Check for updates

Original Article

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2023; 10(2): 152-164
10.34172/jhbmi.2023.21

Early Prediction of Alzheimer's Disease Using Interpretable Machine Learning Algorithms

Sorayaie Azar Amir[1], Rezaei Samaneh[1], Bagherzadeh Mohasefi Jamshid[2], Niazkhani Zahra[3], Pirnejad Habibollah[4]*

• **Received:** 7 Feb 2023 • **Accepted:** 24 May 2023

Introduction: Alzheimer's disease is one of the most common neurodegenerative diseases in adults. The progressive nature of Alzheimer's disease causes widespread damage to the brain, and early diagnosis can manage the disease and slow down its progression effectively.
Method: In this study, a dataset related to the early prediction of Alzheimer's was used. Spark framework was used for data management and three machine learning algorithms including Naïve Bayes, Decision Tree, and Artificial Neural Networks with the best hyperparameters were implemented and compared. To prevent overfitting and measure the efficiency of the models, a 5-fold cross-validation method was utilized. Furthermore, a method was used for interpreting machine learning black box models.
Results: The decision tree and artificial neural network models obtained 98.61% accuracy and 98.60% F1-Score in the Spark framework including one or three computers. Important features in the decision-making process of the artificial neural network were identified using the interpretability technique. In addition, the computational time required for training the proposed models was calculated through different approaches, and the use of multiple computers was 35.95% faster than a single computer.
Conclusion: With increasing the number of Alzheimer's disease patients around the world, the need for a decision support system using machine learning algorithms, which can predict the disease early in a huge amount of data, is felt more. Therefore, the machine learning models proposed in this study for early prediction of Alzheimer's disease as an interpretable auxiliary tool in the decision-making process can help clinicians.

Keywords: Early Prediction, Alzheimer's Disease, Machine Learning, Interpretability, Big Data

• **Citation:** Sorayaie Azar A, Rezaei S, Bagherzadeh Mohasefi J, Niazkhani Z, Pirnejad H. Early Prediction of Alzheimer's Disease Using Interpretable Machine Learning Algorithms. Journal of Health and Biomedical Informatics 2023; 10(2): 152-64. [In Persian] doi: 10.34172/jhbmi.2023.21

***Amir Sorayaie Azar and Samaneh Rezaei both contributed equally in this study as the first authors.**

1. Ph.D. Student in Computer Engineering, Faculty of Electrical and Computer Engineering, Urmia University, Urmia, Iran
2. Ph.D. in Software Engineering, Professor, Faculty of Electrical and Computer Engineering, Urmia University, Urmia, Iran
3. Ph.D. in Medical Informatics, Associate Professor, Nephrology and Kidney Transplant Research Center, Clinical Research Institute, Urmia University of Medical Sciences, Urmia, Iran
4. Ph.D. in Medical Informatics, Associate Professor, Patient Safety Research Center, Clinical Research Institute, Urmia University of Medical Sciences, Urmia, Iran
***Corresponding Author:** Jamshid Bagherzadeh Mohasefi
Address: Faculty of Electrical and Computer Engineering, Urmia University, Urmia, Iran
• **Tel:** 044-33438840 • **Email:** j.bagherzadeh@urmia.ac.ir

© 2023 The Author(s); Published by Kerman University of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cite

[Downloaded from jhbmi.com on 2024-10-12]
[DOI: 10.34172/jhbmi.2023.21]

مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره دهم شماره دوم. ص ۱۶۴-۱۵۲، ۱۴۰۲
doi 10.34172/jhbmi.2023.21

مقاله پژوهشی

پیش‌بینی زودهنگام بیماری آلزایمر توسط الگوریتم‌های یادگیری ماشین قابل تفسیر

امیر ثریائی آذر[۱]، سمانه رضائی[۱]، جمشید باقرزاده محاسفی[۲]*، زهرا نیازخانی[۳]، حبیب‌الله پیرنژاد[۴]

• دریافت مقاله: ۱۴۰۱/۱۱/۱۸ • پذیرش مقاله: ۱۴۰۲/۳/۳

مقدمه: بیماری آلزایمر یکی از شایع‌ترین اختلال‌های زوال مغز در بزرگسالان است. ماهیت پیش‌رونده آلزایمر باعث بروز آسیب گسترده به قسمت‌های مغز می‌شود و تشخیص زودرس می‌تواند در مدیریت بیماری و کاهش سرعت پیشرفت آن مؤثر باشد.

روش: در این مطالعه، از یک مجموعه داده مربوط به پیش‌بینی زودهنگام بیماری آلزایمر در بستر اسپارک (Spark Framework) به منظور مدیریت داده‌ها استفاده شد و با سه الگوریتم یادگیری ماشین از جمله بیز ساده، درخت تصمیم، شبکه عصبی مصنوعی با بهترین فراپارمترها پیاده‌سازی و مقایسه گردید. به منظور جلوگیری از بیش برازش و سنجش کارایی مدل‌ها از ۵-تکرار اعتبارسنجی متقابل استفاده شد. در آخر، یک روش برای تفسیر مدل‌های جعبه سیاه یادگیری ماشین در این مطالعه در نظر گرفته شد.

نتایج: مدل‌های درخت تصمیم و شبکه عصبی مصنوعی دقتی برابر ۹۸/۶۱٪ و امتیاز-اف۱ ۹۸/۶۰٪ را در بستر اسپارک شامل یک و یا سه کامپیوتر به دست آوردند. ویژگی‌های تأثیرگذار در تصمیم‌گیری شبکه عصبی مصنوعی با استفاده از تکنیک تفسیرپذیری مشخص شد. علاوه بر این، زمان محاسباتی لازم برای یادگیری مدل‌های پیشنهادی در رویکردهای مختلف محاسبه و استفاده از چندین کامپیوتر در مقایسه با تک کامپیوتر ۳۵/۹۵٪ سریع‌تر بود.

نتیجه‌گیری: با افزایش تعداد بیماران آلزایمر در جهان، لزوم وجود سیستم پشتیبان تصمیم که بتواند در حجم عظیمی از داده‌ها با استفاده از الگوریتم‌های یادگیری ماشین، توانایی پیش‌بینی زودهنگام بیماری را داشته باشد بیشتر احساس می‌شود. بدین منظور، مدل‌های یادگیری ماشین پیشنهادی این مطالعه، می‌توانند به عنوان یک ابزار کمکی با قابلیت تفسیرپذیری در تصمیم‌گیری پزشکان به جهت پیش‌بینی زودهنگام بیماری آلزایمر کمک نمایند.

کلیدواژه‌ها: تشخیص زودهنگام، بیماری آلزایمر، یادگیری ماشین، تفسیرپذیری، داده‌های حجیم

• **ارجاع:** ثریائی آذر امیر، رضائی سمانه، باقرزاده محاسفی جمشید، نیازخانی زهرا، پیرنژاد حبیب‌الله. پیش‌بینی زودهنگام بیماری آلزایمر توسط الگوریتم‌های یادگیری ماشین قابل تفسیر. مجله انفورماتیک سلامت و زیست پزشکی ۱۴۰۲؛ ۱۰(۲): ۶۴-۱۵۲. doi: 10.34172/jhbmi.2023.21

*** امیر ثریایی آذر و سمانه رضایی به صورت مشترک نویسنده اول این مقاله می باشند.**

۱. دانشجوی دکتری تخصصی مهندسی کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه ارومیه، ارومیه، ایران
۲. دکتری تخصصی مهندسی نرم‌افزار، استاد، دانشکده مهندسی برق و کامپیوتر، دانشگاه ارومیه، ارومیه، ایران
۳. دکتری تخصصی انفورماتیک پزشکی، دانشیار، مرکز تحقیقات نفرولوژی و پیوند کلیه، پژوهشکده تحقیقات بالینی، دانشگاه علوم پزشکی ارومیه، ارومیه، ایران
۴. دکتری تخصصی انفورماتیک پزشکی، دانشیار، مرکز تحقیقات ایمنی بیمار، پژوهشکده تحقیقات بالینی، دانشگاه علوم پزشکی ارومیه، ارومیه، ایران

