



Research Article

Predicting Alzheimer's disease: A machine learning approach using advanced feature selection techniques

Hamed SabbaghGol¹ , * Hamid Saadatfar² , Mahdi Khazaiepoor³

1. PHD Candidate, Department of Computer Engineering, Birjand Branch, Islamic Azad University, Birjand, Iran.
2. PHD, Computer Engineering, Department of Computer Engineering, University of Birjand, Birjand, Iran.
3. PHD, Computer Engineering, Department of Computer Engineering, Birjand Branch, Islamic Azad University, Birjand, Iran.

Use your device to scan
and read the article online



Citation: SabbaghGol H, Saadatfar H, Khazaiepoor M. Predicting Alzheimer's disease: A machine learning approach using advanced feature selection techniques. *Journal of Modern Medical Information Science*. 2024; 10(3):307-324.

Article Info:

Received: 16 Sep 2024

Accepted: 14 Dec 2024

Available Online: 20

Dec 2024

ABSTRACT

Introduction: Alzheimer's disease, a progressive and debilitating neurological disorder, significantly impacts the quality of life, particularly in the elderly. Given the increasing prevalence of this disease, developing accurate methods for early prediction and diagnosis is crucial. This study aims to identify key factors influencing Alzheimer's disease prediction using novel feature selection techniques and machine learning models. The primary objective of this study is to contribute to the development of more accurate diagnostic tools, thereby improving the management and treatment of this disease.

Methods: In this study, we employed ten wrapper-based feature selection methods to identify the most accurate and relevant features of Alzheimer's disease. The performance of these models was evaluated using popular machine learning algorithms and standard evaluation metrics such as accuracy, precision, recall, F1-score, and ROC curve analysis. All evaluations were conducted on the ADNI standard Alzheimer's disease dataset.

Results: The influential features included cognitive test results (e.g., Mini-Mental State Examination), functional assessments, patient-reported memory and behavioral problems, and activities of daily living scores, which were identified as key indicators for Alzheimer's disease diagnosis.

Discussion: The results demonstrate that employing novel feature selection techniques and machine learning algorithms can lead to the development of more accurate models for predicting Alzheimer's disease. These findings can contribute to improving early diagnosis and management of this disease.

Key Words:

Dimensionality Reduction,
Feature Selection,
Alzheimer's Disease,
Machine Learning.

* Corresponding Author:

Dr Hamid Saadatfar

Address: University of Birjand, Birjand, Iran.

Tel: +98 9155611399

E-mail: saadatfar@birjand.ac.ir



Copyright © 2024 The Author[s];
This is an open access article distributed under the terms of the Creative Commons Attribution License [CC-BY-NC: <https://creativecommons.org/licenses/by-nc/4.0/legalcode>], which permits use, distribution, and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.



Extended Abstract

Introduction:

Alzheimer's disease (AD) is a progressive neurological disorder marked by harmful protein Alzheimer's Disease (AD) is a neurodegenerative disorder characterized by the progressive accumulation of amyloid plaques and neurofibrillary tangles, leading to the gradual degeneration of brain cells. As the leading cause of dementia in older adults, AD significantly impairs cognitive and functional abilities, resulting in a substantial decline in quality of life for both patients and their caregivers.

While aging is the primary risk factor, a multitude of factors contribute to the development and progression of AD. These include genetic predispositions, such as the APOE gene, as well as lifestyle factors such as physical inactivity, poor diet, smoking, alcohol consumption, and depression. Furthermore, the presence of comorbidities such as diabetes, hypertension, and cardiovascular diseases can significantly increase the risk of developing AD.

The global burden of AD is expected to increase dramatically in the coming decades. With the rapidly aging global population, the prevalence of AD is projected to rise significantly. According to the World Health Organization (WHO), the number of people aged 60 and above is projected to double from 1 billion in 2020 to 2.1 billion by 2050, while the number of individuals aged 80 and older is expected to triple to 426 million. Concurrently, the global prevalence of dementia is projected to increase from 55 million in 2019 to 139 million by 2050.

The lack of a definitive cure for AD underscores the critical need for early diagnosis and effective disease management strategies. Machine learning (ML) offers a powerful approach to analyzing complex medical datasets, enabling the identification of subtle patterns and relationships that may not be readily apparent through traditional analytical methods. This study focuses on the application of advanced feature selection methods, particularly Wrapper-based approaches, in conjunction with ML algorithms. The primary objective is to identify key risk factors, improve diagnostic accuracy, and gain deeper insights into the underlying mechanisms of AD, ultimately paving the way for the development of more effective

diagnostic tools and therapeutic interventions.

Methods:

This study employed a systematic four-stage approach to analyze data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The ADNI dataset, a widely recognized resource in Alzheimer's disease research, encompasses comprehensive clinical, imaging, and laboratory data from individuals with Alzheimer's disease, mild cognitive impairment (MCI), and healthy controls.

The initial stage involved rigorous data preprocessing. Missing data points were addressed through appropriate imputation techniques, such as mean imputation or K-Nearest Neighbors (KNN) imputation. Furthermore, numerical data underwent normalization using the Z-score method to ensure comparability and enhance the accuracy of subsequent analyses.

The second stage focused on feature selection using ten prominent Wrapper-based methods, including Sequential Forward Selection (SFS), Recursive Feature Elimination (RFE), Evolutionary Recursive Forward Selection (ERSFS), Sequential Forward Selection (SFE), Sequential Floating Forward Selection (SFFS), Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Whale Optimization Algorithm (WOA), Bat Algorithm (BAA), and Firefly Algorithm (FA). These methods systematically evaluated different combinations of features, optimizing for model performance and reducing dimensionality. In the third stage, five machine learning algorithms – K-Nearest Neighbors (KNN), Support Vector Machine (SVM), J48 (a decision tree algorithm), Random Forest (RF), and Multilayer Perceptron (MLP) – were employed to classify subjects based on the selected features. The performance of each model was rigorously evaluated using five-fold cross-validation and a comprehensive set of performance metrics, including accuracy, sensitivity, specificity, precision, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

This multi-stage approach, integrating advanced feature selection techniques and robust machine learning algorithms, provides a comprehensive framework for analyzing the ADNI dataset and gaining deeper insights into the complex interplay of factors contributing to Alzheimer's disease. The findings of this study have the potential to significantly



advance our understanding of Alzheimer's disease, improve diagnostic accuracy, and ultimately inform the development of more effective prevention and treatment strategies.

Results:

Model performance was optimized by evaluating the results of ten independent runs and selecting the combination of features and machine learning algorithms that yielded the highest average F1-score. This approach not only enhanced diagnostic accuracy but also facilitated a comparative analysis of the effectiveness of different feature selection and machine learning methods.

The analysis revealed several key features consistently identified as significant predictors of Alzheimer's disease, including the Mini-Mental State Examination (MMSE) score, functional assessment scores, memory complaints, and behavioral problems. Additionally, features such as Activities of Daily Living (ADL) score, triglyceride cholesterol levels, confusion, and smoking status were also frequently identified as important predictors. Furthermore, the study highlighted the potential importance of age, sleep quality, and family history of Alzheimer's disease, which should be further investigated in future research.

The Evolutionary Recursive Forward Selection (ERSFS) method demonstrated high efficacy in identifying key features, including MMSE score, functional assessment score, memory complaints, behavioral problems, and ADL score. Recursive Feature Elimination (RFE) and Particle Swarm Optimization (PSO) also exhibited strong performance in feature selection, with some overlap in identified features and the addition of smoking status and confusion as significant predictors.

In terms of model performance, Random Forest and Decision Tree algorithms demonstrated superior predictive accuracy.

Conclusion:

This research investigated the impact of Wrapper-based feature selection methods on the performance of machine learning algorithms in diagnosing Alzheimer's Disease (AD). The findings demonstrate the superior performance of the Evolutionary Recursive Forward Selection (ERSFS)

method in identifying key features and improving diagnostic accuracy. These results corroborate previous research that highlights the synergistic effects of feature interactions on the performance of machine learning models. Key features identified as significant predictors of AD included the Mini-Mental State Examination (MMSE) score, functional assessment scores, memory complaints, behavioral issues, and Activities of Daily Living (ADL) score. These findings align with existing literature emphasizing the clinical significance of MMSE and ADL in assessing cognitive decline and functional impairment in AD. Furthermore, the study highlights the potential importance of other factors, including triglyceride cholesterol levels, disorientation, smoking status, age, sleep quality, and family history of Alzheimer's disease, particularly in the later stages of the disease.

Limitations and Future Research Directions:

This study has certain limitations, including the use of a specific dataset (ADNI) with inherent limitations in terms of diversity. The inclusion of more diverse datasets, incorporating genetic data, and the exploration of advanced deep learning models are crucial for enhancing the generalizability and robustness of the findings.

Future research should focus on:

- Incorporating genetic data: Integrating genetic data into the analysis can provide a more comprehensive understanding of the genetic and environmental factors contributing to AD.
- Exploring advanced deep learning models: Utilizing deep learning architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), may further improve diagnostic accuracy and provide deeper insights into the underlying disease mechanisms.
- Addressing the limitations of current datasets: Utilizing more diverse and representative datasets to enhance the generalizability of findings and improve the clinical applicability of the developed models.

By addressing these limitations and expanding the scope of future research, we can further refine our understanding of Alzheimer's disease and develop more effective diagnostic and therapeutic strategies.



مقاله پژوهشی

شناسایی عوامل مؤثر بر پیش‌بینی ابتلا به بیماری آلزایمر با استفاده از روش‌های نوین انتخاب ویژگی و الگوریتم‌های یادگیری ماشین

حامد صباغ گل^۱ ID، * حمید سعادت‌فر^۲ ID، مهدی خزاعی‌پور^۳ ID

۱. دانشجوی دکتری، مهندسی کامپیوتر، گروه مهندسی کامپیوتر، واحد بیرجند، دانشگاه آزاد اسلامی، بیرجند، ایران.
۲. دکتری تخصصی، مهندسی کامپیوتر، گروه مهندسی کامپیوتر، دانشگاه بیرجند، بیرجند، ایران.
۳. دکتری تخصصی، مهندسی کامپیوتر، گروه مهندسی کامپیوتر، واحد بیرجند، دانشگاه آزاد اسلامی، بیرجند.

Use your device to scan and read the article online



Citation: SabbaghGol H, Saadatfar H, Khazaeipoor M. Predicting Alzheimer's disease: A machine learning approach using advanced feature selection techniques. *Journal of Modern Medical Information Science*. 2024; 10(3):307-324.

