General Psychiatry

Explainable and externally validated machine learning for neurocognitive diagnosis via ECGs

Juan Miguel Lopez Alcaraz ,[1] Ebenezer Oloyede,[2,3] David Taylor,[2,4] Wilhelm Haverkamp,[5] Nils Strodthoff[1]

To cite: Lopez Alcaraz JM, Oloyede E, Taylor D, *et al*. Explainable and externally validated machine learning for neurocognitive diagnosis via ECGs. *General Psychiatry* 2025;**38**:e102107. doi:10.1136/gpsych-2025-102107

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/gpsych-2025-102107).

Received 04 March 2025
Accepted 14 September 2025

[1]AI4Health Division, Carl von Ossietzky Universitat Oldenburg, Oldenburg, Germany
[2]Pharmacy Department, Maudsley Hospital, London, UK
[3]Department of Psychiatry, University of Oxford, Oxford, UK
[4]Institute of Pharmaceutical Sciences, King's College London, London, UK
[5]Department of Cardiology, German Heart Center of the Charité, Berlin, Germany

Correspondence to
Professor Nils Strodthoff; nils.strodthoff@uol.de

ABSTRACT

Background Electrocardiogram (ECG) analysis has emerged as a promising tool for detecting physiological changes linked to non-cardiac disorders. Given the close connection between cardiovascular and neurocognitive health, ECG abnormalities may be present in individuals with co-occurring neurocognitive conditions. This highlights the potential of ECG as a biomarker to improve detection, therapy monitoring and risk stratification in patients with neurocognitive disorders, an area that remains underexplored.

Aims We aimed to demonstrate the feasibility of predicting neurocognitive disorders from ECG features across diverse patient populations.

Methods ECG features and demographic data were used to predict neurocognitive disorders, as defined by the International Classification of Diseases 10th revision, focusing on dementia, delirium and Parkinson's disease. Internal and external validations were performed using the Medical Information Mart for Intensive Care IV and ECG-View datasets. Predictive performance was assessed by the area under the receiver operating characteristic curve (AUROC) scores, and Shapley values were used to interpret feature contributions.

Results Significant predictive performance was observed for several neurocognitive disorders. The highest predictive performance was observed for F03: dementia, with an internal AUROC of 0.848 (95% confidence interval (CI) 0.848 to 0.848) and an external AUROC of 0.865 (95% CI 0.864 to 0.965), followed by G30: Alzheimer's disease, with an internal AUROC of 0.809 (95% CI 0.808 to 0.810) and an external AUROC of 0.863 (95% CI 0.863 to 0.864). Feature importance analysis revealed both established and novel ECG correlates.

Conclusions These findings suggest that ECG holds promise as a non-invasive, explainable biomarker for selected neurocognitive disorders. This study demonstrates robust performance across cohorts and lays the groundwork for future clinical applications, including early detection and personalised monitoring.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Electrocardiogram (ECG) abnormalities have been explored as potential biomarkers for neurocognitive disorders, but their diagnostic value has not been thoroughly investigated. Recent studies have only begun to examine these abnormalities in exploratory analyses to identify potential associations.

WHAT THIS STUDY ADDS

⇒ This study demonstrates that ECG features can effectively predict neurocognitive disorders, extending beyond their traditional use as biomarkers for cardiac conditions. The findings show high predictive performance across both internal and external datasets. Additionally, the use of an explainable approach allows for a deeper understanding of the associations between specific ECG features and neurocognitive disorders, confirming existing knowledge while also providing new insights. This research sheds light on the physiological changes that may underlie the interplay between cardiac features and neurocognitive disorders.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This work lays the foundation for integrating ECG into the diagnosis and monitoring of neurocognitive disorders. It presents a non-invasive, cost-effective tool that could aid as a companion tool in more efficient detection and personalised therapy management in clinical settings, with the potential to transform both research and clinical practices.

INTRODUCTION

Clinical relevance

Neurocognitive disorders are among the most challenging diseases, disrupting the lives of those affected and placing a considerable burden on caregivers and healthcare systems.[1] Conditions such as Alzheimer's disease[2] and Parkinson's disease cause progressive cognitive and motor decline, severely limiting patients' independence and quality of life.[3] Other disorders like symptomatic dementia and delirium further impair mental health, often causing confusion, memory loss and significant behavioural changes.[4] Early diagnosis is crucial to effectively manage these diseases. However, current diagnostic methods can be complex, costly and sometimes inaccessible.[5] Differential diagnosis is further complicated by the

fact that these disorders are closely inter-related, as they share common neurobiological pathways, genetic factors and overlapping cognitive, affective and behavioural symptoms. These overlaps also challenge patient management, highlighting the need for integrated, multidisciplinary approaches in neurocognitive care.[6]

Emerging role of Electrocardiogram (ECG) beyond cardiology

ECG is well-established as a critical tool in cardiology, used primarily to detect and monitor cardiac disorders by measuring the electrical activity of the heart. ECGs provide vital information about heart rhythms, helping to identify issues such as arrhythmias, ischaemia and other cardiac abnormalities.[7] Recent advancements in data analysis and machine learning have enabled novel uses of ECGs in predictive modelling applications across fields beyond cardiology, such as estimating laboratory values[8] and predicting patient deterioration in emergency departments.[9]

Emerging role of neurocardiology

The potential link between ECG abnormalities and neurocognitive disorders is an emerging area of study. Growing evidence highlights the complex interplay between cardiovascular and mental health, underscoring the importance of early detection and integrated care strategies. For example, depression is often accompanied by cardiovascular manifestations, and treatment responses can differ depending on the underlying physiological patterns.[10] In parallel, substantial gaps persist between clinical guidelines and practice in the monitoring of cardiometabolic risk among patients treated with antipsychotics, despite cardiovascular disease being a leading cause of mortality in this population.[11] Novel approaches such as virtual reality-based psychological interventions have also demonstrated benefits in reducing anxiety following acute myocardial infarction, illustrating how digital solutions can complement cardiovascular care.[12] Furthermore, neurodegenerative disorders, particularly Parkinson's disease, can impair autonomic regulation,[13] leading to measurable alterations in heart rate variability and ECG profiles. Relatedly, psychiatric syndromes such as dementia and delirium have been linked with cardiac dysfunction,[14] suggesting that disturbances in neural and cardiovascular systems may share common physiological pathways.

