Article

Not peer-reviewed version

A Comparative Evaluation of Artificial Intelligence and Physician Based Approaches in the Early Diagnosis of Cognitive Aging in Patients with Subjective Cognitive Decline and Risk of Alzheimer’s Disease

Davis Kannenieks * , Zanda Priede , Andrejs Millers , Karlis Kristofers Velins

Posted Date: 13 May 2026

doi: 10.20944/preprints202605.0817.v1

Keywords: subjective cognitive decline; mild cognitive impairment; early dementia detection; machine learning algorithm; NACC dataset; SHAP analysis; XGBoost classification

Article

A Comparative Evaluation of Artificial Intelligence and Physician Based Approaches in the Early Diagnosis of Cognitive Aging in Patients with Subjective Cognitive Decline and Risk of Alzheimer's Disease

Davis Kannenieks [1,*], Zanda Priede [2], Andrejs Millers [2] and Karlis Kristofers Velins [3]

[1] Faculty of Medicine, Riga Stradins University, LV-1007, Riga
[2] Department of Neurology and Neurosurgery, Riga Stradins University, LV-1002, Riga
[3] Department of Mathematics, Imperial College London, London, United Kingdom
* Correspondence: davis.kannenieks@gmail.com

Abstract

Background: As the society ages, the number of patients with early cognitive impairment that can progress to Alzheimer's disease also increases. Early diagnosis and risk assessment allows effectively initiate the necessary lifestyle changes and monitoring. The use of artificial intelligence (AI), when analyzing medical histories, enables more productive evaluation of large datasets and identify patterns that may go unnoticed in clinical practice. This kind of approach can improve early screening, reduce physicians' workload and develop bigger support for personalized treatment. *The aim of the study:* To compare the performance of machine learning (ML) algorithm with a physician (neurologist) in assessing patient's subjective cognitive decline and Alzheimer's disease risk in early stages. *Research methods:* The research was designed as a retrospective, comparative cohort study that used two data sources. Firstly, the National Alzheimer's Coordination Center (NACC) longitudinal dataset to train the ML model. Secondly, medical records gathered from Pauls Stradins Clinical University Hospital dating from 2020 till May 2025 to evaluate the algorithm's precision. *Results:* The research included 154 patients, predominantly women (68.8%), with a mean age of 80.3 years. Class distribution consisted of dementia (n=139); mild cognitive impairment (MCI) (n=13); subjective cognitive decline (SCD) (n=2). Dementia was identified the best – 128/139 (accuracy – 92.1%) with errors tending towards MCI. MCI was correct in 9/13 cases (accuracy – 69.2%) All SCD cases were classified as dementia. Overall model's accuracy was 91.6% (141/154). *Conclusions:* ML algorithm can match to neurologist made diagnoses with high precision but is struggles to separate adjacent early-stage diagnoses. At this moment, ML models are great decision supporters, but no yet alone diagnosticians. Nevertheless, this technology has high potential to being integrated in the future to aid triage and early screening, especially when advanced diagnostics are limited.

Keywords: subjective cognitive decline; mild cognitive impairment; early dementia detection; machine learning algorithm; NACC dataset; SHAP analysis; XGBoost classification

1. Introduction

The cognitive function is a mental process that enables individuals to acquire, understand and utilize information through thinking, experience and sensory input [1]. It is necessary for a better quality of life, independence, and social participation. Although cognitive aging is a natural part of human lifespan, the cognitive associated diagnoses are increasingly becoming more important in both medicine and public health. According to the World Health Organization (WHO), the global population aged 60 years and older reached 1 billion in 2020 and is projected to increase to 2.1 billion by 2050 [2]. Consequently, the age associated neurodegenerative diseases are expected to rise substantially.

Alzheimer's disease (AD) is the most common cause of dementia and is characterized by progressive, age-related cognitive decline that affects not only the patient but also the surrounding people, family, friends and caregivers. Importantly, the pathogenesis of AD begins many years, often decades, before the first clinical symptoms become apparent. In this preclinical period, biochemical and structural brain changes accumulate, which do not necessarily correlate with clinical symptoms at first. As a result, the diagnosis is frequently delayed until measurable objective data can be obtained for cognitive deficit, at which point treatment options are more limited [3,4].

Clinically relevant stage that may improve early detection, therefore guide patients to faster lifestyle modifications and closer clinical monitoring, is the period before objectively measurable cognitive impairment, commonly referred to as a subjective cognitive decline (SCD) [5]. SCD describes a condition in which individuals report worsening memory, orientation, or concentration, while objective cognitive testing does not yet demonstrate clear abnormalities [6,7]. Historically, such complaints were often considered as part of normal aging. However, more recent evidence suggests that SCD may be associated with early manifestations of AD, particularly in individuals with a relevant family history or other risk factors [6].