چکیده

اطلاعات مقاله

تاریخ دریافت: ۱۶ شهریور ۱۴۰۳

تاریخ پذیرش: ۲۴ آذر ۱۴۰۳

تاریخ انتشار: ۳۰ آذر ۱۴۰۳

هدف: بیماری آلزایمر به عنوان یک اختلال عصبی پیش‌رونده و تخریب‌کننده، به دلیل تأثیر قابل توجه بر کیفیت زندگی افراد، به ویژه سالمندان، از اهمیت بالایی برخوردار است. با توجه به افزایش روزافزون شیوع این بیماری، توسعه روش‌های دقیق برای پیش‌بینی و تشخیص زودهنگام آن ضروری است. در این پژوهش، با بهره‌گیری از روش‌های نوین انتخاب ویژگی و مدل‌های یادگیری ماشین، به دنبال شناسایی عوامل کلیدی مؤثر در پیش‌بینی بیماری آلزایمر هستیم. هدف اصلی این مطالعه، کمک به توسعه ابزارهای تشخیصی دقیق‌تر و در نتیجه بهبود مدیریت و درمان این بیماری است.

روش‌ها: در این پژوهش، با بهره‌گیری از ده روش انتخاب ویژگی مبتنی بر Wrapper، به شناسایی دقیق‌ترین و مرتبط‌ترین ویژگی‌های بیماری آلزایمر پرداخته‌ایم. کارایی این مدل‌ها با استفاده از الگوریتم‌های یادگیری ماشین پرکاربرد و معیارهای ارزیابی استاندارد نظیر دقت، ویژگی، حساسیت، اندازه‌گیری F1 و تحلیل منحنی ROC مورد ارزیابی قرار گرفته و نتایج حاصل با یکدیگر مقایسه می‌شوند. تمامی ارزیابی‌ها بر روی مجموعه داده استاندارد بیماران آلزایمر ADNI انجام شده است.

نتایج: ویژگی‌های اثرگذار شامل نتایج آزمون‌های شناختی (مانند آزمون کوتاه وضعیت ذهنی)، ارزیابی‌های عملکردی، گزارش‌های بیماران درباره مشکلات حافظه و رفتاری، و همچنین امتیاز فعالیت‌های روزمره، به عنوان شاخص‌های کلیدی در تشخیص بیماری آلزایمر، شناسایی شدند.

نتیجه‌گیری: نتایج نشان می‌دهد که با استفاده از روش‌های نوین انتخاب ویژگی و الگوریتم‌های یادگیری ماشین، می‌توان مدل‌های دقیق‌تری برای پیش‌بینی بیماری آلزایمر توسعه داد. این یافته‌ها می‌تواند در بهبود تشخیص زودهنگام و مدیریت این بیماری مؤثر باشد.

کلیدواژه‌ها:

کاهش ابعاد، انتخاب ویژگی، بیماری آلزایمر، یادگیری ماشین.

*نویسنده مسئول:

دکتر حمید سعادت‌فر

نشانی: دانشگاه بیرجند، بیرجند، ایران.

تلفن: +98 9155611399

پست الکترونیک: saadatfar@birjand.ac.ir



Copyright © 2024 The Author[s].

This is an open access article distributed under the terms of the Creative Commons Attribution License [CC-BY-NC: <https://creativecommons.org/licenses/by-nc/4.0/legal-code.en>], which permits use, distribution, and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

مقدمه:

بیماری آلزایمر یک اختلال پیش‌رونده مغزی و یکی از شایع‌ترین علل زوال عقل در افراد مسن است که با تجمع پروتئین‌های مضر و تخریب تدریجی سلول‌های عصبی، به کاهش توانایی‌های شناختی و عملکردی فرد منجر می‌شود [۱]. افزایش سن مهم‌ترین عامل خطر این بیماری است، اما عوامل ژنتیکی (مانند ژن APOE)، سبک زندگی نامناسب، بیماری‌های همزمان (دیابت، فشار خون بالا) و عوامل محیطی نیز در پیشرفت آن نقش دارند [۲]. با توجه به افزایش جمعیت سالمند در سطح جهان، تعداد مبتلایان به این بیماری نیز به طور چشمگیری در حال افزایش است [۳]. طبق گزارش سازمان بهداشت جهانی، انتظار می‌رود تعداد افراد مبتلا به زوال عقل از ۵۵ میلیون نفر در سال ۲۰۱۹ به ۱۳۹ میلیون نفر در سال ۲۰۵۰ افزایش یابد [۴]. همچنین هزینه‌های مرتبط با زوال عقل نیز از ۱/۳ تریلیون دلار در سال ۲۰۱۹ به ۲/۸ تریلیون دلار تا سال ۲۰۳۰ خواهد رسید [۵]. در ایران نیز مطالعه‌ای نشان می‌دهد بیش از ۲/۳ درصد از سالمندان به زوال عقل مبتلا هستند [۶]. با توجه به نبود درمان قطعی، شناسایی عوامل خطر، توسعه روش‌های تشخیص زودهنگام و درمان‌های مؤثر از اهمیت بالایی برخوردار است و درک بهتر مکانیسم‌های بیماری می‌تواند گامی اساسی در بهبود کیفیت زندگی بیماران و خانواده‌های آن‌ها باشد [۲].

یکی از چالش‌های اصلی در تشخیص آلزایمر، حجم زیاد داده‌های پزشکی و وجود ویژگی‌های غیرمرتبط یا نامربوط است که می‌تواند دقت مدل‌های تشخیصی را کاهش دهد [۷]. در سال‌های اخیر، توجه به استفاده از یادگیری ماشین و رویکردهای مبتنی بر داده برای بهبود درک بیماری آلزایمر به طور چشمگیری افزایش یافته است. با تحلیل مجموعه داده‌های بزرگ و حاوی اطلاعات دقیق سلامت، محققان قادر به شناسایی الگوها و روابطی هستند که ممکن است در روش‌های بالینی سنتی قابل تشخیص نباشند [۷]. در این زمینه، روش‌های کاهش ابعاد به عنوان ابزارهای قدرتمندی مطرح شده‌اند. این روش‌ها با هدف شناسایی مهم‌ترین و مرتبط‌ترین ویژگی‌ها در یک مجموعه داده، بهبود دقت و قابلیت تفسیر مدل‌های پیش‌بینی را فراهم می‌کنند [۸].

روش‌های کاهش ابعاد به دو دسته اصلی تقسیم می‌شوند: استخراج ویژگی و انتخاب ویژگی؛ در استخراج ویژگی، داده‌ها به فضای جدیدی با ابعاد کمتر تبدیل می‌شوند، در حالی که انتخاب ویژگی، زیرمجموعه‌ای بهینه از ویژگی‌های اصلی را بدون تغییر ساختار آن‌ها برمی‌گزیند. این روش با حذف ویژگی‌های اضافی، دقت مدل‌های یادگیری ماشین را افزایش داده و هزینه‌های محاسباتی را کاهش می‌دهد [۸]. روش‌های انتخاب ویژگی شامل سه گروه فیلتر، Wrapper و تعبیه‌شده هستند؛ فیلترها از معیارهای آماری، Wrapperها از عملکرد مدل، و تعبیه‌شده‌ها انتخاب ویژگی را در فرآیند ساخت مدل انجام می‌دهند [۹].

الگوریتم‌های انتخاب ویژگی Wrapper به دلیل توانایی در شناسایی تعاملات بین ویژگی‌ها و انتخاب بهترین زیرمجموعه‌های مؤثر، بسیار محبوب و کارآمد هستند [۸]. برخلاف روش‌های فیلتر، Wrapperها با دریافت بازخورد از مدل یادگیری، زیرمجموعه‌های ویژگی‌ها را بر اساس عملکرد مدل ارزیابی می‌کنند و معمولاً ویژگی‌های بهتری را انتخاب می‌نمایند [۹].

از جمله الگوریتم‌های ترتیبی Wrapper معروف می‌توان به SBS، FFS، SFS و SFBS اشاره کرد که با اضافه یا حذف متوالی ویژگی‌ها، به دنبال بهبود دقت طبقه‌بندی هستند [۱۱]. در سال ۲۰۲۴، صباغ‌گل و همکاران الگوریتم Wrapper پیشرفته‌ای ارائه کردند که با ترکیب روش‌های جستجوی تصادفی و تکاملی توسعه یافته است (با نام ERSFS) [۱۰]. از دیگر روش‌های نوین، الگوریتم‌های SFE و SFE-PSO هستند که ویژگی‌های نامربوط را حذف کرده و جستجوی زیرمجموعه‌های بهینه را ارتقاء می‌دهند [۱۰، ۱۱]. آکمن و همکاران نیز در سال ۲۰۲۳ الگوریتم SPFSR را معرفی کردند که با تلفیق جستجوی تصادفی و بهینه‌سازی، بهترین زیرمجموعه ویژگی‌ها را شناسایی می‌کند. همچنین در مطالعه ترک اوغلو و همکاران، الگوریتم دودویی جلبک مصنوعی (AAA) معرفی شده است. از دیگر تحقیقات انجام شده در این زمینه می‌توان به الگوریتم انتخاب ویژگی بر اساس الگوریتم ژنتیک (GA)، بهینه‌سازی ازدحام ذرات، بهینه‌سازی نهنگ‌ها (WOA) و کرم شب

تاب (FA) اشاره کرد [۱۷-۱۲].

داده است [۱۸،۲۳]. پژوهش‌ها نشان می‌دهند که انتخاب ویژگی می‌تواند به‌طور قابل توجهی عملکرد الگوریتم‌های یادگیری ماشین را در کاربردهای پزشکی بهبود بخشد و راه‌های جدیدی برای توسعه ابزارهای مؤثر در پیشگیری و درمان فراهم کند.

استفاده از الگوریتم‌های یادگیری ماشین همراه با روش‌های انتخاب ویژگی این امکان را فراهم می‌آورد که ویژگی‌های مؤثر در پیش‌بینی و تشخیص بیماری شناسایی شوند. این رویکردها نه تنها موجب کاهش پیچیدگی محاسباتی و افزایش دقت مدل‌ها می‌شوند، بلکه برای کشف الگوهای پنهان در داده‌های بالینی و تصویربرداری ضروری هستند. بکارگیری این تکنیک‌های پیشرفته، می‌تواند به بهبود کارایی سیستم‌های تشخیصی و درک بهتر عوامل مؤثر در بروز AD کمک کند [۱۸،۲۱]. در ادامه به مرور پژوهش‌های انجام‌شده در این حوزه پرداخته می‌شود.