Current diagnostic practices

To date, few studies have explored the diagnostic utility of ECG for neurocognitive disorders. A retrospective study linked specific ECG features to incident dementia and modestly improved risk prediction, but it lacked machine learning to model complex patterns and did not validate findings across independent cohorts, limiting generalisability.[15] Another study showed that ECG-derived heart rate variability could detect autonomic dysfunction in patients with rapid eye movement sleep behaviour disorder, an early sign of Parkinson's disease, but it was limited by its small sample size, lacked broader validation and did not confirm progression to Parkinson's.[16] Similarly, a recent study used machine learning to predict neurological outcomes after cardiac arrest; however, it relied on a single dataset, required EEG for strong performance since ECG-only models underperformed and used synthetic data that may not reflect real-world populations.[17]

Contributions

Based on these observations, this study investigates whether machine learning models trained on ECG data and basic demographic information can accurately identify specific neurocognitive disorders, including dementia (Alzheimer's disease and unspecified dementia), delirium and Parkinson's disease. We developed and validated models capable of distinguishing these conditions, demonstrating strong performance across both internal and external datasets. In addition, we provided transparent and explainable insights into the ECG features contributing to each diagnostic category. Our findings highlight the potential of ECG as a clinically useful, low-cost and non-invasive diagnostic support tool in neurocognitive care, either as a standalone method or, more conservatively, as a complement to existing diagnostic practices. To our knowledge, this is the first study to apply explainable machine learning to ECG data for diagnosing these disorders, with validation across diverse clinical populations confirming its applicability.

METHODS

Dataset

The Medical Information Mart for Intensive Care IV-ECG (MIMIC-IV-ECG) dataset served as our internal dataset, providing demographic and ECG features used as input for training a tree-based model to diagnose various neurocognitive disorders. For external validation, we used a second patient cohort from the ECG-View II dataset, extracting the same set of features and neurocognitive targets. The disorders are defined based on the International Classification of Diseases, 10th revision, Clinical Modification (ICD-10-CM) codes.

Figure 1 shows a schematic representation of our proposed methodology. The primary dataset for training and internal evaluation was obtained from the MIMIC-IV-ECG database,[18 19] a subset of a large-scale critical care dataset collected at the Beth Israel Deaconess Medical Center (BIDMC) in Boston, Massachusetts, USA. It includes data from patients admitted to the emergency department and intensive care unit. We included patients aged 18 years and older with at

