ISBN: 978-628-96613-1-6. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2025.1.1.547

Deep learning and Machine learning predictive models for neurodegenerative disease detection: A Systematic Review of the Literature

Cataño-Portocarrero Frank[1]; Soria-Quijaite Juan[2]; Yapo-Cáceres Carlos[3]
[1,2,3]Universidad Tecnológica del Perú, Perú, *U20215523@utp.edu.pe, C20723@utp.edu.pe, C18236@utp.edu.pe*

Abstract*– In recent years, artificial intelligence has revolutionized various fields of medicine, particularly in the early detection of neurodegenerative diseases (ND). This study analyzes Machine Learning (ML) and Deep Learning (DL) models applied to the detection of neurodegenerative diseases. To achieve this, a systematic literature review (SLR) was conducted following the PICO method to guide the search for relevant studies in the SCOPUS and Web of Science databases, and the PRISMA statement was used for the final screening. Performance metrics such as accuracy, sensitivity, and area under the curve (AUC) were evaluated. Additionally, the Bibliometrix tool in R was used for an in-depth bibliometric analysis of the selected studies. Among the 30 articles finally included in the review, the most frequent approaches included Support Vector Machines (SVM) in ML, with 6 main studies, and Convolutional Neural Networks (CNN) in DL, with 11 prominent works. Notably, one SVM model achieved 100% accuracy, while the CNN-InceptionV4 architecture stood out in DL with 99% accuracy. DL models such as Graph Convolutional Networks (GCN), and hybrid approaches like CNN-GCN, proved more robust in managing complex data, whereas ML models offer advantages in terms of lower computational requirements. In conclusion, DL- and ML-based models represent a promising tool for the early detection of ND. However, their adoption in clinical practice requires further optimization to overcome technical limitations and ensure real-world applicability.

Keywords*— Deep Learning, Machine Learning, Sensitivity and Specificity, Neurodegenerative Diseases, Disease Prediction

I. Introduction

Neurodegenerative diseases (ND) represent a major challenge worldwide due to their slow progression and significant impact on quality of life. Among the best known at present are Alzheimer's disease and Parkinson's disease, both of which currently have no known cure[1]. The development of Deep Learning (DL) and Machine Learning (ML) has revealed great potential in early detection of these diseases, either by analyzing medical images and clinical data [1], [2], [3]. Models such as convolutional neural networks (CNNs) have demonstrated effectiveness in identifying patterns in medical images, allowing prediction of the evolution of diseases even without symptoms [2]. In addition, multimodal data have been integrated to improve diagnostic accuracy. [4], [5], [6].

Despite significant advances, the implementation of these technologies faces several obstacles. In this regard, current models have limited generalizability, as they are trained on specific datasets that do not adequately represent the diversity of clinical populations [1], [6]. Added to this is the poor interpretability of these models, which often operate as "black boxes". These problems generate great distrust among health professionals, who, not being immersed in these technologies, do not understand how to reach this diagnosis [3], [4]. Recent studies have begun to explore how ML systems can be integrated into clinical workflows to improve decision-making and operational efficiency. For instance, Sandhu et al. [44] examined the implementation of a real-time ML model for early sepsis detection in a hospital environment, highlighting the importance of clinician trust and workflow alignment. Similarly, Miller et al. [45] discussed the use of ML approaches in neurology-related clinical trials, including Alzheimer's and Parkinson's disease, emphasizing their utility in patient selection, remote assessments, and outcome prediction. The direction of magnetic resonance imaging (MRI) also affects the accuracy of the models, which raises the need for a careful procedure [6]. In addition, a number of new models, such as conversational interactions captured by chatbots, have not yet been fully explored, which opens up a possibility for further research [7].

The need to synthesize and evaluate developments in the application of DL and ML for early detection of ND supports the conduct of this review. By analyzing studies that integrate a variety of data and new learning architectures, we see a potential to provide comprehensive guidance for future developments in ND diagnosis [4], [8], [9]. The results of this review may serve to improve the implementation of these models in clinical settings, facilitating more accurate diagnoses. [7]. The exploration of multitask models, such as reinforcement learning, has underscored the importance of investigating approaches that address the aforementioned problem in order to accommodate the diversity of clinical populations. [9], [10].

The focus of this research lies in the identification and synthesis of the newest studies on how to use DL and ML to detect ND at an early stage. We have sought to evaluate the most effective techniques, find current limitations, and suggest future lines of research that address the challenges of generalizability and interpretability [10], [11], [12], [13]. This review is intended to be a valuable resource for future research in implementing these models in clinical practice.

II. Methodology

In this review, the PICO method was used to search for articles, as it allows formulating questions based on the problem addressed through its components: Problem (P), Intervention (I), Comparison (C), and Outcomes (O). This approach enables the precise selection of relevant articles.

A. Formulation of PICOC

During the first phase, the components of PICOC shown in Table I were identified, and the Context (C) variable was added for a more precise search. Additionally, the main topic of the review was presented to align the PICOC components with the research topic.

TABLE I
IDENTIFICATION OF PICOC COMPONENTS

P	Limitations in the early detection of neurodegenerative diseases (ND) using conventional methods.
I	DL or ML models for the early detection of ND.
C	Comparison between different DL and ML models in terms of accuracy, efficiency, and scalability.
O	Performance of the models in terms of accuracy, efficiency, and scalability in the early detection of ND.
C	Development and testing environments on specialized software platforms for ND detection.

B. Question Formulation

Having identified the PICOC components, the formulation of the general research question (RQ) was carried out:

"What DL or ML models have been developed for the early detection of ND, and how have they performed in terms of accuracy, efficiency, and scalability in development and testing environments on specialized platforms?"

Specific questions for each PICOC component, shown in Table II, were subsequently developed to select the most relevant articles for this systematic review.

TABLE II
PICOC SPECIFIC QUESTIONS

P	What are the limitations of conventional methods in early detection of ND?
I	What DL and ML models have been developed for early detection of ND?
C	What are the differences between DL and ML models in terms of accuracy, efficiency, and scalability?
O	What accuracy, efficiency, and scalability statistics have the models reported in early ND detection?
C	In what environments have the DL and ML models been implemented and evaluated?

C. Keyword Identification

The keywords of each PICOC component shown in Table III were identified and the Boolean operators "OR" and double quotation marks ("") were used to facilitate the search in the databases.