At the same time, SCD assessment and diagnosis remains challenging in clinical settings, due to evaluation that relies on patient reported symptoms and physician interpretations, both of which introduce high subjectivity and possibility of errors [5]. Although a broad range of diagnostic tools are available, for instance, biomarkers and neuroimaging [8], they are not always accessible and may not be clinically or economically justified for an early-stage diagnostic evaluation.

Recent advances in clinical technologies and data analytics have created new possibilities to support physicians in their diagnosis, risk stratification and prognosis. Artificial intelligence (AI), more specifically machine learning (ML), is one of these technologies that can integrate and analyze structured data (cognitive test results, laboratory findings) and unstructured data (radiological images, medical history records, and clinical notes) to identify patterns that can reveal clinically meaningful associations that are difficult to detect through standard assessments [9, 10]. This technology also comes with its limitations and challenges – trusting the algorithm outputs/results, "black box" phenomenon, data interpretability and quality [11]. Producing a reasonable question, whether ML should be viewed as a replacement or a supporter to physician in clinical judgments and early cognitive assessments. Therefore, direct comparison between ML-based assessments and neurologist evaluations would help to identify the strengths and limitations of each approach and provide a clearer analysis of how effective this collaboration in cognitive healthcare is.

Aim of the Research

To compare the performance of an ML-based algorithm with that of a neurologist in evaluating subjective cognitive decline and the risk of Alzheimer's disease development at an early stage without the use of biomarkers or neuroimaging data.

Research Objectives

1. To train the ML algorithm using a prospective international cohort dataset from the National Alzheimer's Coordinating Center [12].
2. To evaluate and compare the accuracy between an ML algorithm and neurologist clinical assessment in patients with subjective cognitive decline and risk of Alzheimer's disease.
3. To identify the clinical and cognitive parameters that strongly contribute to ML-based prediction of cognitive decline in patients with subjective cognitive decline and risk of Alzheimer's disease.
4. To assess the accuracy and clinical applicability of machine learning models in clinical practice.
5. To identify cases in which ML-based assessment differs substantially from the physician's clinical conclusion and to analyze the possible causes of these variations.

2. Materials and Methods

2.1. Study Design

The study was designed as a retrospective comparative cohort study. The main purpose was to compare the diagnostic category produced by the machine learning (ML) model with the final diagnosis established by a neurologist, and to evaluate the model's potential to support early identification of diagnoses associated with cognitive change. The model was developed so it could be used in clinical practice and realistic situations, using only subjective and objective clinical data, without biomarker testing or radiological imaging findings.

Two datasets were used. A prospective international cohort longitudinal dataset from the National Alzheimer's Coordinating Center (NACC) was used to develop and train the ML model. The data are contributed by the NIA (National Institute of Aging)-funded Alzheimer's Disease Research Centers (ADRCs) [12]. A retrospective Latvian clinical cohort from Pauls Stradins Clinical University Hospital (PSCUH) was used for external evaluation against physician diagnosis. The cohort design approach was selected because it allows structured comparison of diagnostic outcomes and clinical profiles across different cognitive-status categories in a way that reflects real clinical pathways.

2.2. Data Sources

For the model development, data were obtained from NACC, which provides standardized longitudinal clinical information. The data includes repeated clinical visits, structured cognitive and functional assessments, demographical, clinical and psychometrical information, supporting the model's learning across diverse patient populations. Access to NACC data is restricted due to the NACC Data Use Agreement, and the dataset cannot be redistributed by the authors; qualified researchers may obtain access by submitting a data request to NACC and completing the required agreements.

To evaluate the model's accuracy, the Latvian retrospective dataset was obtained from PSCUH and covered January 2020 to May 2025. Patients were included if a diagnosis matched the cognitive status category (subjective cognitive decline; mild cognitive impairment; dementia) of this research and was diagnosed within this period. The last clinical visit at which the diagnosis was established and recorded, was used to assess the model's accuracy. Patients whose diagnostic endpoint fell outside the period were excluded from the research.

2.3. Ethical Approval

The research was approved by the Riga Stradins University Ethics Committee (No. 2-PĒK-4/952/2025) and authorized by the PSCUH Scientific Institute. Patient confidentiality and privacy were maintained according to institutional requirements, and analyses were performed using encrypted data.

2.4. Diagnostic Categories

The machine learning (ML) model was configured to output four diagnostic categories: normal cognition; subjective cognitive decline (SCD); mild cognitive impairment (MCI), and dementia. In the Latvian cohort study, reference labels were derived from the International Classification of Diseases (ICD) codes recorded in the clinical documentation. The model did not attempt to distinguish dementia types, so all dementia etiologies were grouped as "dementia".