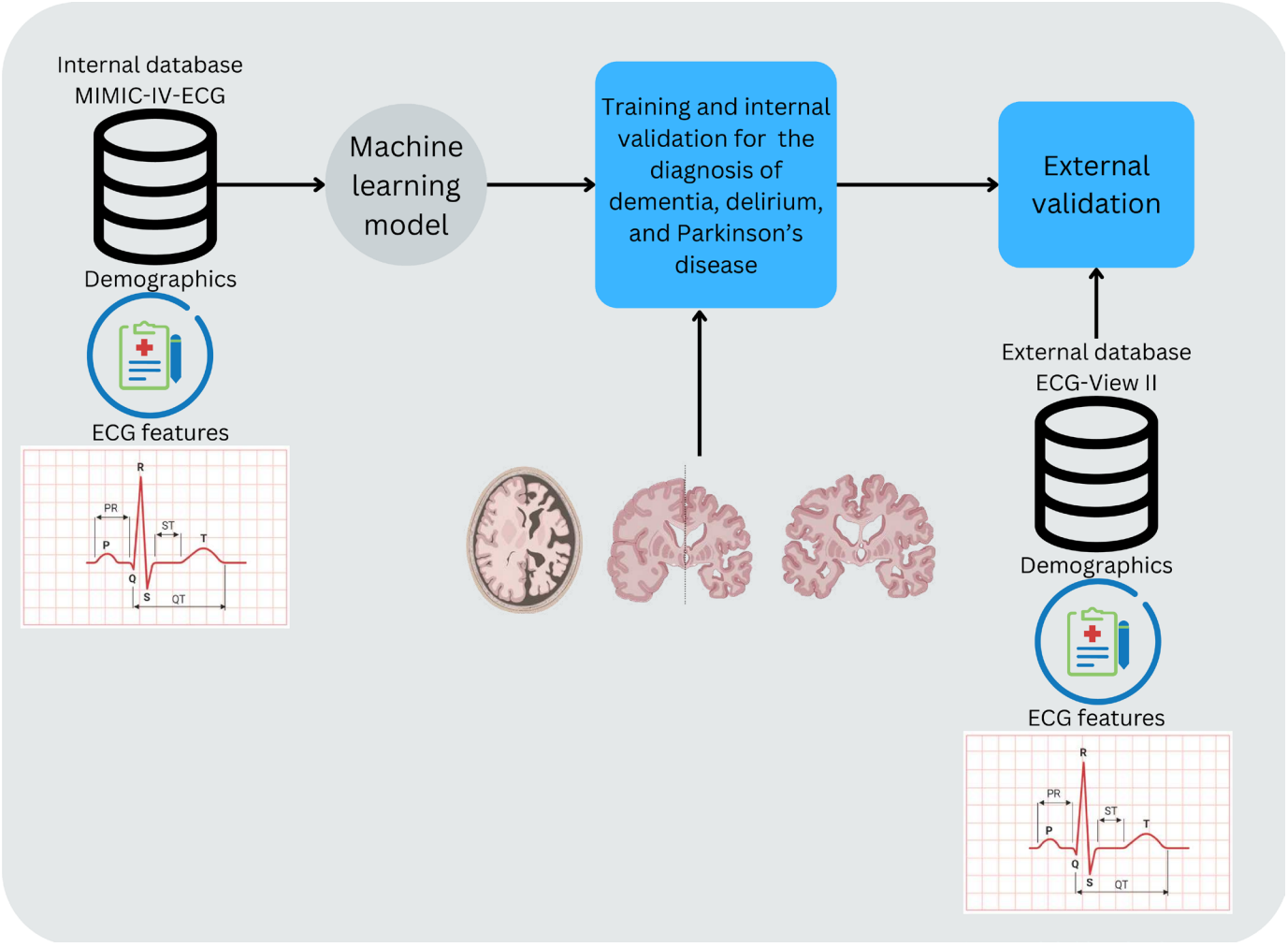

Figure 1 Diagrammatic illustration of our proposed methodology. ECG, electrocardiogram; MIMIC-IV-ECG, Medical Information Mart for Intensive Care IV-ECG.

least one ECG record and corresponding discharge diagnoses recorded using ICD-10-CM codes between 2008 and 2019. MIMIC-IV-ECG is a publicly available, de-identified dataset and thus individual patient consent was not required. The use of this dataset was approved by the institutional review boards of the Massachusetts Institute of Technology and BIDMC through Physionet.

External validation was performed using the ECG-VIEW-II database,[20] which includes data from patients at a South Korean tertiary teaching hospital. To create a comprehensive and harmonised feature set, ECG features from the MIMIC-IV dataset were aligned with those from the ECG-VIEW-II database, where the standardised feature set consists of ECG-derived measurements such as RR interval (time between two consecutive R-waves), PR interval (time from the onset of the P-wave to the start of the QRS complex), QRS duration (time for ventricular depolarisation), QT interval (time from the start of the Q-wave to the end of the T-wave), corrected QT interval (QTc) in milliseconds, as well as P-wave axis, QRS axis, and T-wave axis in degrees, along with demographic attributes (binary sex and age as a continuous variable) which served as our secondary dataset for external validation. Inclusion criteria for ECG-VIEW-II followed the same logic: adult patients (18+) with at least one ECG record and corresponding ICD-10-CM diagnosis between 1994 and 2015. ECG-VIEW-II is also a publicly available, de-identified dataset released under institutional review board approval and thus patient consent was not applicable. Target variables are based on discharge diagnoses encoded using the ICD-10-CM.[21] We primarily investigated the following disorders: G30: Alzheimer's disease, G20: Parkinson's disease, F01: vascular dementia, F03: dementia and F05: delirium (physiological).

For the internal dataset, stratified folds were created based on diagnoses, age and gender distributions, using an 18:1:1 split as described in previous research.[22] The external evaluation was performed on the complete cohort. The training process emphasised MIMIC-IV-ECG due to its broader ethnic diversity compared with ECG-VIEW-II, which improved the model's generalisation across various populations, as shown in prior studies on cardiac disorders,[23] liver

disorders[24] and neoplasms.[25] This strategy ensures robust internal training and reliable external validation across ethnically and geographically diverse cohorts.

Models and evaluation

In this study, we developed individual tree-based models using Extreme Gradient Boosting (XGBoost) to tackle binary classification tasks, with a distinct model for each selected ICD-10-CM code. To avoid overfitting, we applied early stopping with a patience of 10 iterations on the validation fold during training. No imputation was performed, as XGBoost inherently handles missing data. Model performance was assessed using the area under the receiver operating characteristic curve (AUROC) on the validation fold. We report AUROC scores for both the internal test set and the external dataset, along with 95% confidence intervals (CIs) calculated through empirical bootstrapping with 1000 iterations. Model calibration was assessed using calibration curves on the internal test set. We applied a model-agnostic calibration approach by fitting isotonic regression models on the validation set and reporting the calibrated results on the test set. Additionally, we evaluated clinical utility using decision curve analysis, comparing our model's net benefit to standard strategies such as 'refer all' and 'refer none'.[26]

Explainability

Our objective goes beyond merely assessing model performance. To gain deeper insights into the trained models, we integrated SHapley Additive exPlanations (SHAP) values into our workflow.[27] These values provide a method for evaluating feature importance by measuring the individual contribution of each feature to the model's predictions.

Study protocol and reporting standards

This study follows a comprehensive protocol aligned with established guidelines, including Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis, Standards for Reporting Diagnostic Accuracy Studies and Minimum Information About Clinical Artificial Intelligence Modelling. Detailed documentation is provided in the online supplemental material.

Statistical analysis

Analyses were conducted in Python (V.3.10) with XGBoost (V.1.7.4) and scikit-learn (V.1.1.3). Model performance was evaluated using AUROC with 95% CIs from 1000 bootstrap iterations. Model training included an early stopping strategy of 10 iterations on the validation fold, while calibration was assessed via isotonic regression and visualised with calibration curves. Decision curve analysis was applied to quantify clinical utility, and no imputation was required given the model's handling of missing values.

