OPEN ACCESS JOURNALS

Indiana Journal of Multidisciplinary Research

Abbriviate Tittle- Ind J Multi Res

ISSN *(Online)*- 2583-3820

Journal Homepage Link- https://indianapublications.com/Journals/IJMR

DOI: https://doi.org/10.5281/zenodo.12686196

INDIANA PUBLICATIONS
PRODUCTIVE AND QUALITY RESEARCH

Research Article

Volume-04|Issue-03|2024

Early Detection of Alzheimer's Disease Using Cognitive Features A Voting-Based Ensemble Machine Learning Approach

Aadhya Sriranga D R*[1], Devaryan Jain[2], Rashmi S[3], Sahil Rajput[4], Dr. Shruthi P[5]

[1,2,3,4,5]Information Science & Engineering, RV Institute of Technology & Management/Visvesvaraya Technological University, India.

Article History
Received: 20.05.2024
Accepted: 05.06.2024
Published: 30.06.2024

Citation
Sriranga, A. D. R., Jain, D., Rashmi, S., Rajput, S., & Shruthi, P. (2024). Early Detection of Alzheimer's Disease Using Cognitive Features A Voting-Based Ensemble Machine Learning Approach. *Indiana Journal of Multidisciplinary Research*, *4*(3), 284-288.

Abstract: Detecting Alzheimer's disease (AD) early is crucial for effective management, with machine learning techniques increasingly utilised for their efficacy in predicting AD using cognitive tests. Ensemble machine learning models are particularly valuable for enhancing system robustness by combining multiple models. This article introduces a novel ensemble machine learning approach for early AD detection. Firstly, a novel feature selection technique, Neighbourhood Component Analysis and Correlation-based Filtration (NCA-F), is proposed to identify key cognitive features from a dataset. Subsequently, various machine learning classifiers are trained using NCA-F. The top classifiers are chosen for voting based on their performance. Voting employs an adaptive weight matrix process, where the output label of a model is multiplied by its F1score to determine its weight. Results demonstrate that adaptive voting achieves an accuracy of 93.92%, surpassing the 90.53% accuracy achieved by traditional artificial neural networks. Furthermore, the proposed technique enhances accuracy by 12.12% using the same features.

Keywords: Machine learning, network management, intrusion detection system (IDS), software defined networking.

INTRODUCTION

Alzheimer's disease (AD) unfolds gradually, primarily impacting memory initially and progressing to confusion and disorientation later on. Given its higher prevalence among older individuals, early detectionbecomes pivotal for timely intervention. While predicting AD subtype remains intricate, it constitutes a considerable portion of dementia cases. Concerns over its increasing prevalence fuel research employing mathematical models to forecast its trajectory, revealing concerning projections. Nonetheless, some countries with advanced healthcare systems report a decline in AD cases, suggesting the potential impact of early detection and intervention.Machine learning (ML) methods harness neuropsychological tests for AD prediction, employing sophisticated techniques like stacked autoencoders and convolutional neural networks. This article introduces an adaptive voting-based ML ensemble model for early AD prediction. Trained on cognitive features extracted via the innovative Neighbourhood Component Analysis and Correlation-based Filtration (NCA-F) method, this approach amalgamates various classifiers' strengths, bolstering generalisation and robustness. Results underscore significant enhancements in AD prediction, particularly in its early stages, by incorporating cognitive features into ML models.

PROBLEM STATEMENT

Developing an innovative healthcare solution to detect Alzheimer's disease at an Early stage and enhance medical procedures to delay the occurrence.

EXISTING SYSTEM

In an existing research they proposed content-based image retrieval (CBIR) systems have been widely researched and applied in many medical applications. Combining an automated image classification system and the radiologist's professional knowledge, to increase the accuracy of prediction and diagnosis, were the main motives. In this paper, a multistage classifier using machine learning, including Naive Bayes classifier, support vector machine (SVM), and K-nearest neighbor (KNN), was used to classify Alzheimer's disease more acceptably and efficiently. For this, MRI (Magnetic resonance imaging) scans were processed by FreeSurfer, a powerful software tool suitable for processing and normalizing brain MRI images. We also applied a feature selection technique - PSO (particle swarm optimization) to many feature vectors in order to obtain the best features that represent the salient characteristics of AD.

PROPOSED SYSTEM

This article proposes an adaptive voting-based ML ensemble model using cognitive features to predict AD in the early stages. The proposed technique includes a feature selection technique called Neighborhood Component Analysis and Correlation-based Filtration (NCA-F) to select vital cognitive features from a given dataset. Various machine learning classifiers like Support Vector Machine, Random Forest, AdaBoost, KNN, LR,Decision Tree, Artificial Neural networks, are trained using the proposed NCA-F method, and the top classifiers are selected for voting based on their

***Corresponding Author:** Aadhya Sriranga D R

performance results using an adaptive weight matrix process. The model produced comparatively better results using the ensemble learning classifier. The ensemble learning combines the performance of predictions, rather than discrete labels, of all the base classifiers to improve the generalization process and the robustness over a single estimator.

