the dataset, which is used in the research work, has risk factors of Alzheimer disease, along with the general information of the patients. In this specific implementation, the association rules of risk factors are generated for predicting whether the patient will be demented or not, in the given limited tree depths as shown in Figure 2. In the below shown example, level 1 shows the chance of diabetes(D) if the association of cholesterol (Ch) and cardiovascular diseases(C), based on the probability of occurrence. Similarly, possibility of cholesterol, with pressure (P) and kidney (K), the tree will be grown further.

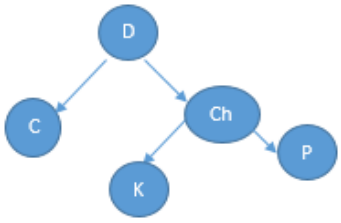

Figure 2. Tree representation of rule generation

2.2.2. Refining and filtering

Evaluate the selected rules by refining and filtering, apriori algorithm is used for mining most specific and frequently reported symptoms and devising association rules from a transactional database. If a patient is having risk factor 1 then the chances of the same patient having risk factor 2 or risk factor 3. So, the medical practitioners can observe the co-occurrence pattern/association followed by several symptoms and advise/notify the patients about the diseases that they can have in future. Medical practitioners can start medication and other necessary precautions to dodge the occurrence of other predicted diseases. Here, we consider the symptoms/risk factors as the antecedents and the disease occurs due to antecedents as consequent. The three matrices that helps us to refine and filter the rules are support, confidence, and lift. So, support calculates the frequency of occurrence of two symptoms and rule out the combinations with low frequency.

$$\text{Support} = \frac{freq(A,B)}{N} \tag{2}$$

Confidence calculated how often the combination of symptoms occurs and can be represented as (3):

$$\text{Confidence} = \frac{freq(A,B)}{freq(A)} \tag{3}$$

Lift is basically the strength of a rule designed. Lift calculates the Independence occurrence probability of symptom A and symptom B.

$$\text{Lift} = \frac{Support}{Supp(A)xSupp(B)} \tag{4}$$

Table 2 shows the metrics calculation. Assumed mi_support is 40% and min_support_count is min_support x item set which will be 1.6. So, in the above example we will discard everything less than 1.6. The algorithm for the above-mentioned methodology is given in Algorithm 1.

Algorithm 1. ApDeC Algorithm

1. if predictionParameters is empty:
2. return [[]] // base case, an empty set has one combination (empty set itself)
3. else:

```
        currentParameter=predictionParameters[0]
        restParameters=predictionParameters[1]:
        withoutCurrent=generatePermutationsAndCombinationsHelper(restParameters)
        withCurrent=[]
        for combination in withoutCurrent:
          withCurrent.append(combination+[currentParameter])
            withCurrent.append(combination) // include the current parameter as a separate combination
        return withCurrent
#Calculate Probabilities
probabilities=calculateProbabilities(combinations)
#Extract Association Rules using Decision Trees
decisionTreeModel=trainDecisionTreeModel(dataset)
associationRules=extractAssociationRulesFromDecisionTree(decisionTreeModel)
# Apply Apriori Algorithm for Best Combinations
bestCombinations=applyAprioriAlgorithm(dataset)
# Evaluate Rules using Support, Confidence, and Lift
supportMatrix=calculateSupportMatrix(dataset, associationRules)
confidenceMatrix=calculateConfidenceMatrix(dataset, associationRules)
liftMatrix=calculateLiftMatrix(supportMatrix, confidenceMatrix)
#Shortlist Rules with Specific Risk Factors
rulesWithDiabetes=filterRulesWithRiskFactor(associationRules, 'Diabetes')
#Define Function to Get Rules for Diabetes
diabetesRules=getRulesForDiabetes(decisionTreeModel)
#Calculate Support, Confidence, and Lift for Diabetes Rules
supportDiabetesRules=calculateSupportMatrix(dataset, diabetesRules)
confidenceDiabetesRules=calculateConfidenceMatrix(dataset, diabetesRules)
liftDiabetesRules=calculateLiftMatrix(supportDiabetesRules, confidenceDiabetesRules)
#Fit Models with Maximum Items in the If Condition
fitModelWithMaxItems(dataset, maxItems=4)
```

Table 2. Metrics calculation

Risk factor combination	Support count
Glyhb, MMSE	4
Glyhb, HDL	1
Glyhb, Chol	2
Glyhb, BMI	2
MMSE, hdl	1
MMSE, Chol	2
MMSE, BMI	2
HDL, Chol	1
HDL, BMI	1
Chol, BMI	1