TABLE III
KEYWORDS WITH OPERATOR

P	"Alzheimer's disease" OR "Limitations" OR "conventional methods" OR "early detection" OR "neurodegenerative diseases" OR "traditional diagnosis" OR "diagnostic challenges" OR "image-based diagnosis" OR "Parkinson's disease"
I	"Deep Learning models" OR "Machine Learning models" OR "early detection" OR "predictive algorithms" OR "automated diagnosis" OR "disease classification" OR "medical artificial intelligence" OR "neural networks" OR "Alzheimer detection" OR "Parkinson detection" OR "supervised learning"
C	"Model comparison" OR "diagnostic accuracy" OR "model efficiency" OR "model scalability" OR "Deep Learning vs Machine Learning" OR "model performance" OR "algorithm comparison" OR "cross-validation" OR "evaluation metrics" OR "sensitivity" OR "specificity"
O	"Accuracy" OR "efficiency" OR "scalability" OR "predictive performance" OR "classification results" OR "precision statistics" OR "model validation" OR "ROC curve" OR "AUC" OR "clinical performance" OR "error rate" OR "diagnostic accuracy"
C	"Development environments" OR "testing platforms" OR "model evaluation" OR "clinical settings" OR "medical software" OR "computational infrastructure" OR "healthcare systems" OR "model implementation" OR "AI applications" OR "AI frameworks"

D. PICOC Formula Syntax

The PICOC method was finalized, joining all the previously identified keywords using the Boolean operator "AND" to perform the correct search in the "Scopus" and "Web of Science" databases, finding the search equations shown in Table IV.

TABLE IV
SEARCH EQUATION

	Scopus	Web of Sciencie
Search Equation	(TITLE-ABS-KEY ("Alzheimer's disease" OR "Limitations" OR "conventional methods" OR "early detection" OR "neurodegenerative diseases" OR "traditional diagnosis" OR "diagnostic challenges" OR "image-based diagnosis" OR "Parkinson's disease") AND TITLE-ABS-KEY ("Deep Learning models" OR "Machine Learning models" OR "early detection" OR "predictive algorithms" OR "automated diagnosis" OR "disease classification" OR "medical artificial intelligence" OR "neural networks" OR "Alzheimer detection" OR "Parkinson detection" OR "supervised learning") AND TITLE-ABS-KEY ("Model comparison" OR "diagnostic accuracy" OR "model efficiency" OR "model scalability" OR "Deep Learning vs Machine Learning" OR "model performance" OR "algorithm comparison" OR "cross-validation" OR "evaluation metrics" OR "sensitivity" OR "specificity") AND TITLE-ABS-KEY ("Accuracy" OR "efficiency" OR "scalability" OR "predictive performance" OR "classification results" OR "precision statistics" OR "model validation" OR "ROC curve" OR "AUC" OR "clinical performance" OR "error rate" OR "diagnostic accuracy") AND TITLE-ABS-KEY ("Development environments" OR "testing platforms" OR "model evaluation" OR "clinical settings" OR "medical software" OR "computational infrastructure" OR "healthcare systems" OR "model implementation" OR "AI applications" OR "AI	"Alzheimer's disease"OR "Limitations"OR"conventional methods"OR"early detection"OR"neurodegenerative diseases"OR"traditional diagnosis"OR"diagnostic challenges"OR"image-based diagnosis"OR "Parkinson's disease" (Topic) and "Deep Learning models"OR"Machine Learning models"OR"early detection"OR"predictive algorithms"OR"automated diagnosis"OR"disease classification"OR"medical artificial intelligence"OR"neural networks"OR"Alzheimer detection"OR"Parkinson detection"OR"supervised learning" (Topic) and "Model comparison"OR"diagnostic accuracy"OR"model efficiency"OR"model scalability"OR"Deep Learning vs Machine Learning"OR"model performance"OR"algorithm comparison"OR"cross-validation"OR"evaluation metrics"OR"sensitivity"OR"specificity" (Topic) and "Accuracy"OR"efficiency"OR"scalability"OR"predictive performance"OR"classification results"OR"precision statistics"OR"model validation"OR"ROC curve"OR"AUC"OR"clinical performance"OR"error rate"OR"diagnostic accuracy" (Topic) and "Development environments"OR"testing platforms"OR"model evaluation"OR"clinical settings"OR"medical software"OR"computational infrastructure"OR"healthcare systems"OR"model

	frameworks")) AND PUBYEAR > 2020 AND PUBYEAR < 2026 AND (LIMIT-TO (OA , "all")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "ch") OR LIMIT-TO (DOCTYPE , "cp")) AND (LIMIT-TO (LANGUAGE , "English") OR LIMIT-TO (LANGUAGE , "Spanish"))	implementation"OR"AI applications"OR"AI frameworks" (Topic) and Open Access and 2024 or 2023 or 2022 or 2021 (Publication Years) and Article (Document Types) and English (Languages)

Once the PICOC method had been completed and the information had been searched, a total of 302 articles were obtained, 235 from Scopus and 67 from the Web of Science (WOS) database.

E. Specification of Inclusion and Exclusion Criteria

In order to ensure the validity and quality of the selected studies, inclusion and exclusion criteria were defined. These criteria were established to ensure that the selected articles were relevant to the study objectives and accurately addressed the use of ML and DL in the detection of ND. These criteria allowed selection of the most relevant studies, maximizing data quality and consistency.

As for the inclusion criteria, journal articles, conference papers and book chapters were accepted. In addition, these should include performance metrics such as sensitivity, specificity accuracy, area under the curve (AUC), AIC, BIC or RMSA. Similarly, studies should focus on ML and DL models applied for ND detection. Even, studies with real clinical data for training or validation of the models, such as MRI, Electroencephalogram (EEG) or biomarkers.

Regarding exclusion criteria, initially, articles that turned out to be duplicates or earlier versions of the same articles found in WOS and Scopus were excluded. Likewise, articles published before 2021 were discarded, as well as those that did not focus directly on ND or the use of ML and DL. In addition, articles that did not report performance metrics were discarded. Articles that are not available in English or Spanish, or that do not offer open access to the full text were also discarded.

The inclusion and exclusion criteria previously described were essential to ensure that the selected studies met quality standards and aligned with the objectives of the systematic review. After implementation, we were able to eliminate studies that were not relevant or did not provide meaningful data on the use of ML and DL in the setting of ND. The application of these criteria allowed an accurate selection of articles for analysis, ensuring the validity and relevance of the findings.

F. PRISMA Declaration

The PRISMA statement allowed a choice and analysis of articles, as well as the establishment of an extraction process for the articles to be screened in this literature review.

G. PRISMA Process

The development was carried out in five stages. First, 302 articles were identified (235 from Scopus and 67 from WOS) and 55 duplicates were removed using Mendeley. Then, 214 studies were excluded after analyzing unrelated titles and abstracts. One additional article was discarded for lack of access to the full text and two more for focusing on tumors instead of ND. Finally, 30 relevant articles were selected for the review.

Figure 1 shows the comprehensive way the article selection process was carried out, using the PRISMA diagram where 30 were screened.

