2024 10th International Conference on Control, Decision and Information Technologies (CoDIT) | 979-8-3503-7397-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/CoDIT62066.2024.10708434

Analysis of Clinical, Genetic, and Demographic Data for Prediction of Alzheimer's Disease with Machine Learning

Anita Petreska, Sasho Nikolovski, Gabriela Novotni, Blagoj Ristevski IEEE Senior Member

***Abstract*— In the context of the ageing of the global population and the increasing prevalence of Alzheimer's disease (AD), early and accurate diagnosis is crucial for effective management and treatment. Using Exploratory Data Analysis (EDA) we dissect the complex relationships between various risk factors and disease progression, establishing a basis for our predictive modelling. Uncovering critical insights, and emphasizing the importance of adopting a multidimensional approach to analyze diverse datasets effectively, our study highlights the critical role of data quality and diversity in improving model performance. The fundamental aspect of our analysis focuses on the predictive power of combining different data types, which traditionally include clinical parameters, genetic markers, and demographic and lifestyle data.**

The research highlights the application of machine learning (ML) techniques for early detection and predictive analysis of Alzheimer's disease, demonstrating the enormous potential of artificial inelegance in transforming healthcare diagnostics. The study conducted a comparative analysis of various ML algorithms and evaluated their efficiency in disease detection.

This research contributes to the academic discourse on the diagnosis of Alzheimer's disease and provides practical insights for the application of artificial intelligence and machine learning in clinical practice.

Keywords: Alzheimer's Disease, Machine Learning, Predictive Analysis, Exploratory Data Analysis.

I. Introduction

Using traditional diagnostic methods due to the wide time frames often leads to AD patients receiving the diagnosis at a late stage of the disease.

A. The importance of early detection of AD

AD leads to significant disability and dependency in the elderly, and the challenges it poses apply not only to people with AD but also to their families. In the modern world, although there is a solid level of awareness about this disease, the lack of widespread awareness and understanding of AD can still be observed, which leads to stigmatisation and obstacles in the early diagnosis and appropriate care, and the consequences are reflected through the physical, psychological, social and economic dimensions, affecting caregivers, families and society at large [1].

B. AI vs. Traditional Methods in AD Diagnostics

Diagnosing AD using traditional diagnostics is an expensive, error-prone method due to the influence of factors such as human fatigue, cognitive biases [18], as well as systematic errors. On the contrary, diagnostic systems based on artificial intelligence [1][2][9][12] represent promising solutions, more reliable and less prone to errors, of course, provided that they are correctly set up and dimensioned according to the requirements and the data material that is the subject of analysis.

C. Objectives of the paper

The primary goal of the research is the improvement of a complex model for predicting AD [3][4][15][25][34], with the implementation of techniques for exploratory data analysis to obtain a deeper insight into the database. The study included three different groups of patients: the first group - patients with Alzheimer's disease (IG1), the second group - adult descendants of these patients (IG2), and the third group – the control group of patients (CG).

By using advanced machine learning techniques [11][14][28], this study aims to create promising high-accuracy predictive models [20] that can help assess the development of a familial disease where genetic predisposition plays a significant role [16].provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

II. Methodology

This research includes multi-layered data analysis [13] aimed at uncovering the complexities of Alzheimer's disease by:

- delves into descriptive statistics, providing a comprehensive insight into the characteristics of each group, including demographic, social, and clinical parameters.
- comparative analysis thoroughly examines these groups across various parameters: genetic data, lifestyle factors, and clinical parameters, offering insight into potential differentiating factors [21].
- correlation analysis investigated associations between specific traits and Alzheimer's disease within the IG1 group, looking for statistically significant differences between the second study group and the control group.
- predictive modelling uses advanced machine learning algorithms to predict Alzheimer's risk [17][30].