*** نویسنده مسئول:** جمشید باقرزاده محاسفی
آدرس: ارومیه، کیلومتر ۱۱ جاده سرو، دانشگاه ارومیه، دانشکده مهندسی برق و کامپیوتر
• **شماره تماس:** ۰۴۴-۳۳۴۳۸۸۴۰ • **Email:** j.bagherzadeh@urmia.ac.ir

[Downloaded from jhbmi.com on 2024-10-12]
[DOI: 10.34172/jhbmi.2023.21]

مقدمه

بیماری آلزایمر از جمله شایع‌ترین اختلال‌های مغزی بوده که در آمریکا به عنوان ششمین عامل منجر به مرگ انسان شناخته شده است [۱]. اختلال زوال مغز برای توصیف شرایطی که به سبب آسیب در مغز باعث مشکلاتی از جمله در حافظه، توانایی صحبت، فکر کردن و انجام کارهای روزمره بیمار می‌شود، استفاده می‌گردد [۲]. بر اساس گزارش سازمان بهداشت جهانی (World Health Organization) در سال ۲۰۱۹، حدود ۵۰ میلیون نفر دچار اختلال زوال مغز بوده و هر سال حدود ۱۰ میلیون نفر بر این تعداد افزوده می‌شود [۳]. در آلزایمر، قشر خارجی مغز مانند هیپوکامپ، شکمچه مغز و غشاء مغزی دچار آسیب می‌شود. بدین ترتیب، بیمار به درستی قادر به یادآوری وقایع جدید نبوده و در نتیجه در انجام کارهای روزمره خود دچار مشکلات جدی می‌شود [۲].

آلزایمر نوعی بیماری پیش‌رونده است که به سبب ماهیت نوروژنیک آن، تا مدت‌های طولانی بعد از شروع بیماری، هیچ‌گونه علائم بالینی بروز نمی‌دهد [۴]. علاوه بر این، با وجود مطالعات گسترده در این حوزه، تاکنون درمان قطعی برای بهبود علائم بیماری یا جلوگیری از پیشرفت این بیماری گزارش نشده است [۴،۵]. به طور مثال، داروی Aduhelm که توسط سازمان غذا و داروی ایالات متحده مورد تأیید می‌باشد؛ فقط پلاک آمیلوئید را کاهش می‌دهد و برای تأیید اثربخشی این دارو در درمان آلزایمر، شواهد زیادی نیاز است [۶].

علائم بالینی این بیماری در مراحل غیر اولیه آن بروز می‌کنند و به سبب گستردگی میزان آسیب وارده به مغز، علائم پایدارتر می‌باشند [۷]. با پیشرفت بیماری، مشکلات شدید تکلم، راه رفتن، زندگی بیمار و اطرافیان او را به شدت تحت تأثیر قرار می‌دهد [۵]. علاوه براین، هزینه‌های نگهداری از بیمار، فشار اقتصادی زیادی به دولت و خانواده‌های بیماران تحمیل می‌کند. به طوری‌که، در سال ۲۰۱۹، حدود ۲۹۰ میلیارد دلار هزینه تخمینی این بیماری برای ایالات متحده گزارش شده است. همچنین، انتظار می‌رود که تا سال ۲۰۵۰، با افزایش تعداد بیماران، هزینه‌ها برای سیستم بهداشت و درمان در ایالات متحده به بیش از یک تریلیون دلار افزایش یابد [۵]. با این وجود، تشخیص بیماری در مراحل اولیه آن باعث جلوگیری از آسیب به نواحی گسترده مغز شده و امکان مدیریت موفق روند بیماری را افزایش می‌دهد [۷،۸].

در سال‌های اخیر، کاربرد روش‌های یادگیری ماشین در پزشکی از جمله بیماری‌های قلبی، تشخیص سرطان، و تشخیص کرونا پیشرفت چشمگیری داشته است [۹-۱۱، ۶]. برای یک مسئله خاص، الگوریتم‌های یادگیری ماشین با استفاده از حجم بالایی از داده مناسب آموزش دیده می‌شود و سپس، قادر به برچسب‌گذاری نمونه‌های جدید با دقت بالایی خواهند بود. هر چه تعداد داده‌های آموزشی جهت آموزش مدل بیشتر باشد، کارایی مدل‌های یادگیری ماشین افزایش می‌یابد. در طول سال‌های آتی و افزایش تعداد داده‌های بالینی برای آموزش مدل‌های یادگیری ماشین، با استفاده از این مدل‌ها، می‌توان انتظار داشت که خطای انسانی کاهش یافته و دقت تصمیم‌گیری پزشکان افزایش یابد [۹].

هدف اصلی استفاده از یادگیری ماشین در پزشکی، کمک به پزشکان با افزایش دقت در تشخیص مشکلات طبی است. با این وجود، عدم شفافیت روند کارکرد الگوریتم‌های یادگیری ماشین در تشخیص مسائل پزشکی، مانعی بر اعتماد و استفاده از این فناوری توسط پزشکان به شمار می‌رود [۱۲]. از این‌رو، شفاف‌سازی روند تشخیص توسط فراهم آوردن امکان تفسیرپذیری مدل‌های یادگیری ماشین، می‌تواند سبب ایجاد اعتماد پزشکان گردد. تفسیرپذیری به عنوان تکنیکی جهت توصیف فرآیند مورد استفاده در الگوریتم‌های یادگیری ماشین برای تشخیص مسائل استفاده می‌شود [۱۲].

با توجه به اهمیت تشخیص آلزایمر در مراحل اولیه آن، استفاده از یادگیری ماشین سبب کاهش زمان تشخیص این بیماری خواهد شد [۷]. با تشخیص زود هنگام بیماری، احتمال مدیریت روند بیماری افزایش می‌یابد. علاوه بر این، نیاز به نیروی انسانی کاهش یافته و از میزان هزینه‌های بیماری کاسته خواهد شد [۷].

از جمله داده‌های موجود جهت تشخیص آلزایمر، تصویربرداری رزونانس مغناطیسی (Magnetic Resonance Imaging)MRI است. Rabeh و همکاران [۱۲] تصاویر خاکستری ۵۰۰×۵۰۰، را در سه سطح نواحی هیپوکامپ، جسم پینه‌ای (Corpus Callosum) و قشر (Cortex)، ناحیه‌بندی (Segmentation) نمودند. در ابتدا، تصاویر ناحیه‌بندی شده هیپوکامپ توسط مدل ماشین بردار پشتیبان (Support Vector Machine) برای دسته‌بندی مورد استفاده قرار دادند. درصورتی که نتیجه استفاده از مدل بر روی نواحی هیپوکامپ به عنوان غیرنرمال دسته‌بندی می‌شد، نواحی جسم پینه‌ای توسط مدل ماشین بردار پشتیبان بررسی می‌گردید. در ادامه، در صورتی

[Downloaded from jhbmi.com on 2024-10-12]

[DOI: 10.34172/jhbmi.2023.21]

که نتایج حاصل از این بررسی غیر نرمال بود، نواحی قشر توسط مدل بررسی می‌شد و در نهایت، در صورتی که نواحی قشر بیمار در دسته غیرطبیعی کلاسه‌بندی شد، از درخت تصمیم جهت تشخیص زودهنگام آلزایمر استفاده می‌گردید [۱۲].

در مطالعه Pavalarajan و همکاران [۱۳] با استفاده از روش‌های طبقه‌بندی موجود در یادگیری ماشین از جمله رگرسیون لجستیک، ماشین بردار پشتیبان، درخت تصمیم و جنگل تصادفی برای تشخیص آلزایمر در مراحل اولیه آن استفاده شده بود. برای یادگیری مدل‌ها از مجموعه داده تصویری شامل تصاویر تصویربرداری پرتو مغناطیسی (ام آر آی) (Magnetic resonance imaging) MRI مغز استفاده شده و پارامترهای بهینه هر مدل توسط روش K-تکرار اعتبارسنجی متقابل به دست آمده بود. بر طبق نتایج، روش جنگل تصادفی در مقایسه با سایر الگوریتم‌های یادگیری ماشین ذکر شده، نتیجه بهتری داشته است.