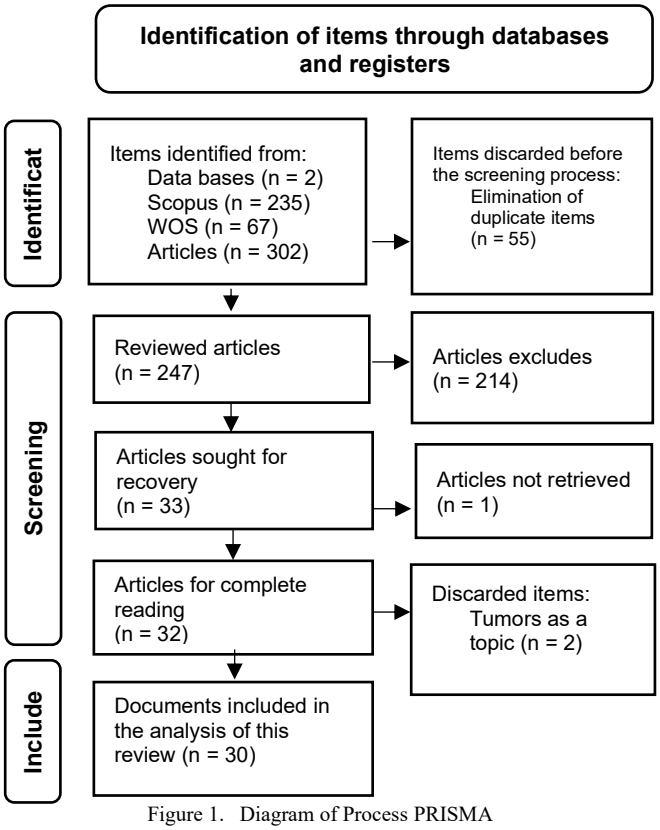

Figure 1. Diagram of Process PRISMA

III. RESULTS

In the results, tables and graphs were created to visually represent the different categories of the characteristics of the screened items. In the case of the graphs, Excel spreadsheet tools were used. In addition, the results address the PICOC questions formulated during the writing of the research.

A. Origin of the Articles

The genesis of the selected articles is shown in Figure 2, showing the volume of articles that were published in each country.

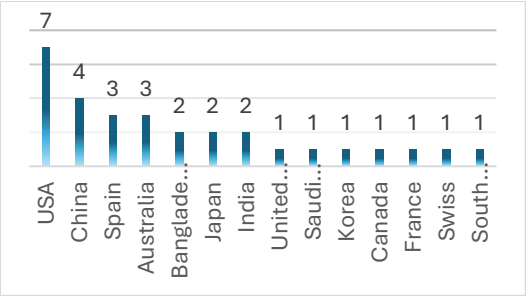

Figure 2. Frequency of Articles by Country

Figure 3, which was elaborated with R Studio's Bibliometrix software and based on all the results obtained in this review, presents the most cited countries in research related to the topic, highlighting China and the United States with 481 and 444 citations respectively, followed by Australia (283) and Korea (228). These data reflect a strong scientific contribution from these regions in the construction and use of DL and ML models for the early detection of ND, marking important trends at the global level.

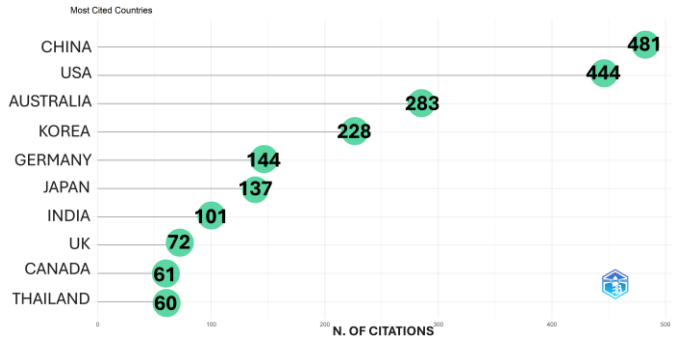

Figure 3. Most cited countries

B. Limitations of Conventional Methods

RQ1: "What are the limitations of conventional methods in the early detection of ND?"

In this section, the limitations and problems inherent in the conventional methods used for early detection of ND have been identified. A thorough review of previous studies was conducted with the purpose of pointing out the most relevant shortcomings and understanding how the methods studied attempt to address these difficulties. All the limitations reviewed in the studies are shown in Table V. The authors state which are the problems they want to overcome with the ML or DL model studied.

TABLE V
LIMITATIONS ENCOUNTERED PER ITEM

Author	Limitations Encountered
Patel J. et al. [14]	Documentation in free text, lack of oral health integration, habits and social factors not considered, late diagnosis, insufficient preventive treatments, lack of precision, lack of accessibility.
Mahmud T. et al [15]	Variability of MRI images, similarity between Alzheimer's phases, lack of interpretability, imbalance in data sets, bias towards majority classes, need for large amount of labeled data, high cost and time.
Kimura N. et al [16]	Costly, invasive, specialized infrastructure, low prevalence of amyloid positivity, high costs, patient burden, cost-effective solution, non-invasive solution, improved early identification.
Eguchi K. et all [17]	Costly, time consuming, unfeasible in pandemics, subjective, assessor dependent, remote assessment, video analysis, objective assessments, reduces burden.
Hasan M. et al [18]	Insufficient accuracy, unbalanced data handling, lack of robustness, failure to classify minority classes, inadequate pre-trained models.
Kourtzi Z. et al [19]	Lack of generalizability, reliance on invasive and costly data, insufficient demographic representativeness, heterogeneous data management, incorrect or late diagnosis, lack of interpretability.
Alhudhaif A. [20]	Unequal number of images, data magnification, rotation techniques, vertical flipping, center cropping, vertical translation, contrast and brightness adjustments, need to improve MRI resolution, deep super-resolution neural network.
Hossen M. et al [21]	Risk of overfitting, need to improve interpretability, RFE technique, selection of relevant features, complexity reduction, SHAP, XAI, increased transparency.
Tsuang D. et al [22]	Low representation of women, limited generalization, mostly male population, reliance on retrospective clinical records, dementia criteria not applied, underreporting of prevalence, high indicators of negative prediction, low in positive prediction.
Agarwal D. et al [23]	Overcome important limitations, improve accuracy, reduce computational complexity, learn knowledge transfer, optimize the use of MRI data, develop advanced preprocessing and normalization techniques, increase consistency and precision.
Khan S. et al [24]	scarcity of labeled data, variability of medical images, high MRI complexity, and risk of overfitting.
Kim K. et al [25]	Non-representative data sets, overestimating model performance, lack of confirmation of Aβ, exaggerating model accuracy.
Liu Y. et al [26]	Lack of accessible and specific biomarkers, invasive and expensive methods, clinical applicability, multiscale characteristics, reliance on single and low-resolution cohort data, need for more diverse cohorts.
Gayathri P. et al [27]	Class imbalance in medical data, training bias, synthetic sample generation, generalization challenges.
Cheung C. et al [28]	Data distribution discrepancy, unsupervised domain adaptation techniques, data availability, image consistency.
Lin A. et al [29]	High computational cost, overfitting, small sample size, lack of robust 3D pre-trained models.
Kofman J. et al [30]	Generalization, leave-one-participant-out (LOPO) validation, data processing, real-time, latency, computational memory, portable devices, low power consumption.
Tsai K. et al [31]	Multi-stage classification, automatic classification, speed, reduced evaluation burden, clinical accuracy.
Wang X. et al [32]	. Uncontrolled environments, cluttered backgrounds, manual intervention reduction, advanced neural networks, detection without controlled environment.
Gaurav R. et al [33]	Inter-rater variability, manual segmentation, scalability in large studies, automatic parameter adjustment, reproducible quantification, neuromelanin.
Crary J. et al [34]	Low sensitivity, inter-observer variability, early stage detection, automated analysis, replicability.
Kumar D. et al [35]	Uncontrolled environments, background noise, telemedicine, language independence and language ability, validation in diverse populations.