Anita Petreska is with the Faculty of Information and Communication Technologies, University "St. Kliment Ohridski" - Bitola, North Macedonia, e-mail: petreska.anita@uklo.edu.mk

Sasho Nikolovski is with the Faculty of Information and Communication Technologies, University "St. Kliment Ohridski" - Bitola, North Macedonia, e-mail: sasnik@gmail.com

Gabriela Novotni is with the University Clinic for Neurology, Faculty of Medicine, Ss Cyril and Methodius University of Skopje, Skopje, North Macedonia

Blagoj Ristevski is with the Faculty of Information and Communication Technologies, University "St. Kliment Ohridski" - Bitola, Macedonia
e-mail: blagoj.ristevski@uklo.edu.mk

From a methodological perspective, one of the primary questions addressed by the research concerns the existence of a genetic predisposition [1] to Alzheimer's disease. Another significant research question addressed in the study focuses on the influence of lifestyle factors on the risk of developing the disease.

The research outlined in this study endeavors to identify early indicators of Alzheimer's disease [6][8][15][26][32][36], leading to the expectation that the cohort of offspring comprising the IG2 cohort may exhibit early biomarkers or health patterns that resemble those observed in the cohort of Alzheimer's patients from the IG1 group [5][8].

A. Data Collection

The study included 289 participants (144 patients diagnosed with AD at the Dementia outpatient clinic at the University Clinic of Neurology - Skopje and University Clinic of Psychiatry - Skopje, in the period from 2016-2018, 55 middle-aged descendants of patients with AD and 90 cognitively unimpaired control subjects).

B. Insights in Alzheimer's Study Groups

Demographic analysis as one of the key aspects of the data analysis gives the following results: the average age in IG1 is approximately 71.6 years, which indicates a predominance of older individuals, compared to 49.0 years in IG2, the average age of CG is about 68.5 years. The distribution of Mini-Mental State Examination (MMSE) scores in IG1 shows a mean score of 15.1, indicating cognitive impairment, a clear contrast to the missing or unreported MMSE data for IG2 and CG. Analysis of health indicators such as glycemia, cholesterol, and triglycerides across groups revealed variability in mean levels, with IG1 generally displaying lower triglyceride levels compared to IG2 and CG, highlighting the impact of health and dietary habits unique to each group. The gender composition reveals a significant bias towards female CG participants (78.9%), in IG1 (60.4% female) and IG2 (58.2% female). The urban residence is especially dominant among all groups, especially in CG where predominantly 98.8% of participants live in urban areas. Education levels show diversity across the spectrum, with secondary education being most prevalent among participants in IG1 (36.3%), and CG (48.7%) and (48.9%) IG2. In terms of lifestyle factors, smoking rates are relatively similar across groups, with IG1 having the highest reported rate at 26.39%, IG2 at 23.64%, and CG at 13.33%. Alcohol consumption shows some variation, with IG2 at 40%, IG1 at 31.94%, and CG at 32.22%, possibly indicating different demographic or cultural trends in these groups. Hypertension is more prevalent in IG1, with a rate of 47.92%, compared to IG2 with 16.36% and CG with 30%. Diabetes mellitus shows the least variation, with CG having the least reported cases at 14.44%, while IG1 reported 17.36% and IG2 reported no cases of diabetes mellitus. Rates of dyslipidemia and obesity show differences between groups. IG1 had higher rates, with 34.72% reporting dyslipidemia and 12.5% reporting obesity. On the other hand, IG2 has 10.91% for dyslipidemia and 7.27% for obesity, while CG has 17.78% for dyslipidemia and 10% for obesity. In terms of family history, IG1 have a higher prevalence of Alzheimer's disease in their families, with a significant majority reported by one family member, on the other hand, CG reports fewer cases of familial Alzheimer's disease, 30%, indicating a lower familial history incidence. Relative to other familial diseases, vascular disease appears to be evenly distributed among the three groups, suggesting a consistent risk across the dataset. Reports of "other" diseases are highest in IG1, at 44.44%, followed by IG2 at 36.36%, and CG at 52.22%, which may be related to the age and health profiles of the groups. Fewer cases of cancers and neurodegenerative diseases are observed in CG compared to IG1.

C. Data Preprocessing

Data preprocessing is a fundamental step in data analysis [7], including tasks such as handling missing data, addressing outliers, ensuring data integrity, and transforming raw data into appropriate formats for further analysis. There are many missing values in the database (mostly for Vascular Cell Adhesion Molecule-1 (VCAM-1) - even 54.7% are missing). Dealing with missing values is done with Mean/Median Imputation and Mode Imputation. In the field of clinical data, outliers can carry significant clinical implications, potential data errors, or extreme cases [10]. In the dataset, ordered columns like glycemia, cholesterol, triglycerides, and VCAM-1, MMSE display a notable presence of outliers. Therefore, advanced outlier detection techniques are used, aiming to strike a balance between preserving valuable clinical insights and maintaining data quality. Data preprocessing also includes feature scaling [24][29], categorical variable coding, dimensionality reduction, and data integration and visualization [27].