Table 1 Descriptive statistics of the two ECG datasets, MIMIC-IV-ECG and ECG-View II

Variable	MIMIC-IV-ECG	ECG-View II
Gender (%)		
Female samples	226 892 (48.509)	375 733 (48.448)
Male samples	240 837 (51.491)	399 802 (51.552)
Age (%)		
Median years (IQR)	66 (25)	52 (25)
Quantile 1 age range (%)	18–53 (23.830)	18–40 (24.030)
Quantile 2 age range (%)	53–66 (25.160)	40–52 (25.750)
Quantile 3 age range (%)	66–78 (25.600)	52–65 (24.940)
Quantile 4 age range (%)	78–101 (25.400)	65–109 (25.280)
ECG features median (IQR)		
RR interval (ms)	769 (264)	857 (227)
PR interval (ms)	158 (38)	158 (28)
QRS duration (ms)	94 (23)	90 (14)
QT interval (ms)	394 (68)	392 (48)
QTc interval (ms)	447 (47)	421 (37)
P-wave axis (°)	51 (32)	53 (28)
QRS axis (°)	13 (61)	48 (49)
T-wave axis (°)	42 (58)	44 (33)

The table summarises demographic variables: gender (absolute sample count and percentage) and age (median and IQR), along with age group distributions by quantiles. ECG features are reported as median (IQR), with temporal intervals (RR interval, PR interval, QRS duration, QT interval, QTc interval) given in milliseconds (ms) and electrical axes (P-wave, QRS axis, T-wave) in degrees (°).
ECG, electrocardiogram; IQR, interquartile range; MIMIC-IV-ECG, Medical Information Mart for Intensive Care IV-ECG; ms, milliseconds; QTc, corrected QT interval.

RESULTS

Descriptive statistics

Table 1 highlights key demographic and ECG feature differences between MIMIC-IV-ECG and ECG-View II. Both datasets have a nearly identical gender distribution, with slightly more male than female samples. However, MIMIC-IV-ECG includes an older population (median age: 66 vs 52 years), which may influence ECG characteristics. Notable differences include a shorter RR interval in MIMIC-IV-ECG (769 ms vs 857 ms) and a higher QTc interval (447 ms vs 421 ms), suggesting potential variations in heart rate and repolarisation patterns. Additionally, the QRS axis shows a substantial shift (13° vs 48°), which may reflect the differences in patient populations. A detailed comparison of ECG features between diagnostic-positive and diagnostic-negative groups across the MIMIC-IV and ECG-VIEW II datasets is presented in the online

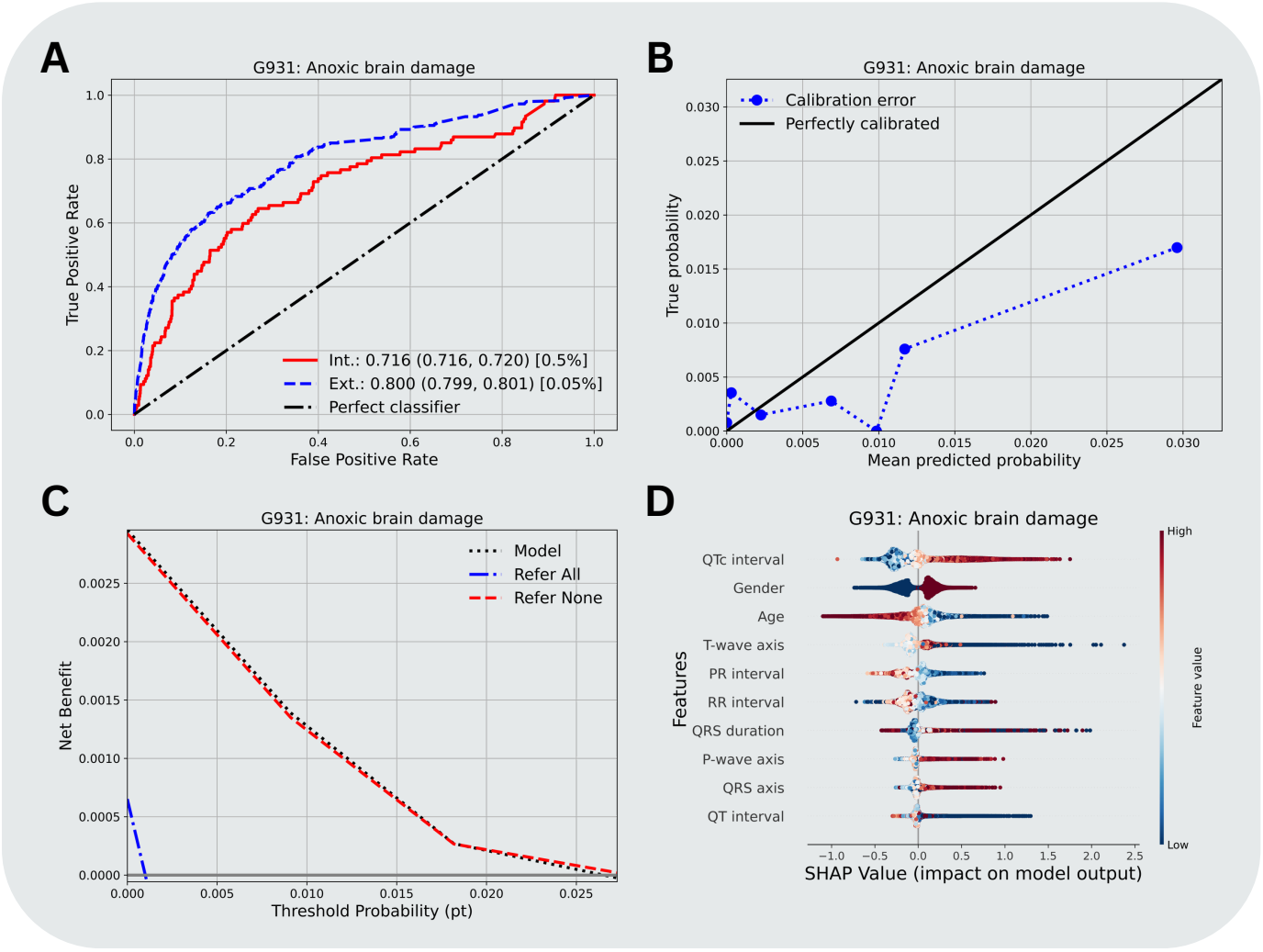

Figure 2 Model evaluation results across five neurocognitive disorders. (A) Discriminative performance (area under the receiver operating characteristic curve with 95% CI); (B) Calibration plots; (C) Net benefit decision curves and (D) SHAP-based feature importance. Rows correspond to the following International Classification of Diseases 10th revision codes: G30: Alzheimer's disease, G20: Parkinson's disease, F01: vascular dementia, F03: unspecified dementia and F05: delirium due to physiological condition. CI, confidence interval; Ext, external; Int, internal; QTc, corrected QT interval; SHAP, SHapley Additive exPlanations.

supplemental table 3. The analysis reveals modest yet consistent feature differences, suggesting potential disease-related alterations in ECG signals in an exploratory manner.