Kim S. et al [36]	Hyperparameter selection, automatic optimization, Harris Hawk (HHO), IoT scalability, data quality validation, further optimization (PSO, GA).
Panetsos F. et al [37]	Variability in laughter (emotional, psychological), lack of data in various conditions, static, dynamic representation of laughter.
Wang C. et al [38]	Multimodal data availability, knowledge distillation, complex patterns in MRI, additional monitoring, performance gap.
Qasim H. et al [39]	Class unbalancing, feature redundancy, noise reduction, feature selection (RFE, PCA).
Kang D. et al [40]	Variability in data acquisition (PET), spatial and count normalization, generalization in external data, consistency in diagnosis, AI for objectivity.
Youze X. et al [41]	Dependence on specialized equipment, cost, convenience, detection of minor steps, frequent monitoring, reduced burden on specialists.
Suarez C. et al [42]	Low accuracy in ICM detection, accessibility of data, minimization of clinical bias, standardization of diagnoses, speed of early detection.
Ghoraani B. et al [43]	Dependence on cognitive tests, cultural and educational influence, subjectivity, variability in clinical evaluations, cost reduction, simplicity, integrable devices.

A detailed chart illustrating the ten most common limiting keywords identified in analytical research was developed. These keywords represent recurring challenges in the scientific literature and are carefully grouped by thematic similarity, allowing for a deeper understanding of the main barriers and limitations facing this field of study. This diagram not only provides a clear visual representation of the key constraints, but also offers a precise quantitative reference of the critical aspects to be considered when developing and applying ML and DL models shown in Figure 4. By visualizing these challenges in a grouped manner, it facilitates the identification of areas that require further attention and improvement, thus improving the effectiveness and applicability of the models in real research and application contexts.

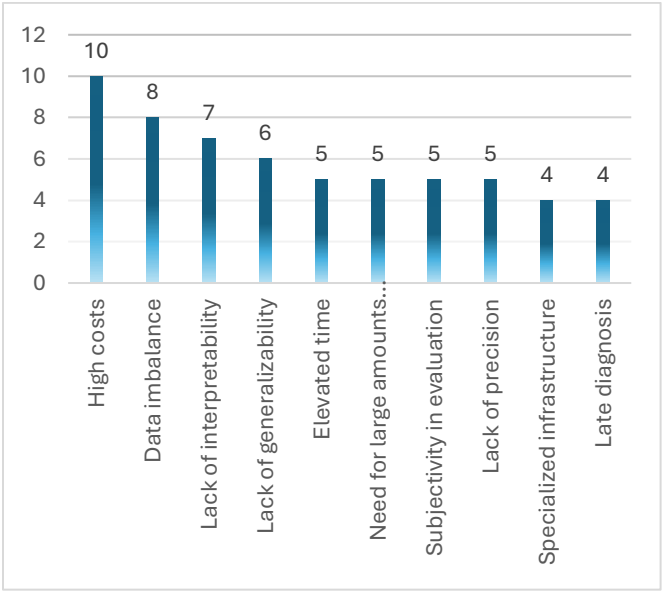

Figure 4. Main Limitations

C. Developed Models

The main DL and ML models developed for the early detection of neurodegenerative diseases (ND) have been extracted and comprehensively analyzed. A detailed analysis has been performed to highlight the most significant techniques and architectures employed throughout the study, emphasizing the key innovations and methodologies that have contributed to advances in this field.

The results are presented in detail in Table VI, which includes a list of authors, the models developed, the citation score, and the reference score. The citation score refers to the number of times each article has been cited by other scholarly works, reflecting its impact on the scientific community. The reference score, on the other hand, indicates the total number of bibliographic references used within each article, demonstrating the depth of its literature support.

This table provides a clear and concise overview of the most advanced techniques and their scientific influence based on both external citations and internal reference usage. These data are interesting and relevant as they provide a comparative view of the different approaches and models used in research on the early detection of neurodegenerative diseases. They highlight the most cited and, therefore, possibly most influential techniques in the field, providing a quantitative reference on which methods have been most effective and recognized in the scientific literature.

TABLE VI
MODELS STUDIED BY ARTICLE

	Author	Model	Cite Score	Reference Score
ML	Patel J. et al (2023) [14]	Extreme Gradient Boosting (XGBoost)	0	11
	Kimura N. et al (2023) [16], [19]	Support Vector Machine SVM, Elastic Net, Regresión Logística	1	82
	Kourtzi Z. et al (2024) [19]	Generalized Metric Learning Vector Quantization (GMLVQ)	0	30
	Hossen M. et al (2024) [21]	Regresión Logística (LR) combinada con Eliminación Recursiva de Características (RFE)	0	45
	Tsuang D. et al (2023) [22]	SVM	0	39
	Liu Y. et al (2023) [26]	Individual Brain-Related Abnormalities In Neurodegeneration (IBRAIN)	8	53
	Lin A. et al (2023) [29]	CNN en 3D y 2D, Learnable Weighted Pooling (LWP), ResNet34 (ResNet34), módulo de doble atención	12	38
	Kumar D. et al (2022) [35]	SVM	11	50
	Qasim H. et al (2021) [39]	SVM	19	36