در مطالعه سودارسان و همکاران، به چالش تشخیص زودهنگام بیماری آلزایمر با استفاده از تکنیک‌های یادگیری ماشین و تحلیل مؤلفه‌های اصلی (PCA) پرداخته شده است [۲۴]. با تمرکز بر طبقه‌بندی بیماری با داده‌های MRI ساختاری، این پژوهش نشان می‌دهد که ترکیب PCA با روش‌های سنتی یادگیری ماشین، با کاهش ابعاد و انتخاب ویژگی‌های مرتبط، عملکرد مدل را بهبود می‌بخشد و دقت طبقه‌بندی را افزایش می‌دهد. در تحقیق دیگری فیصل و همکاران رویکرد نوآورانه‌ای ارائه می‌دهد که PCA را با حذف بازگشتی ویژگی (RFE) ترکیب می‌کند تا تشخیص زودهنگام آلزایمر را بهبود بخشد [۲۵]. این روش در داده‌های MRI از مجموعه داده ADNI برای طبقه‌بندی‌های آلزایمر در مقابل MCI اولیه، MCI دیررس، و کنترل‌های سالم استفاده می‌شود و با کاهش ابعاد، دقت مدل‌های SVM را افزایش می‌دهد و بر اثربخشی این رویکرد در تشخیص آلزایمر تأکید دارد.

در پژوهش حاضر، به‌منظور شناسایی ریسک‌فاکتورهای مؤثر بر بیماری آلزایمر، از مجموعه‌ای از روش‌های پیشرفته انتخاب ویژگی مبتنی بر یادگیری ماشین استفاده شده است. این روش‌ها با هدف ارائه نتایج قابل اعتمادتر، توانایی

مطالعات متعددی به بررسی عوامل خطر بیماری آلزایمر (AD) پرداخته‌اند که این عوامل به دو گروه قابل تغییر و غیرقابل تغییر تقسیم می‌شوند [۱۸]. سن و استعداد ژنتیکی، به‌ویژه وجود آلل APOEε4، از مهم‌ترین عوامل غیرقابل تغییر به شمار می‌روند که با افزایش سن و به‌ویژه پس از ۶۵ سالگی، خطر ابتلا به AD را به‌طور قابل توجهی افزایش می‌دهند [۱۹،۲۰]. از سوی دیگر، عوامل قابل تغییر ذکرشده در پژوهش‌ها، مانند سیگار کشیدن، رژیم غذایی ناسالم، فعالیت بدنی ناکافی، و بیماری‌های مزمن نظیر دیابت، فشار خون، و چاقی، تأثیر چشمگیری بر خطر ابتلا به آلزایمر دارند [۱۸،۲۱]. در کنار این‌ها، عوامل محیطی مانند آسیب‌های جدی به سر و مواجهه با سموم نیز در افزایش خطر بیماری نقش دارند. با این حال، چالش‌هایی مانند محدودیت در دسترسی به ابزارهای پیشرفته تشخیصی و نابرابری در آگاهی عمومی و تخصصی درباره AD، به‌ویژه در کشورهای در حال توسعه، موانع مهمی در مسیر تشخیص زودهنگام و پیشگیری مؤثر ایجاد کرده‌اند [۲۱].

علاوه بر عوامل خطر، مطالعات تصویربرداری عصبی مانند PET و MRI، با تحلیل بیومارکرهایی نظیر بتا-آمیلوئید و تغییرات پروتئین تاو، به پیشرفت‌های چشمگیری در تشخیص زودهنگام AD منجر شده‌اند. این روش‌ها با ارائه تصویری دقیق از تغییرات پاتولوژیک مغز، به درک بهتر مکانیسم‌های بیماری کمک کرده‌اند [۲۲]. در حوزه‌های دیگر، مطالعات ژنتیکی نشان داده‌اند که زمینه‌های ژنتیکی نقش برجسته‌ای در ابتلا به AD دارند، درحالی‌که تحقیقات روی سبک زندگی سالم، تأثیر مثبت عواملی مانند فعالیت فیزیکی و رژیم غذایی متعادل را بر کاهش خطر بیماری نشان داده‌اند [۱۸].

ابزارهای یادگیری ماشین با تحلیل داده‌های بزرگ و شناسایی الگوهای پیچیده مرتبط با بیماری آلزایمر (AD)، نقش مهمی در توسعه ابزارهای تشخیصی دقیق و پیش‌بینی خطر ابتلا ایفا کرده‌اند. با این حال، تنوع و حجم بالای داده‌های مرتبط با عوامل خطر و اثرات متقابل آن‌ها، نیازمند بهره‌گیری از تکنیک‌های پیشرفته تحلیل

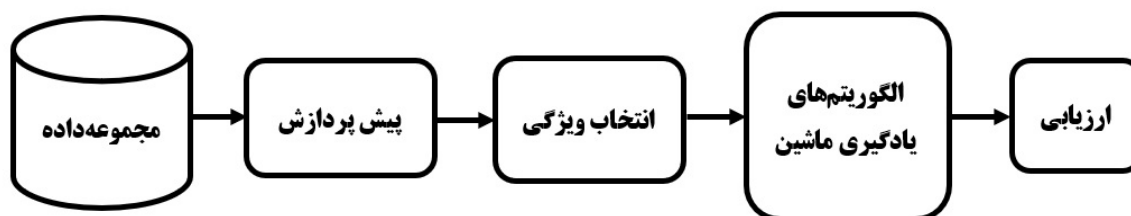
نتیجه‌گیری قرار می‌دهد و مقاله را با بررسی تحقیقات آینده به پایان می‌رساند.

مواد و روش‌ها:

این پژوهش، همان‌طور که در فلوچارت شکل ۱ نمایش داده شده است، یک فرآیند سیستماتیک چهارمرحله‌ای را دنبال می‌کند: (۱) پیش‌پردازش داده‌ها؛ در این مرحله داده‌های خام برای تحلیل آماده می‌شوند، شامل پاک‌سازی و استانداردسازی. (۲) اعمال تکنیک‌های انتخاب ویژگی: از مجموعه‌ای از روش‌های پیشرفته انتخاب ویژگی استفاده شده تا ویژگی‌های مرتبط‌تر شناسایی شوند. (۳) پیاده‌سازی الگوریتم‌های یادگیری ماشین: بهترین زیرمجموعه ویژگی‌ها به‌عنوان ورودی الگوریتم‌های یادگیری ماشین برای مدل‌سازی مورد استفاده قرار می‌گیرند. (۴) ارزیابی عملکرد: دقت، حساسیت، و سایر معیارهای ارزیابی برای تحلیل و مقایسه عملکرد روش‌ها استفاده می‌شوند.

تحلیل اثرات متقابل ویژگی‌ها و شناسایی الگوهای پیچیده درون داده‌ها را دارد، که در روش‌های سنتی آماری اغلب نادیده گرفته می‌شوند. همچنین، بهره‌گیری از مجموعه‌ای از روش‌های انتخاب ویژگی، موجب افزایش دقت و قابلیت اعتماد نتایج شده و احتمال شناسایی ویژگی‌های مهمی که ممکن است در رویکردهای مرسوم نادیده گرفته شوند را بیشتر می‌کند. این پژوهش به‌طور خاص، بر استفاده از یادگیری ماشین و روش‌های پیشرفته انتخاب ویژگی متمرکز است تا با ارائه یک رویکرد جامع به درک عمیق‌تری از عوامل کلیدی مرتبط با آلزایمر دست یابد و ابزارهای پیش‌بینی و تشخیصی مؤثرتری را توسعه دهد.

این مقاله به‌صورت زیر ساختار یافته است: بخش ۲ به توضیح مجموعه‌داده و روش‌شناسی انتخاب ویژگی اختصاص دارد. در بخش ۳، ارزیابی روش‌های پیشنهادی بر روی مجموعه‌داده آلزایمر انجام شده و نتایج ارائه می‌شود. بخش ۴ اهمیت ویژگی‌های انتخاب‌شده را مورد بحث و



شکل ۱: فلوچارت روش پیشنهادی

در این پژوهش، از زیرمجموعه‌ای شامل ۹۷۵۱ رکورد و ۲۸ ویژگی مختلف استفاده شده است. این ویژگی‌ها طیف گسترده‌ای از اطلاعات را پوشش می‌دهند، از جمله ویژگی‌های دموگرافیک (مانند سن و جنسیت)، عوامل مربوط به سبک زندگی (مانند وضعیت مصرف دخانیات، سطح فعالیت فیزیکی و کیفیت خواب)، سوابق پزشکی (مانند سابقه خانوادگی آلزایمر)، اندازه‌گیری‌های بالینی (مانند فشار خون، کلسترول و تری‌گلیسریدها)، ارزیابی‌های شناختی و عملکردی (مانند امتیازات MMSE و ADL)، و علائم بالینی (مانند سردرگمی، ناهماهنگی و فراموشی).

در این پژوهش از مجموعه داده‌های استاندارد و معتبر ADNI^۱ استفاده شده است. این پروژه که در سال ۲۰۰۳ راه‌اندازی شد، به‌عنوان یک بنچمارک بین‌المللی شناخته می‌شود و شامل داده‌های مرتبط با بیماری آلزایمر، از جمله اطلاعات بالینی، تصویربرداری و آزمایشگاهی است. هدف اصلی پروژه ADNI بررسی تغییرات بیومارکری مرتبط با پیشرفت آلزایمر بوده و داده‌های آن از مراکز متعدد درمانی و تحقیقاتی جمع‌آوری شده است. این مجموعه داده شامل نمونه‌هایی از بیماران مبتلا به آلزایمر، افراد با اختلال شناختی خفیف (MCI) و افراد سالم است.

تمام داده‌ها از طریق وبسایت رسمی ADNI

1 Alzheimer's Disease Neuroimaging Initiative

SBS، و SFBS با اضافه یا حذف متوالی ویژگی‌ها، به دنبال بهبود دقت طبقه‌بندی هستند. SFS با اضافه کردن تدریجی ویژگی‌ها و SBS با حذف تدریجی آن‌ها کار می‌کنند، در حالی که SFBS و SFFS ترکیبی از این دو رویکرد هستند و امکان جابجایی ویژگی‌ها را فراهم می‌کنند. هدف این الگوریتم‌ها یافتن کوچک‌ترین زیرمجموعه‌ای از ویژگی‌ها است که بیشترین دقت را در مدل یادگیری ماشین ایجاد کند [۱۱].