2.5. Data and Variables Used

The data and variables were selected to reflect clinically accessible information that would be meaningful for early diagnostic assessment [5,13]. The full variables include: sociodemographic (sex, age, education, marital status, living status, economic situation); patient/member of patient's family reported cognitive complaints; objective cognitive measures (MoCA (*The Montreal Cognitive Assessment*), MMSE (*The Mini-Mental State Examination*), CDR (*The Clinical Dementia Rating)* tests and CDR domains; psychiatric (depression and anxiety related measures); vascular/somatic risk factors (hypertension, diabetes mellitus, stroke and heart attack history); comorbidities and the use of medication (type of medication).

In the NACC dataset, the final feature list used to train the model was determined by selecting only the predefined variables (mentioned above) that were also present in the Latvian cohort study, ensuring consistent data output without including additional variables that could influence the model's accuracy.

2.6. Data Processing and Handling of Missing Values

Data preparation and modeling were implemented using a structured machine-learning pipeline. The predictor matrix was constructed from the selected feature set, and the target variable was defined as "Diagnosis." Medical records missing the target diagnosis label were removed prior to training.

In both datasets, many entries contained missing or unknown data. In the NACC dataset, originally missing or unknown values were encoded as random numbers. For example, for a binary column having values of 0 and 1, the value 9 was assigned in the case of 'unknown'. Adjustments for all variables that contained such numbers were made as the model would otherwise treat it as a continuous value which would be 9 times larger than the answer 'yes' coded as 1 which would cause noise injected as a signal into the model and reduce performance. Adjustments were made on a case-by-case basis, either setting the missing values to NaN (Not a Number) for numerical features and by distinguishing another categorical class such as 'unknown' for the categorical features [14].

To accommodate different data types, preprocessing was performed using a *ColumnTransformer* with separate pipelines for numerical and categorical features. The numerical variables were imputed using the median value and then standardized. Categorical variables were imputed using the most frequent value and transformed using one-hot encoding with handling of previously unseen categories [15].

In the NACC dataset the 4 classes were heavily imbalanced. While normal cognition made up 48.9% of all entries, the SCD accounted for only 4.4% of the total dataset. MCI 17.5% and Dementia 29.2%. To adjust for the imbalance and to reduce bias, multiple bias-reducing approaches such as SMOTE (Synthetic Minority Over-sampling Technique), inverse-frequency sample weights and under sampling the majority classes were tested on the model [14,16].

2.7. Model Development

A supervised multiclass classification model based on gradient-boosted decision trees was trained using the *XGBoost* architecture. This choice was made as *XGBoost* works well on tabular, mixed-type datasets and can handle non-linear data [17-19].

The NACC dataset was split into training, validation and test subsets using a stratified 80/10/10 split to handle class imbalance and with a fixed random seed (random_state = 42) to support reproducibility. L1 and L2 regularization was added to reduce overfitting. Extensive hyperparameter tuning was performed using coarse-to-fine grid search. 55 features were used in total.

2.8. Statistical Analysis and Model Performance Metrics

Both the loss and accuracy were tracked on the training, validation and test sets to monitor model performance and to adjust hyperparameters. Model performance was monitored using multiple metrics such as accuracy, precision, recall, f1 score for each class. SHAP (Shapley Additive Explanations) values were calculated to analyze the most impactful features of the model [20].

doi:10.20944/preprints202605.0817.v1

A large focus was set on reducing true-positive and false-negative errors as this is a critical concern for models when making medical diagnosis [20]. Therefore, the weighted macro-F1 score was chosen as the key evaluation metric for the model.

Additionally, to track the model's performance, the prediction confidence levels were tracked together with the predictions. The ECE (Expected Calibration Error) score was used to analyze whether the model becomes overconfident at certain confidence levels and to make sure that the model is well-calibrated [21].

2.9. Computational Environment and Software

All analyses were performed in the Google Colab environment (runtime 2025.10) using Python 3.12.13. Data handling and modelling were implemented with NumPy (v2.0.2) and pandas (v2.2.2), while preprocessing, splitting procedures, and performance metrics as well as model interpretability analyses were carried out using scikit-learn (v1.6.1). Model training was conducted with XGBoost (v3.2.0).

2.10. Data/Code Availability and Restrictions

NACC data are available only through the official NACC request process and are provided under a Data Use Agreement [22], therefore, the authors cannot share the dataset. The Latvian retrospective dataset is not publicly available due to ethics committee requirements and patient confidentiality restrictions.

The authors intend to secure the code due to ongoing patent protection process. Therefore, the full analysis code is not publicly released at the time of the submission. Methodological details, model development, data variables used, preprocessing as well as processing information are reported in the manuscript to support transparency.

2.11. The Use of Generative Artificial Intelligence

This manuscript was translated from the original Latvian research materials into English with the assistance of ChatGPT (model: GPT-5.2 Thinking) for translation and language editing. Generative AI was not used to assist in study design, data collection, analysis, or interpretation. The research content, methods, results and conclusions are based on the original study materials and the author's work.