LITERATURE REVIEW

The dynamic ensemble selection of classifiers is an effective approach for processing label-imbalanced data classifications. However, such a technique is prone to overfitting, owing to the lack of regularization methods and the dependence of the aforementioned technique on local geometry. In this study, focusing on binary imbalanced dataclassification, a novel dynamic ensemble method, namely adaptive ensemble of classifiers with regularization (AER), is proposed, to overcome the stated limitations. The method solves the overfitting problem through implicit regularization. Specifically, it leverages the properties of stochastic gradient descent to obtain the solution with the minimum norm, thereby achieving regularization; furthermore, it interpolates the ensemble weights by exploiting the global geometry of data to further prevent overfitting. According to our theoretical proofs, the seemingly complicated AER paradigm, in addition to its regularization capabilities, can actually reduce the asymptotic time and memory complexities of several other algorithms. We evaluate the proposed AER method on seven benchmark imbalanced datasets from the UCI machine learning repository and one artificially generated GMM-based dataset with five variations. The results show that the proposed algorithm outperforms the major existing algorithms based on multiple metrics inmost cases, and two hypothesis tests (McNemar's and Wilcoxon tests) verify the statistical significance further. In addition, the proposed method has other preferred properties such as special advantages in dealing with highly imbalanced data, and it pioneers the research on the regularization for dynamic ensemble methods.Many classical machine learning techniques have been used to explore Alzheimer's disease (AD), evolving from image decomposition techniques such as principal component analysis toward higher complexity, non-linear decomposition algorithms.

With the arrival of the deep learning paradigm, it has become possible to extract high-level abstract features directly from MRI images that internally describe the distribution of data in low-dimensional manifolds. In this work, we try a new exploratory data analysis of AD based on deep convolutional autoencoders. We aim at finding links between cognitive symptoms and the underlying neurodegeneration process by fusing the information of neuropsychological test outcomes, diagnoses, and other clinical data with the imaging features extracted solely via a data-driven decomposition of MRI.

The distribution of the extracted features in different combinations is then analyzed and visualized using regression and classification analysis, and the influence of each coordinate of the autoencoder manifold over the brain is estimated. The imaging-derived markers could then predict clinical variables with correlations above 0.6 in the case of neuropsychological evaluation variables such as the MMSE or the ADAS11 scores,achieving classification accuracy over 80% for the diagnosis of AD.

The reliable diagnosis remains a challenging issue in the early stages of dementia. We aimed to develop and validate a new method based on machine learning to help the preliminary diagnosis of normal, mild cognitive impairment (MCI), very mild dementia (VMD), and dementia using an informant-based questionnaire. Methods. We enrolled 5,272 individuals who filled out a 37-item questionnaire. In order to select the most important features, three different techniques of feature selection were tested. Then, the top features combined with six classification algorithms were used to develop the diagnostic models. Results. Information Gain was the most effective among the three feature selection methods. The Naive Bayes algorithm performed the best (accuracy = 0.81, precision = 0.82, recall = 0.81, and F-measure = 0.81) among the six classification models. Conclusion. The diagnostic model proposed in this paper provides a powerful tool for clinicians to diagnose the early stages of dementia.

Machine learning (ML) is a promising technique for patient-specific prediction of mild cognitive impairment (MCI) and dementia development. Neuropsychiatric symptoms (NPS)might improve the accuracy of ML models but have barely been used for this purpose.

Objectives: To investigate if baseline mild behavioral impairment (MBI) status used for NPS quantification along with brain morphology features are predictive of follow-up diagnosis, median 40 months later in patients with normal cognition (NC) or MCI. Method: Baseline neuroimaging, neuropsychiatric, and clinical data from 102 individuals with NC and 239 with MCI were extracted from the Alzheimer's Disease Neuroimaging Initiative database. Neuropsychiatric inventory questionnaire items were transformed to MBI domains using a published algorithm. Diagnosis at latest follow-up was used as the outcome variable and ground truth classification. A logistic model tree classifier combined with information gain feature selection was trained to predict follow-up diagnosis.

Results: In the binary classification (NC versus MCI/AD), the optimal ML model required only two features from over 200, MBI total score and left hippocampal volume. These features correctly classified participants as remaining normal or developing cognitive impairment with 84.4% accuracy (area under the receiver

***Corresponding Author:** Aadhya Sriranga D R

operating characteristics curve [ROC-AUC] = 0.86). Seven features were selected for the three-class model (NC versus MCI versus dementia) achieving an accuracy of 58.8% (ROC-AUC=0.73).