الگوریتم انتخاب تصادفی زیرمجموعه ویژگی (RSFS) یکی از روش‌های محبوب برای کاهش ابعاد داده است که از طبقه‌بندهایی نظیر KNN و معیارهای مبتنی بر نتایج طبقه‌بندی استفاده می‌کند. این روش با حفظ روابط بین داده‌ها و کاهش ابعاد، عملکرد مدل‌های یادگیری ماشین را بهبود می‌بخشد، اما در شناسایی تعاملات و هم‌افزایی بین ویژگی‌ها محدودیت دارد که ممکن است منجر به گیرکردن در بهینه‌های محلی شود [۳۱]. برای رفع این مشکل، در سال ۲۰۲۴، صباغ‌گل و همکاران الگوریتم توسعه‌یافته‌ی ERSFS را معرفی کردند. این نسخه با ارزیابی مستمر ویژگی‌ها، تقویت فاز تکامل، و در نظر گرفتن هم‌افزایی، از گیر افتادن در بهینه محلی جلوگیری کرده و همگرایی سریع‌تر و دقت بالاتری را با انتخاب تعداد کمتری ویژگی فراهم می‌کند [۱۰].

الگوریتم انتخاب ویژگی SFE، با استفاده از یک عامل جستجو و دو اپراتور غیرانتخابی و انتخابی، ویژگی‌های نامربوط، زائد و نویزی را شناسایی و حذف می‌کند [۱۱]. این الگوریتم با تمرکز بر ویژگی‌های مؤثر بر طبقه‌بندی، به انتخاب بهینه ویژگی‌ها کمک می‌کند. الگوریتم SPFSR نیز با تلفیق جستجوی تصادفی و بهینه‌سازی، بهترین زیرمجموعه ویژگی‌ها را شناسایی می‌کند و در کاهش ابعاد داده و بهبود دقت مدل‌ها مؤثر است. همچنین الگوریتم دودویی جلبک مصنوعی (AAA)، که از رفتار میکرو جلبک‌ها الهام گرفته، عملکرد موفق‌تری در حل مسائل مختلف دارد و با جستجوی کارآمد در فضای ویژگی‌ها، زیرمجموعه‌های بهینه را برای طبقه‌بندی شناسایی می‌کند [۱۲، ۱۳]. تطبیق‌های دودویی AAA با محدود کردن فضای پیوسته‌ی اولیه از

دریافت شده‌اند و مراحل جمع‌آوری و انتشار آن‌ها طبق استانداردهای اخلاقی و پژوهشی انجام گرفته است. استفاده از این دیتاست به دلیل جامعیت و کیفیت بالای آن، امکان تحلیل دقیق‌تر عوامل مرتبط با بیماری آلزایمر و ارائه نتایج قابل اعتمادتر را فراهم کرده است [۲۶].

مرحله پیش‌پردازش این داده‌ها شامل چندین گام کلیدی بوده که به هدف بهبود کیفیت و قابلیت تحلیل داده‌ها انجام شده است. ابتدا، داده‌های مفقود شناسایی و مدیریت شدند. برای این منظور، مقادیر مفقود در ویژگی‌های عددی با میانگین و در ویژگی‌های دسته‌بندی شده با مقادیر پرتکرار جایگزین شدند. همچنین از تکنیک‌های پیشرفته‌ای مانند KNN برای برخی فیلدها بهره گرفته شد [۲۷]. پس از آن، برای نرمال‌سازی داده‌های عددی و کاهش تأثیر مقیاس‌های مختلف بر عملکرد الگوریتم‌ها، از روش Z-score استفاده شد [۲۸]. این روش با تبدیل مقادیر به مقیاس استاندارد، به بهبود دقت و کارایی مدل‌های یادگیری ماشین کمک می‌کند. این مراحل پیش‌پردازش، دقت و اطمینان‌پذیری تحلیل‌ها و مدل‌های یادگیری ماشین را به‌طور قابل‌توجهی بهبود بخشیده است.

مرحله انتخاب ویژگی پس از پیش‌پردازش داده‌ها، با هدف شناسایی و انتخاب زیرمجموعه‌ای از ویژگی‌های کلیدی برای بهبود عملکرد مدل‌های طبقه‌بندی، انجام می‌شود. در این پژوهش، از روش‌های انتخاب ویژگی مبتنی بر Wrapper استفاده شده است. این روش‌ها با بهره‌گیری از الگوریتم‌های یادگیری ماشین، به ارزیابی زیرمجموعه‌های مختلف ویژگی‌ها پرداخته و به دقت و انعطاف‌پذیری بیشتری نسبت به روش‌های فیلتر دست می‌یابند. هرچند که این روش‌ها هزینه محاسباتی بالاتری دارند اما تأثیرات متقابل بین ویژگی‌ها و وابستگی به الگوریتم یادگیری را در نظر می‌گیرند [۲۳]. در این مطالعه، روش‌های مختلفی شامل انتخاب رو به جلو، حذف معکوس، SFE، ERSFS، BAAA، الگوریتم ژنتیک (GA)، بهینه‌سازی ذرات، بهینه‌سازی نهنگ‌ها و کرم شب تاب برای انتخاب ویژگی‌ها به کار گرفته شده است [۹-۱۸، ۲۹، ۳۰].

الگوریتم‌های ترتیبی Wrapper شامل SFS، SFFS

[۳۳]. جنگل تصادفی، که ترکیبی از چندین درخت تصمیم است، به بهبود دقت پیش‌بینی و کاهش خطر بیش‌برازش کمک می‌کند. شبکه‌های عصبی پرسپترون چندلایه نیز با توانایی یادگیری غیرخطی و پردازش ویژگی‌های پیچیده، به مدل‌سازی دقیق‌تر و پیش‌بینی‌های قوی‌تر منجر می‌شوند [۳۴]. استفاده از این مدل‌های مختلف به پژوهشگران امکان می‌دهد تا از جوانب مختلف داده‌ها را بررسی کرده و نتایج دقیق‌تری به دست آورند.

در این پژوهش برای ارزیابی عملکرد مدل‌های پیشنهادی، از روش اعتبارسنجی متقاطع ۵ تایی استفاده شد. معیارهای عملکردی همچون دقت، حساسیت، ویژگی، صحت و F1-score بر اساس ماتریس آشفتگی طبق جدول ۱ محاسبه شده‌اند. ماتریس آشفتگی مقادیر واقعی کلاس‌ها (بیمار یا سالم) را با پیش‌بینی‌های مدل مقایسه می‌کند و به ما امکان می‌دهد تا تعداد موارد مثبت واقعی (TP)، مثبت کاذب (FP)، منفی کاذب (FN) و منفی واقعی (TN) را مشخص کنیم. همچنین، از شاخص AUC-ROC برای ارزیابی جامع‌تر عملکرد مدل‌ها استفاده شده است. این شاخص توانایی مدل در تمایز بین کلاس‌های مختلف را نشان می‌دهد [۳۵]. تمامی مراحل این پژوهش، از جمله پیش‌پردازش داده‌ها، انتخاب ویژگی‌ها، آموزش مدل‌ها و ارزیابی آن‌ها، با استفاده از زبان برنامه‌نویسی پایتون پیاده‌سازی شده است.

در این مطالعه، به‌منظور ارزیابی عملکرد روش‌های انتخاب ویژگی در تشخیص بیماری آلزایمر، از مجموعه داده استاندارد بیماران آلزایمر ADNI استفاده شد. ابتدا ده روش انتخاب ویژگی شامل Wrapper، SFS، RFE، ERSFS، SFE، GA، WOA، BAAA، SPFSR، PSO و FA بر روی این مجموعه داده اعمال شدند تا ویژگی‌های مهم برای تشخیص بیماری استخراج شوند. سپس، ویژگی‌های انتخاب‌شده توسط هر یک از این روش‌ها با استفاده از پنج الگوریتم مختلف یادگیری ماشین شامل RF، J48، SVM، KNN و MLP مورد ارزیابی قرار گرفتند. این ارزیابی با استفاده از تکنیک اعتبارسنجی متقاطع پنج‌برابری انجام شد تا عملکرد مدل‌ها به‌صورت دقیق‌تری سنجیده شود.

طریق توابع آستانه، در کاهش تعداد ویژگی‌ها مؤثر است و در عین حال، دقت طبقه‌بندی را به حداکثر می‌رساند.

روش‌های فراابتکاری متعددی برای انتخاب ویژگی به‌کار گرفته می‌شوند که از شبیه‌سازی رفتار طبیعی الهام گرفته‌اند. الگوریتم ژنتیک (GA) با الهام از فرآیند تکامل طبیعی، با استفاده از اپراتورهای مانند جهش و کراس‌اور، بهترین زیرمجموعه ویژگی‌ها را جستجو می‌کند [۱۴]. الگوریتم بهینه‌سازی ازدحام ذرات (PSO) با شبیه‌سازی رفتار جمعی ذرات در فضای جستجو، به دنبال موقعیت‌های بهینه است [۱۵]. بهینه‌سازی نهنگ‌ها (WOA) از الگوی شکار نهنگ‌ها برای یافتن راه‌حل‌های بهینه استفاده می‌کند و به‌ویژه در مسائلی با فضای جستجوی بزرگ کارایی دارد [۱۶]. همچنین، الگوریتم کرم شب‌تاب (FA) با شبیه‌سازی رفتار جذب کرم‌های شب‌تاب به منابع نور، زیرمجموعه‌های بهینه ویژگی را جستجو می‌کند [۱۷]. این الگوریتم‌ها با تقویت فرآیند جستجو و در نظر گرفتن هم‌افزایی بین ویژگی‌ها، به انتخاب مؤثرترین زیرمجموعه‌ها برای بهبود عملکرد مدل‌های یادگیری ماشین کمک می‌کنند.

برای انجام یک تحلیل جامع‌تر، از مجموعه‌ای از مدل‌های یادگیری ماشین استفاده می‌شود که شامل روش‌های متنوع و قدرتمند است. این مدل‌ها عبارتند از: K نزدیک‌ترین همسایه، ماشین بردار پشتیبان (SVM)، درخت تصمیم J48، جنگل تصادفی و شبکه‌های عصبی پرسپترون چندلایه (MLP). هر یک از این الگوریتم‌ها ویژگی‌ها و مزایای خاص خود را دارند که می‌توانند به تحلیل داده‌ها و پیش‌بینی بیماری آلزایمر کمک کنند.