	Ghoraani B. et al (2021) [43]	SVM	56	33
DL	Mahmud T. et al (2024) [15]	CNN	13	41
	Eguchi K. et all (2023) [17]	CNN based on the architecture ECO-Lite	3	30
	Hasan M. et al (2024) [18]	CNNs built from scratch, pre-trained VGG16 network with additional convolutional layers, Graph Convolutional Networks (GCNs), combined CNN-GCN architecture	4	33
	Alhudhaif A. (2024) [20]	DenseNet201, DarkNet53, Xception	2	46
	Agarwal D. et al (2023) [23]	Three-dimensional Convolutional Neural Network (3D CNN) based on EfficientNet-B0 architecture	12	
	Khan S. et al (2023) [24]	DenseNet201	6	71
	Kim K. et al (2023) [25]	Inception-V4	1	37
	Gayathri P. et al (2024) [27]	CNN, Minority Synthetic Oversampling Technique (SMOTE), Spider Monkey Optimization (SMO)	0	21
	Cheung C. et al (2022) [28]	EfficientNet-b2	100	29
	Kofman J. et al (2021) [30]	Long Short-Term Memory (LSTM)	47	51
	Tsai K. et al (2023) [31]	artificial neural networks (ANN)	6	52
	Wang X. et al (2021) [32]	CNN combined LSTM (CNN-LSTM)	28	28
	Gaurav R. et al (2022) [33]	NigraNet	8	78
	Crary J. et al (2022) [34]	CNN-InceptionV4	9	53
	Kim S. et al (2022) [36]	CNN - HHO	6	50
	Panetsos F. et al (2022) [37]	Ceptral Coeficients at Frequencies of Mel (MFCC)	4	46
	Wang C. et al (2021) [38]	CNN-profesor	23	
	Kang D. et al (2021) [40]	3D Visual Geometry Group Network (VGG3D)	13	29
	Youze X. et al (2021) [41]	3D CNN	14	28
	Suarez C. et al (2021) [42]	Counterpropagation Network (CPN)	12	32

Table 7, which indicates the number of models that are based on or built from other models, was developed to provide a generalization of the basis for their development. This table classifies models into two main categories: ML and DL, and lists different types of models along with the number of times each has been studied.

The data provide a comprehensive view of how models have been developed and studied within the context of ND diagnostics, allowing for the identification of common trends and approaches and highlighting the influence of previously established models on the development of new solutions. Table VII, therefore, is critical to understanding the basis on which these models have been built and their evolution over time.

TABLE VII
GENERAL MODELS STUDIED

ML	SVMCNN	6
	Logistic Regression (LR)	2
	Extreme Gradient Boosting (XGBoost)	1
	Generalized Metric Learning	1
	Elastic Net	1
DL	CNN	11
	CNN-based networks (DenseNet, VGG, etc.)	5
	CNN + Combinations (CNN-LSTM, CNN-GCN, etc.)	3
	Networks based on Inception	2
	3D CNN	2
	EfficientNet	2
	Counterpropagation Network (CPN)	1
	LSTM	1
	Other specific architectures	1

D. Model Metrics

RQ4: "What precision, accuracy, and sensitivity statistics have the models reported on early ND detection?"

In this section four metrics were extracted from all articles: Accuracy, Precision, Accuracy, F1 and Recall. Some articles have studied several models or several samples, so it is necessary to specify which model or sample obtained those results. These details are presented in Table VIII.

TABLE VIII
MODEL METRICS PER ITEM

Autor	Precisión	Accuracy	F1	Recall
Patel J. et al. [14]	83,56%	77%	70,54%	61%
Mahmud T. et al [15]	89%	96%	91%	93%
Kimura N. et al [16]	Kernel SVM	-	Kernel SVM	Kernel SVM

	49%; Elastic Net 51%; Logistic Regresion 51%		56%; Elastic Net 55%; Logistic Regresion 51%	69%; Elastic Net 63%; Logistic Regresion 58%
Eguchi K. et all [17]	-	71%	55%	63%
Hasan M. et al [18]	CNN 52%; VGG16 91%; GCN 100%; CNN-GCN 100%	CNN 43.83%; VGG16 71.17%; GCN 99.06%; CNN-GCN 100%	CNN 59%; VGG16 83%; GCN 100%; CNN-GCN 100%	CNN 69%; VGG16 76%; GCN 100%; CNN-GCN 100%
Kourtzi Z. et al [19]	-	88.66%	-	82.38%
Alhudhaif A. [20]	98.65%	99.11%	98.70%	98.75%
Hossen M. et al [21]	100%	99.5%	99.5%	100%
Tsuang D. et al [22]	-	Afroame_ ricans 91%; Cauca_ sian 89%	-	Afroame_ ricans 61%; Cauca_ sian 43%
Agarwal D. et al [23]	86.38%	87.38%	86.43%	87.51%
Khan S. et al [24]	80.14%	90.01%	80.22%	70.69%
Kim K. et al [25]	-	87.7%	-	85.6%
Liu Y. et al [26]	-	95%	-	-
Gayathri P. et al [27]	89%	91%	91%	92%
Cheung C. et al [28]	79% - 92%	83.6%	-	93.2%
Lin A. et al [29]	88.27%	88.71%	86.93%	85.63%
Kofman J. et al [30]	25.3%	-	35%	82.1%
Tsai K. et al [31]	93.72%	-	-	-
Wang X. et al [32]	81%	-	80%	79%
Gaurav R. et al [33]	80%	-	-	-
Crary J. et al [34]	81%	99%	89%	99%
Kumar D. et al [35]	-	100%	100%	100%
Kim S. et al [36]	88.06%	96.21%	94.68%	92.54%
Panetsos F. et al [37]	83%	-	-	83%
Wang C. et al [38]	-	80.10%	-	80.3%
Qasim H. et al [39]	99%	98.2%	97%	97%
Kang D. et al [40]	97.7%	97.7%	83.1%	94.5%
Youze X. et al [41]	-	90.8%	-	84.2%
Suarez C. et al [42]	-	86.84%	-	84.78%
Ghoraani B. et al [43]	-	78%	77%	-

E. Developed Environments

RQ5: "In which environments have the DL and ML models been implemented and evaluated?"

This section has described the settings in which the DL and ML models developed during the study have been implemented and evaluated. A detailed analysis of the clinical and experimental settings in which these models have been applied has been conducted, highlighting their efficacy and feasibility in different scenarios. Table IX provides a detailed overview of the environments used, including author names, years of publication and the specific platforms employed, such as Google Colab with Tesla-K80 GPU, Python, PyTorch, MATLAB, among others. This information is crucial in order to examine the effectiveness and application of the models in different scenarios.