3. RESULTS AND DISCUSSION

The research work introduced a novel strategy of ApDec that generates association rule, to identify the correlation between the risk factors which can trigger Alzheimer disease and thereby helping in early prediction of the same. Starting from permutation generation followed by probability calculation and utilizing decision trees to extract the association rules, finally, rule metrics analysis is done by evaluating support, confidence and lift. The derived association rules shed light on intricate relationships between different medical conditions. The application of the proposed methodology yielded noteworthy outcomes as shown in Tables 3 and 4.

From the result analysis, patients diagnosed with blood pressure, diabetes, cholesterol and kidney diseases have a perfect association with Alzheimer's, indicating a robust relationship between these conditions as shown in Table 3. Even though this research concentrated on diseases associated with Alzheimer diseases, the results obtained few very interesting associations between the risk factors, as shown in Table 4. Understanding these complex relationships can assist healthcare professionals in diagnosing and managing multiple co-occurring conditions more effectively. The graphical representation of the association between the above-mentioned diseases with Alzheimer is shown in Figures 3 and 4.

Table 3. Result analysis

Rules	Observations
Rule 16: Blood Pressure ➡ (Alzheimers, Diabetes) Support: 5.26% Confidence: 100% Lift: 19.00	This rule suggests that if a patient has Blood Pressure, there's a strong association with both Alzheimers and Diabetes, with a perfect confidence and lift.
Rule 9: Blood Pressure ➡ Alzheimers Support: 5.26% Confidence: 100% Lift: 4.75	Blood Pressure conditions are strongly associated with Alzheimers, emphasizing a potential link between these medical conditions.
Rule 0: Cholesterol and Kidney ➡ Alzheimers Support: 5.26% Confidence: 100% Lift: 4.75	Cholesterol and Kidney issues somewhat relate to the presence of Alzheimers.
Rule 3: Diabetes and Alzheimers ➡ Blood Pressure Support: 5.26% Confidence: 100% Lift: 19.00	If a patient has both Diabetes and Alzheimers, there's a strong association with the presence of Blood Pressure.
Rule 12: Alzheimers, kidney ➡ Cholesterol Support: 5.26% Confidence: 100% Lift: 9.50	The presence of Alzheimers and kidney conditions exhibits a strong association with Cholesterol-related issues, showcasing a significant relationship.
Rule 5: Alzheimers, Blood Pressure ➡ Diabetes Support: 5.26% Confidence: 100% Lift: 3.80	A combination of Alzheimers and Blood Pressure suggests a significant association with Diabetes, highlighting a complex relationship among these conditions.
Rule 2: Alzheimers and Cholesterol ➡ Kidney Support: 5.26% Confidence: 100% lift: 9.50	Alzheimers and Cholesterol together show a strong association with Kidney-related conditions.
Rule 5: Alzheimers and Kidney ➡ Cholesterol Support: 5.26% Confidence: 100% Lift: 9.50	Alzheimers and Kidney conditions often associate with Cholesterol-related issues.

Table 4. Observational results

Rules	Observations
Rule6: cardiovascular disease ➡ cancer Support: 5.26% Confidence: 100% Lift: 19.00	Similarly, cardiovascular disease often co-occurs with cancer, demonstrating a reciprocal relationship.
Rule 17: cancer ➡ (cardiovascular disease, diabetes) Support: 5.26% Confidence: 100% Lift: 19.00	Patients diagnosed with cancer show a perfect association with both cardiovascular disease and diabetes, indicating a robust relationship among these conditions.
Rule 11: cardiovascular disease ➡ cancer Support: 5.26% Confidence: 100% Lift: 19.00	Conversely, cardiovascular disease exhibits a perfect association with cancer, highlighting a reciprocal strong relationship.
Rule 0: blood pressure ➡ diabetes Support: 5.26% Confidence: 100% Lift: 3.80	Patients with blood pressure conditions show a strong association with diabetes, emphasizing a significant correlation between these medical conditions.

The apriori algorithm and decision tree classifier can enhance predictive performance and interpretability in data mining and machine learning tasks, particularly in disease prediction. The apriori algorithm helps reduce the dimensionality of the dataset by selecting only the most significant features, reducing computational complexity. This results in a more manageable dataset for the decision tree classifier to process, potentially improving model performance. The decision tree classifier can focus on the most relevant variables, leading to more accurate and reliable predictions. Both algorithms are inherently interpretable, providing insights into data relationships and patterns [14]–[16] are few papers on apriori algorithm for disease prediction and the results and discussions of these papers helped to identify the research gaps for ApDeC algorithm.