D. Exploratory Data Analysis

The integration between EDA and ML for the detection of Alzheimer's disease (AD) involves a systematic approach to understanding and preparing the data, followed by the development and evaluation of ML models. EDA uses visualizations and statistical techniques to detect patterns and anomalies, acknowledge trends and evaluate hypotheses that cover various elements including examining the distribution of key variables through histograms, correlations between variables using a matrix heatmap of correlation and identifying missing data patterns and outliers.

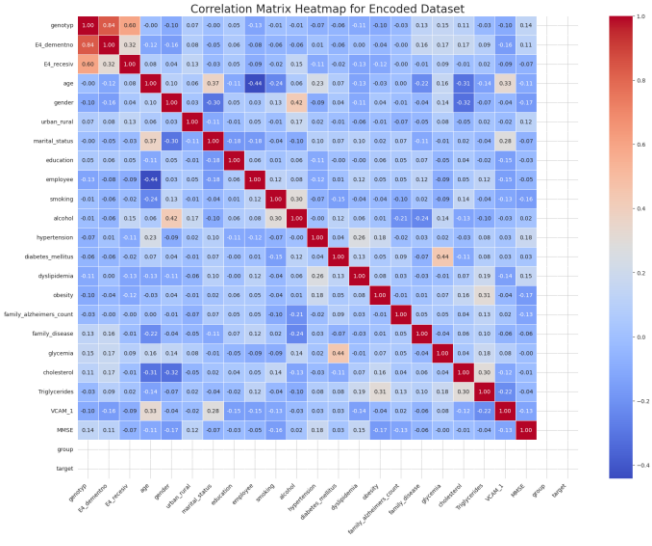

Figure 1 Correlation matrix heatmap (IG1)

As part of EDA, heatmaps serve as useful tools for visually illustrating relationships and correlations among variables within a database. Additionally, they provide a visual representation of data patterns, aiding the identification of potential insights for further analysis.

The analysis of the correlation matrix of IG1 (Figure 1), showed strong correlations between genotype and E4 dementia, and moderate correlations between E4_dementno and E4_recessiv. In contrast, demographic factors such as age and sex, lifestyle factors such as smoking and alcohol, health indicators such as glycemia and cholesterol and cognitive measures such as the MMSE have shown varying levels of association with genetic and health outcomes.

Genetic factors: the E4/E4 genotype is highly correlated with dementia (0.84), suggesting that this genotype is probably related to the presence of the E4 allele, which is associated with an increased risk for Alzheimer's disease [35].

Demographic factors: Gender has a negligible correlation with the genotype variables, but shows a stronger correlation with alcohol (0.42), which may reflect gender differences in alcohol consumption patterns.

Lifestyle factors: Smoking shows a negative correlation with employed (-0.24), possibly indicating lower employment rates among smokers. Urban/rural living conditions show very little correlation with genetic factors but have some association with employee status (0.13) and alcohol consumption (0.17). Alcohol consumption showed little or no correlation with genotypic factors, but moderate correlations with gender and smoking (0.42 and 0.30, respectively), suggesting potential lifestyle clusters.

Health indicators: Glycemia and cholesterol have very low correlations with genotype variables, indicating that these metabolic factors may be independent of genetic risk factors for the target outcome. Triglycerides have a moderate negative correlation with VCAM-1 (-0.22), indicating an inverse relationship between these two markers in this dataset.

Cognitive function (MMSE): MMSE, a measure of cognitive function, shows a positive correlation with genotype (0.14) and a stronger one with E4 dominant (0.29), strengthening the relationship between this genotype and cognitive health. There is also a strong positive correlation between the MMSE and the target (0.43), possibly because the target variable is related to cognitive outcomes.