CONCLUSION

Baseline NPS, categorized for MBI domain and duration, have prognostic utility in addition to brain morphology measures for predicting diagnosis change using ML. MBI total score, followed by impulse dyscontrol and affective dysregulation were most predictive of future diagnosis. Incidence rates of dementia appear to be declining in high-income countries according to several large epidemiological studies. We aimed to describe declining incident dementia rates across successive birth cohorts in a U.S. population-based sample and to explore the influences of sex and education on these trends. Methods: We pooled data from two community-sampled prospective cohort studies with similar study aims and contiguous sampling regions: the Monongahela Valley Independent Elders Survey (1987-2001) and the Monongahela-Youghiogheny Healthy Aging Team (2006-Ongoing). We identified four decade-long birth cohorts spanning birth years 1902-1941. In an analysis sample of 3,010 participants (61% women, mean baseline age = 75.7 years, mean follow-up = 7.1 years), we identified 257 cases of incident dementia indicated by a Clinical Dementia Rating of 1.0 or higher.

METHODOLOGY

Methodology involves these phases:

1. Data Exploration: Dataset loading and initial analysis.
2. Preprocessing: Data preparation including cleansing, feature transformation, normalization, and shuffling.
3. Train/Test Split: Dataset division for model training and evaluation.
4. Model Generation: Training diverse machine learning algorithms (KNN, DT, RF, NB, LR, AdaBoost, XGBoost) and potentially ensemble techniques for optimal results.
5. User Signup/Login: Authentication system for access control.
6. User Input: Interface for gathering data required for predictions.
7. Prediction: The trained model predicts whether the incoming traffic represents a cyberattack.

Benefits

- Security Focus: directly addresses the security vulnerabilities often found in cloud- centric systems, enhancing the protection of sensitive data.
- Comprehensive Threat Detection: The use of machine learning allows for the detection of a wide array of threats, providing broad protection.
- Real-Time Efficiency: The focus on real-time threat detection supports rapid response, minimizing potential harm in healthcare settings.
- Resource Optimization: The system's design prioritizes minimal resource usage, making it suitable for environments with potential limitations.

Drawbacks

- Dataset Dependence: , like other machine learning systems,relies on the quality and representativeness of the training dataset.
- Computational Overhead: While optimized, the system may have some computational resource requirements compared to simpler network monitoring tools.
- Evolving Attacks: Continuous adaptation is necessary to keep pace with the changing landscape of cyberattacks.
- Potential for False Positives: There is a possibility, albeit minimized through careful algorithm selection, of flagging benign traffic as malicious.

SYSTEM ARCHITECTURE

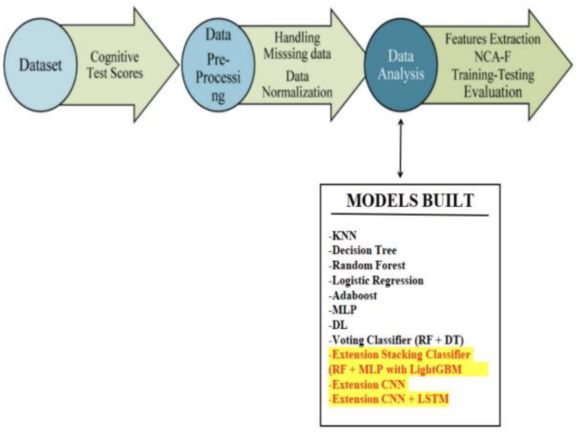

Figure 1: System Architecture

ALGORITHMS

KNN: The okay-nearest associates algorithm (KNN or okay-NN) is a non-parametric supervised mastering classifier that is predicated on proximity to classify or are expecting the grouping of character statistics points.

DT: Decision tree is a non-parametric supervised mastering algorithm used for category and regression responsibilities. It follows a hierarchical tree shape comprising root, branch, inner, and leaf nodes.

RF: Random wooded area is a extensively-used machine studying set of rules that combines outputs from a couple of selection bushes to produce a single result. Known for its ease of use and versatility, it handles each classification and regression troubles effectively.

NB: Naïve Bayes classifier is a supervised getting to know algorithm normally used for category tasks like textual content class. It belongs to the family of generative getting to know algorithms, aiming to model the distribution of inputs for a given magnificence.

LR: Logistic regression is a supervised gaining knowledge of set of rules typically hired for type responsibilities, predicting the chance of an example belonging to a particular elegance. It's a statistical algorithm analyzing the connection between independent and established binary variables, assisting decision-making.

AdaBoost: AdaBoost, or Adaptive Boosting, is an Ensemble Method in Machine Learning. It commonly employs choice timber with handiest one level (Decision Stumps) as estimators to reinforce overall performance.