Comparing the numerous classifiers and hybrid methods for predicting various diseases, highlighting the evolving nature of machine learning techniques in medical diagnostics. Naïve Bayes and

decision tree classifiers [17]–[19] have been shown to be outperformed by more sophisticated models like decision trees, studies [20], [21] suggesting that more complex algorithms might be necessary for higher accuracy in predictive models. Hybrid [22] and ensemble methods, such as Gaussian naïve Bayes and random forest classifiers, have shown significant advantages in enhancing prediction accuracy. Feature selection methods have also been shown to improve naïve Bayes classifiers. Bayesian decision tree algorithm [23], combining Bayes' theorem with decision tree structures, has shown promise in improving prediction accuracy. Ensemble methods [16], [24], [25], particularly stacking and voting classifiers, have shown promise in improving prediction accuracy.

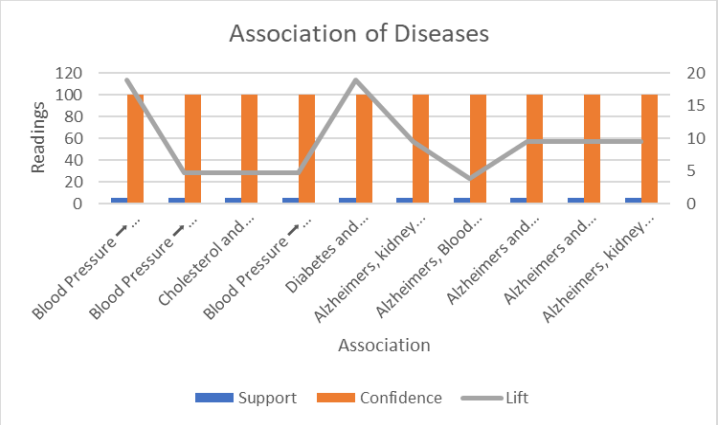

Figure 3. Association rule analysis for Alzheimer disease

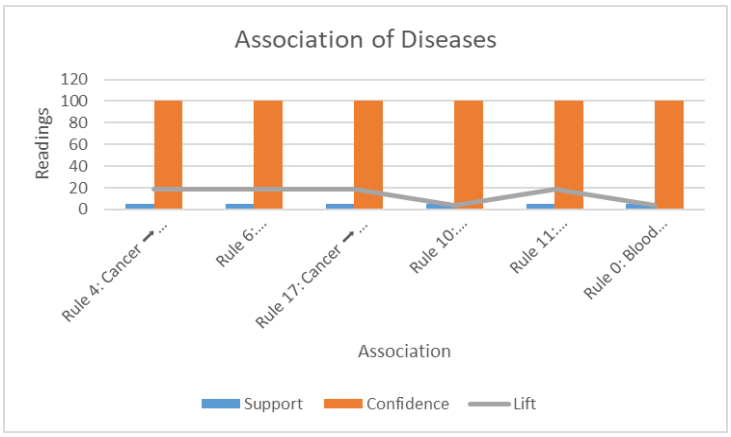

Figure 4. Association rule analysis for other disease

One limitation of our research work is that the model works only with the given risk factors, the number of combinations of risk factors is limited. The predictive capability of models can be improved by including sophisticated feature selection techniques like apriori algorithms for feature selection and machine learning classifiers for classification. By using the most pertinent features, this two-step method guarantees an improvement in overall model correctness. The combination of apriori and decision trees can also reduce overfitting, as the apriori algorithm reduces the complexity of the model and enhances the generalization capability of the decision tree.

The combination also offers scalability and efficiency, making it suitable for large datasets in healthcare. The combination of these strengths creates a more holistic and effective predictive model. Future research should focus on refining hybrid models, exploring parallelization techniques, and employing advanced feature selection methods to further enhance predictive performance.