Outcome variable (target): The target variable, which represents the risk of AD, showed the strongest correlation with age (0.23) and MMSE (0.43), suggesting that older age and lower cognitive function are related to the outcome. Purpose also has a moderate negative correlation with employed (−0.30), which may suggest that employed people are less likely to experience the outcome, perhaps because they are younger or healthier.

Heatmap analysis of the correlation matrix (Figure 2) shows that cognitive performance, age, gender, employment status, and lifestyle factors such as smoking and alcohol consumption are associated with the outcome variable and health indicators. Each reveals various links to genetic and health outcomes.

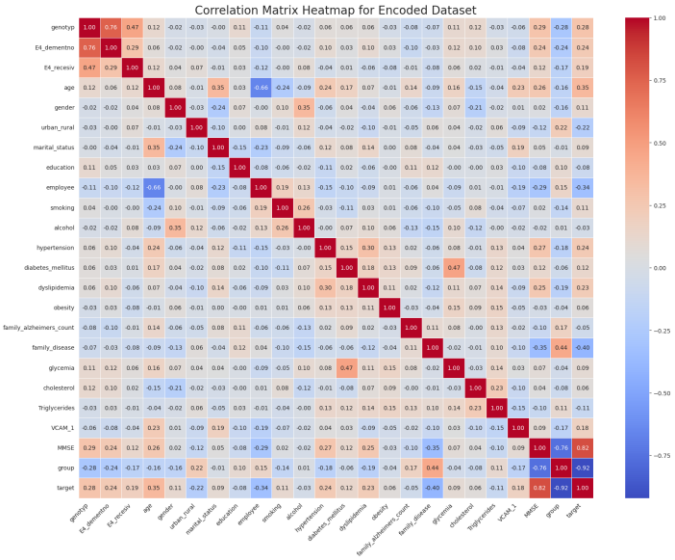

Figure 2 Correlation matrix heatmap

E4 has a strong positive correlation with MMSE (0.239353) and the target (0.237504), suggesting that the presence of the E4 allele is closely related to cognitive performance and the outcome variable.

Age: A strong positive correlation with the target (0.405698), indicating that as age increases, so does the likelihood of the outcome variable, which may include age-related conditions.

Moderate positive correlation with VCAM_1 (0.340840), a marker associated with ageing.

Gender: Moderate positive correlation with alcohol (0.351610), potentially reflecting differences in alcohol consumption between genders.

Urban/rural: Weak correlations with most variables, indicating minimal influence on genetic factors, health conditions, and the outcome variable. Employment status has a very strong negative correlation with age (-0.694233), highlighting the relationship between employment and age. A strong negative correlation with the target (-0.335365), suggests that employment status is significantly related to the outcome variable.

Smoking has a moderate positive correlation with alcohol (0.263633), indicating a lifestyle pattern where smoking and alcohol consumption are associated. Alcohol has a moderate positive correlation with gender (0.351610), again suggesting gender differences in alcohol consumption. Obesity has a moderate positive correlation with glycemia (0.151196), which is expected because obesity can affect blood sugar levels.

Glycemia has a weak to moderate positive correlation with obesity (0.151196) and triglycerides (0.138446), suggesting an association with metabolic health, and cholesterol has a moderate positive correlation with triglycerides (0.230540), which is typical because both are lipid profiles.

A weak negative correlation with VCAM_1 (-0.165996), indicates a complex relationship with inflammatory markers. Triglycerides have a moderate positive correlation with cholesterol (0.230540), consistent with lipid metabolism, and a moderate negative correlation with VCAM_1 (-0.217215), which may indicate an inverse relationship between lipid levels and certain vascular markers. VCAM_1 has a moderate positive correlation with age (0.340840) and target (0.240554), indicating that its role increases with age and is related to the outcome variable.

There is a moderate negative correlation with cholesterol (-0.165996) and triglycerides (-0.217215) and there is a moderate to strong positive correlation with age (0.405698), indicating that age likely influences the outcome.

Individuals with genotypes E2E4, E3E4, and E4E4 are classified as high-risk, while all other genotypes fall into the low/intermediate risk categories. The variable indicating a family history of AD was positively correlated with AD, suggesting that a family history of Alzheimer's disease increases the risk. Cramér V analysis shows that the E4E4 genotype has a value of 0.178, indicating an association with AD. Other genotypes, such as E3E3 and E3E4, also show some degree of association. A family history of Alzheimer's disease shows moderate associations with AD. ANOVA analysis found no significant differences in age, glycemia, cholesterol, or triglycerides among different genotypes. Genotypes, particularly E4E4, as well as family history of AD should be focal points when assessing the risk of Alzheimer's disease.