Stacking Classifier (RF MLP with Light GBM): Stacking classifier is an ensemble method wherein outputs from a couple of classifiers function inputs to a meta-classifier for very last class. This approach effectively tackles multi-category troubles.

Voting Classifier (RF DT): Voting classifier is a gadget getting to know estimator that trains numerous base fashions or estimators and aggregates their findings for prediction. It combines selection outputs from each base estimator to make very last predictions.

RESULTS

The project aims to enhance the security of healthcare systems implemented within software-defined networks (SDNs)utilizing a layer 3 (L3) learning switch application to collect and analyze normal and abnormal network traffic, deployed on the Ryu controller. Rigorous testing involving multiple machine learning algorithms and cyberattack scenarios was conducted to provide a comprehensive performance evaluation.

The results suggest that Adaptive -shield machine learning defence for healthcare SDN's demonstrates robust performance, achieving a high F1-score for both normal and attack training, indicating reliability. This enhances data security and network resilience.

In summary, the project effectively tackles the critical task of safeguarding sensitive patient data within SDNs, offering valuable insights into the efficacy of machine learning-based approaches for cybersecurity in healthcare settings.

CONCLUSION

In the pursuit of early Alzheimer's Disease (AD) detection, a cutting-edge approach, referred to as the NCA-F-based novel feature selection technique, has been introduced. This innovative method focuses on selecting crucial cognitive features that are instrumental in identifying the onset of AD. To attain this, several machine learning (ML) classifiers are trained based on the selected features, culminating in the generation of groundbreaking results in a relatively short time frame. The heart of this method lies in the application of an adaptive voting technique, utilizing different weights to achieve optimal detection results across multiple performance metrics, including accuracy, precision, recall, F1-score, AUC (Area Under the Curve), and ROC (Receiver Operating Characteristic). Remarkably, the experimental outcomes strongly suggest that this approach outperforms other existing methods across various performance measures, signifying a significant leap forward in early AD detection. Central to the proposed approach is the concept of ensemble learning, which capitalizes on the strengths of various ML models to enhance AD detection. The ensemble models, particularly when coupled with the adaptive voting technique, exhibit a substantial improvement in accuracy compared to other models. Moreover, the AUC, a critical indicator of classification performance, also experiences a notable boost. The experiments further underscore the indispensable role of cognitive features derived from clinical tests in the early detection of AD using ensemble ML models. This method offers an efficient alternative to time-consuming clinical trials, making it a promising avenue for accelerating the AD detection process. In essence, the NCA-F-based feature selection technique and adaptive voting with ensemble learning represent a pivotal breakthrough in the endeavor to identify AD in its early stages.

REFERENCES

1. Wang, C., Deng, C., Yu, Z., Hui, D., Gong, X., & Luo, R. (2021). Adaptive ensemble of classifiers with regularization for imbalanced data classification. *Information Fusion*, *69*, 81–102.
2. Martinez-Murcia, F. J., Ortiz, A., Gorriz, J. M., Ramirez, J., & Castillo-Barnes, D. (2020). Studying the manifold structure of Alzheimer's disease: A deep learning approach using convolutional autoencoders. *IEEE Journal of Biomedical and Health Informatics*, *24*(1), 17–26.
3. Zhu, F., Li, X., Tang, H., He, Z., Zhang, C., Hung, G. U., ... & Zhou, W. (2020). Machine learning for the preliminary diagnosis of dementia. *Scientific Programming*, *2020*(1), 5629090.
4. Gill, S., Mouches, P., Hu, S., Rajashekar, D., MacMaster, F. P., Smith, E. E., ... & Alzheimer's Disease Neuroimaging Initiative. (2020). Using machine learning to predict dementia from neuropsychiatric symptom and neuroimaging data. *Journal of Alzheimer's Disease*, *75*(1), 277-288.
5. Sullivan, K. J., Dodge, H. H., Hughes, T. F., Chang, C. C. H., Zhu, X., Liu, A., & Ganguli, M. (2019). Declining incident dementia rates across four

***Corresponding Author:** Aadhya Sriranga D R

population-based birth cohorts. *The Journals of Gerontology: Series A*, *74*(9), 1439-1445.

6. Cruz, R. M. O., Sabourin, R., & Cavalcanti, G. D. C. (2018). Dynamic classifier selection: Recent advances and perspectives. *Information Fusion, 41*, 195–216.
7. Barber, D. (2012). *Bayesian reasoning and machine learning*. Cambridge University Press.
8. Reynolds, D. A., & Rose, R. C. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, *3*(1), 72–83.
9. Gardner, W. A. (1984). Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique. *Signal Processing, 6*(2), 113–133.
10. Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2016). Understanding deep learning requires rethinking generalization. *arXiv preprint arXiv:1611.03530*.

***Corresponding Author:** Aadhya Sriranga D R