مطالعه Rajab و همکاران [۱۴] با استفاده از روش‌های یادگیری ماشین، ۲۲ ویژگی آسیب‌شناسی عصبی (Neuropathology) برتر مربوط به داده‌های آلزایمر را انتخاب کردند. در ادامه، هفت مدل یادگیری ماشین شامل درخت تصمیم، رگرسیون لجستیک، K-نزدیک‌ترین همسایه، بیز ساده گاوسی (Gaussian Naive Bayes)، ماشین بردار پشتیبان با تابع کرنل غیرخطی و خطی، تجزیه و تحلیل تفکیک خطی (Linear Discriminant Analysis) برای طبقه‌بندی استفاده شد.

در مطالعه Neelaveni و همکاران [۱۵] از پارامترهای روانشناسی مهم در مجموعه داده مورد استفاده مانند سن، میزان تحصیلات شخص، آزمون کوتاه وضعیت ذهنی MMSE (Mini Mental State Examination) که نتایج آن به صورت دوره‌ای بررسی می‌شوند، برای آموزش مدل‌های یادگیری ماشین استفاده کردند. در ادامه، از دو مدل یادگیری ماشین شامل کلاسه‌بند ماشین بردار پشتیبان و درخت تصمیم برای پیش‌بینی بیماری آلزایمر استفاده شد. نتایج بیانگر برتری دقت مدل ماشین بردار پشتیبان در مقایسه با دقت مدل درخت تصمیم بود.

روش‌های ارائه شده در تحقیقات فعلی به جهت عدم شفافیت فرآیند تشخیص بیماری توسط مدل‌های یادگیری ماشین به سختی مورد اعتماد پزشکان قرار می‌گیرد. یکی از اهداف اصلی در این مطالعه علاوه بر مقایسه نتایج چندین الگوریتم یادگیری ماشین، افزایش اعتماد پزشکان نسبت به نتایج به دست آمده از آن‌ها در تشخیص زودهنگام بیماری آلزایمر در دنیای واقعی بود.

از این رو، در این مطالعه از یک تکنیک تفسیرپذیری به منظور ایجاد شفافیت و تعیین میزان تأثیرگذاری ویژگی‌های مجموعه داده در فرآیند تصمیم‌گیری مدل‌های یادگیری ماشین استفاده گردید. به جهت افزایش کارایی مدل‌های یادگیری ماشین، معمولاً آموزش مدل‌ها با استفاده از داده‌های آموزشی با حجم بالا ضروری است. به منظور مدیریت و پردازش داده‌های حجیم از بستر اسپارک (Spark Framework) استفاده شد. در این مطالعه، برای اولین بار در این زمینه، از دو رویکرد متفاوت در بستر اسپارک شامل گروهی از کامپیوترها و تک کامپیوتر برای پیاده‌سازی مدل‌ها استفاده شد. نتایج این مطالعه می‌تواند به پزشکان در تشخیص زود هنگام بیماری آلزایمر کمک شایانی کند.

روش

در این مطالعه، از یک مجموعه داده بالینی بیماری آلزایمر استفاده شد. تعداد بیماران هر کلاس مجموعه داده با استفاده از یک روش متعادل‌سازی داده‌ها، متعادل شد. روابط بین ویژگی‌های مجموعه داده بررسی و از بستر اسپارک به منظور مدیریت داده‌ها استفاده شد و با استفاده از کتابخانه‌های یادگیری ماشین موجود در آن، الگوریتم یادگیری ماشین بیز ساده (Naïve Bayes)، درخت تصمیم (Decision Tree) و شبکه عصبی مصنوعی (Artificial Neural Network) به منظور پیش‌بینی زودهنگام بیماری آلزایمر در نظر گرفته شد. در آخر، به منظور ایجاد شفافیت و افزایش اعتماد پزشکان به مدل‌های جعبه سیاه یادگیری ماشین، از یک روش تفسیرپذیری الگوریتم‌های یادگیری ماشین استفاده و نمودار درخت تصمیم نیز رسم شد.

مجموعه داده

در این مطالعه از مجموعه داده بالینی بیماران آلزایمر مربوط به مجموعه مطالعات تصویربرداری با دسترسی آزاد (Open Access Series of Imaging Studies) OASIS به منظور پیش‌بینی زودهنگام بیماری آلزایمر استفاده شد [۱۶]. مجموعه داده حاضر شامل مجموعه‌ای از ۳۷۳ نمونه در دو وضعیت بیماران مبتلا به آلزایمر یا غیر مبتلا به آلزایمر است. وضعیت غیر دمانس (کلاس ۰) و احتمال بیماری آلزایمر (کلاس ۱) در نمونه‌ها با استفاده از رتبه‌بندی دمانس بالینی (Clinical Dementia Rating) مشخص شد. در این مجموعه داده همه بیماران راست دست بودند و از محاسبه خودکار حجم کل مـغز

[Downloaded from jhbmi.com on 2024-10-12]

[DOI: 10.34172/jhbmi.2023.21]

برای ایجاد تمائز مرتبط با پیری طبیعی و بیماری آلزایمر استفاده شد. در جدول ۱ اطلاعات آماری از جمله کم‌ترین مقدار، میانه و بیشترین مقدار هر ویژگی مجموعه داده تشریح شده است.

جدول ۱: اطلاعات آماری ویژگی‌های مجموعه داده بیماران آلزایمر

نام ویژگی	توضیحات	کمترین مقدار	میانه	بیشترین مقدار
Age in years (Age)	سن بیمار در زمان ثبت	۶۰	۷۷	۹۸
Gender	جنسیت	–	–	–
Years of education (EDUC)	سال جمع‌آوری داده	۷	۱۴/۲	۲۲
Socio Economic Status (SES)	وضعیت اجتماعی – اقتصادی که توسط شاخص موقعیت اجتماعی هالینگزهد ارزیابی شده	۲	۲/۳	۶
Mini Mental State Examination (MMSE)	امتیاز آزمون کوتاه وضعیت ذهنی	۱۶	۲۶/۲	۳۰
Atlas Scaling Factor (ASF)	ضریب مقیاس پذیری اطلس	۰/۸۷	۱/۳	۱/۴۳
Estimated Total Intracranial Volume (eTIV)	حجم کل داخل جمجمه تخمین‌زده شده	۱/۱۲۰	۱/۴۵۰	۱/۹۹۰
Normalize Whole Brain Volume (nWBV)	حجم کل مغز نرمال‌شده	۰/۵۵	۰/۸۱	۰/۷
Clinical Dementia Rating (CDR)	رتبه‌بندی دمانس بالینی	–	–	–

پیش‌پردازش مجموعه داده

با استفاده از زبان برنامه‌نویسی پایتون نسخه ۳.۷.۵ نمونه‌های مجموعه داده که مقادیر گم‌شده (Missing Values) داشتند کنار گذاشته شدند. همچنین به دلیل متعادل نبودن تعداد نمونه‌های هر کلاس، از تکنیک بیش‌نمونه‌گیری اقلیت مصنوعی (Synthetic Minority Over Sampling Technique) استفاده شد [۱۷]. در این روش، نمونه داده‌های جدید مصنوعی در همسایگی نمونه‌های موجود در کلاس‌ها تولید می‌شوند. با استفاده از این روش مجموعه داده‌ای جدید با ۳۸۰ (هر کلاس از ۱۹۰ نمونه) نمونه تشکیل گردید. همچنین، مطابق معادله (۱)، آزمون ضریب چولگی پیرسون دوم (Pearson Second Coefficient of Skewness) برای مجموعه داده برای نشان دادن تقارن توزیع داده استفاده شد [۱۸]. در معادله (۱)، $\bar{x}$ میانگین، m میانه، و s انحراف معیار است. بازه ضریب چولگی بین ۱- و ۱ می‌باشد، که مقادیر (۰/۵-, ۱-) به معنای چولگی به سمت چپ و مقادیر (۰/۵,۱) به معنای چولگی به سمت راست می‌باشد.