3. Results

3.1. XGBoost Model Results on the NACC Dataset

The XGBoost model achieved a test set accuracy of 88.3% with a weighted macro F1 score of 86.6% on the NACC dataset. Table 1 shows that the F1-scores of normal cognition and dementia classes are much higher (95.3% and 92.4% respectively). As mentioned above, the heavy class-imbalance present in the dataset can be noticed in the results as the model completely fails to classify SCD, with an F1 score of 2.2% on 885 samples.

Table 1. Classification Report of the XGBoost model.

Class	Precision, %	Recall, %	F1-score, %	Support
Dementia	93.1	91.7	92.4	5875
MCI	67.9	81.7	74.2	3530
SCD	83.3	1.1	2.2	885
Normal cognition	94.1	96.4	95.3	9846
Accuracy			88.3	20136
Macro avg	84.6	67.7	66.0	20136
Weighted avg	88.7	88.3	86.6	20136

Looking at Table 2, the model very rarely mistakes normal cognition from dementia with a misclassification rate of 0.14% between the two classes which is a strong result. As already visible from the F1-score, the model fails to properly diagnose SCD. In 94.9% of cases, the model classifies SCD as one of its neighboring classes – normal cognition or MCI. This is expected as SCD in practice is difficult to diagnose and additionally, being the minority class in the dataset, adds more challenge to diagnose the SCD class itself [23].

The most misclassifications come from comparing MCI vs dementia with 835 misclassifications. Overall, the results suggest that neighboring class diagnosis are not strictly drawn lines that can be fully separated by the data and that more research and data is required to identify patterns that can truly diagnose each of the 4 states.

Table 2. Confusion Matrix of the XGBoost model.

	pred: Normal cognition	**pred: SCD**	**pred: MCI**	**pred: Dementia**
true: Normal cognition	9495	0	342	9
true: SCD	298	10	542	35
true: MCI	285	2	2885	358
true: Dementia	12	0	477	5386

The model was trained on 55 features derived from the original dataset. Performing a SHAP analysis on the NACC test set shows that CDRSUM (Standard CDR sum), NACCOGF (predominant changes in cognition) and MEMORY (Uniform Data Set to assess memory) are the leading features for diagnosis. These features are strong indicators because they can capture the main memory related cognition changes and the degree as well as severity of how it affects person’s everyday life [24].

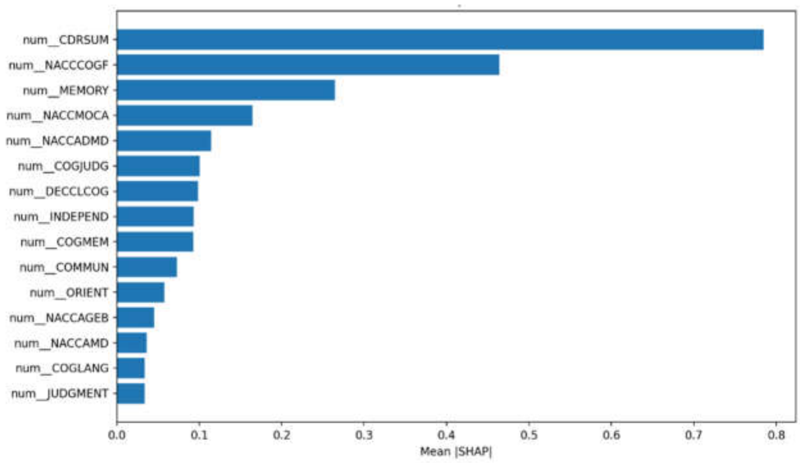

Figure 1. SHAP analysis for the XGBoost model on the NACC test set.

In the case of a medical diagnosis, the model must be certain of its decisions. Therefore, in Figure 2 a visual of the model’s confidence is compared to the actual accuracy of the model on the test set. As can be seen from Figure 2 that the model is near perfectly calibrated. It holds an ECE score of 0.0043 and 68.2% of the test set fall into the two most confident ranges of 90- 95% and 95-100% with accuracies of 91.9% and 98.1% respectively.

doi:10.20944/preprints202605.0817.v1

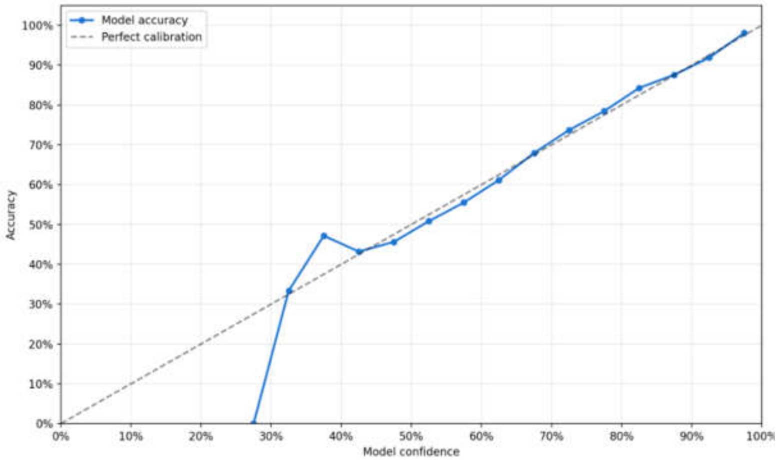

Figure 2. XGBoost model's confidence vs accuracy on the NACC test set.