Discriminative, calibration, decision and interpretability analysis

Evaluation of the predictive performance of artificial intelligence algorithms requires assessment covering multiple domains, including discrimination and clinical utility.[28] We assessed discrimination in terms of receiver operating curves/AUROC scores, calibration through calibration plots and clinical utility through decision curve analysis/net benefit.[26] This is complemented through an explainability analysis by means of SHAP values. We showcase each of the four categories as columns in figure 2.

Figure 2A presents the predictive performance of our model for various neurocognitive disorders, evaluated using AUROC scores on both internal (MIMIC-IV) and external (ECG-View) test sets. The 95% prediction intervals offer insight into the reliability of these scores. Additionally, the class prevalence for each condition within its respective dataset is provided, offering context on the population distribution and demonstrating predictive performance robustness due to their prevalence differences. In the MIMIC-IV cohort, prevalences range from 0.50% to 3.41%, while the ECG-View cohort exhibits much lower prevalences, from 0.05% to 0.21%. Among the disorders evaluated, F03: dementia exhibited the highest predictive performance, with an internal AUROC of 0.848 (95% CI 0.848 to 0.848) and an external AUROC of 0.865 (95% CI 0.864 to 0.865), followed by G30: Alzheimer's disease, with an internal AUROC of 0.811 (95% CI 0.811 to 0.811) and an external AUROC of 0.863 (95% CI 0.863 to 0.864). These findings highlight the reliability of our model

in predicting neurocognitive disorders, despite differences in prevalence across datasets.

Figure 2B illustrates the calibration of the model predictions for each condition, comparing predicted probabilities with observed outcomes. Across all conditions, our models demonstrate good calibration, indicating that the predicted probabilities are consistent with actual event rates. This is particularly relevant in clinical applications, where overestimation or underestimation of risk could have critical consequences. The reliability of these probability estimates further supports the model's utility in risk stratification and decision support.

Figure 2C presents decision curve analysespresents decision curve analyses, assessing the clinical utility of our models. For all examined disorders, the models provide a higher net benefit across a range of threshold probabilities compared with the 'refer-all' and 'refer-none' strategies. This suggests that using our models to guide referral decisions could lead to better patient outcomes while minimising unnecessary evaluations. The consistent superiority of our models in net benefit reinforces their practical value in aiding early detection and triage for neurocognitive disorders.

Figure 2D represents the result of the interpretability analysis. Across all disorders investigated, age is the most important predictive feature, with increasing age contributing positively. For Alzheimer's disease, decreased values of QRS duration and PR interval contribute positively the most. For Parkinson's disease, decreased values of QTc interval and T-wave axis contribute positively the most. For vascular dementia, decreased values of QRS axis and QRS duration contribute positively the most. For dementia, decreased values of RR interval as well as increased values of PR interval contribute positively the most. Finally, for delirium, decreased values of RR interval as well as increased values of QTc interval contribute positively the most.

We also investigated ICD-10 codes with strong internal discriminative performance on the MIMIC-IV-ECG dataset that could not be externally evaluated on the ECG-View II. Some of the high-performing categories include eating disorders such as F502: bulimia nervosa (AUROC=0.953) and F509: eating disorder, unspecified (0.916), personality disorders such as F602: antisocial personality disorder (0.952) and conduct disorders such as F91: conduct disorders (0.878). Additionally, several substance-related disorders showed strong internal performance, including F1510: other stimulant abuse, uncomplicated (0.903), F1220: cannabis dependence, uncomplicated (0.891), F1310: sedative, hypnotic or anxiolytic abuse, uncomplicated (0.884), F1910: other psychoactive substance abuse, uncomplicated (0.895) and F1012: alcohol abuse with intoxication (0.873). The full list of these codes is provided in the online supplemental table1, including eight codes with AUROCs above 0.9 and 40 codes between 0.8 and 0.9, representing promising directions for future investigation. Additionally, online supplemental table 2 lists conditions that showed good performance on the internal test set but failed to generalise externally. For instance, G931: anoxic brain damage was excluded from the final analysis despite strong internal and external AUROC values due to poor calibration and limited clinical utility (see online supplemental figure 1).

DISCUSSION

Main findings

Detecting neurocognitive disorders through ECG features may initially seem unconventional, as ECG is traditionally used for cardiovascular diagnoses. However, the physiological interplay between the cardiovascular and neurocognitive systems offers a unique opportunity for diagnostic innovation. While the mechanisms linking neurocognitive disorders to ECG abnormalities remain incompletely understood, some known factors include autonomic nervous system dysfunction, neurocognitive changes, vascular alterations, electrophysiological disruptions and hormonal or metabolic factors. For example, Alzheimer's disease, Parkinson's disease and brain damage share common ECG abnormalities, primarily reflecting autonomic dysfunction and brain-heart communication.

Across the disorders investigated, reduced heart rate variability, bradycardia due to parasympathetic dominance and arrhythmias like atrial fibrillation[29] are frequently observed. Medication effects, such as those from cholinesterase inhibitors in Alzheimer's or dopamine agonists in Parkinson's, can further influence conduction and structural heart changes.[30] Despite these insights, the specific relationship to certain abnormalities remains unclear. Our findings identified distinctive ECG patterns for these disorders using machine learning. Beyond diagnosis, ECG features hold promise for guiding therapy management by monitoring treatment responses, risk stratification for medication-related adverse events, early intervention to identify stress or relapse and longitudinal monitoring for chronic disorders. This interdisciplinary approach bridges neurology, psychiatry and cardiology, advancing novel diagnostic and therapeutic strategies.