These correlations provide insights that could be valuable for further research, interventions, and understanding factors associated with the outcome variable and highlight the complex interplay between genetics, lifestyle, metabolic health, and cognitive function.

III. CHOOSING THE RIGHT MACHINE LEARNING MODEL FOR DEMENTIA PREDICTION

Selecting an appropriate machine learning (ML) model is a key step in the dementia prediction workflow [22]. The choice of model directly affects the prediction performance and generalizability of the resulting prediction tool, especially when dealing with the diverse, heterogeneous datasets used to predict Alzheimer's disease (AD). The training data is fed into the selected ML algorithm during the model-building phase to develop a predictive model. This training data is the foundation upon which the model learns the complex patterns and correlations between the various clinical, genetic, demographic, and lifestyle variables that contribute to Alzheimer's risk.

After the initial training, the validation dataset serves as a key tool for optimizing the hyperparameters, which control the learning process of the model. The determination of optimal hyperparameter values to improve the model's ability to handle data complexity and variability.

Effective tuning ensures that the model avoids both underfitting and overfitting, achieving a balance that allows accurate predictions for new data.

A. Analysis of Machine Learning Algorithms for Alzheimer's Disease Prediction

In this study, different ML algorithms were compared to assess their effectiveness in detecting Alzheimer's disease. Supervised and unsupervised learning models were also explored: Logistic regression, K-Nearest Neighbors, Support Vector Machine, Naïve Bayes, Perceptron, Ridge classifier, and Extra Trees classifier. The choice of the ML algorithm must consider factors such as the nature and quality of the data, computational efficiency, interpretability, and the specific goals of the predictive modeling task.

Using logistic regression to detect AD involves applying this statistical modeling technique to predict the probability that an individual has AD based on one or more variables, well suited for binary classification tasks [23] because it estimates the probability that an individual belongs to a particular class (AD or non-AD). K-Nearest Neighbors classifies data points based on the majority class of their k-nearest neighbors, useful for identifying patterns in dementia data. SVMs aim to find a hyperplane that maximizes the margin between dementia and non-dementia cases in high-dimensional spaces.

Naïve Bayes, a probabilistic algorithm, is useful for predicting dementia from categorical or textual data such as medical reports. Using a perceptron to detect (AD) makes it possible to analyze relevant targets or data points by classifying individuals into AD and non-AD groups for resource-limited environments. Neural networks effectively predict dementia with complex, high-dimensional data such as images or time series data [19][31]. The Ridge classifier is a variant of logistic regression that introduces L2 regularization in the linear classification model and can effectively handle binary classification tasks.

The incremental tree classifier is an ensemble learning method that builds a forest of decision trees to reduce overfitting and improve prediction accuracy. It is used when dealing with small datasets because it can handle noise and variances efficiently, making it a robust choice for AD detection.

The selection of an appropriate dementia prediction algorithm relies on factors like the nature of the database, data size, computational resources, and research goals. Hence, it is useful to experiment with multiple algorithms, tune the hyperparameters, and evaluate their performance accordingly.

B. Metrics-Based Evaluation of AD Predictive Models

To analyze and compare the given algorithms, we will consider the performance of each algorithm based on the analyzed metrics: accuracy, precision, recall, F1-score, and ROC-AUC [38], which are essential for the selection of an appropriate model (Table I).

TABLE I DIFFERENT ALGORITHMS IN MODELS ANALYSIS

Model	accuracy	precision	recall	F1-score	ROC-AUC
Logistic Regression	0.988506	1.00	0.977778	0.988764	0.999471
K-Nearest Neighbors	0.931034	0.92	0.927678	0.920000	0.978182
Support Vector Machine	0.965517	0.96	0.960000	0.960000	0.997576
Naive Bayes	0.982759	1.00	0.977778	0.979592	1.000000
Perceptron	0.954023	0.97	0.933333	0.954545	NaN
Ridge Classifier	0.977011	1.00	0.955556	0.977273	NaN
Extra Trees Classifier	0.988506	1.00	0.977778	0.988764	1.000000

Logistic regression has shown high precision (0.988506) and perfect precision (1.00), showing no false positives in predictions, recall is also high (0.977778), showing that it correctly identifies most positive cases. An excellent F1-Score (0.988764) indicates a balance between precision and recall, and a very high ROC-AUC (0.999471) indicates an excellent ability to discriminate classes. Since its recall is slightly lower than its precision, it may suggest a very small tendency to miss some positive examples.