TABLE IX
MODEL DEVELOPMENT ENVIRONMENTS

Author	Environments
Patel J. et al. [14]	-
Mahmud T. et al [15]	Google Colab with GPU Tesla-K80
Kimura N. et al [16]	Google Colab with GPU Tesla-K80
Eguchi K. et all [17]	Python
Hasan M. et al [18]	Python with Adagrad Optimizer
Kourtzi Z. et al [19]	-
Alhudhaif A. [20]	Python with GPUs NVIDIA Tesla V100
Hossen M. et al [21]	Google Colab
Tsuang D. et al [22]	Python
Agarwal D. et al [23]	Google Colab Pro+
Khan S. et al [24]	Google Colab, TensorFlow, Keras
Kim K. et al [25]	VUNO Med-DeepBrain AD
Liu Y. et al [26]	-
Gayathri P. et al [27]	Python
Cheung C. et al [28]	-
Lin A. et al [29]	PyTorch
Kofman J. et al [30]	MATLAB
Tsai K. et al [31]	Microsoft Excel
Wang X. et al [32]	Python
Gaurav R. et al [33]	Python
Crary J. et al [34]	Pytorch
Kumar D. et al [35]	MATLAB - Python
Kim S. et al [36]	Python
Panetsos F. et al [37]	MATLAB
Wang C. et al [38]	Pytorch
Qasim H. et al [39]	-
Kang D. et al [40]	MATLAB
Youze X. et al [41]	Pytorch
Suarez C. et al [42]	-
Ghoraani B. et al [43]	MATLAB

F. Data Bases

The following analysis has detailed the frequency of use of various databases. Fig. 5 presents a bar chart illustrating the volume of queries made in each of the situations.

The percentage distribution of queries in different databases has been analyzed and is represented in the bar chart. The most frequently used databases have been identified, providing a clear and precise view of their relevance in the

context of the study. The databases are divided percentage-wise as follows: Own (27%), ADNI (23%), Medical Center (20%), Kaggle (17%), OASIS (4%), TIM - Tremor (3%), S4 (3%) and iPhone 6s (3%).

The graphical representation is of great relevance, since it facilitates the visualization of the predominance of certain databases over others. This visualization helps to efficiently understand the distribution and frequency of use, highlighting the most consulted databases and, therefore, the most influential in the context of the analysis performed.

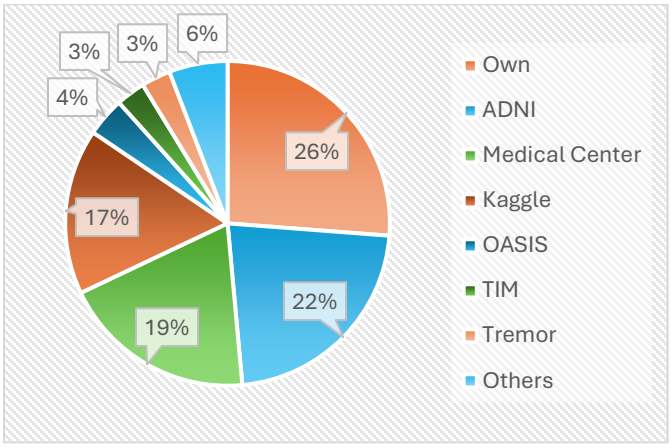

Figure 5. Databases Used

G. World Cloud

Fig. 6 performed with Bibliometrix shows the N-gram of the most frequent words used in the research, highlighting terms such as machine learning, deep learning, diagnostic accuracy, artificial intelligence and diagnostic imaging, which reflect the focus on the use of advanced models for diagnostic accuracy in human populations, with emphasis on neurodegenerative and oncological diseases.

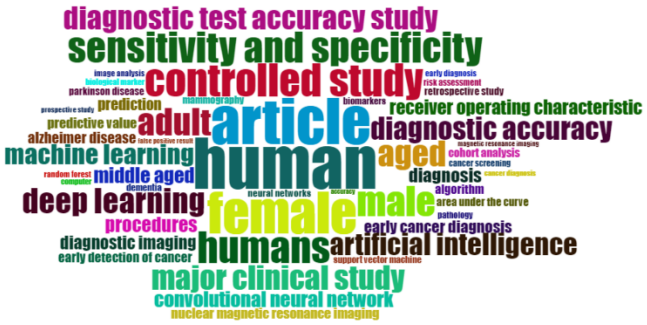

Figure 6. Worl cloud of frequent words

IV. DISCUSSION

In this review analysis, the metrics obtained for traditional ML models, such as SVMs, show limited sensitivity to more advanced DL models. As described by Tsuang D. et al. [22], SVM achieves a sensitivity of 61% in Afroamericans, but only 43% in Caucasians. In contrast, CNN-based models, such as those reported by Mahmud T. et al. [15], achieve much higher sensitivities, reaching 93%. These differences could be attributed to the limitations of SVMs to process complex data, unlike deep models that capture more sophisticated patterns.

In addition, differently from what was reported by Kimura N. et al.[16], where models such as logistic regression or Elastic Net show sensitivities of no more than 63%, more advanced architectures, such as CNNs combined with GCNs, achieve perfect sensitivities of 100%, according to Hasan M. et al. [18]. This is evidence that, in tasks that demand higher accuracy and robustness, deep models are not only superior, but also more consistent in various scenarios.

In addition, three-dimensional models, such as those based on 3D CNN, achieve high sensitivities. Agarwal D. et al. [23], reports that 3D CNNs have a sensitivity of 87.51%, unlike traditional approaches, these architectures achieve better performance in more complex environments.

V. CONCLUSIONS

Recent advances in the early diagnosis of neurodegenerative diseases (ND) have been remarkable, especially with the incorporation of DL and ML models. Despite the achievements made by conventional methods, such as visual observation or diagnostic approaches based on clinical parameters, these still have significant limitations in terms of accuracy and variability. In comparison, DL and ML models, such as CNNs and the combined use of CNGs, have demonstrated greater efficiency, achieving accuracy and sensitivity rates above 90%. For example, an SVM-based model achieved 100% accuracy, while CNN-InceptionV4 stood out with 99% accuracy and 89% sensitivity. DenseNet201 obtained outstanding metrics, with 98.65% accuracy, 98.75% sensitivity and an F1-score of 98.70%, while CNN-GCN presented perfect performance, with 100% in all metrics evaluated. These results highlight the superiority of these architectures for handling large volumes of complex clinical data and improving diagnostic accuracy.

Indeed, models implemented in various research and clinical practice settings have shown high performances in diagnostic tasks. However, although state-of-the-art models such as DenseNet201 and CNN-GCN offer outstanding performance, their practical implementation still requires further evaluation under real-world conditions to optimize their use in clinical diagnostics and ensure their large-scale applicability. This scenario suggests that while DL and ML models represent promising tools, validation and fine-tuning of technical parameters will be key to their adoption to maximize their impact on early detection of ND.