مدل KNN به دلیل سادگی و قابلیت فهم بالا، به راحتی می‌تواند برای طبقه‌بندی داده‌ها بر اساس شباهت‌های آن‌ها به کار رود [۳۲]. ماشین بردار پشتیبان به دلیل قدرت بالا در جداسازی کلاس‌ها در فضاهای چندبعدی و توانایی برخورد با داده‌های غیرخطی بسیار مؤثر است. درخت تصمیم J48 که یک الگوریتم پرکاربرد در طبقه‌بندی است، با ایجاد مدل‌های تصمیم‌گیری بر اساس ویژگی‌ها، به تحلیل ساختاری داده‌ها کمک می‌کند

جدول ۱: شاخص‌های ارزیابی	
معیار عملکرد	محاسبه
دقت (Accuracy)	$(TP + TN) / (TP + FP + TN + FN)$
صحت (Precision)	$TP / (TP + FP)$
حساسیت (Sensitivity)	$P / (TP + FN)$
ویژگی (Specificity)	$TN / (FP + TN)$
امتیاز F1 (F1-score)	$(Precision * Sensitivity) / (Precision + Sensitivity) * 2$

یافته‌ها:

ویژگی‌های امتیاز آزمون کوتاه وضعیت ذهنی (MMSE)، امتیاز ارزیابی عملکردی، وجود شکایت‌های حافظه و وجود مشکلات رفتاری دارای بیشترین فراوانی بوده و به عنوان مهم‌ترین عوامل تشخیص زودهنگام بیماری آلزایمر شناسایی می‌شوند. همچنین، نتایج نشان می‌دهند که پس از این عوامل، ویژگی‌های امتیاز فعالیت‌های زندگی روزمره (ADL)، کلسترول تری‌گلیسرید، سردرگمی و وضعیت سیگار کشیدن، فراوانی نسبتاً بالایی را در انتخاب ویژگی‌ها به خود اختصاص داده‌اند، که نشان‌دهنده نقش مؤثر آن‌ها در تشخیص بیماری آلزایمر است. علاوه بر این، نقش ویژگی‌های سن، کیفیت خواب و سابقه خانوادگی بیماری آلزایمر نیز باید در مراحل بعدی مورد توجه قرار گیرد.

پس از اجرای روش پیشنهادی و ارزیابی نتایج بر اساس معیار F1، بهترین ترکیب از ویژگی‌های انتخاب‌شده و الگوریتم‌های یادگیری ماشین، با محاسبه میانگین عملکرد در ۱۰ اجرای مستقل، شناسایی و نتایج در جدول ۲ گزارش شده است. این رویکرد علاوه بر افزایش دقت در تشخیص بیماری، امکان مقایسه عملکرد روش‌های مختلف انتخاب ویژگی و الگوریتم‌های یادگیری ماشین در ترکیب با یکدیگر را فراهم می‌کند. تحلیل نتایج نشان می‌دهد که این فرآیند نه تنها به شناسایی ویژگی‌های مؤثر کمک می‌کند، بلکه کارایی کلی مدل‌ها را نیز بهبود می‌بخشد.

همان‌طور که در جدول ۲ نشان داده شده است،

جدول ۲ - نتایج انتخاب ویژگی برای مجموعه داده بیماران آلزایمر ADNI										
ردیف	روش انتخاب ویژگی	سن	استعمال دخانیات	کیفیت خواب	سابقه خانوادگی	کلسترول	MMSE	ارزیابی عملکردی	شکایات حافظه	مشکلات رفتاری
۱	ERSFS	*	*	*	*	*	*	*	*	*
۲	SFE	*	*	*	*	*	*	*	*	*
۳	SPFSR	*	*	*	*	*	*	*	*	*
۴	SFS	*	*	*	*	*	*	*	*	*
۵	RFE	*	*	*	*	*	*	*	*	*
۶	PSO	*	*	*	*	*	*	*	*	*
۷	WOA	*	*	*	*	*	*	*	*	*
۸	GA	*	*	*	*	*	*	*	*	*
۹	BAAA	*	*	*	*	*	*	*	*	*
۱۰	FA	*	*	*	*	*	*	*	*	*
فراوانی	۱	۲	۱	۱	۱	۲	۸	۱۰	۸	۹

ویژگی‌های سن، کیفیت خواب و سابقه خانوادگی بیماری آلزایمر نیز باید در مراحل بعدی مورد توجه قرار گیرد.

شکل ۲ عملکرد روش‌های یادگیری ماشین پس از اعمال ده روش مختلف انتخاب ویژگی بر روی مجموعه داده آلزایمر ADNI را نشان می‌دهد. همان‌طور که در این شکل مشاهده می‌شود، الگوریتم ERSFS نسبت به سایر روش‌ها بهترین عملکرد را داشته و توانسته بهترین نتایج را ارائه دهد. این نتایج نشان‌دهنده تأثیر مثبت این روش انتخاب ویژگی در بهبود عملکرد مدل‌های یادگیری ماشین است.

همان‌طور که در جدول ۲ نشان داده شده است، ویژگی‌های امتیاز آزمون کوتاه وضعیت ذهنی (MMSE)، امتیاز ارزیابی عملکردی، وجود شکایت‌های حافظه و وجود مشکلات رفتاری دارای بیشترین فراوانی بوده و به عنوان مهم‌ترین عوامل تشخیص زودهنگام بیماری آلزایمر شناسایی می‌شوند. همچنین، نتایج نشان می‌دهند که پس از این عوامل، ویژگی‌های امتیاز فعالیت‌های زندگی روزمره (ADL)، کلسترول تری‌گلیسرید، سردردگی و وضعیت سیگار کشیدن، فراوانی نسبتاً بالایی را در انتخاب ویژگی‌ها به خود اختصاص داده‌اند، که نشان‌دهنده نقش مؤثر آن‌ها در تشخیص بیماری آلزایمر است. علاوه بر این، نقش



شکل ۲: عملکرد روش‌های یادگیری ماشین پس از اعمال ده روش مختلف انتخاب ویژگی بر روی مجموعه داده آلزایمر

یادگیری ماشینی جنگل تصادفی و درخت تصمیم بدست آمده است.

جدول ۳ عملکرد معیارهای مختلف ارزیابی با استفاده از مدل‌های مختلف یادگیری ماشینی را پس از اعمال روش‌های مختلف انتخاب ویژگی نشان می‌دهد.

همانطور که در این شکل مشخص است، روش ERSFS، بهترین نتایج را بر روی کلیه معیارهای ارزیابی نشان می‌دهد. بنابراین طبق جدول ۳ ویژگی‌های انتخابی توسط این روش نیز بهترین عملکرد را ارائه می‌دهد.

ویژگی‌های انتخابی توسط روش ERSFS شامل، امتیاز آزمون کوتاه وضعیت ذهنی، امتیاز ارزیابی عملکردی، وجود شکایت‌های حافظه، وجود مشکلات رفتاری و امتیاز فعالیت‌های زندگی روزمره می‌باشند. نتایج حاصل از جدول ۲ و شکل ۲ اهمیت ویژگی‌های انتخابی را تایید می‌کنند. همچنین طبق شکل ۲، پس از ERSFS، روش‌های انتخاب ویژگی RFE و PSO نتایج خوبی نسبت به سایر روش‌ها در اکثر مدل‌های یادگیری ماشینی ارائه می‌دهند. بنابراین علاوه بر چهار ویژگی مشترک آن‌ها با روش ERSFS، ویژگی‌های استعمال دخانیات و سردرگمی نیز دارای اهمیت می‌باشند. همچنین طبق شکل ۲ بهترین نتایج، با مدل‌های