$$Skewness = \frac{۳\times(\bar{x}-m)}{s} \quad (۱)$$

سپس از روش حداقل–حداکثر (Min-Max) از کتابخانه Sklearn پایتون به منظور نرمال‌سازی داده‌ها استفاده شد. با انجام این کار، داده‌ها با توجه به معادله (۲) به محدوده صفر و یک نگاشت داده شده‌اند. در معادله (۲)، X_{min} حداقل مقدار و X_{max} حداکثر مقدار یک ویژگی در مجموعه داده است.

$$X_{normal} = \frac{X-X_{min}}{X_{max}-X_{min}} \quad (۲)$$

ویژگی‌های مهم مجموعه داده

برای بررسی ارتباط بین تمامی ویژگی‌های مجموعه داده از ضریب همبستگی پیرسون (Pearson Correlation Coefficient) استفاده شد. ضریب همبستگی درجه و نوع ارتباط‌های بین ویژگی‌ها را تعیین می‌کند [۱۹]. این ضرایب مطابق معادله (۳)، مقادیری بین ۱- و ۱ را کسب می‌کنند که ۱ نشان‌دهنده حداکثر همبستگی است. در معادله (۳)، $\bar{x}$ میانگین مقادیر x_i ها و $\bar{y}$ میانگین مقادیر y_i ها است.

[Downloaded from jhbmi.com on 2024-10-12]

[DOI: 10.34172/jhbmi.2023.21]

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 (y_i - \bar{y})}} \quad (۳)$$

از کتابخانه توضیحات افزودنی شیپلی (شپ) SHAP (Shapley Additive exPlanations) برای تفسیر مدل استفاده شد. این روش مبتنی بر یک رویکرد نظریه بازی برای توضیح خروجی هر مدل یادگیری ماشین است. شپ اهمیت و اثربخشی هر ویژگی را در تصمیم گیری در مدل تعیین می‌کند [۲۰].

اسپارک

اسپارک این قابلیت را دارد که مجموعه داده‌ها را با دو رویکرد، بر روی چندین کامپیوتر یا یک کامپیوتر، توزیع کرده و پردازش کند. این ویژگی‌های اسپارک موجب شده است که الگوریتم‌های یادگیری ماشین به خوبی بر روی آن پیکربندی شده و با سرعت بالا بر روی داده‌ها پردازش‌ها را انجام دهند [۲۱].

در رویکرد اول، استفاده از چندین کامپیوتر (سه کامپیوتر) پیکربندی شده در بستر اسپارک، مقادیر حافظه با دسترسی تصادفی (رم) RAM(Random Access Memory) برابر ۲۴ گیگابایت (Gigabytes) و توان پردازشی برابر ۲۴ هسته (Core) بود. در رویکرد دوم، استفاده از یک کامپیوتر پیکربندی شده، مقادیر رم برابر ۸ گیگابایت و توان پردازشی برابر ۸ هسته بود. در هر دو رویکرد، به صورت یکسان و همزمان مدل‌های یادگیری ماشین با شرایط کاملاً یکسان در کنار هم اجرا و مورد ارزیابی قرار گرفته‌اند.

اسپارک به عنوان یک بستر قدرتمند برای پردازش داده‌ها شناخته می‌شود. اسپارک طوری طراحی شده است که بهینه بوده و عملیات پردازش را در حافظه انجام دهد. با انجام عملیات بر روی حافظه، سرعت پردازش داده‌ها نسبت به ذخیره‌سازی و فراخوانی داده‌ها از روی دیسک افزایش پیدا می‌کند.

مدل‌های یادگیری ماشین مورد مقایسه در مطالعه حاضر

بیز ساده

بیز ساده یک الگوریتم یادگیری ماشین سریع با پیاده‌سازی آسان است [۲۲]. اساس این الگوریتم مبتنی بر قضیه بیز مطابق معادله (۴)، است.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (۴)$$

با توجه به معادله (۴)، احتمال وقوع رخداد A در صورت وقوع B در نظر گرفته می‌شود. در اینجا، B اصل و A فرضیه است.

درخت تصمیم

درخت تصمیم یکی از پرکاربردترین الگوریتم‌های طبقه‌بندی یادگیری ماشین است [۲۲]. این الگوریتم مانند درخت از برگ، ریشه و شاخه تشکیل شده است. الگوریتم درخت تصمیم از معیار جینی (Gini) و یا آنتروپی (Entropy) استفاده کرده تا تعیین نماید کدام استراتژی بیشترین احتمال را برای حل موفق مسئله دارد.

شبکه عصبی مصنوعی

این الگوریتم یکی از روش‌های مورد استفاده برای طبقه‌بندی می‌باشد که برای حل کردن مسائل از شبکه‌های عصبی بیولوژیکی الهام گرفته است [۲۳]. در این الگوریتم مجموعه‌ای از گره‌ها به نام نورون‌های مصنوعی وجود دارند. هر یک از این نورون‌ها به یکدیگر متصل بوده و سیگنال دریافتی را با وزن مخصوص منتقل می‌کنند. شبکه عصبی مصنوعی در لایه‌های مختلف خود نورون‌ها را سازمان‌دهی کرده و تبدیل‌های متفاوتی را می‌تواند به داده ورودی اعمال نماید.

از روش ۵-تکرار اعتبارسنجی متقابل (5-Fold Cross Validation) برای پیشگیری از بروز بیش برازش (Over-Fitting) و ارزیابی کارایی مدل‌های پیشنهادی در این مطالعه استفاده شد. در هر تکرار، ۸۰ درصد از مجموعه داده به منظور آموزش و ۲۰ درصد باقی‌مانده برای آزمایش در نظر گرفته می‌شود.

در این مطالعه بهترین مقادیر فراپارامترها (Hyperparameter) ممکن برای هر مدل به منظور پیاده‌سازی در نظر گرفته شد. برای انجام این کار، از یک روش جستجو (Grid Search) به منظور یافتن بهترین مقادیر بهینه استفاده شد. بهترین مقادیر فراپارامترهای ممکن این مطالعه در جدول ۲ نشان داده شده‌اند.

[Downloaded from jhbmi.com on 2024-10-12]

[DOI: 10.34172/jhbmi.2023.21]

جدول ۲: بهترین فراپارمترهای یافته شده الگوریتم‌های پیشنهادی

مدل	فراپارامترها
بیز ساده	var_smoothing: 1e-3
درخت تصمیم	spilter: best, criterion: gini, max_depth: none, min_samples_split: 2
شبکه عصبی مصنوعی	learning_rate: adaptive, learning_rate_init: 0.0001, solver: adam, hidden_layer_sizes: (8, 16, 32, 64, 16, 4, 1)

ارزیابی مدل‌های یادگیری ماشین

چندین معیار بسیار مهم برای ارزیابی مدل‌های یادگیری ماشین در کاربردهای پزشکی مورد استفاده قرار گرفته‌اند. پس از پایان یافتن مرحله آموزش مدل‌ها، سه معیار اساسی دقت (Accuracy)، امتیاز–اف۱ (F1-Score) و مساحت زیر نمودار منحنی عملیاتی گیرنده (Area Under Curve) به منظور ارزیابی مدل‌های پیاده‌سازی شده مورد استفاده قرار گرفته‌اند [۱۰].

دقت نسبت بین پیش‌بینی‌های صحیح به تعداد کل پیش‌بینی‌ها است. این معیار برای ارزیابی مدل‌های طبقه‌بندی یادگیری ماشین استفاده می‌شود. معادله (۵)، دقت را نشان می‌دهد.

$$\text{دقت} = \frac{TP + TN}{TP + FP + FN + TN} \quad (۵)$$

امتیاز–اف۱ یکی دیگر از معیارهای ارزیابی یادگیری ماشین است. امتیاز–اف۱ بالا نشان دهنده مثبت کاذب و منفی کاذب کم است. معادله (۶)، امتیاز–اف۱ را نشان داده است.

$$\text{امتیاز} - \text{اف۱} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (۶)$$

در معادلات (۵، ۶)، TP نشان دهنده مثبت حقیقی، FP نشان دهنده مثبت کاذب، FN به معنی منفی کاذب و TN به معنی منفی حقیقی است.