3.2. SCD Class Detection Using Alternate Approaches

As discussed in 3.1. the model fails to classify the SCD class. To tackle this problem multiple approaches were tested and the results summarized as seen in Table 3. Looking at the per-class F1 scores, in each of the imputation techniques, it was possible to increase the model's ability to predict the SCD class by oversampling the dataset using SMOTE or by adding inverse-weights to make the least-represented classes perform better or by dropping the majority classes so that each class had the same number of rows. The model managed to improve the F-1 Score of SCD in all 3 cases from 2.2% to 22.3%, 28.9% and 51.9% respectively. Important to note, that in all 3 cases this ended up resulting in a worse overall accuracy of the model of 88.3%.

Important to note that by looking at the 'Balanced' test's results from Table 3., the SCD and MCI classes were more difficult to predict than normal cognition or dementia. This is a key finding as it proves that when removing biases in the data, under equal conditions, SCD and MCI are still much more difficult to diagnose - 51,9% and 63.6% respectively, compared to normal cognition 82.9% and dementia 87.8%.

Finally, from Table 3., the "No SCD" column shows that testing out a more radical approach by taking out the SCD class from the model entirely showed great improvements in the model's performance by increasing the test accuracy from 88.28% to 92.3% and the weighted F1 score from 86.6% to 92.3%.

Table 3. Alternative models to tackle the SCD class performance issues.

Overall accuracy	SMOTE	Inverse-Weighted	Balanced	No SCD
Overall Accuracy	87.5%	84.2%	72.3%	92.3%
Weighted F1	86.9%	85.7%	71.5%	92.3%
Per-class F1				
Normal cognition	95.2%	93.8%	82.9%	96.6%
Dementia	91.7%	91.5%	87.8%	92.6%
MCI	72.1%	67.5%	63.6%	79.8%
SCD	22.3%	28.9%	51.9%	—

3.3. XGBoost Model Results on the PSCUH Dataset

After training the model on the NACC dataset, the model was tested using the PSCUH dataset consisting of 154 patients. Class distribution consisted of dementia (n=139); mild cognitive impairment (MCI) (n=13); subjective cognitive decline (SCD) (n=2). The model achieved a 91.56% accuracy on the Latvian dataset making 141 out of 154 correct classifications. No patients with cognitive decline were diagnosed as normal cognition. Dementia was diagnosed correctly in 92% of cases (128/139) and MCI 69.2% of cases (9/13). As expected, SCD was incorrectly diagnosed as dementia in both cases. Of course, one must note that more data is required to properly evaluate the performance of the model on each of the classes, but by looking at all the dementia cases from Table 4, the results suggest that the model is capable of generalizing from the NACC dataset to the PSCUH dataset well.

Table 4. Confusion Matrix of the PSCUH results.

	Pred:Dementia	**Pred:MCI**	**Pred:SCD**
Pred:Dementia	128	11	0
Pred:MCI	4	9	0
Pred:SCD	2	0	0

Analyzing the prediction confidences for each of the 154 patients from Figure 3, when the model was more certain (>82% confidence) on its predictions in the case of dementia, it managed to classify all dementia cases correctly. In the misclassified cases, the confidence was lower and in turn the accuracy dropped significantly. In the case of the other two classes, the model was less certain in its predictions. No prediction for either class was above 90% certainty. As expected, the model failed to classify the two SCD cases. For MCI the model managed to classify 9/13 patients correctly.