جدول ۳: عملکرد معیارهای مختلف ارزیابی با استفاده از مدل‌های مختلف یادگیری ماشینی

		دقت										حساسیت										ویژگی	
ردیف	روش	KNN	SVM	D-Tree	R-F	MLP	KNN	SVM	D-Tree	R-F	MLP	KNN	SVM	D-Tree	R-F	MLP	KNN	SVM	D-Tree	R-F	MLP		
۱	ERSFS	۰/۸۷۴	۰/۸۷۸	۰/۹۰۷	۰/۹۵۱	۰/۸۸۴	۰/۸۴	۰/۸۴	۰/۸۹۷	۰/۹۴۱	۰/۸۶۶	۰/۸۴	۰/۸۳	۰/۸۹۷	۰/۹۴۱	۰/۸۶۶	۰/۸۴	۰/۸۳	۰/۸۹۷	۰/۹۴۱	۰/۸۶۶		
۲	SFE	۰/۷۰۲	۰/۶۴۶	۰/۷۶۱	۰/۷۸۴	۰/۷۸۷	۰/۶۵۶	۰/۶	۰/۷۳۸	۰/۷۶۲	۰/۷۵۲	۰/۶۵۶	۰/۶	۰/۷۳۸	۰/۷۶۲	۰/۷۵۲	۰/۶۵۶	۰/۶	۰/۷۳۸	۰/۷۶۲	۰/۷۵۲		
۳	SPFSR	۰/۷۰۲	۰/۶۴۶	۰/۷۶۱	۰/۷۸۴	۰/۷۷۹	۰/۶۵۶	۰/۶	۰/۷۳۸	۰/۷۶۲	۰/۷۴۸	۰/۶۵۶	۰/۶	۰/۷۳۸	۰/۷۶۲	۰/۷۴۸	۰/۶۵۶	۰/۶	۰/۷۳۸	۰/۷۶۲	۰/۷۴۸		
۴	SFS	۰/۷۶۵	۰/۷۷۹	۰/۷۷۴	۰/۷۸۲	۰/۷۷۹	۰/۷۲۱	۰/۷۳۹	۰/۷۴۹	۰/۷۵۷	۰/۷۴۸	۰/۷۲۱	۰/۷۳۹	۰/۷۴۹	۰/۷۵۷	۰/۷۴۸	۰/۷۲۱	۰/۷۳۹	۰/۷۴۹	۰/۷۵۷	۰/۷۴۸		
۵	RFE	۰/۷۸۷	۰/۷۶۵	۰/۷۴	۰/۷۹۱	۰/۷۹۱	۰/۷۲۱	۰/۷۳۶	۰/۷۴۹	۰/۷۵۷	۰/۷۴۸	۰/۷۲۱	۰/۷۳۶	۰/۷۴۹	۰/۷۵۷	۰/۷۴۸	۰/۷۲۱	۰/۷۳۶	۰/۷۴۹	۰/۷۵۷	۰/۷۴۸		
۶	PSO	۰/۷۸۱	۰/۷۷۹	۰/۷۷۶	۰/۷۹۵	۰/۷۸۹	۰/۷۴	۰/۷۳۸	۰/۷۵۲	۰/۷۶۹	۰/۷۵۶	۰/۷۴	۰/۷۳۸	۰/۷۵۲	۰/۷۶۹	۰/۷۵۶	۰/۷۴	۰/۷۳۸	۰/۷۵۲	۰/۷۶۹	۰/۷۵۶		
۷	WOA	۰/۷۳۸	۰/۶۴۶	۰/۶۷۲	۰/۷۶۹	۰/۷۶۳	۰/۶۹۷	۰/۶۴	۰/۷۲۶	۰/۷۱۷	۰/۶۹۷	۰/۶۴	۰/۷۲۶	۰/۷۱۷	۰/۶۹۷	۰/۶۴	۰/۷۲۶	۰/۷۱۷	۰/۶۹۷	۰/۶۴	۰/۷۲۶		
۸	GA	۰/۷۱۶	۰/۷۳۷	۰/۷۶۲	۰/۷۹۲	۰/۷۸۹	۰/۶۷۳	۰/۶۷۳	۰/۷۳۷	۰/۷۶۸	۰/۷۵۶	۰/۷۱	۰/۷۵۶	۰/۷۶۸	۰/۷۳۷	۰/۷۶۸	۰/۷۵۶	۰/۷۱	۰/۷۵۶	۰/۷۶۸	۰/۷۳۷		
۹	BAAA	۰/۷۶۸	۰/۷۸۱	۰/۷۶۹	۰/۷۸۲	۰/۷۸۲	۰/۷۲۵	۰/۷۴۲	۰/۷۴۳	۰/۷۵۶	۰/۷۵۱	۰/۷۴۳	۰/۷۴۲	۰/۷۵۶	۰/۷۵۱	۰/۷۴۳	۰/۷۴۲	۰/۷۵۶	۰/۷۵۱	۰/۷۴۳	۰/۷۴۲		
۱۰	FA	۰/۷۴۳	۰/۷۵۶	۰/۷۱۴	۰/۷۶	۰/۷۵۴	۰/۷۰۸	۰/۷۱۱	۰/۶۸۹	۰/۷۳۴	۰/۷۲	۰/۷۰۸	۰/۷۱۱	۰/۶۸۹	۰/۷۳۴	۰/۷۲	۰/۷۰۸	۰/۷۱۱	۰/۶۸۹	۰/۷۳۴	۰/۷۲		
ردیف	روش	صحت										اندازه F1										ROC-Area	
		KNN	SVM	D-Tree	R-F	MLP	KNN	SVM	D-Tree	R-F	MLP	KNN	SVM	D-Tree	R-F	MLP	KNN	SVM	D-Tree	R-F	MLP		
۱	ERSFS	۰/۸۴۸	۰/۸۴۵	۰/۸۹۹	۰/۹۵	۰/۸۷۸	۰/۸۳	۰/۸۳۳	۰/۸۹۷	۰/۹۴۱	۰/۸۶۶	۰/۸۳۳	۰/۸۱	۰/۸۹۸	۰/۹۴۵	۰/۸۷۱	۰/۸۳۳	۰/۸۱	۰/۸۹۸	۰/۹۴۵	۰/۸۷۱		
۲	SFE	۰/۶۷۱	۰/۶۲۳	۰/۷۳۸	۰/۷۶۴	۰/۷۷۱	۰/۶۵۶	۰/۶	۰/۷۳۸	۰/۷۶۲	۰/۷۵۲	۰/۶۶	۰/۶۲	۰/۷۳۸	۰/۷۶۲	۰/۷۵۲	۰/۶۶	۰/۶۲	۰/۷۳۸	۰/۷۶۲	۰/۷۵۲		
۳	SPFSR	۰/۶۷۱	۰/۶۲۳	۰/۷۳۸	۰/۷۶۴	۰/۷۵۹	۰/۷۲۱	۰/۷۳۹	۰/۷۴۹	۰/۷۵۷	۰/۷۴۸	۰/۷۲۱	۰/۷۳۹	۰/۷۴۹	۰/۷۵۷	۰/۷۴۸	۰/۷۲۱	۰/۷۳۹	۰/۷۴۹	۰/۷۵۷	۰/۷۴۸		
۴	SFS	۰/۷۴۷	۰/۷۶۳	۰/۷۵۲	۰/۷۶۲	۰/۷۵۹	۰/۷۲۱	۰/۷۳۹	۰/۷۴۹	۰/۷۵۷	۰/۷۴۸	۰/۷۲۱	۰/۷۳۹	۰/۷۴۹	۰/۷۵۷	۰/۷۴۸	۰/۷۲۱	۰/۷۳۹	۰/۷۴۹	۰/۷۵۷	۰/۷۴۸		
۵	RFE	۰/۷۶	۰/۷۵۴	۰/۷۲۲	۰/۷۷۴	۰/۷۷۴	۰/۷۴	۰/۷۳۶	۰/۷۱۹	۰/۷۸۴	۰/۷۸۲	۰/۷۴	۰/۷۳۶	۰/۷۱۹	۰/۷۸۴	۰/۷۸۲	۰/۷۴	۰/۷۳۶	۰/۷۱۹	۰/۷۸۴	۰/۷۸۲		
۶	PSO	۰/۷۶۶	۰/۷۶۳	۰/۷۵۳	۰/۷۷۶	۰/۷۷۲	۰/۷۴	۰/۷۳۸	۰/۷۵۲	۰/۷۶۹	۰/۷۵۶	۰/۷۴	۰/۷۳۸	۰/۷۵۲	۰/۷۶۹	۰/۷۵۶	۰/۷۴	۰/۷۳۸	۰/۷۵۲	۰/۷۶۹	۰/۷۵۶		
۷	WOA	۰/۷۱۴	۰/۶۳۲	۰/۶۴۵	۰/۷۵۳	۰/۷۴۷	۰/۶۹۷	۰/۶۴	۰/۷۲۶	۰/۷۱۷	۰/۶۹۷	۰/۶۴	۰/۷۲۶	۰/۷۱۷	۰/۶۹۷	۰/۶۴	۰/۷۲۶	۰/۷۱۷	۰/۶۹۷	۰/۶۴	۰/۷۲۶		
۸	GA	۰/۶۸۷	۰/۷۲۴	۰/۷۳۹	۰/۷۷۳	۰/۷۷۲	۰/۶۷۱	۰/۶۷۳	۰/۷۳۷	۰/۷۶۸	۰/۷۵۶	۰/۶۷۳	۰/۶۷۳	۰/۷۳۷	۰/۷۶۸	۰/۷۵۶	۰/۶۷۳	۰/۶۷۳	۰/۷۳۷	۰/۷۶۸	۰/۷۵۶		
۹	BAAA	۰/۷۵۱	۰/۷۶۵	۰/۷۴۶	۰/۷۶۱	۰/۷۶۳	۰/۷۲۵	۰/۷۴۲	۰/۷۴۳	۰/۷۵۶	۰/۷۵۱	۰/۷۴۳	۰/۷۴۲	۰/۷۵۶	۰/۷۵۱	۰/۷۴۳	۰/۷۴۲	۰/۷۵۶	۰/۷۵۱	۰/۷۴۳	۰/۷۴۲		
۱۰	FA	۰/۷۱۸	۰/۷۳۸	۰/۶۸۷	۰/۷۳۲	۰/۷۳۲	۰/۷۰۸	۰/۷۱۱	۰/۶۸۹	۰/۷۳۴	۰/۷۲	۰/۷۰۸	۰/۷۱۱	۰/۶۸۹	۰/۷۳۴	۰/۷۲	۰/۷۰۸	۰/۷۱۱	۰/۶۸۹	۰/۷۳۴	۰/۷۲		

این پژوهش تأثیر روش‌های انتخاب ویژگی مبتنی بر Wrapper بر عملکرد الگوریتم‌های یادگیری ماشین در

بحث و نتیجه‌گیری:

امتیاز، نشانه‌ای از نیاز به مراقبت بیشتر و افت توانایی‌های عملکردی است، که این موضوع به‌ویژه در بیماران آلزایمر قابل توجه است. این یافته با پژوهش‌های هم‌سو است [۳۶،۳۹].

وجود شکایت‌های حافظه، یکی از نخستین نشانه‌هایی است که معمولاً توسط خود بیماران یا اطرافیان آن‌ها گزارش می‌شود. این شکایات می‌توانند به عنوان هشدار اولیه برای شروع اختلالات شناختی تلقی شوند. در کنار آن، مشکلات رفتاری مانند تحریک‌پذیری، افسردگی، و تغییرات شخصیتی نیز معمولاً در مراحل ابتدایی بیماری ظاهر می‌شوند و بر کیفیت زندگی بیمار تأثیر منفی می‌گذارند. این مشکلات رفتاری ممکن است به دلیل تغییرات در ساختارهای مغزی مرتبط با کنترل احساسات و رفتارها ایجاد شوند. نتایج ما در این خصوص با مطالعات مطابقت دارد [۵،۴۰،۴۱].

امتیاز فعالیت‌های زندگی روزمره نیز ابزاری است که برای ارزیابی میزان استقلال فرد در انجام وظایف روزمره مانند لباس پوشیدن، تغذیه، و بهداشت شخصی استفاده می‌شود. در بیماران آلزایمر، کاهش نمره ADL معمولاً با پیشرفت بیماری و نیاز بیشتر به مراقبت‌های حمایتی مرتبط است. این کاهش می‌تواند نشانه‌ای از کاهش توانایی‌های جسمی و شناختی باشد که از ویژگی‌های بارز آلزایمر در مراحل پیشرفته‌تر است. این یافته با مطالعات کیستودیو و همکاران، لی و همکاران، و فولیانگ و همکاران، هم‌راستا است [۳۹،۴۲،۴۳].

علاوه بر این ویژگی‌های کلیدی، نتایج این پژوهش نشان می‌دهد که عواملی نظیر کلاسترول تری‌گلیسرید، سردرگمی مکانی و زمانی، وضعیت سیگار کشیدن، سن، کیفیت خواب، و سابقه خانوادگی بیماری آلزایمر نیز در مراحل پیشرفته‌تر نقش مهمی ایفا می‌کنند، که توسط مطالعات تایید شده‌اند [۴۳-۴۶].