The predictive power of specific ECG features highlights their ability to reliably identify neurocognitive disorders from a single ECG recording. Consistently high AUROC scores in both internal and external validations demonstrate the robustness of these features across diverse cohorts. The distinct patterns observed across physiological systems underscore the intricate connection between cardiac and neurocognitive health. In this study, age was identified as a key factor, with older patients contributing more to most disorders. This is consistent with well-documented evidence of brain and structural degeneration associated with ageing. For example, previous research has reported

abnormal electroencephalograms in patients over 45 years old with disorders such as senile and arteriosclerotic psychosis, involutional psychosis, psychosis with mental deficiency, manic-depressive states, psychoneurosis and schizophrenia.[31]

Previous studies support our findings regarding ECG features in neurocognitive disorders. For example, a previous study has demonstrated a link between low QRS duration, low PR interval and abnormally high P-wave axes with increased Alzheimer's disease risk.[32] Also, a meta-analysis reported a clear preponderance of higher incidence and prevalence of Parkinson's disease in males,[33] which also aligns with our cohort observations. Similarly, a study including 981 patients demonstrated that high QT interval values are associated with Parkinson's disease, which again matches our findings. However, the authors also reported high QTc interval values in their cohort.[34] In contrast, our analysis, based on nearly 4000 samples, suggests that while the raw QT interval is elevated in patients with Parkinson's disease, the QTc interval, when corrected for heart rate, appears to be lower. This discrepancy may be due to differences in sample size, patient characteristics or the methodologies used to correct the QT interval, as well as the larger heterogeneity within our cohort. Left QRS axis deviation has been associated with all-cause neurocognitive conditions,[35] which aligns with low or negative values found in our study. Shorter QRS duration has been associated with non-Alzheimer dementia,[15] and increased QTc intervals have been reported in vascular dementia populations,[36] both of which match our findings. Finally, elevated QTc intervals in patients with delirium are consistent with prior reports in delirium tremens, where tachyarrhythmias resolved on appropriate treatment.[37]

Certain psychiatric and behavioural diagnoses, including eating disorders, substance use disorders and personality disorders, showed high predictive internal performance that cannot be externally validated due to missing labels in the external dataset. This raises important clinical considerations about the physiological manifestations of these conditions. Traditionally, the diagnosis of these disorders relies on behavioural assessments, structured interviews and, in cases of substance use, confirmatory tests like toxicology screenings. However, the ability of ECG-based models to predict these diagnoses suggests that these conditions may also leave detectable and specific cardiophysiological footprints, possibly through chronic stress, autonomic dysfunction or cardiotoxic effects of certain substances.

The results presented in this work suggest potentially new ECG features for neurocognitive disorders. This includes elevated RR interval for Alzheimer's disease, reduced QRS axis for Parkinson's disease and shortened QRS duration for vascular dementia. We hypothesise that the elevated RR interval in Alzheimer's may reflect altered autonomic regulation due to parasympathetic dysfunction.[38] The reduced QRS axis in Parkinson's disease may result from basal ganglia dysfunction affecting autonomic control of the heart.[39] Similarly, the shortened QRS duration in vascular dementia might result from cerebral ischaemia and microvascular changes affecting cardiac conduction.[40] These hypotheses warrant further investigation.

Limitations

ECG is a valuable tool for identifying electrical abnormalities that may correlate with neurocognitive disorders; however, it does not provide direct diagnostic confirmation or insights into specific neurophysiological changes underlying these disorders. Many ECG alterations are non-specific and can be attributed to factors such as stress, medication effects or comorbid disorders, making it challenging to isolate patterns uniquely associated with neurocognitive disorders. Additionally, the causal mechanisms linking ECG abnormalities to neurocognitive disorders remain poorly understood, highlighting the need for further investigation into these complex relationships. Evidence from ICD-10 diagnoses in MIMIC-IV suggests minimal significant label correlations,[22] indicating that confounding from co-occurring diseases may be limited. Similarly, ECG feature abnormalities have been shown to be detectable in treatment-naive patients with incident dementia,[15] further supporting the value of this work.

Future research should explore how ECG abnormalities differ across diverse demographic groups and distinguish them from normal variations, such as age-related changes.[41] Investigating the causal connections between ECG features and neurocognitive disorders will be essential for advancing our understanding.[42] Further studies should focus on using raw ECG waveforms and validating findings across external datasets to improve diagnostic accuracy.[9 22] The demonstrated potential of raw ECG waveforms to outperform traditional ECG features emphasises the importance of refining these approaches to enhance precision and reliability in detecting neurocognitive disorders.

Implications

Our approach naturally encompasses both therapy-naive patients and those attending follow-up visits, as ICD-10 codes capture both newly assigned diagnoses and ongoing treatment cases. This creates potential confounding effects, since predictions might reflect therapy-related cardiotoxic changes rather than the neurocognitive condition. Notably, some medications for Alzheimer's disease carry cardiac risks: acetylcholinesterase inhibitors (donepezil, rivastigmine, galantamine) can, in rare cases, cause bradycardia or QT interval prolongation, particularly in patients with pre-existing cardiac arrhythmias.[43] In contrast, memantine has minimal cardiovascular side effects. Similarly, some Parkinson's disease medications have been associated with cardiac complications. Certain dopamine agonists,

particularly ergot-derived agents such as pergolide and cabergoline, can increase the risk of valvular heart disease by stimulating serotonin 5-hydroxytryptamine receptor 2B receptors, leading to fibrotic changes in heart valves.[44] In addition, antipsychotic medication is commonly used in neurocognitive disorders, which are associated with tachycardia, bradycardia and QTc prolongation.[45]

These therapy-related cardiac effects further complicate the distinction between disease progression and treatment influence, underscoring the need for careful monitoring, especially in patients with known cardiac disorders.[43] Future research should aim to stratify newly diagnosed cases from follow-ups, as such differentiation would enable a more precise evaluation of the model's ability to distinguish between new diagnoses and therapy-induced patterns.