K-Nearest Neighbors (KNN) gives good recall (0.927678) suggesting that it is reasonably good at identifying positive cases. Lower accuracy (0.931034), precision (0.92) and F1-Score (0.920000) compared to other models shows that it is less effective in accurately predicting positive examples and balancing precision and recall.

Support Vector Machine (SVM) shows high precision (0.96) and recall (0.960000) leading to a balanced F1-score (0.960000), indicating good performance in identifying positive cases and maintaining a balance between precision and recall. The ROC-AUC score (0.997576) is excellent, indicating strong class differentiation ability.

Naïve Bayes has very high precision (0.982759) and perfect precision (1.00), with high recall (0.977778), excellent F1-score (0.979592), ROC-AUC score is perfect (1.000000), which indicates a superior ability to distinguish between classes. Although minimal, the difference between precision and recall suggests that it may slightly favour precision over recall.

Perceptron provides good accuracy (0.97) suggesting a high rate of correct positive predictions. Lower recall (0.933333) and F1-Score (0.954545) compared to other models, indicating challenges in identifying all positive examples and balancing precision and recall.

The Ridge classifier has high accuracy (0.977011) and perfect precision (1.00), showing no false positives, F1-Score (0.977273) suggests a good balance between precision and recall. A slightly lower recall (0.955556) compared to the best performers and the absence of the ROC-AUC score means that we cannot fully evaluate its ability to discriminate between classes.

The complementary tree classifier yields precision (0.988506) and has perfect precision (1.00), with high recall (0.977778), excellent F1-score (0.988764) and its ROC-AUC score (1.000000), indicating the ability for supreme distinction.

IV. OVERCOMING CHALLENGES IN ML MEDICAL RESEARCH

This research faces various challenges, including dealing with data heterogeneity, providing larger and more representative datasets to improve model accuracy, developing techniques for handling unbalanced data, ensuring the validity and applicability of research findings, ethical considerations, and the need for long-term control and development.

ML models require large datasets to achieve optimal performance. In AD research, obtaining extensive data sets is limited, which affects the model's predictive capabilities. Ensuring that algorithms work reliably with different demographic and geographic parameters is a constant challenge for successfully implementing strategies to address the difference between healthy and affected individuals and building unbiased models. Developing interpretability techniques and ensuring model generalizability remains an ongoing challenge. Training of ML models, especially complex neural networks, requires significant computational resources. They have to undergo rigorous validation and reproducibility testing to establish their position in the field of clinical decision-making and earn the trust of healthcare professionals and patients.

The collection and processing of sensitive patient data raises significant privacy concerns. Researchers and healthcare institutions must prioritize data privacy by implementing strong data protection measures that include de-identification, strict access controls and encryption, informed consent practices, and transparent data handling to ensure that individuals are aware of how their data will be utilized.

V. CONCLUSION

Given the new era of Alzheimer's disease-modifying drugs, it is becoming increasingly important to reduce the risk of its onset and progression, through early diagnosis and timely treatment, with proper and timely assessment of symptoms and the risk factors.

The primary objective of this study is the analysis of clinical, genetic, demographic and lifestyle data in the prediction of Alzheimer's disease through precision EDA and application of various ML algorithms. The used methodology helps in the development of predictive models for assessing the risk of developing Alzheimer's disease, especially in assessing the possibility of developing familial Alzheimer's disease, where genetic predisposition plays a significant role. Using a multidimensional approach enables us to gain valuable insights into this complex disease. The application of advanced statistical and ML techniques allowed us to achieve higher accuracy and improved performance in classifying Alzheimer's dementia based on the available data in the attached database.