این نتایج نشان می‌دهد که استفاده از روش‌های مناسب انتخاب ویژگی می‌تواند بهبود قابل توجهی در دقت و سرعت تشخیص بیماری آلزایمر ایجاد کند و بر اهمیت

تشخیص بیماری آلزایمر را مورد بررسی قرار داده است. نتایج نشان می‌دهد که روش ERSFS عملکرد بهتری در انتخاب ویژگی‌های کلیدی و بهبود دقت تشخیص بیماری ارائه می‌دهد. این نتایج با یافته‌های مطالعات قبلی که تأثیر هم‌افزایی ویژگی‌ها بر عملکرد مدل‌های یادگیری ماشین را گزارش کرده‌اند، هم‌خوانی دارد. برای مثال، در پژوهش صباغ‌گل و همکاران روش ERSFS به دلیل توجه به تعاملات بین ویژگی‌ها و جلوگیری از گیر افتادن در بهینه‌های محلی، به‌عنوان یک روش مؤثر در انتخاب ویژگی‌های پزشکی معرفی شده است [۱۰].

ویژگی‌های کلیدی شناسایی شده در این مطالعه شامل امتیاز آزمون کوتاه وضعیت ذهنی (MMSE)، ارزیابی عملکردی، شکایت‌های حافظه، مشکلات رفتاری و امتیاز فعالیت‌های روزمره زندگی (ADL) هستند. این یافته‌ها با پژوهش‌های دیگر که اهمیت MMSE و ADL را در تشخیص زودهنگام آلزایمر برجسته کرده‌اند، مانند مطالعات چوور و همکاران و پارلاک و همکاران، هم‌سو هستند [۳۶،۳۷]. با این حال، نتایج این پژوهش نشان می‌دهد که در ترکیب روش‌های انتخاب ویژگی و یادگیری ماشین، می‌توان دقت تشخیص را با استفاده از تعداد کمتری از ویژگی‌ها افزایش داد که این دستاورد در مقایسه با روش‌های سنتی، مزیت بزرگی است.

این مطالعه تأکید می‌کند که آزمون کوتاه وضعیت ذهنی (MMSE) یکی از ابزارهای معتبر و مؤثر در تشخیص زودهنگام بیماری آلزایمر است، چرا که به‌طور دقیق توانایی‌های شناختی مانند حافظه، زبان، توجه، و جهت‌یابی را ارزیابی می‌کند. یافته‌های ما نشان می‌دهد که کاهش نمره MMSE با پیشرفت اختلالات شناختی و شدت بیماری ارتباط مستقیم دارد. این نتیجه با مطالعات چوور و همکاران، پارلاک و همکاران، و بونفا و همکاران هم‌خوانی دارد که همگی اهمیت این شاخص را برجسته کرده‌اند [۳۶-۳۸].

امتیاز ارزیابی عملکردی نیز به عنوان یک شاخص کلیدی برای سنجش توانایی‌های عملکردی و استقلال فرد در انجام فعالیت‌های روزمره شناخته شده است. کاهش این

این، گسترش مطالعه به دیگر جوامع و گروه‌های سنی می‌تواند به تعمیم‌پذیری نتایج و افزایش کارایی مدل‌ها کمک کند. این امر به ویژه در مورد جوامع با تنوع ژنتیکی و عوامل محیطی متفاوت اهمیت دارد.

ملاحظات اخلاقی:

پیروی از اصول اخلاق در پژوهش

این مطالعه دارای تاییدیه اخلاقی به شماره IR.IAU. BIRJAND.REC.1403.001 از دانشگاه آزاد اسلامی واحد بیرجند است.

حامی مالی

این مقاله از طرف هیچ گونه نهاد یا موسسه ای حمایت مالی نشده و تمام منابع مالی آن از طرف نویسنده اول یا نویسندگان تامین شده است.

مشارکت نویسندگان

حامد صباغ گل: نگارش، پیش نویس اصلی، اعتبارسنجی، نرم افزار، تحقیق؛ حمید سعادت فر: نقد و ویرایش، نظارت، روش‌شناسی، مفهوم‌سازی؛ مهدی خزاعی‌پور: منابع، ویرایش.

تعارض منافع

هیچ گونه تضاد منافعی از سوی پژوهشگران گزارش نشده است.

تشکر و قدردانی

از دانشگاه آزاد اسلامی واحد بیرجند برای حمایت و فراهم آوردن امکانات علمی جهت انجام این پژوهش صمیمانه سپاسگزاریم.

استفاده از این روش‌ها در ترکیب با الگوریتم‌های یادگیری ماشین تأکید می‌کند. انتخاب صحیح ویژگی‌های مرتبط می‌تواند نقش حیاتی در توسعه سیستم‌های تشخیص زودهنگام آلزایمر داشته باشد و به بهبود کیفیت زندگی بیماران کمک کند.

نتیجه گیری:

این پژوهش، با وجود دستاوردهای قابل توجه در بهبود دقت و کارایی تشخیص آلزایمر، با محدودیت‌هایی مواجه بوده است. نخست، تنوع محدود مجموعه داده‌ها و کمبود نمونه‌هایی با ابعاد جغرافیایی و فرهنگی گسترده‌تر، قابلیت تعمیم نتایج را کاهش می‌دهد. همچنین، فقدان اطلاعات دقیق‌تر درباره شاخص‌های ژنتیکی و تاریخیچه کامل خانوادگی، تحلیل عوامل وراثتی را محدود کرده است. در حوزه داده‌های بالینی، عدم دسترسی به بیومارکرهای پیشرفته و دقیق‌تر، فرصت‌هایی برای بهبود نتایج تشخیصی را از دست داده است. در نهایت، اطلاعات جامع‌تر درباره سبک زندگی و تأثیر محیط می‌تواند تحلیل عوامل خطر مرتبط با آلزایمر را ارتقا دهد.

همچنین در آینده، پژوهش‌های بیشتری می‌توانند بر روی بهبود روش‌های انتخاب ویژگی و ارزیابی مدل‌های تشخیص آلزایمر با استفاده از داده‌های مختلف تمرکز کنند. به ویژه، توسعه روش‌های انتخاب ویژگی که بتوانند با حجم بالای داده‌های چندوجهی مقابله کنند و مدل‌های یادگیری عمیق که بتوانند الگوهای پیچیده و غیرخطی را شناسایی کنند، از اهمیت بالایی برخوردار است. ترکیب داده‌های تصویربرداری مغزی، به ویژه تصاویر MRI و PET، با این ویژگی‌های کلینیکی می‌تواند به دقت بیشتری در تشخیص زودهنگام بیماری منجر شود. این امر به شناسایی تغییرات ساختاری و عملکردی مغز در مراحل اولیه بیماری کمک خواهد کرد. توسعه مدل‌های یادگیری عمیق، مانند شبکه‌های عصبی کانولوشنال و شبکه‌های عصبی بازگشتی، برای تحلیل تصاویر پزشکی، پردازش زبان طبیعی و استخراج اطلاعات از داده‌های الکترونیکی سلامت می‌تواند به پیشرفت‌های مهمی در این حوزه منجر شود. علاوه بر

References

1. Prince M, Wimo A, Guerchet M, Ali GC, Wu YT, Prina M. World Alzheimer report 2015. The global impact of dementia: An analysis of prevalence, incidence, cost and trends. London: Alzheimer's Disease International; 2015. [Link](#)
2. Madnani RS. Alzheimer's disease: A mini-review for the clinician. *Front Neurol*. 2023; 14:1178588. DOI: [10.3389/fneur.2023.1178588](#) PMID: [37426432](#)
3. Alzheimer's Association. 2019 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*. 2019; 15(3):321-87. DOI: [10.1016/j.jalz.2019.01.010](#)
4. World Health Organization. WHO health workforce support and safeguards list 2023. World Health Organization; 2023. [Link](#)
5. Alzheimer's Disease International. World Alzheimer report 2023: Reducing Dementia Risk: Never too early, never too late. London: Alzheimer's Disease International; 2023. [Link](#)
6. Gholamzadeh S, Heshmati B, Mani A, Petramfar P, Bagheri Z. The prevalence of alzheimer's disease; its risk and protective factors among the elderly population in Iran. *Shiraz E-Medical Journal*. 2017; 18(9):e57576. DOI: [10.5812/semj.57576](#)
7. Rathore S, Habes M, Iftikhar MA, Shacklett A, Davatzikos C. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *Neuroimage*. 2017; 155:530-548. DOI: [10.1016/j.neuroimage.2017.03.057](#) PMID: [28414186](#)
8. Crespo Márquez A. The curse of dimensionality. In *digital maintenance management: Guiding digital transformation in mintenance 2022* (pp. 67-86). Cham: Springer International Publishing; 2022. [Link](#)
9. Bouchlaghem Y, Akhiat Y, Amjad S. Feature selection: A review and comparative study. In *E3S web of Conferences 2022* (Vol. 351, p. 01046). EDP Sciences. DOI: [10.1051/e3s-conf/202235101046](#)
10. SabbaghGol H, Saadatfar H, Khazaiepoor M. Evolution of the random subset feature selection algorithm for classification problem. *KBS*. 2024; 285:111352. DOI: [10.1016/j.kno-sys.2023.111352](#)
11. Ahadzadeh B, Abdar M, Safara F, Khosravi A, Menhaj MB, Suganthan PN. SFE: A simple, fast, and efficient feature selection algorithm for high-dimensional data. *IEEE Transactions on Evolutionary Computation*. 2023; 27(6):1896-911. DOI: [10.1109/tevc.2023.3238420](#)
12. Akman DV, Malekipirbazari M, Yenice ZD, Yeo A, Adhikari N, Wong YK, et al. K-best feature selection and ranking via stochastic approximation. *Expert Systems with Applications*. 2023; 213:118864. DOI: [10.1016/j.eswa.2022.118864](#)
13. Turkoglu B, Uymaz SA, Kaya E. Binary artificial algae algorithm for feature selection. *Applied Soft Computing*. 2022; 120:108630. DOI: [10.1016/j.asoc.2022.108630](#)
14. Babatunde OH, Armstrong L, Leng J, Diepeveen D. A genetic algorithm-based feature selection. *IJECCE*. 2014; 5(4):899-905. [Link](#)
15. Tran B, Xue B, Zhang M. Improved PSO for feature selection on high-dimensional datasets. In *Simulated evolution and learning: 10th International Conference, SEAL 2014, Dunedin, New Zealand, December 15-18, 2014. Proceedings 10 2014* (pp. 503-515). Springer International Publishing; 2014. pp 503-15. [Link](#)
16. Mafarja M, Mirjalili S. Whale optimization approaches for wrapper feature selection. *Applied Soft Computing*. 2018; 62:441-53. DOI: [10.1016/j.asoc.2017.11.006](#)
17. Bacanin N, Venkatachalam K, Bezdan T, Zivkovic M, Abouhawwash M. A novel firefly algorithm approach for efficient feature selection with COVID-19 dataset. *Microprocess Microsyst*. 2023; 98:104778. DOI: [10.1016/j.micpro.2023.104778](#) PMID: [36785847](#)
18. Yang T. The application of machine learning in Alzheimer's disease. *Highl Sci Eng Technol*. 2024; 92:248-51. DOI: [10.54097/dqsnj347](#)
19. Alzheimer's disease facts and figures, 2021. *Alzheimer's & Dementia*. 17(3):327-4006. DOI: [10.1002/alz.12328](#)
20. Alshamlan H, Omar S, Aljurayyad R, Alabduljabbar R. Identifying effective feature selection methods for Alzheimer's disease biomarker Gene detection using machine learning. *Diagnostics (Basel)*. 2023; 13(10):1771. DOI: [10.3390/diagnostics13101771](#) PMID: [37238255](#)