نتایج

با استفاده از ضریب همبستگی پیرسون ویژگی جنسیت با ویژگی ضریب مقیاس‌پذیری اطلس با ضریبی بیش از ۰/۵ با یکدیگر رابطه داشته و در نتیجه ارتباط قوی بین آن‌ها برقرار بود. سایر ویژگی‌ها، مقادیر ضریب همبستگی کمتر از ۰/۵ با یکدیگر داشتند؛ بنابراین در این مطالعه هیچ ارتباط کاملی بین ویژگی‌ها وجود نداشت.

آزمون ضریب چولگی در این مطالعه انجام یافته که مقدار اولیه چولگی پیش از متعادل‌سازی مجموعه داده برابر ۰/۴۷– به دست آمد که هرچند قابل قبول بود، اما برای بهبود آن از تکنیک بیش‌نمونه‌گیری اقلیت مصنوعی بر روی مجموعه‌داده اعمال شده و مقدار ضریب چولگی مجدداً برای مجموعه‌داده محاسبه گردید. ضریب چولگی برای مجموعه‌داده متعادل‌سازی شده برابر ۰/۴۲– به دست آمد که به نسبت به مقدار ۰ نزدیک‌تر بوده و نشان‌دهنده توزیع متقارن و متعادل‌تر بود.

در این مطالعه با استفاده از بستر اسپارک، دو رویکرد، سه کامپیوتر و یک کامپیوتر، به منظور انجام پردازش‌ها در نظر گرفته شد. میانگین ارزیابی تمام تکرارها برای تمام مدل‌ها در دو رویکرد، بر روی چندین کامپیوتر و یک کامپیوتر محاسبه شد. مدل بیز ساده دقت ۹۳/۰۵٪ و امتیاز–اف۱ ۹۲/۹۹٪، مدل‌های درخت تصمیم و شبکه عصبی مصنوعی، دقت ۹۸/۶۱٪ و امتیاز–اف ۱ ۹۸/۶۰٪ را در هر دو رویکرد توانسته‌اند کسب کنند. علاوه بر این، در شکل ۱ مقادیر مساحت زیر نمودار عملیاتی گیرنده برای مدل‌های پیشنهاد شده در این مطالعه در تکرار پنجم نشان داده شده‌اند. مطابق این شکل، دو مدل درخت تصمیم و شبکه عصبی مصنوعی در تکرار پنجم از روش اعتبارسنجی متقابل بیشترین مقادیر را به دست آورده‌اند.

با توجه به استفاده از بستر اسپارک، به منظور مدیریت و پردازش داده‌ها، زمان محاسباتی لازم برای پردازش داده‌ها در رویکرد سه کامپیوتر سریع‌تر از رویکرد تک کامپیوتر بود که در جدول ۳ به تفصیل آورده شد. مدل‌های درخت تصمیم و شبکه

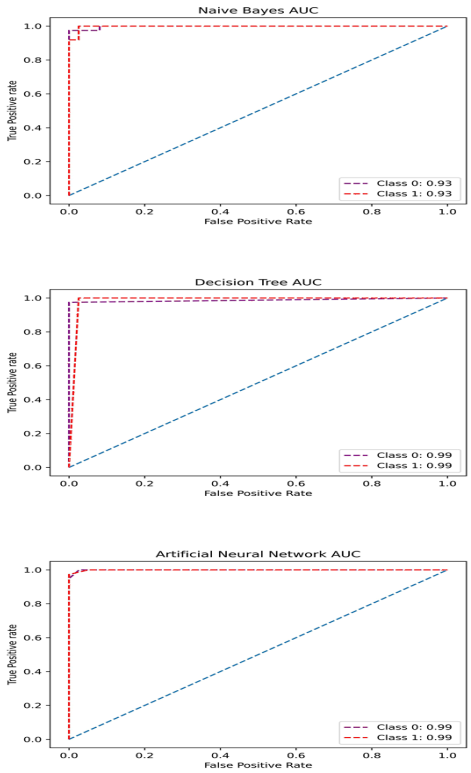
[Downloaded from jhbmi.com on 2024-10-12]
[DOI: 10.34172/jhbmi.2023.21]

عصبی مصنوعی بهترین کارایی را در پیش‌بینی زود هنگام بیماری آلزایمر در هر دو رویکرد این مطالعه از خود نشان داده‌اند. نمودار مدل درخت تصمیم این مطالعه جهت نمایش فرآیند تصمیم‌گیری و اثرگذاری ویژگی‌های این مدل در شکل ۲ نشان داده شد. مطابق این نمودار، چگونگی تصمیم‌گیری مدل درخت تصمیم بر اساس اهمیت ویژگی‌ها و تأثیر آن‌ها نشان داده شده و نحوه انتخاب شدن هر کلاس نیز مشخص شده است.

همچنین، مهم‌ترین ویژگی‌های تأثیرگذار و اثر هر یک از آن‌ها در نحوه پیش‌بینی مدل شبکه عصبی مصنوعی با روش شپ در شکل ۳ نشان داده شده‌اند. بر اساس این شکل، سه ویژگی شامل آزمون کوتاه وضعیت ذهنی، حجم نرمال شده کل مغز و جنسیت بیمار تأثیرگذارترین و مهم‌ترین ویژگی‌های مجموعه داده در تصمیم‌گیری مدل شبکه عصبی مصنوعی بوده است.

جدول ۳: زمان محاسباتی پردازش‌ها در رویکردهای متفاوت این مطالعه

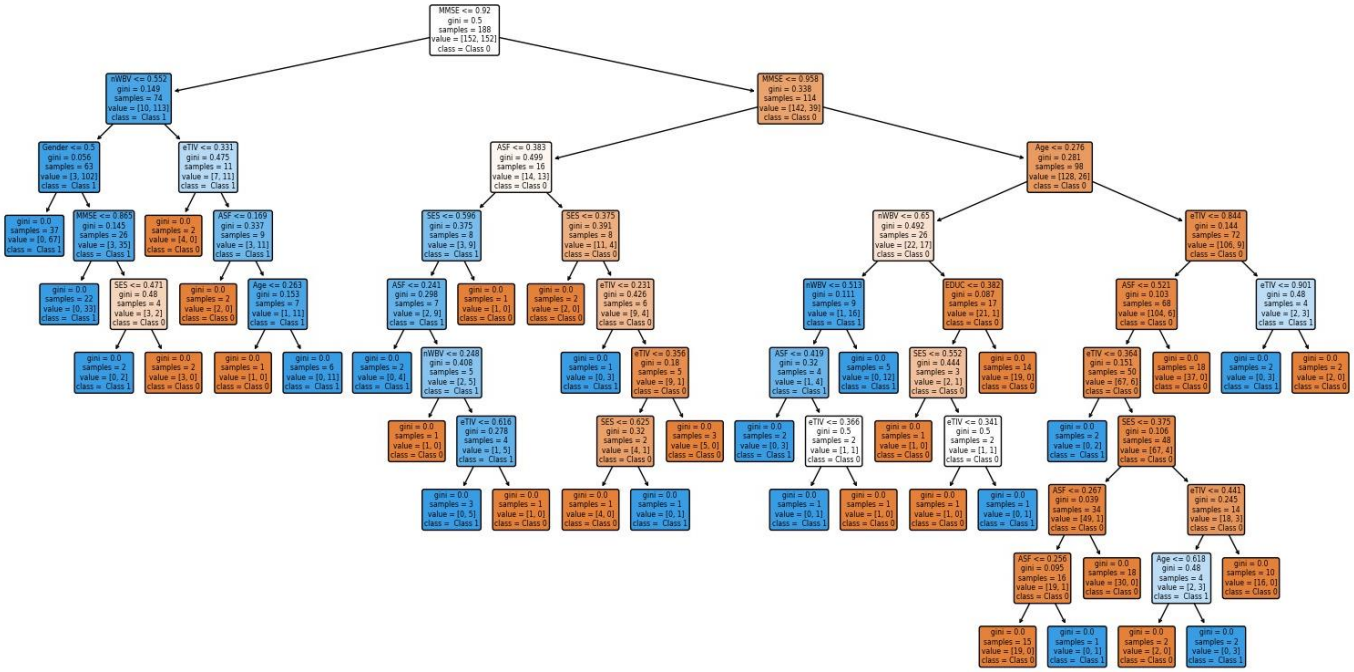
واحد زمانی	رویکرد اول (سه کامپیوتر)	رویکرد دوم (یک کامپیوتر)
ثانیه	۵۷	۸۹



شکل ۱: مساحت زیر نمودار عملیاتی گیرنده در تکرار پنجم از روش اعتبارسنجی متقابل برای مدل‌های پیشنهادی این مطالعه (الگوریتم‌های درخت تصمیم و شبکه عصبی مصنوعی مقدار ۰/۹۹ و الگوریتم بیزساده مقدار ۰/۹۳ را به دست آورده‌اند).