REFERENCES

[1] Z. Zhou *et al.*, "A novel graph neural network method for Alzheimer's disease classification," *Comput Biol Med*, vol. 180, Sep. 2024, doi: 10.1016/j.compbiomed.2024.108869.

[2] S. K. Tripathy *et al.*, "Alzheimer's Disease Detection via Multiscale Feature Modelling Using Improved Spatial Attention Guided Depth Separable CNN," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, Dec. 2024, doi: 10.1007/s44196-024-00502-y.

[3] M. A. Zayene, H. Basly, and F. E. Sayadi, "Multi-View Separable Residual convolution neural Network for detecting Alzheimer's disease progression," *Biomed Signal Process Control*, vol. 95, Sep. 2024, doi: 10.1016/j.bspc.2024.106375.

[4] O. Topsakal and S. Lenkala, "Enhancing Alzheimer's Disease Detection through Ensemble Learning of Fine-Tuned Pre-Trained Neural Networks," *Electronics (Switzerland)*, vol. 13, no. 17, Sep. 2024, doi: 10.3390/electronics13173452.

[5] R. F. Tian, J. N. Li, and S. W. Zhang, "MSCLK: Multi-scale fully separable convolution neural network with large kernels for early diagnosis of Alzheimer's disease," *Expert Syst Appl*, vol. 252, Oct. 2024, doi: 10.1016/j.eswa.2024.124241.

[6] B. A. C. Ramalho *et al.*, "The impact of the orientation of MRI slices on the accuracy of Alzheimer's disease classification using convolutional neural networks (CNNs)," *J Med Artif Intell*, vol. 7, Dec. 2024, doi: 10.21037/jmai-24-51.

[7] M. S. Kadafi, A. K. Yaqubi, Purbandini, and S. D. Astuti, "Alzheimer's prediction via CNN-SVM on chatbot platform with MRI," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 36, no. 1, pp. 64–73, Oct. 2024, doi: 10.11591/ijeecs.v36.i1.pp64-73.

[8] J. S. Bazargani, N. Rahim, A. Sadeghi-Niaraki, T. Abuhmed, H. Song, and S. M. Choi, "Alzheimer's disease diagnosis in the metaverse," *Comput Methods Programs Biomed*, vol. 255, Oct. 2024, doi: 10.1016/j.cmpb.2024.108348.

[9] M. Hatami, F. Yaghmaee, and R. Ebrahimpour, "Investigating the potential of reinforcement learning and deep learning in improving Alzheimer's disease classification," *Neurocomputing*, vol. 597, Sep. 2024, doi: 10.1016/j.neucom.2024.128119.

[10] F. Liu *et al.*, "Multi-task joint learning network based on adaptive patch pruning for Alzheimer's disease diagnosis and clinical score prediction," *Biomed Signal Process Control*, vol. 95, Sep. 2024, doi: 10.1016/j.bspc.2024.106398.

[11] I. Ahammad *et al.*, "AITeQ: a machine learning framework for Alzheimer's prediction using a distinctive five-gene signature," *Brief Bioinform*, vol. 25, no. 4, Jul. 2024, doi: 10.1093/bib/bbae291.

[12] C. Ozdemir and Y. Dogan, "Advancing early diagnosis of Alzheimer's disease with next-generation deep learning methods," *Biomed Signal Process Control*, vol. 96, Oct. 2024, doi: 10.1016/j.bspc.2024.106614.

[13] M. Heenaye-Mamode Khan, P. Reesaul, M. M. Auzine, and A. Taylor, "Detection of Alzheimer's disease using pre-trained deep learning models through transfer learning: a review," *Artif Intell Rev*, vol. 57, no. 10, Oct. 2024, doi: 10.1007/s10462-024-10914-z.

[14] J. Patel and H. Wu, "Utilizing Electronic Dental Records to Predict Neuro-Degenerative Diseases in a Dental Setting: A Pilot Study," in *Studies in Health Technology and Informatics*, IOS Press BV, Jan. 2024, pp. 1322–1326. doi: 10.3233/SHTI231179.

[15] T. Mahmud, K. Barua, S. U. Habiba, N. Sharmen, M. S. Hossain, and K. Andersson, "An Explainable AI Paradigm for Alzheimer's Diagnosis Using Deep Transfer Learning," *Diagnostics*, vol. 14, no. 3, Feb. 2024, doi: 10.3390/diagnostics14030345.

[16] N. Kimura *et al.*, "Predicting positron emission tomography brain amyloid positivity using interpretable machine learning models with wearable sensor data and lifestyle factors," *Alzheimers Res Ther*, vol. 15, no. 1, Dec. 2023, doi: 10.1186/s13195-023-01363-x.

[17] K. Eguchi *et al.*, "Gait video-based prediction of unified Parkinson's disease rating scale score: a retrospective study," *BMC Neurol*, vol. 23, no. 1, Dec. 2023, doi: 10.1186/s12883-023-03385-2.

[18] M. E. Hasan and A. Wagler, "New Convolutional Neural Network and Graph Convolutional Network-Based Architecture for AI Applications in Alzheimer's Disease and Dementia-Stage Classification," *AI (Switzerland)*, vol. 5, no. 1, pp. 342–363, Mar. 2024, doi: 10.3390/ai5010017.

[19] L. Y. Lee *et al.*, "Robust and interpretable AI-guided marker for early dementia prediction in real-world clinical settings," *EClinicalMedicine*, vol. 74, Aug. 2024, doi: 10.1016/j.eclinm.2024.102725.

[20] A. Alhudhaif, "A novel approach to recognition of Alzheimer's and Parkinson's diseases: random subspace ensemble classifier based on deep hybrid features with a super-resolution image," *PeerJ Comput Sci*, vol. 10, 2024, doi: 10.7717/peerj-cs.1862.

[21] R. Ahmed *et al.*, "A novel integrated logistic regression model enhanced with recursive feature elimination and explainable artificial intelligence for dementia prediction," *Healthcare Analytics*, vol. 6, Dec. 2024, doi: 10.1016/j.health.2024.100362.

[22] Y. Shao *et al.*, "Identifying Probable Dementia in Undiagnosed Black and White Americans Using Machine Learning in Veterans Health Administration Electronic Health Records," *Big Data and Cognitive Computing*, vol. 7, no. 4, Dec. 2023, doi: 10.3390/bdcc7040167.

[23] D. Agarwal, M. Á. Berbís, A. Luna, V. Lipari, J. B. Ballester, and I. de la Torre-Díez, "Automated Medical Diagnosis of Alzheimer´s Disease Using an Efficient Net Convolutional Neural Network," *J Med Syst*, vol. 47, no. 1, Dec. 2023, doi: 10.1007/s10916-023-01941-4.