21. Uwishema O, Kassahun Bekele B, Nazir A, Filbert Luta E, Abdalnaser Al-Saab E, Jacques Desire I, et al. Breaking barriers: Addressing inequities in Alzheimer's disease diagnosis and treatment in Africa. *Ann Med Surg (Lond)*. 2024; 86(9):5299-5303. DOI: [10.1097/MS9.0000000000002344](https://doi.org/10.1097/MS9.0000000000002344) PMID: [39239000](https://pubmed.ncbi.nlm.nih.gov/39239000/)
22. Milà-Alomà M, Brinkmalm A, Ashton NJ, Kvartsberg H, Shekari M, Operto G, et al. ALFA Study. CSF synaptic biomarkers in the preclinical stage of Alzheimer disease and their association with MRI and PET: A cross-sectional study. *Neurology*. 2021; 97(21):2065-78. DOI: [10.1212/WNL.00000000000012853](https://doi.org/10.1212/WNL.00000000000012853) PMID: [34556565](https://pubmed.ncbi.nlm.nih.gov/34556565/)
23. Dara OA, Lopez-Guede JM, Raheem HI, Rahebi J, Zulueta E, Fernandez-Gamiz U. Alzheimer's disease diagnosis using machine learning: A survey. *Appl Sci*. 2023; 13(14):8298. DOI: [10.3390/app13148298](https://doi.org/10.3390/app13148298)
24. Sudharsan M, Thailambal G. Alzheimer's disease prediction using machine learning techniques and principal component analysis (PCA). *Materials Today: Proceedings*. 2023; 81:182-90. DOI: [10.1016/j.matpr.2021.03.061](https://doi.org/10.1016/j.matpr.2021.03.061)
25. Faisal FU, Khatri U, Kwon GR. Diagnosis of Alzheimer's disease using combined feature selection method. *Journal of Korea Multimedia Society*. 2021; 24(5):667-75. DOI: [10.9717/kmms.2021.24.5.667](https://doi.org/10.9717/kmms.2021.24.5.667)
26. ADNI. Alzheimer's Disease Neuroimaging Initiative. 2003 [cited 2024] Available from: [Link](https://adni.loni.usc.edu/)
27. Pudjihartono N, Fadason T, Kempa-Liehr AW, O'Sullivan JM. A review of feature selection methods for machine learning-based disease risk prediction. *front bioinform*. 2022; 2:927312. DOI: [10.3389/fbinf.2022.927312](https://doi.org/10.3389/fbinf.2022.927312) PMID: [36304293](https://pubmed.ncbi.nlm.nih.gov/36304293/)
28. Patro S, Sahu KK. Normalization: A preprocessing stage. *arXiv preprint arXiv*; 2015. [Link](https://arxiv.org/abs/1508.06416)
29. Zhou J, Lu J, Wu Q, Wen J. High-dimensional feature selection in classification: a length-adaptive evolutionary approach. In *2022 IEEE International Conference on Networking, Sensing and Control (ICNSC) 2022*. United States: Institute of Electrical and Electronics Engineers; 2022 pp. 1-5. [Link](https://doi.org/10.1109/ICNSC52621.2022.9920000)
30. Guha R, Chatterjee B, Khalid Hassan SK, Ahmed S, Bhattacharyya T, Sarkar R. Py_FS: A python package for feature selection using meta-heuristic optimization algorithms. *Incomputational intelligence in pattern recognition: Proceedings of CIPR 2021*. Springer Singapore; 2022. pp. 495-504. [Link](https://doi.org/10.1007/978-98-1-19-8000-0_23)
31. Lakshmipadmaja D, Vishnuvardhan B. Classification performance improvement using random subset feature selection algorithm for data mining. *Big Data Research*. 2018; 12:1-2. DOI: [10.1016/j.bdr.2018.02.007](https://doi.org/10.1016/j.bdr.2018.02.007)
32. Zhang Z. Introduction to machine learning: k-nearest neighbors. *Ann Transl Med*. 2016; 4(11):218. DOI: [10.21037/atm.2016.03.37](https://doi.org/10.21037/atm.2016.03.37) PMID: [27386492](https://pubmed.ncbi.nlm.nih.gov/27386492/)
33. Maimon OZ, Rokach L. Data mining with decision trees: Theory and applications. *World scientific*; 2014. [Link](https://doi.org/10.1142/9789814613318)
34. Goodfellow I, Bengio Y, Courville A. *Example Notation for Deep Learning*; 2016. [Link](https://arxiv.org/abs/1603.04467)
35. Sanyal D, Bosch N, Paquette L. Feature selection metrics: Similarities, differences, and characteristics of the selected models. *International Educational Data Mining Society*; 2020. [Link](https://doi.org/10.1016/j.iedm.2020.100000)
36. Chéour S, Chéour C, Gendreau T, Bouazizi M, Singh KP, Saeidi A, et al. Remediation of cognitive and motor functions in Tunisian elderly patients with mild Alzheimer's disease: Implications of music therapy and/or physical rehabilitation. *Front Aging Neurosci*. 2023; 15:1216052. DOI: [10.3389/fnagi.2023.1216052](https://doi.org/10.3389/fnagi.2023.1216052) PMID: [37539345](https://pubmed.ncbi.nlm.nih.gov/37539345/)
37. Parlak MM, Bizbinar Ö, Köse A. The effect of holistic therapy in Alzheimer's disease. *Altern Ther Health Med*. 2023; 29(3):52-9. PMID: [36689358](https://pubmed.ncbi.nlm.nih.gov/36689358/)
38. Bonfa A, Zhang W, Keller Y, Barbisch F, Schürmann P, Moguel LF, et al. 0840 Exploring the relationship between sleep disturbances and Alzheimer's disease using machine learning. *Sleep*. 2023; 46(Supplement_1):A370. DOI: [10.1093/sleep/zsad077.0841](https://doi.org/10.1093/sleep/zsad077.0841)
39. Custodio N, Montesinos R, Chambergo-Michilot D, Herrera-Perez E, Pintado-Caipa M, Seminario GW, et al. A functional assessment tool to distinguish controls from Alzheimer's disease in Lima, Peru. *Am J Alzheimers Dis Other Demen*. 2022; 37:15333175221104354. DOI: [10.1177/15333175221104354](https://doi.org/10.1177/15333175221104354) PMID: [35656899](https://pubmed.ncbi.nlm.nih.gov/35656899/)
40. Snitz BE, Lopez OL, McDade E, Becker JT, Co-

hen AD, Price JC, et al. Amyloid- β imaging in older adults presenting to a memory clinic with subjective cognitive decline: A pilot study. *J Alzheimers Dis.* 2015; 48 Suppl 1(0 1):S151-9. DOI: [10.3233/JAD-150113](https://doi.org/10.3233/JAD-150113) PMID: [26402082](https://pubmed.ncbi.nlm.nih.gov/26402082/)

41. Chatterjee P, Pedrini S, Doecke JD, Dore V, Gozette KG, Taddei K, et al. The diagnostic and prognostic potential of plasma GFAP in older adults with subjective memory complaints at risk of Alzheimer's disease. *Alzheimer's & Dementia.* 2023; 19:e072740. DOI: [10.1002/alz.072740](https://doi.org/10.1002/alz.072740)
42. Li J, Wu X, Tan X, Wang S, Qu R, Wu X, et al. The efficacy and safety of anti-A β agents for delaying cognitive decline in Alzheimer's disease: A meta-analysis. *Front Aging Neurosci.* 2023; 15:1257973. DOI: [10.3389/fnagi.2023.1257973](https://doi.org/10.3389/fnagi.2023.1257973) PMID: [38020763](https://pubmed.ncbi.nlm.nih.gov/38020763/)
43. Yi F, Yang H, Chen D, Qin Y, Han H, Cui J, et al. XGBoost-SHAP-based interpretable diagnostic framework for alzheimer's disease. *BMC Med Inform Decis Mak.* 2023; 23(1):137. DOI: [10.1186/s12911-023-02238-9](https://doi.org/10.1186/s12911-023-02238-9) PMID: [37491248](https://pubmed.ncbi.nlm.nih.gov/37491248/)
44. Loughman A, Adler CJ, Macpherson H. Unlocking modifiable risk factors for Alzheimer's disease: does the oral microbiome hold some of the keys? *J Alzheimers Dis.* 2023; 92(4):1111-29. DOI: [10.3233/JAD-220760](https://doi.org/10.3233/JAD-220760) PMID: [36872775](https://pubmed.ncbi.nlm.nih.gov/36872775/)
45. Deng K, Gupta D, Shu XO, Lipworth L, Zheng W, Yu D. Abstract P449: Life's essential 8 and risk of coronary heart disease: A prospective study among racially and geographically diverse populations. *Circulation.* 2023; 147(Suppl_1):AP449. DOI: [10.1161/circ.147.suppl_1.P449](https://doi.org/10.1161/circ.147.suppl_1.P449)
46. Liu B, Zhao Y, Zhao G, Jin L, Li W. Association of hearing loss with cognitive decline in Alzheimer's disease and vascular dementia. DOI: [10.21203/rs.3.rs-1075876/v1](https://doi.org/10.21203/rs.3.rs-1075876/v1)