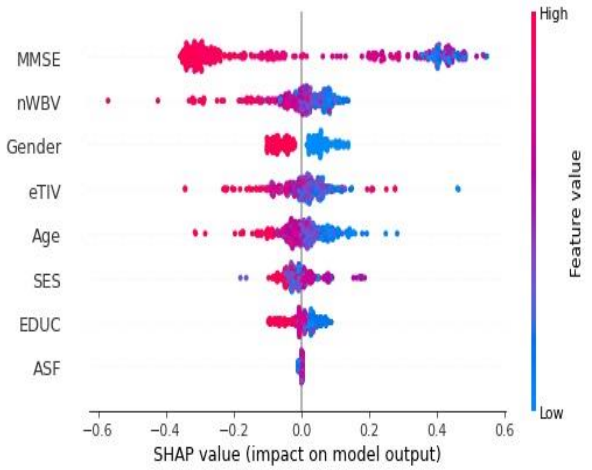
[Downloaded from jhbmi.com on 2024-10-12]

[DOI: 10.34172/jhbmi.2023.21]



شکل ۲: نمودار فرآیند تصمیم‌گیری و انتخاب کلاس هدف در مدل درخت تصمیم‌گیری

(مطابق این شکل سه ویژگی آزمون کوتاه وضعیت ذهنی، حجم نرمال شده کل مغز و جنسیت بیمار در ریشه‌های اولیه مدل بوده و پر اهمیت‌ترین ویژگی‌ها هستند).



شکل ۳: مهم‌ترین ویژگی‌ها در پیش‌بینی زودهنگام بیماری آلزایمر توسط مدل شبکه عصبی مصنوعی با استفاده از روش شپ

(سه ویژگی آزمون کوتاه وضعیت ذهنی، حجم نرمال شده کل مغز و جنسیت بیمار تأثیرگذارترین ویژگی‌ها در فرآیند تصمیم‌گیری مدل شبکه عصبی مصنوعی است).

[Downloaded from jhbmi.com on 2024-10-12]

[DOI: 10.34172/jhbmi.2023.21]

بحث و نتیجه‌گیری

با افزایش موارد درگیری به بیماری آلزایمر در جهان، حجم عظیمی از داده‌های بالینی بیماران در بیمارستان‌ها در این زمینه در حال ذخیره‌سازی است. بدین منظور، در مطالعه حاضر از بستر اسپارک با امکان مدیریت کلان داده‌ها و قابلیت پیکربندی با مدل‌های یادگیری ماشین با اساس ساختاری متفاوت شامل بیزساده، درخت تصمیم و شبکه عصبی مصنوعی توسعه داده شد و در دو رویکرد چندین کامپیوتر و تک کامپیوتر کارایی مدل‌های پیشنهادی این مطالعه جهت تشخیص دقیق بیماری آلزایمر ارزیابی شدند. در هر دو رویکرد این مطالعه در بستر اسپارک، مدل‌های درخت تصمیم و شبکه عصبی مصنوعی، در مجموعه داده از پیش پردازش نشده توانسته‌اند دقت برابر ۸۶/۱۱٪ و امتیاز–اف۱ برابر ۸۶/۰۲٪ را به دست آوردند. همچنین، هر دو مدل درخت تصمیم و شبکه عصبی مصنوعی در مجموعه داده پیش پردازش شده در هر دو رویکرد دقت برابر ۹۸/۶۱٪ و امتیاز–اف۱ برابر ۹۸/۶۰٪ را به دست آوردند. این نتایج حاکی از پیش پردازش دقیق و متعادل‌سازی صحیح تعداد نمونه‌های مجموعه داده است. استفاده از تکنیک بیش‌نمونه‌گیری اقلیت مصنوعی امری رایج در مطالعات بالینی بوده و در مقالات مشابه پیشین نیز از این روش برای متعادل‌سازی تعداد نمونه‌های هر کلاس از مجموعه داده مرتبط با بیماری آلزایمر استفاده شده است [۲۴]. علاوه بر این، در مقایسه رویکرد اول با رویکرد دوم، عملکرد الگوریتم‌های این مطالعه در رویکرد اول ۳۵/۹۵٪ سریع‌تر از رویکرد دوم بوده است. از این رو، با در نظر گرفتن شرایط کاملاً یکسان در هر دو رویکرد، رویکرد اول به خوبی توانسته با پیکربندی دقیق و پردازش توزیع‌شده در مدت زمان کمتر به بهترین کارایی ممکن دست یابد.

Rabeh و همکاران [۱۲]، از مدل‌های ماشین بردار پشتیبان و درخت تصمیم بر روی مجموعه داده تصاویر پزشکی به منظور تشخیص زود هنگام بیماری آلزایمر استفاده کردند. آن‌ها حالت‌های مختلفی برای این کار در نظر گرفته و کارایی سیستم پیشنهادی آن‌ها در بهترین حالت دقت ۹۰٪/۶۶ را داشته است. Pavalarajan و همکاران [۱۳] بر روی مجموعه داده تصاویر ام آر آی مغز، از مدل‌های رگرسیون لجستیک، ماشین بردار پشتیبان، درخت تصمیم و جنگل تصادفی استفاده کردند تا بیماری آلزایمر را در مراحل اولیه تشخیص دهند. نتایج آن‌ها نشان داد که مدل جنگل تصادفی توسعه داده شده آن‌ها توانسته دقت و مساحت زیر نمودار منحنی عملیاتی گیرنده ۸۳٪ را به دست آورد. Basheer و همکاران [۲۵] یک شبکه عصبی عمیق به منظور تشخیص زودهنگام بیماری آلزایمر با استفاده از مجموعه داده تصاویر ام آر آی مغز ارائه کرده‌اند. مدل پیشنهادی آن‌ها توانسته دقت ۹۲/۳۹٪ را به دست آورده و به عنوان یک ابزار کمکی جهت تشخیص سریع این بیماری معرفی کردند. Saratxaga و همکاران [۲۶] یک روش جدید یادگیری عمیق به منظور تشخیص بیماری آلزایمر با استفاده از مجموعه داده تصاویر ام آر آی مغز بیماران توسعه داده‌اند. آن‌ها این روش خود را به عنوان یک دستیار کمکی در تصویربرداری از مغز بیماران معرفی کرده و مدل پیشنهادی دقت ۰/۹۳ را به دست آورده است.

اغلب مطالعه‌های پیشین به منظور تشخیص زودهنگام بیماری آلزایمر از تصاویر پزشکی بیماران استفاده کردند. استفاده از تصاویر پزشکی بیماران برای تشخیص زود هنگام بیماری آلزایمر با کمک الگوریتم‌های یادگیری ماشین هزینه و توان پردازشی بسیار زیادی به همراه دارد. مطالعات معدودی فقط از مجموعه داده‌های بالینی بیماران برای تشخیص بیماری آلزایمر با استفاده مدل‌های یادگیری ماشین استفاده کردند [۱۶]. Kavitha و همکاران [۱۶]، مدل‌های مختلف یادگیری ماشین را با مجموعه داده بالینی بیماران آلزایمر پیاده‌سازی و آزمایش کردند. براساس نتایج مطالعه، الگوریتم جنگل تصادفی پیشنهادی آن‌ها بیشترین دقت را برابر ۸۶/۹۲٪ داشته است. بهترین الگوریتم‌های این مطالعه در مقایسه با مطالعه گذشته اخیر کارایی بسیار بالاتری را بر روی مجموعه داده یکسان از خود نشان داده‌اند [۱۶].