4. CONCLUSION

This research work delves deeper into the associations between various Alzheimer medical conditions, elucidating intricate relationships identified through association rule mining techniques. The EHR was analyzed to identify frequent item sets and generate association rules based on statistically significant metrics, like support, confidence, and lift metrics. By introducing the ApDeC machine learning model, this research offers a solution to the challenge of early prediction of Alzheimer's disease. Through the analysis of EHR, the risk factors are identified, and the rules are generated with combinations to classify the risk factors which identifies potential combinations leading to Alzheimer's and the combinations which triggers lesser probability towards Alzheimer disease. The performance metrics in this research work clearly shows the rules that triggers the Alzheimer disease. The frequent growth algorithm can also be tried with the decision trees as future work to do an analysis on the presence of frequents sets in the dataset.

FUNDING INFORMATION

Authors state no funding is involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sonam Vayaliparambil Maju	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Gnana Prakasi Oliver Siryapushpam	✓	✓		✓	✓					✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/jboysen/mri-and-alzheimers and https://hbiostat.org/data.

REFERENCES

[1] M. L. Kolling *et al*., "Data mining in healthcare: applying strategic intelligence techniques to depict 25 years of research development," *International Journal of Environmental Research and Public Health*, vol. 18, no. 6, Mar. 2021, doi: 10.3390/ijerph18063099.
[2] N. G. Nia, E. Kaplanoglu, and A. Nasab, "Evaluation of artificial intelligence techniques in disease diagnosis and prediction," *Discover Artificial Intelligence*, vol. 3, no. 1, Jan. 2023, doi: 10.1007/s44163-023-00049-5.
[3] Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz, "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 7, pp. 8459–8486, Jul. 2023, doi: 10.1007/s12652-021-03612-z.
[4] M. W. Bondi, E. C. Edmonds, and D. P. Salmon, "Alzheimer's disease: past, present, and future," *Journal of the International Neuropsychological Society*, vol. 23, no. 9–10, pp. 818–831, Oct. 2017, doi: 10.1017/S135561771700100X.
[5] Z. Breijyeh and R. Karaman, "Comprehensive review on alzheimer's disease: causes and treatment," *Molecules*, vol. 25, no. 24, Dec. 2020, doi: 10.3390/molecules25245789.
[6] R. Kost, B. Littenberg, and E. S. Chen, "Exploring generalized association rule mining for disease co-occurrences," in *AMIA Annual Symposium Proceedings 2012*, American Medical Informatics Association, 2012, pp. 1284–1293.
[7] T. Sarwar *et al.*, "The secondary use of electronic health records for data mining: data characteristics and challenges," *ACM Computing Surveys*, vol. 55, no. 2, pp. 1–40, Feb. 2023, doi: 10.1145/3490234.
[8] K. Lakshmi and G. Vadivu, "Extracting association rules from medical health records using multi-criteria decision analysis," *Procedia Computer Science*, vol. 115, pp. 290–295, 2017, doi: 10.1016/j.procs.2017.09.137.
[9] S. Ramasamy and K. Nirmala, "Disease prediction in data mining using association rule mining and keyword based clustering algorithms," *International Journal of Computers and Applications*, vol. 42, no. 1, pp. 1–8, Jan. 2020, doi: 10.1080/1206212X.2017.1396415.
[10] F. D. M. F. M. Inamdar, N. JayaKumar, and B. Bari, "Heart disease predictive analysis using association rule mining," in *Intelligent Communication Technologies and Virtual Mobile Network*, Springer Nature Singapore, pp. 111–125, 2023, doi: 10.1007/978-981-19-1844-5_10.