Our study highlights the essential role of data quality and diversity in improving model performance, advocating the integration of extensive datasets spanning diverse demographic and genetic parameters [37]. It is worth noting that the study offers significant insight into the application of ML for dementia prediction, providing a basis for future

research focused on enhancing and broadening the suggested models.

Transforming machine learning models from research labs to clinical practice, and bridging the gap between innovative research findings and real-world healthcare applications requires collaboration and integration efforts.

REFERENCES

[1] García-Gutierrez, Fernando, et al. "GA-MADRID: Design and validation of a machine learning tool for the diagnosis of Alzheimer's disease and frontotemporal dementia using genetic algorithms." Medical & Biological Engineering & Computing 60.9 (2022): 2737-2756.

[2] Merkin, Alexander, Rita Krishnamurthi, and Oleg N. Medvedev. "Machine learning, artificial intelligence and the prediction of dementia." Current Opinion in Psychiatry 35.2 (2022): 123-129.

[3] Battineni, Gopi, et al. "Artificial intelligence models in the diagnosis of adult-onset dementia disorders: a review." Bioengineering 9.8 (2022): 370.

[4] Mirzaei, Golrokh, and Hojjat Adeli. "Machine learning techniques for diagnosis of Alzheimer disease, mild cognitive disorder, and other types of dementia." Biomedical Signal Processing and Control 72 (2022): 103293.

[5] Winchester, Laura M., et al. "Artificial intelligence for biomarker discovery in Alzheimer's disease and dementia." Alzheimer's & Dementia 19.12 (2023): 5860-5871.

[6] Verma, Rohit K., et al. "An insight into the role of Artificial Intelligence in the early diagnosis of Alzheimer's disease." CNS & Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders) 21.10 (2022): 901-912.

[7] Aditya Shastry, K., and H. A. Sanjay. "Artificial Intelligence Techniques for the effective diagnosis of Alzheimer's Disease: A Review." Multimedia Tools and Applications (2023): 1-36.

[8] Vrahatis, Aristidis G., et al. "Revolutionizing the Early Detection of Alzheimer's Disease through Non-Invasive Biomarkers: The Role of Artificial Intelligence and Deep Learning." Sensors 23.9 (2023): 4184.

[9] Newby, Danielle, et al. "Artificial intelligence for dementia prevention." Alzheimer's & Dementia (2023).

[10] Petreska, Anita, Blagoj Ristevski, Daniela Slavkovska, Saso Nikolovski, Pero Spirov, Nikola Rendevski, and Snezana Savoska. "Machine Learning Algorithms for Heart Disease Prognosis using IoMT Devices." (2023): 141-150.

[11] Javeed, Ashir, et al. "Machine Learning for Dementia Prediction: A Systematic Review and Future Research Directions." Journal of medical systems 47.1 (2023): 17.

[12] Borchert, Robin J., et al. "Artificial intelligence for diagnostic and prognostic neuroimaging in dementia: A systematic review." Alzheimer's & Dementia 19.12 (2023): 5885-5904.

[13] Soares Dias Portela, Ariana, et al. "Role of Artificial Intelligence in Multinomial Decisions and Preventative Nutrition in Alzheimer's Disease." Molecular Nutrition & Food Research (2024): 2300605.

[14] Kasula, Balaram Yadav. "A Machine Learning Approach for Differential Diagnosis and Prognostic Prediction in Alzheimer's Disease." International Journal of Sustainable Development in Computing Science 5.4 (2023): 1-8.

[15] Kavitha, C., et al. "Early-stage Alzheimer's disease prediction using machine learning models." Frontiers in Public Health 10 (2022): 853294.

[16] Balaji, Chetan, and D. S. Suresh. "Multiclass recognition of Alzheimer's and Parkinson's disease using various machine learning techniques: A study." International Journal of Modeling, Simulation, and Scientific Computing 13.01 (2022): 2250008.

[17] Veena, K. C., Mrs R. Kavi Priya, and D. Sumathi. "Predictive Diagnostic Analysis for Early Detection of Alzheimer's disease Using Machine Learning." Journal of Algebraic Statistics 13.1 (2022): 586-592.

[18] Galety, Mohammad Gouse, and Shweta Gupta. "Artificial Intelligence in the Detection of Alzheimer's Disease." Bio-Inspired Algorithms and Devices for Treatment of Cognitive Diseases Using Future Technologies. IGI Global, 2022. 136-155.