ECG innovations for neurocognitive disorders hold great promise across diagnosis, therapy management and personalised care. Distinctive ECG patterns can serve as features for more efficient, non-invasive detection of neurocognitive disorders. Beyond diagnosis, ECG monitoring can track treatment responses and detect early physiological changes related to medication efficacy or adverse effects. Risk stratification is another vital application, enabling clinicians to predict adverse events such as arrhythmias linked to neurocognitive treatments or stress-related cardiovascular complications.

Longitudinal monitoring with periodic or wearable ECG devices provides valuable insights into disease progression, relapse or episodes prediction, facilitating real-time assessments of mental and physical health. ECG-derived features also support personalised medicine by guiding treatments tailored to individual patients. Subtle ECG changes can signal stress, anxiety or relapse, allowing timely clinical intervention. Recent perspectives emphasise that early detection of cognitive impairment in primary care enables earlier interventions, including lifestyle changes and management of comorbid conditions, that can slow cognitive decline and improve overall outcomes for patients and caregivers.[46] Similarly, predictive models for neurocognitive disorders demonstrate that early detection and intervention significantly enhance patient outcomes by enabling timely, individualised therapeutic strategies.[47]

Contributors JMLA, WH and NS conceived and designed the project. JMLA conducted the full experimental analyses, with WH and NS supervising them, and WH, EO and DT providing critical revision of clinical intellectual content. JMLA produced the first draft and the rest of the authors revised it. All authors critically revised the content and approved the final version for publication. NS is the guarantor author. AI was used for grammar correction only.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. The datasets can be found in raw format in their appropriate source codes: https://physionet.org/content/mimic-iv-ecg/1.0/, and http://ecgview.org/. The code for dataset preprocessing and reproducing experiments is available at https://github.com/AI4HealthUOL/CardioDiag.

ORCID iD

Juan Miguel Lopez Alcaraz https://orcid.org/0009-0008-3056-3521

REFERENCES

1 Eichel H von, Heine J, Wegner F, *et al*. Neuropsychiatric symptoms in Parkinson's disease patients are associated with reduced health-related quality of life and increased caregiver burden. *Brain Sci* 2022;12:89.
2 Karantzoulis S, Galvin JE. Distinguishing Alzheimer's disease from other major forms of dementia. *Expert Rev Neurother* 2011;11:1579–91.
3 Aarsland D, Batzu L, Halliday GM, *et al*. Parkinson disease-associated cognitive impairment. *Nat Rev Dis Primers* 2021;7:47.
4 Fong TG, Inouye SK. The inter-relationship between delirium and dementia: the importance of delirium prevention. *Nat Rev Neurol* 2022;18:579–96.
5 Cummings J. The role of neuropsychiatric symptoms in research diagnostic criteria for neurodegenerative diseases. *Am J Geriatr Psychiatry* 2021;29:375–83.
6 Taslim S, Shadmani S, Saleem AR, *et al*. Neuropsychiatric disorders: bridging the gap between neurology and psychiatry. *Cureus* 2024;16:e51655.
7 Siontis KC, Noseworthy PA, Attia ZI, *et al*. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol* 2021;18:465–78.
8 Alcaraz JML, Strodthoff N. Cardiolab: laboratory values estimation from electrocardiogram features–an exploratory study. *arXiv* 2024.
9 Alcaraz JML, Bouma H, Strodthoff N. Enhancing clinical decision support with physiological waveforms - a multimodal benchmark in emergency care. *Comput Biol Med* 2025;192:110196.
10 Liu L, Wang J, Li W, *et al*. Guidelines for the diagnosis and treatment of depressive disorders by integrating Chinese and Western medicine (English edition). *Gen Psychiatr* 2025;38:e101747.
11 Azfr Ali RS, Jalal Z, Paudyal V. Guidelines versus practice in screening and monitoring of cardiometabolic risks in patients taking antipsychotic medications: where do we stand? *Gen Psychiatr* 2021;34:e100561.
12 Li Y, Peng J, Yang P, *et al*. Virtual reality-based cognitive–behavioural therapy for the treatment of anxiety in patients with acute myocardial infarction: a randomised clinical trial. *Gen Psychiatr* 2024;37:e101434.
13 Gonçalves VC, Cuenca-Bermejo L, Fernandez-Villalba E, *et al*. Heart matters: cardiac dysfunction and other autonomic changes in Parkinson's disease. *Neuroscientist* 2022;28:530–42.
14 Raberi VS, Qazi MSK, GhorbaniNia R, *et al*. Postoperative delirium and dementia in patients undergoing cardiac surgery: a review of randomized controlled trials. *Galen Medical Journal* 2023;12:e3045.
15 Isaksen JL, Ghouse J, Skov MW, *et al*. Associations between primary care electrocardiography and non-Alzheimer dementia. *J Stroke Cerebrovasc Dis* 2022;31:106640.
16 Valappil RA, Black JE, Broderick MJ, *et al*. Exploring the electrocardiogram as a potential tool to screen for premotor Parkinson's disease. *Mov Disord* 2010;25:2296–303.