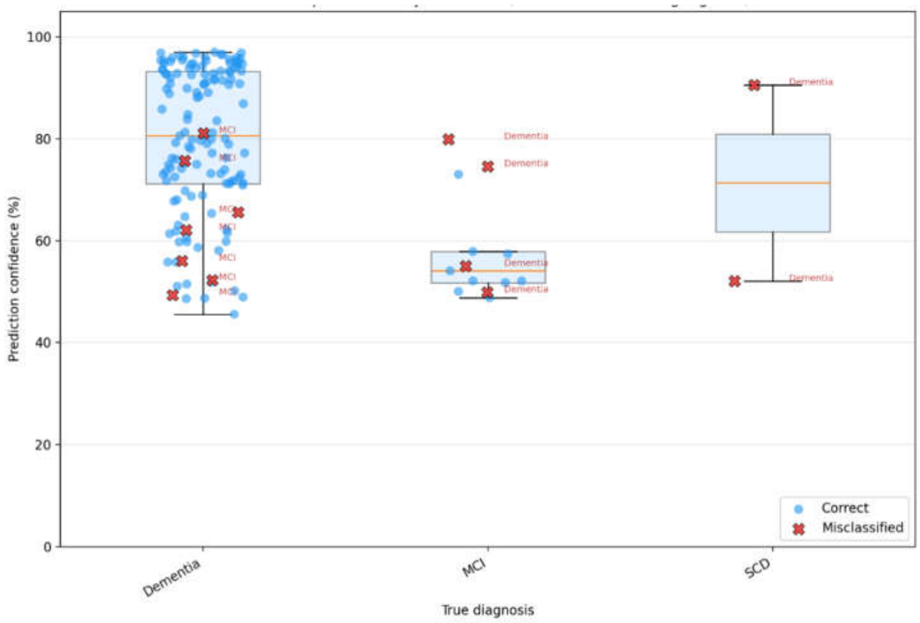

Figure 3. PSCUH dataset true diagnosis vs prediction confidence .

4. Discussion

The aim of the study was to evaluate the ML model algorithm in patients with SCD and increased risk of Alzheimer's disease at an early diagnostic stage and to compare the model's performance with neurologist made clinical diagnosis. The results provide deeper insight into the potential role of ML models in healthcare systems, especially when early pathogenic changes, diagnostic uncertainty, and heterogeneity of cognitive functioning make it difficult for the physician to formulate the diagnosis.

One of the most important findings was the model's overall high diagnostic accuracy, despite the limited clinical information. However, its ability to distinguish between two clinically similar and adjacent diagnosis, such as subjective cognitive decline and mild cognitive impairment, was lower due to exclusion of radiological information, biomarkers and by being the minority class in the dataset, indicating that a larger datasets group needs to be implemented to increase the accuracy of

doi:10.20944/preprints202605.0817.v1

the model when overlapping conditions must be differentiated. The most appropriate role of ML at this current stage of technological development may be as a supportive tool for the physician rather than an independent diagnostic instrument. Such role could be used in outpatient or primary care settings, where patients with subjective or objective cognitive complaints can fill out a structured questionnaire/test before consultation, for example, a Quick Response (QR) code-based test. This could provide the physician and the model with an initial overview of the cognitive status of the patient before the visit, thereby reducing differential diagnosis and diagnostic delay.

When comparing the assessments of both parties, an interesting pattern was observed. The algorithm more often classified patients into more severe cognitive category, suggesting a cautious strategy aimed to reduce the risk of incorrectly classifying findings as normal. In contrast, the neurologist reasoning was based more on the observation of symptom dynamics overt time and the possible progression of the diagnosis. This often delayed the recognition of the diagnosis, thereby the patients transitioned from early cognitive complaints to more pronounced symptoms, such as memory loss and social functioning difficulties.

A particularly meaningful finding relates to the pharmacological treatment and the lack of execution by the patients. Analysis of the results showed that 82 of 154 patients (53.3%) did not use the prescribed medication, including cholinesterase inhibitors, N-methyl-D-aspartate (NMDA) receptor antagonists, antidepressants, and related agents [25], issued by the neurologist in the last visit of informing the patient about the diagnosis. Whereas 72 patients (46.7%) were using at least one of such medication.

The overall distribution of dementia types was also analyzed. The data showed ratio of vascular dementia to Alzheimer's dementia as 5.1:1. Out of 139 dementia diagnoses, 107 (77.0%) were vascular dementia and 21 (15.1%) were Alzheimer's dementia. This does not correspond to World Health Organization data that states the most common cause of dementia as Alzheimer's dementia (60 - 70%), with vascular dementia as the second most frequent type (15 – 20%) [26,27]. This finding should not necessarily be interpreted as a contraindication to WHO data, but rather as a necessity to reevaluate the assessment of Latvian clinical settings of diagnosing dementia.

Several limitations should also be acknowledged. The number of patients and large dataset class-imbalance between diagnostic categories, particularly SCD and MCI, likely contributed to the model's limited ability to distinguish seemingly adjacent diagnoses. The retrospective nature of the research provides a useful overview of how diagnoses are established in practice; however prospective studies could provide more targeted questions and obtain specific information to improve the ML model's accuracy as well as to reduce the proportion of missing values. Another limitation is the "black box" limitation for it is difficult for the researchers to fully interpret and represent the model's internal reasoning strategy and decision-making logic in a completely transparent way.

5. Conclusions

This research compared the performance of an ML algorithm with neurologist made assessments in the early evaluation of patients with subjective cognitive decline and risk of Alzheimer's disease using only subjective and objective clinical data, excluding the utilization of biomarkers or neuroimaging. The results suggest that the aims and objectives of this research were successfully accomplished. The ML model's algorithm demonstrated the potential clinical embracement of such technology as a supportive tool, rather than standalone diagnostic system, in cognitive assessment and diagnostic decision-making, especially when used in combination with physicians. The SHAP analysis indicated that the primary driven factors for successful diagnostic predictions were driven by cognitive test results and other clinically relevant factors, supporting the model's capability of interpreting clinical data.