این مطالعه نتایج چشم‌گیری داشته و بهترین مدل‌های پیشنهاد شده در مطالعه حاضر، خروجی‌های صحیح را به درستی پیش‌بینی کرده و خطای کمی دارند. استفاده از تکنیک‌های دقیق در هنگام پیش‌پردازش داده‌ها، متعادل‌سازی مجموعه داده را می‌توان از علل این موفقیت عنوان کرد. به‌طورکلی، نوآوری‌های این مطالعه عبارت‌اند از:

- استفاده از یک تکنیک برای متعادل‌سازی تعداد نمونه‌های آموزشی در کلاس‌های مختلف بر روی این مجموعه داده بیماری آلزایمر برای اولین بار در این مطالعه صورت گرفته است.
- مطالعه حاضر برای اولین بار به معرفی و استفاده از یک بستر جدید به منظور مدیریت داده‌های حجیم در مجموعه داده بالینی

[Downloaded from jhbmi.com on 2024-10-12]

[DOI: 10.34172/jhbmi.2023.21]

بیماران آلزایمر پرداخته که کاهش زمان محاسباتی برای پردازش‌ها به همراه داشته است.

- برای اولین بار در مجموعه داده بیماری آلزایمر حاضر، از یک روش جستجو برای یافتن بهترین فراپارامترهای مدل‌های یادگیری ماشین مختلف استفاده شد.
- برای اولین بار در این مطالعه، از یک تکنیک تفسیرپذیری به منظور شفاف‌سازی عملکرد مدل یادگیری ماشین جهت افزایش اعتماد پزشکان استفاده شد.

پیش‌بینی زودهنگام بیماری آلزایمر برای پزشکان بسیار حائز اهمیت است چرا که آن‌ها می‌توانند سریع‌تر پروسه درمانی را برای بیماران آغاز کنند [۲۷]. برخی مطالعات پیشین (از جمله [۱۳-۱۶])، سعی نموده‌اند که به صورت زودهنگام بیماری آلزایمر را به کمک داده‌های بالینی و یا تصویربرداری از مغز بیماران تشخیص دهند. با این حال، در این زمینه شکاف‌های تحقیقاتی وجود دارد که نیازمند تحقیقات بیشتر است. در این مطالعه، برخی از این شکاف‌های تحقیقاتی شناسایی شده و سعی شد که به آن‌ها پاسخ داده شود.

همه‌گیری بیماری کروناویروس ۲۰۱۹ (کووید-۱۹)، COVID-19(Coronavirus Disease 2019) باعث تغییر اولویت‌های پزشکی و درمانی شد. از این رو، شناسایی و تشخیص سایر بیماری‌ها مانند بیماری آلزایمر در طول این مدت کمتر مورد توجه محققان و پزشکان قرار گرفت. علاوه بر این، بیماران مبتلا به بیماری آلزایمر طبق تحقیقات نشان داده شده که بیشتر از حد در معرض خطر بیماری کووید-۱۹ هستند [۲۸]. بنابراین، فراهم‌سازی یک ابزار دقیق برای پیش‌بینی زودهنگام بیماری آلزایمر به صورت خودکار و با تعداد نمونه‌های بالا می‌تواند به پزشکان در شناسایی بیماران و آغاز هرچه سریع‌تر پروسه درمان آنان کمک نماید. بهترین مدل‌های یادگیری ماشین پیشنهادی این مطالعه عملکرد چشم‌گیری را در پیش‌بینی زود هنگام بیماری آلزایمر داشته‌اند. علاوه بر این، تفسیرپذیری این مدل‌ها می‌تواند منجر به شناسایی برخی جزئیات جدید در شناسایی ویژگی‌های اثرگذار در بیماری آلزایمر شوند.

در این مطالعه، سه ویژگی آزمون کوتاه وضعیت ذهنی، حجم نرمال شده کل مغز و جنسیت بیمار به ترتیب مهم‌ترین ویژگی‌ها در تصمیم‌گیری مدل شبکه عصبی مصنوعی با امتیاز شپ بالا بوده و به عنوان سه الویت اول در فرآیند تصمیم‌گیری مدل درخت تصمیم نیز مشخص شده‌اند. از این رو، این سه ویژگی در بهترین مدل‌های پیشنهادی به طور مشترک توانسته‌اند اثرگذارترین ویژگی باشند که نشان دهنده اهمیت بالای آن‌ها در پروسه شناسایی زود هنگام بیماری آلزایمر است.

به کارگیری بستر اسپارک با رویکردهای مختلف همراه با توسعه و پیاده‌سازی مدل‌های یادگیری ماشین، می‌تواند حجم عظیمی از مجموعه‌داده‌های بیماران را پردازش و پیش‌بینی لازم را انجام دهند. از این رو، وجود چنین سیستم پشتیبانی تصمیم به عنوان یک ابزار کمکی و کاربردی می‌تواند به پزشکان در شرایط اضطراری بسیار سریع و دقیق در تشخیص زودهنگام بیماری آلزایمر کمک کند. نتایج به دست آمده در این مطالعه، نیاز به وجود مدل‌های یادگیری ماشین و اهمیت تشخیص زودهنگام بیماری آلزایمر توسط آن‌ها را به خوبی نشان داد.

این مطالعه، نقاط قوت بسیاری دارد. برای اولین بار، در مطالعه حاضر دو رویکرد مختلف در بستر اسپارک به منظور پیش‌بینی زودهنگام بیماری آلزایمر در نظر گرفته شد. بهترین مدل‌های این مطالعه نتایج امیدوار کننده داشته و زمان محاسباتی کمی دارند. برای درک بهتر پزشکان از تکنیک‌های مصورسازی و تفسیرپذیری استفاده شده تا با ایجاد شفافیت اعتماد آنان به بهترین مدل‌های پیشنهادی افزایش پیدا کند. با این حال، این مطالعه کاستی‌ها و ضعف‌هایی نیز دارد. اول، به دلیل نبود مجموعه‌داده بالینی مورد تأیید با تعداد نمونه‌های بالا، در حال حاضر امکان ارزیابی بهترین مدل‌های پیشنهادی این مطالعه بر روی مجموعه‌داده دیگر وجود نداشت. دوم، به دلیل کمبود منابع سخت افزاری، حداکثر از سه کامپیوتر برای ایجاد بستر پردازش توزیع شده، استفاده شد. سوم، به دلیل نبودن کتابخانه‌های یادگیری عمیق در کتابخانه‌های خود اسپارک، امکان توسعه و پیاده‌سازی مدل‌های یادگیری عمیق وجود نداشت.

دو مدل درخت تصمیم و شبکه عصبی مصنوعی توسعه داده شده در این مطالعه بهترین کارایی با دقت ۹۸/۶۱٪ و امتیاز اف-۱، ۹۸/۶۰٪ را از خود در شناسایی این بیماری نشان دهند. همچنین، تفسیرپذیری و شفاف‌سازی پروسه تصمیم‌گیری مدل‌های پیشنهادی در این مطالعه برای اولین بار بر روی مجموعه داده بالینی بیماری آلزایمر مورد توجه قرار گرفته است. نتایج به دست آمده این مطالعه بسیار امیدوارکننده بوده و بهترین مدل‌های پیشنهاد شده در این مطالعه می‌توانند به عنوان یک ابزار کمکی با قابلیت تفسیرپذیری به پزشکان به طور دقیق و سریع در شناسایی زود هنگام بیماری آلزایمر در بیماران کمک کنند.