17 Niu Y, Chen X, Fan J, *et al*. Explainable machine learning model based on EEG, ECG, and clinical features for predicting neurological outcomes in cardiac arrest patient. *Sci Rep* 2025;15:11498.
18 Johnson AEW, Bulgarelli L, Shen L, *et al*. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data* 2023;10:1.
19 Gow B, Pollard T, Nathanson LA, *et al*. MIMIC-IV-ECG: diagnostic electrocardiogram matched subset. 2023.
20 Kim YG, Shin D, Park MY, *et al*. ECG-VIEW II, a freely accessible electrocardiogram database. *PLoS One* 2017;12:e0176222.
21 World Health Organization. International statistical classification of diseases and related health problems. Geneva World Health Organization; 1992.
22 Strodthoff N, Lopez Alcaraz JM, Haverkamp W. Prospects for artificial intelligence-enhanced electrocardiogram as a unified screening tool for cardiac and non-cardiac conditions: an explorative study in emergency care. *Eur Heart J Digit Health* 2024;5:454–60.
23 Alcaraz JML, Strodthoff N. Estimation of cardiac and non-cardiac diagnosis from electrocardiogram features. 2024 Computing in Cardiology Conference (CinC); 2024.
24 Lopez Alcaraz JM, Haverkamp W, Strodthoff N. Electrocardiogram-based diagnosis of liver diseases: an externally validated and explainable machine learning approach. *EClinicalMedicine* 2025;84:103242.
25 Alcaraz JML, Haverkamp W, Strodthoff N. Explainable machine learning for neoplasms diagnosis via electrocardiograms: an externally validated study. *Cardiooncology* 2025;11:70.
26 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565–74.
27 Lundberg SM, Erion G, Chen H, *et al*. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2020;2:56–67.
28 Calster BV, Collins GS, Vickers AJ, *et al*. Performance evaluation of predictive AI models to support medical decisions: overview and guidance. *arXiv* 2024.
29 Barkhordarian M, Montazerin SM, Tran HHV, *et al*. Atrial fibrillation and cognitive disorders. *Cardiol Rev* 2024.
30 Huang Y, Alsabbagh MW. Comparative risk of cardiac arrhythmias associated with acetylcholinesterase inhibitors used in treatment of dementias - a narrative review. *Pharmacol Res Perspect* 2020;8:e00622.
31 Greenblatt M. Age and electroencephalographic abnormality in neuropsychiatric patients. *AJP* 1944;101:82–90.
32 Isaksen JL, Ghouse J, Skov MW, *et al*. Association between primary care electrocardiogram markers and Alzheimer's disease. *J Neurol Sci* 2023;447:120581.
33 Georgiev D, Hamberg K, Hariz M, *et al*. Gender differences in Parkinson's disease: a clinical perspective. *Acta Neurol Scand* 2017;136:570–84.
34 Rabkin SW. The QT interval in Parkinson's disease: a systematic review. *J Geriatr Cardiol* 2024;21:855–64.
35 Mao M, Wang C, Hou T, *et al*. Ventricular electrocardiographic signatures associated with dementia and plasma Alzheimer's disease biomarkers in older adults: a population-based study. *JAD* 2023;94:1515–26.
36 Matei D, Corciova C, Matei R, *et al*. QT interval in vascular dementia and Alzheimer's disease. 2015 E-Health and Bioengineering Conference (EHB); Iasi, Romania, 2015
37 Cuculi F, Kobza R, Ehmann T, *et al*. ECG changes amongst patients with alcohol withdrawal seizures and delirium tremens. *Swiss Med Wkly* 2006;136:223–7.
38 Femminella GD, Rengo G, Komici K, *et al*. Autonomic dysfunction in Alzheimer's disease: tools for assessment and review of the literature. *JAD* 2014;42:369–77.
39 Zesiewicz TA, Baker MJ, Wahba M, *et al*. Autonomic nervous system dysfunction in Parkinson's disease. *Curr Treat Options Neurol* 2003;5:149–60.
40 Wang F, Cao Y, Ma L, *et al*. Dysfunction of cerebrovascular endothelial cells: prelude to vascular dementia. *Front Aging Neurosci* 2018;10:376.
41 Ott G, Schaubelt Y, Lopez Alcaraz JM, *et al*. Using explainable AI to investigate electrocardiogram changes during healthy aging- from expert features to raw signals. *PLoS One* 2024;19:e0302024.
42 Alcaraz JML, Strodthoff N. CausalConceptTS: causal attributions for time series classification using high fidelity diffusion models. *arXiv* 2024.
43 Howes LG. Cardiovascular effects of drugs used to treat Alzheimer's disease. *Drug Saf* 2014;37:391–5.
44 Schade R, Andersohn F, Suissa S, *et al*. Dopamine agonists and the risk of cardiac-valve regurgitation. *N Engl J Med* 2007;356:29–38.
45 Buckley NA, Sanders P. Cardiovascular adverse effects of antipsychotic drugs. *Drug Saf* 2000;23:215–28.
46 Fowler NR, Partrick KA, Taylor J, *et al*. Implementing early detection of cognitive impairment in primary care to improve care for older adults. *J Intern Med* 2025;298:31–45.
47 Tsiakiri A, Bakirtzis C, Plakias S, *et al*. Predictive models for the transition from mild neurocognitive disorder to major neurocognitive disorder: insights from clinical, demographic, and neuropsychological data. *Biomedicines* 2024;12:1232.

Juan Miguel Lopez Alcaraz is a researcher specialising in machine learning methodologies and time series analysis, with a focus on healthcare applications. He is currently a Research Assistant and PhD candidate at Oldenburg University, Germany, where his work involves developing methodologies for missing data imputation, explainability, and causal inference, as well as clinical applications for decision support. His research has contributed to new clinical datasets and benchmarks aimed at improving patient care through more reliable predictive models. Juan Miguel holds a Master's degree in Big Data Analytics and Artificial Intelligence from Letterkenny Institute of Technology, Ireland, and a BSc in Finance from UCOL, Mexico. He has industry experience in logistics, digital marketing and holds a Certificate in Investment Strategies from the Mexican Association of Stockbrokers. In addition, he serves as a peer reviewer for several international journals in machine learning and healthcare informatics. His interdisciplinary background supports his commitment to integrating data-driven methods with real-world healthcare challenges.