However, the research also showed that the model performance decreased when trying to distinguish clinically similar and adjacent diagnoses, highlighting the importance of continuous physician assessment of ML based predictions. In addition, the results revealed unpleasant findings of dementia patients selfcare with medications as well as the notable differences of dementia type distributions compared with international patterns, suggesting the need for reevaluation of diagnostic and patient informative practices in Latvia. Overall, the ML seems to be a promising contributing tool for early cognitive assessment in clinical settings, but further prospective research needs to confirm its safety, efficiency and practical value in real time clinical practice.

6. Patents

The authors are undergoing through a patent protection process.

Author Contributions: Conceptualization, D.K. and K.K.V..; methodology, D.K.; software, D.K., K.K.V..; validation, Z.P., D.K. and K.K.V.; formal analysis, D.K.; investigation, D.K; resources, D.K; data curation, D.K. and K.K.V.; writing—original draft preparation, D.K.; writing—review and editing, D.K. and K.K.V.; visualization, D.K. and K.K.V.; supervision, Z.P. and A.M.; project administration, Z.P and A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Riga Stradins University Ethics Committee (No. 2-PĒK-4/952/2025, 24.09.2025.) and authorized by the PSCUH Scientific Institute. Patient confidentiality and privacy were maintained according to institutional requirements, and analyses were performed using encrypted data.

Informed Consent Statement: Not applicable.

Data Availability Statement: NACC data are available only through the official NACC request process and are provided under a Data Use Agreement, therefore, the authors cannot share the dataset. The Latvian retrospective dataset is not publicly available due to ethics committee requirements and patient confidentiality restrictions.

Acknowledgments: NACC data are contributed by the NIA-funded ADRCs: P30 AG062429 (PI James Brewer, MD, PhD), P30 AG066468 (PI Oscar Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30 AG066530 (PI Helena Chui, MD), P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI David Holtzman, MD), P30 AG066518 (PI Lisa Silbert, MD, MCR), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD), P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI Julie A. Schneider, MD, MS), P30 AG072978 (PI Ann McKee, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI Frank LaFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Jessica Langbaum, PhD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD, PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD, FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Glenn Smith, PhD, ABPP), P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD, MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30 AG066546 (PI Sudha Seshadri, MD), P30 AG086401 (PI Erik Roberson, MD, PhD), P30 AG086404 (PI Gary Rosenberg, MD), P20 AG068082 (PI Angela Jefferson, PhD), P30 AG072958 (PI Heather Whitson, MD), P30 AG072959 (PI James Leverenz, MD). This manuscript was translated from the original Latvian research materials into English with the assistance of ChatGPT (model: GPT-5.2 Thinking) for translation and language editing. Generative AI was not used to assist in study design, data collection, analysis, or interpretation. The research content, methods, results and conclusions are based on the original study materials and the author's work. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
AD	Alzheimer's disease
ADRCs	Alzheimer's Disease Research Centers
CDR	Clinical dementia rating
ECR	Expected Calibration Error
ICD	Internation Classification of Diseases
MCI	Mild cognitive impairment
MMSE	Mini-mental state examination
ML	Machine learning
MoCA	Montreal cognitive assessment
NACC	National Alzheimer's Coordinating Center
NIA	National Institute of Aging
NMDA	N-methyl-D-aspartate
NaN	Not a number
PSCUH	Pauls Stradins Clinical University Hospital
QR	Quick response
SCD	Subjective cognitive decline
SHAP	Shapley Additive Explanations
SMOTE	Synthetic Minority Over-sampling Technique
WHO	The World Health Organization