[19] Samhan, Lamis F., Amjad H. Alfarra, and Samy S. Abu-Naser. "Classification of Alzheimer's disease using convolutional neural networks." (2022).

[20] Lyall, Donald M., et al. "Artificial intelligence for dementia—Applied models and digital health." Alzheimer's & Dementia 19.12 (2023): 5872-5884.

[21] Slavkovska, Daniela, Blagoj Ristevski, and Anita Petreska. "Comparative Analysis of ML Algorithms for Breast Cancer Detection." (2023): 151-161.

[22] Shetty, Mangala, Manisha Bhat, and Manisha Devadiga. "Detection of Alzheimer's Disease Using Machine Learning." 2022 International Conference on Artificial Intelligence and Data Engineering (AIDE). IEEE, 2022.

[23] Viswan, Vimbi, et al. "Explainable Artificial Intelligence in Alzheimer's Disease Classification: A Systematic Review." Cognitive Computation (2023): 1-44.

[24] Kapadnis, Manav Nitin, Abhijit Bhattacharyya, and Abdulhamit Subasi. "Artificial intelligence based Alzheimer's disease detection using deep feature extraction." Applications of artificial intelligence in medical imaging. Academic Press, 2023. 333-355.

[25] Niyas, KP Muhammed, and P. Thiyagarajan. "A systematic review on early prediction of Mild cognitive impairment to alzheimers using machine learning algorithms." International Journal of Intelligent Networks (2023).

[26] Uddin, Khandaker Mohammad Mohi, et al. "A Novel Approach Utilizing Machine Learning for the Early Diagnosis of Alzheimer's Disease." Biomedical Materials & Devices (2023): 1-17.

[27] ElZawawi, Nour Saad, et al. "An efficient Hybrid approach for diagnosis High dimensional data for Alzheimer's diseases Using Machine Learning algorithms." International Journal of Intelligent Computing and Information Sciences 22.2 (2022): 97-111.

[28] Bogdanovic, Bojan, Tome Eftimov, and Monika Simjanoska. "In-depth insights into Alzheimer's disease by using explainable machine learning approach." Scientific Reports 12.1 (2022): 6508.

[29] Jahan, Sobhana, et al. "Explainable AI-based Alzheimer's prediction and management using multimodal data." (2022).

[30] Hanke, Sten, et al. "AI-based predictive modelling of the onset and progression of dementia." Smart Cities 5.2 (2022): 700-714.

[31] Mohapatra, Subhasish, Suneeta Satpathy, and Bijay Kumar Paikaray. "A machine learning approach to assist prediction of Alzheimer's disease with convolutional neural network." International Journal of Bioinformatics Research and Applications 19.2 (2023): 141-150.

[32] Hashmi, Arshad, and Omar Barukab. "Dementia classification using deep reinforcement learning for early diagnosis." Applied Sciences 13.3 (2023): 1464.

[33] Arjaria, Siddhartha Kumar, et al. "Performances of Machine Learning Models for Diagnosis of Alzheimer's Disease." Annals of Data Science (2022): 1-29.

[34] Martin, Sophie A., et al. "Interpretable machine learning for dementia: A systematic review." Alzheimer's & Dementia (2023).

[35] Novotni, Gabriela, et al. "Evaluation of APOE genotype and vascular risk factors as prognostic and risk factors for Alzheimer's disease and their influence on age of symptoms onset." Open Access Macedonian Journal of Medical Sciences 7.4 (2019): 516.

[36] Novotni, G., et al. "Early onset dementia with parkinsonism in a patient with pathogenic mutations in CSF1R and ABCD1 gene." Parkinsonism & Related Disorders 113 (2023).

[37] Savoska, Snezana, Blagoj Ristevski, and Vladimir Trajkovik. "Personal health record Data-Driven integration of heterogeneous data." Data-Driven Approach for Bio-medical and Healthcare. Singapore: Springer Nature Singapore, 2022. 1-21.

[38] Beyene, Solomon Shiferaw, Tianyi Ling, Blagoj Ristevski, and Ming Chen. "A novel riboswitch classification based on imbalanced sequences achieved by machine learning." PLoS Computational Biology 16, no. 7 (2020): e1007760.