[11] K. A. Happawana and B. J. Diamond, "Association rule learning in neuropsychological data analysis for alzheimer's disease," *Journal of Neuropsychology*, vol. 16, no. 1, pp. 116–130, Mar. 2022, doi: 10.1111/jnp.12252.
[12] Y. Liu, L. Wang, R. Miao, and H. Ren, "A data mining algorithm for association rules with chronic disease constraints," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–8, Aug. 2022, doi: 10.1155/2022/8526256.
[13] S. V. Maju and G. P. O. S. Pushpam, "A novel two-tier feature selection model for alzheimer's disease prediction," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 1, 2024, doi: 10.11591/ijeecs.v33.i1.pp227-235.
[14] R. J. Amrutha, M. S. Bindu, and K. Saritha, "Forecasting cardiac disease using pam and apriori," in *International Conference on Humans and Technology: A Holistic and Symbiotic Approach to Sustainable Development: ICHT 2022*, AIP Publishing LLC, 2023, doi: 10.1063/5.0138652.
[15] M. Mirmozaffari, A. Alinezhad, and A. Gilanpour, "Data mining apriori algorithm for heart disease prediction," *International Journal of Computing, Communication and Instrumentation Engineering*, vol. 4, no. 1, Jan. 2017, doi: 10.15242/IJCCIE.DIR1116010.
[16] K. Gulzar, M. A. Memon, S. M. Mohsin, S. Aslam, S. M. A. Akber, and M. A. Nadeem, "An efficient healthcare data mining approach using apriori algorithm: a case study of eye disorders in young adults," *Information*, vol. 14, no. 4, Mar. 2023, doi: 10.3390/info14040203.
[17] G. Subbalakshmi, K. Ramesh, and M. C. Rao, "Decision support in heart disease prediction system using naive bayes," *Indian Journal of Computer Science and Engineering*, vol. 2, no. 2, pp. 170-176, 2011.
[18] V. S. K. Reddy, P. Meghana, N. V. S. Reddy, and B. A. Rao, "Prediction on cardiovascular disease using decision tree and naïve Bayes classifiers," *Journal of Physics: Conference Series*, vol. 2161, no. 1, Jan. 2022, doi: 10.1088/1742-6596/2161/1/012015.
[19] V. Jackins, S. Vimal, M. Kaliappan, and M. Y. Lee, "AI-based smart prediction of clinical disease using random forest classifier and naive bayes," *The Journal of Supercomputing*, vol. 77, no. 5, pp. 5198–5219, May 2021, doi: 10.1007/s11227-020-03481-x.
[20] A. Purnomo, M. A. Barata, M. A. Soeleman, and F. Alzami, "Adding feature selection on naïve bayes to increase accuracy on classification heart attack disease," *Journal of Physics: Conference Series*, vol. 1511, no. 1, Apr. 2020, doi: 10.1088/1742-6596/1511/1/012001.
[21] G. Nuti, L. A. J. Rugama, and A.-I. Cross, "An explainable bayesian decision tree algorithm," *Frontiers in Applied Mathematics and Statistics*, vol. 7, Mar. 2021, doi: 10.3389/fams.2021.598833.
[22] N. Renaningtias, J. Endro, and R. Gernowo, "Hybrid decision tree and naïve bayes classifier for predicting study period and predicate of student's graduation," *International Journal of Computer Applications*, vol. 180, no. 49, pp. 28–34, Jun. 2018, doi: 10.5120/ijca2018917329.
[23] A. Niranjan, K. M. Akshobhya, A. S. Chouhan, and P. Tumuluru, "ERDNS: ensemble of random forest, decision tree, and naive Bayes kernel through stacking for efficient cross site scripting attack classification," in *Data Science and Computational Intelligence*, Springer, pp. 353–365, 2021, doi: 10.1007/978-3-030-91244-4_28.
[24] S. S. Raj and M. Nandhini, "Ensemble human movement sequence prediction model with apriori based probability tree classifier (APTC) and bagged J48 on machine learning," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 4, pp. 408–416, May 2021, doi: 10.1016/j.jksuci.2018.04.002.
[25] M. Chen and Z. Yin, "Classification of cardiotocography based on the apriori algorithm and multi-model ensemble classifier," *Frontiers in Cell and Developmental Biology*, vol. 10, May 2022, doi: 10.3389/fcell.2022.888859.

BIOGRAPHIES OF AUTHORS

Sonam Vayaliparambil Maju received the B.Tech. degree in information technology engineering from Mahatma Gandhi University, Kerala, India in 2013 and the M.Tech. degree in computer science and engineering from Mahatma Gandhi University, Kerala, India in 2015. She is currently pursuing the Ph.D. degree in computer science and engineering at Christ University, Bangalore, India. Her research interest includes the machine learning, natural language processing, information management, and clinical research. She had published one Indian patent and several other publications on clinical research. She can be contacted at email: sonam.maju@res.christuniversity.in.

Dr. Gnana Prakasi Oliver Siryapushpam is assistant professor at the CHRIST (Deemed to be University), Department of Computer Science and Engineering, Bangalore, India. She received her B.Tech. degree in information technology and M.E. degree in software engineering from Anna University. Her research interests include mobile Ad hoc networks, IoT, software engineering, and machine learning. Major research focusses in MAC and network layers and published research papers in reputed journals. Currently focusing on analyzing the prediction algorithms using recurrent neural network with clinical data. She has more than 15 publications in refereed national and international journals. She can be contacted at email: gnana.prakasi@christuniversity.in.