References

1. Dhakal, A.; Bobrin, B.D. Cognitive Deficits. In *StatPearls*; StatPearls Publishing: Treasure Island, FL, USA, 2023. Available online: https://www.ncbi.nlm.nih.gov/books/NBK559052/ (accessed on 20 August 2025).
2. World Health Organization. Ageing and Health. Available online: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health (accessed on 15 January 2026).
3. AMBOSS. Alzheimer Disease. Available online: https://next.amboss.com/us/article/D301kf?q=alzheimer+disease (accessed on 11 January 2026).
4. Tiwari, S.; Atluri, V.; Kaushik, A.; Yndart, A.; Nair, M. Alzheimer's disease: Pathogenesis, diagnostics, and therapeutics. *Int. J. Nanomed.* 2019, *14*, 5541–5554. https://doi.org/10.2147/IJN.S200490.
5. Chen, W.; Hu, H.; Ou, Y.-N.; et al. Risk factors for subjective cognitive decline: The CABLE study. *Transl. Psychiatry.* 2021, *11*, 576. https://doi.org/10.1038/s41398-021-01678-z.
6. Jessen, F.; Amariglio, R.E.; van Boxtel, M.; et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimers Dement.* 2014, *10*, 844–852. https://doi.org/10.1016/j.jalz.2014.01.001.
7. Ranson, J.M.; Kuźma, E.; Hamilton, W.; Muniz-Terrera, G.; Licher, S.; Llewellyn, D.J. Estimating prevalence of subjective cognitive decline in and across international cohort studies of aging: A COSMIC study. *Alzheimers Res. Ther.* 2020, *12*, 167. https://doi.org/10.1186/s13195-020-00734-y.
8. Mazzeo, S.; Padiglioni, S.; Bagnoli, S.; et al. Assessing the effectiveness of subjective cognitive decline plus criteria in predicting the progression to Alzheimer's disease: An 11-year follow-up study. *Eur. J. Neurol.* 2020, *27*, 261–268. https://pubmed.ncbi.nlm.nih.gov/32043740/.
9. Kumar, Y.; Koul, A.; Singla, R.; Ijaz, M.F. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. J. Ambient Intell. Humaniz. Comput. **2023**, 14, 8459–8486. https://doi.org/10.1007/s12652-021-03612-z.
10. Kale, M.; Wankhede, N.; Pawar, R.; et al. AI-driven innovations in Alzheimer's disease: Integrating early diagnosis, personalized treatment, and prognostic modelling. *Ageing Res. Rev.* 2024, *98*, 102437. https://pubmed.ncbi.nlm.nih.gov/39293530/.

11. Ahmed, M.I.; Spooner, B.; Isherwood, J.; Lane, M.; Orrock, E.; Dennison, A. A systematic review of the barriers to the implementation of artificial intelligence in healthcare. Cureus 2023, 15, e46454. https://doi.org/10.7759/cureus.46454.
12. National Alzheimer's Coordinating Center. About NACC Data. Available online: https://www.naccdata.org/about-nacc-data/ (accessed on 9 July 2025).
13. Dolcet-Negre, M.M.; Imaz Aguayo, L.; García-de-Eulate, R.; et al. Predicting conversion from subjective cognitive decline to mild cognitive impairment and Alzheimer's disease dementia using ensemble machine learning. *J. Alzheimers Dis.* 2023, *92*, 689–702. doi: 10.3233/JAD-221002. PMID: 36938735.
14. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, **2002**, 16, 321–357. https://doi.org/10.1613/jair.953
15. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E.. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, **2011**, 12, 2825–2830.
16. Rubin, D. B. Inference and missing data. *Biometrika*, **1976**, 63(3), 581–592. https://doi.org/10.1093/biomet/63.3.581
17. Chen, T., & Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* ,**2016**, (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785
18. Shwartz-Ziv, R., & Armon, A. Tabular data: Deep learning is not all you need. *Information Fusion*, **2022**, 81, 84–90. https://doi.org/10.1016/j.inffus.2021.11.011
19. Grinsztajn, L., Oyallon, E., & Varoquaux, G. Why do tree-based models still outperform deep learning on typical tabular data? In *Advances in Neural Information Processing Systems*, **2022**, (NeurIPS 2022, Datasets and Benchmarks Track).
20. Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports*, 12, 5979. https://doi.org/10.1038/s41598-022-09954-8
21. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, **2022**, (ICML 2017), PMLR 70, 1321–1330. https://proceedings.mlr.press/v70/guo17a.html
22. National Alzheimer's Coordinating Center. NACC Data Use Agreement. Available online: https://files.alz.washington.edu/documentation/nacc_data_use_agreement.pdf.
23. Webster-Cordero, F.; Giménez-Llort, L. The challenge of subjective cognitive complaints and executive functions in middle-aged adults as a preclinical stage of dementia: A systematic review. *Geriatrics* 2022, 7, 30. https://doi.org/10.3390/geriatrics7020030
24. Snitz, B.E.; Wang, T.; Cloonan, Y.K.; Jacobsen, E.; Chang, C.H.; Hughes, T.F.; Kamboh, M.I.; Ganguli, M. Risk of progression from subjective cognitive decline to mild cognitive impairment: The role of study setting. *Alzheimers Dement.* 2018, 14, 734–742. https://doi.org/10.1016/j.jalz.2017.12.003.
25. Wu, C.K.; Fuh, J.L. A 2025 update on treatment strategies for the Alzheimer's disease spectrum. *J. Chin. Med. Assoc.* 2025, 88, 495–502. https://doi.org/10.1097/JCMA.0000000000001252.
26. Alzheimer Europe. Vascular Dementia. Available online: https://www.alzheimer-europe.org/dementia/other-dementias/vascular-dementia?language_content_entity=en (accessed on 18 December 2025).
27. World Health Organization. Dementia. Available online: https://www.who.int/news-room/fact-sheets/detail/dementia (accessed on 18 December 2025).