[24] A. W. Saleh, G. Gupta, S. B. Khan, N. A. Alkhaldi, and A. Verma, "An Alzheimer's disease classification model using transfer learning Densenet with embedded healthcare decision support system," *Decision Analytics Journal*, vol. 9, Dec. 2023, doi: 10.1016/j.dajour.2023.100348.

[25] J. Bin Bae *et al.*, "A case-control clinical trial on a deep learning-based classification system for diagnosis of amyloid-positive alzheimer's disease," *Psychiatry Investig*, vol. 20, no. 12, pp. 1195–1203, Dec. 2023, doi: 10.30773/pi.2023.0052.

[26] K. Zhao *et al.*, "A neuroimaging biomarker for Individual Brain-Related Abnormalities In Neurodegeneration (IBRAIN): a cross-sectional study," *EClinicalMedicine*, vol. 65, Nov. 2023, doi: 10.1016/j.eclinm.2023.102276.

[27] P. Gayathri *et al.*, "Deep Learning Augmented with SMOTE for Timely Alzheimer's Disease Detection in MRI Images," 2024. [Online]. Available: www.ijacsa.thesai.org

[28] C. Y. Cheung *et al.*, "A deep learning model for detection of Alzheimer's disease based on retinal photographs: a retrospective, multicentre case-control study," *Lancet Digit Health*, vol. 4, no. 11, pp. e806–e815, Nov. 2022, doi: 10.1016/S2589-7500(22)00169-8.

[29] X. Xing *et al.*, "Efficient Training on Alzheimer's Disease Diagnosis with Learnable Weighted Pooling for 3D PET Brain Image Classification," *Electronics (Switzerland)*, vol. 12, no. 2, Jan. 2023, doi: 10.3390/electronics12020467.

[30] G. Shalin, S. Pardoel, E. D. Lemaire, J. Nantel, and J. Kofman, "Prediction and detection of freezing of gait in Parkinson's disease from plantar pressure data using long short-term memory neural-networks," *J Neuroeng Rehabil*, vol. 18, no. 1, Dec. 2021, doi: 10.1186/s12984-021-00958-5.

[31] S. Y. C. Ho, T. W. Chien, M. L. Lin, and K. T. Tsai, "An app for predicting patient dementia classes using convolutional neural networks (CNN) and artificial neural networks (ANN): Comparison of prediction accuracy in Microsoft Excel," *Medicine (United States)*, vol. 102, no. 4, p. E32670, Jan. 2023, doi: 10.1097/MD.0000000000032670.

[32] X. Wang, S. Garg, S. N. Tran, Q. Bai, and J. Alty, "Hand tremor detection in videos with cluttered background using neural network based approaches," *Health Inf Sci Syst*, vol. 9, no. 1, Dec. 2021, doi: 10.1007/s13755-021-00159-3.

[33] R. Gaurav *et al.*, "NigraNet: An automatic framework to assess nigral neuromelanin content in early Parkinson's disease using convolutional neural network," *Neuroimage Clin*, vol. 36, Jan. 2022, doi: 10.1016/j.nicl.2022.103250.

[34] M. Signaevsky *et al.*, "Antemortem detection of Parkinson's disease pathology in peripheral biopsies using artificial intelligence," *Acta*

Neuropathol Commun, vol. 10, no. 1, Dec. 2022, doi: 10.1186/s40478-022-01318-7.

[35] M. A. Motin, N. D. Pah, S. Raghav, and D. K. Kumar, "Parkinson's Disease Detection Using Smartphone Recorded Phonemes in Real World Conditions," *IEEE Access*, vol. 10, pp. 97600–97609, 2022, doi: 10.1109/ACCESS.2022.3203973.

[36] I. Yousaf, F. Anwar, S. Imtiaz, A. S. Almadhor, F. Ishmanov, and S. W. Kim, "An Optimized Hyperparameter of Convolutional Neural Network Algorithm for Bug Severity Prediction in Alzheimer's-Based IoT System," *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/7210928.

[37] M. Terriza *et al.*, "Use of Laughter for the Detection of Parkinson's Disease: Feasibility Study for Clinical Decision Support Systems, Based on Speech Recognition and Automatic Classification Techniques," *Int J Environ Res Public Health*, vol. 19, no. 17, Sep. 2022, doi: 10.3390/ijerph191710884.

[38] H. Guan, C. Wang, and D. Tao, "MRI-based Alzheimer's disease prediction via distilling the knowledge in multi-modal data," *Neuroimage*, vol. 244, Dec. 2021, doi: 10.1016/j.neuroimage.2021.118586.

[39] H. M. Qasim, O. Ata, M. A. Ansari, M. N. Alomary, S. Alghamdi, and M. Almehmadi, "Hybrid feature selection framework for the parkinson imbalanced dataset prediction problem," *Medicina (Lithuania)*, vol. 57, no. 11, Nov. 2021, doi: 10.3390/medicina57111217.

[40] S. Y. Lee, H. Kang, J. H. Jeong, and D. Y. Kang, "Performance evaluation in [18F]Florbetaben brain PET images classification using 3D Convolutional Neural Network," *PLoS One*, vol. 16, no. 10 October 2021, Oct. 2021, doi: 10.1371/journal.pone.0258214.

[41] X. Cao *et al.*, "Video Based Shuffling Step Detection for Parkinsonian Patients Using 3D Convolution," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 641–649, 2021, doi: 10.1109/TNSRE.2021.3062416.

[42] C. P. Suárez-Araujo, P. García Báez, Y. Cabrera-León, A. Prochazka, N. Rodríguez Espinosa, and C. Fernández Viadero, "A Real-Time Clinical Decision Support System, for Mild Cognitive Impairment Detection, Based on a Hybrid Neural Architecture," *Comput Math Methods Med*, vol. 2021, 2021, doi: 10.1155/2021/5545297.

[43] B. Ghoraani, L. N. Boettcher, M. D. Hssayeni, A. Rosenfeld, M. I. Tolea, and J. E. Galvin, "Detection of mild cognitive impairment and Alzheimer's disease using dual-task gait assessments and machine learning," *Biomed Signal Process Control*, vol. 64, Feb. 2021, doi: 10.1016/j.bspc.2020.102249.

[44] S. Sandhu et al., "A Cross-Sectional Survey of Sepsis Watch: Real-Time Sepsis Alert System," *J. Med. Internet Res.*, vol. 22, no. 8, p. e15182, 2020, doi: 10.2196/15182.

[45] D. D. Miller, D. J. Greig, M. Topol, and E. J. Topol, "Machine Learning and Clinical Trials in Neurology: Opportunities, Challenges, and Future Directions," *Neurology*, vol. 100, no. 3, pp. 101–111, 2023, doi: 10.1212/WNL.0000000000201254.