[Downloaded from jhbmi.com on 2024-10-12]

[DOI: 10.34172/jhbmi.2023.21]

به عنوان کارهای آتی پیشنهاد می‌شود که مجموعه داده‌های بالینی به همراه تصاویر ام آر آی مغز بیماران به عنوان ورودی به مدل‌های یادگیری ماشین چند وجهی داده شده تا این الگوریتم‌ها بتوانند به طور دقیق طیف وسیعی از بیماری‌های مغزی و درجه آن بیماری‌ها را به صورت زودهنگام پیش‌بینی کنند. همچنین، با عنایت به نتایج کاربردی این مطالعه، نویسندگان این مطالعه در حال آماده‌سازی مجموعه‌داده بالینی با تعداد نمونه های بسیار بالا به منظور پیش‌بینی زود هنگام بیماری آلزایمر هستند تا مدل‌های بیشتری را در آینده توسعه داده و ارزیابی شود.

تشکر و قدردانی

محققان این مطالعه هیچ‌گونه کمک هزینه تحقیقاتی از هیچ نهاد و سازمانی برای انجام این مطالعه دریافت نکرده‌اند و از آزمایشگاه یادگیری ماشین دانشگاه ارومیه بابت در اختیار گذاشتن منابع لازم برای پیاده‌سازی این مطالعه تشکر و قدردانی می‌نماییم.

تعارض منافع

بدین‌وسیله نویسندگان تصریح می‌نمایند که در مورد پژوهش حاضر هیچ‌گونه تضاد منافعی وجود ندارد.

References

1. Liu S, Masurkar AV, Rusinek H, Chen J, Zhang B, Zhu W, et al. Generalizable deep learning model for early Alzheimer's disease detection from structural MRIs. Scientific Reports 2022;12(1):17106.

2. RS NN, Priyadarshini J. Machine Learning Algorithms for the diagnosis of Alzheimer's and Parkinson's Disease. International Journal 2020;9(4): 5898-905. https://doi.org/10.30534/ijatcse/2020/252942020

3. World Health Organization(WHO). Dementia .[cited 2023 Feb 05]. Available from: https://www.who.int/news-room/fact-sheets/detail/dementia

4. Gillani N, Arslan T. Intelligent sensing technologies for the diagnosis, monitoring and therapy of Alzheimer's disease: A systematic review. Sensors. 2021;21(12):4249. https://doi.org/10.3390/s21124249

5. Fathi S, Ahmadi M, Dehnad A. Early diagnosis of Alzheimer's disease based on deep learning: A systematic review. Computers in Biology and Medicine 2022;146:105634. https://doi.org/10.1016/j.compbiomed.2022.105634

6. Li Z, Jiang X, Wang Y, Kim Y. Applied machine learning in Alzheimer's disease research: omics, imaging, and clinical data. Emerg Top Life Sci 2021 5(6): 765–77. https://doi.org/10.1042/ETLS20210249

7. Mirabnahrazam G, Ma D, Beaulac C, Lee S, Popuri K, Lee H, et al. Predicting time-to-conversion for dementia of Alzheimer's type using multi-modal deep survival analysis. Neurobiology of Aging 2023;121:139-56. https://doi.org/10.1016/j.neurobiolaging.2022.10.005

8. Al'Aref SJ, Anchouche K, Singh G, Slomka PJ, Kolli KK, Kumar A, et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. Eur Heart J 2019;40(24):1975-86. doi: 10.1093/eurheartj/ehy404.

9. Sorayaie Azar A, Babaei Rikan S, Naemi A, Bagherzadeh Mohasefi J, Pirnejad H, Bagherzadeh Mohasefi M, et al. Application of machine learning techniques for predicting survival in ovarian cancer. BMC Medical Informatics and Decision Making 2022;22(1):345.

10. Rikan SB, Azar AS, Ghafari A, Mohasefi JB, Pirnejad H. COVID-19 diagnosis from routine blood tests using artificial intelligence techniques. Biomedical Signal Processing and Control 2022;72:103263. https://doi.org/10.1016/j.bspc.2021.103263

11. ElShawi R, Sherif Y, Al-Mallah M, Sakr S. Interpretability in healthcare: A comparative study of local machine learning interpretability techniques. Computational Intelligence 2021;37(4):1633-50. https://doi.org/10.1111/coin.12410

12. Rabeh AB, Benzarti F, Amiri H. Diagnosis of Alzheimer diseases in early step using SVM (support vector machine). 13th International conference on computer graphics, imaging and visualization (CGiV); 2016 Mar 29; Beni Mellal, Morocco: IEEE; 2016. p. 364-7. doi: 10.1109/CGiV.2016.76

13. Pavalarajan S, Kumar BA, Hammed SS, Haripriya K, Preethi C, Mohanraj T. Detection of Alzheimer's disease at Early Stage using Machine Learning. International Conference on Advanced Computing Technologies and Applications (ICACTA); 2022 Mar 4-5; Coimbatore, India: IEEE; 2022. p. 1-5. doi: 10.1109/ICACTA54488.2022.9752827

14. Rajab MD, Jammeh E, Taketa T, Brayne C, Matthews FE, Su L, Ince PG, Wharton SB, Wang D. Assessment of Alzheimer-related pathologies of dementia using machine learning feature selection. Alzheimer's Research & Therapy 2023;15(1):1-7.

15. Neelaveni J, Devasana MG. Alzheimer disease prediction using machine learning algorithms. 6th international Conference on Advanced Computing and Communication Systems (ICACCS) 2020 Mar 6-7; Coimbatore, India: IEEE; 2020. p. 101-104. doi: 10.1109/ICACCS48705.2020.9074248

16. Kavitha C, Mani V, Srividhya SR, Khalaf OI, Romero CA. Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models. Frontiers

[Downloaded from jhbmi.com on 2024-10-12]

[DOI: 10.34172/jhbmi.2023.21]

in Public Health 2022;10. doi: 10.3389/fpubh.2022.853294

17. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research 2002;16:321-57.

18. Doane DP, Seward LE. Measuring skewness: a forgotten statistic?. Journal of statistics education 2011;19(2). doi:10.1080/10691898.2011.11889611

19. Sedgwick P. Pearson's correlation coefficient. BMJ 2012;345.

20. Knapič S, Malhi A, Saluja R, Främling K. Explainable artificial intelligence for human decision support system in the medical domain. Machine Learning and Knowledge Extraction 2021;3(3):740-70. https://doi.org/10.48550/arXiv.2105.02357

21. Meng X, Bradley J, Yavuz B, Sparks E, Venkataraman S, Liu D, et al. Mllib: Machine learning in apache spark. The Journal of Machine Learning Research 2016;17(1):1235-41.

22. Jadhav SD, Channe HP. Comparative study of K-NN, naive Bayes and decision tree classification techniques. International Journal of Science and Research 2016;5(1):1842-5.

23. Haglin JM, Jimenez G, Eltorai AE. Artificial neural networks in medicine. Health and Technology. 2019;9(1):1-6. https://doi.org/10.1007/s12553-018-0244-4

24. Liu Y, Tang K, Cai W, Chen A, Zhou G, Li L, et al. MPC-STANet: Alzheimer's disease recognition method based on multiple phantom convolution and spatial transformation attention mechanism. Frontiers in Aging Neuroscience 2022;14:918462. https://doi.org/10.3389/fnagi.2022.918462

25. Basheer S, Bhatia S, Sakri SB. Computational modeling of dementia prediction using deep neural network: analysis on OASIS dataset. IEEE Access 2021;9:42449-62. doi: 10.1109/ACCESS.2021.3066213

26. Saratxaga CL, Moya I, Picón A, Acosta M, Moreno-Fernandez-de-Leceta A, Garrote E, et al. MRI deep learning-based solution for Alzheimer's disease prediction. Journal of Personalized Medicine 2021;11(9):902. https://doi.org/10.3390/jpm11090902

27. Rasmussen J, Langerman H. Alzheimer's disease–why we need early diagnosis. Degenerative Neurological and Neuromuscular Disease 2019:123-30.

28. Golzari-Sorkheh M, Weaver DF, Reed MA. COVID-19 as a Risk Factor for Alzheimer's Disease. Journal of Alzheimer's Disease 2023; 91(1):1-23. doi: 10.3233/JAD-220800

[Downloaded from jhbmi.com on 2024-10-12]

[DOI: 10.34172/jhbmi.2023.21]