



OPEN

Externally validated deep learning model to identify prodromal Parkinson's disease from electrocardiogram

Ibrahim Karabayir¹, Fatma Gunturkun², Liam Butler¹, Samuel M. Goldman^{3✉}, Rishikesan Kamaleswaran⁴, Robert L. Davis⁵, Kalea Colletta⁶, Lokesh Chinthala⁵, John L. Jefferies⁷, Kathleen Bobay⁸, G. Webster Ross⁹, Helen Petrovitch¹⁰, Kamal Masaki^{11,12}, Caroline M. Tanner¹³ & Oguz Akbilgic^{1✉}

Little is known about electrocardiogram (ECG) markers of Parkinson's disease (PD) during the prodromal stage. The aim of the study was to build a generalizable ECG-based fully automatic artificial intelligence (AI) model to predict PD risk during the prodromal stage, up to 5 years before disease diagnosis. This case–control study included samples from Loyola University Chicago (LUC) and University of Tennessee–Methodist Le Bonheur Healthcare (MLH). Cases and controls were matched according to specific characteristics (date, age, sex and race). Clinical data were available from May, 2014 onward at LUC and from January, 2015 onward at MLH, while the ECG data were available as early as 1990 in both institutes. PD was denoted by at least two primary diagnostic codes (ICD9 332.0; ICD10 G20) at least 30 days apart. PD incidence date was defined as the earliest of first PD diagnostic code or PD-related medication prescription. ECGs obtained at least 6 months before PD incidence date were modeled to predict a subsequent diagnosis of PD within three time windows: 6 months–1 year, 6 months–3 years, and 6 months–5 years. We applied a novel deep neural network using standard 10-s 12-lead ECGs to predict PD risk at the prodromal phase. This model was compared to multiple feature engineering-based models. Subgroup analyses for sex, race and age were also performed. Our primary prediction model was a one-dimensional convolutional neural network (1D-CNN) that was built using 131 cases and 1058 controls from MLH, and externally validated on 29 cases and 165 controls from LUC. The model was trained on 90% of the MLH data, internally validated on the remaining 10% and externally validated on LUC data. The best performing model resulted in an external validation AUC of 0.67 when predicting future PD at any time between 6 months and 5 years after the ECG. Accuracy increased when restricted to ECGs obtained within 6 months to 3 years before PD diagnosis (AUC 0.69) and was highest when predicting future PD within 6 months to 1 year (AUC 0.74). The 1D-CNN model based on raw ECG data outperformed multiple models built using more standard ECG feature engineering approaches. These results demonstrate that a predictive model developed in one cohort using only raw 10-s ECGs can effectively classify individuals with prodromal PD in an independent cohort, particularly closer to disease diagnosis. Standard ECGs may help identify individuals with prodromal PD for cost-effective population-level early detection and inclusion in disease-modifying therapeutic trials.

¹Cardiovascular Section, Department of Internal Medicine, Wake Forest School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, USA. ²Quantitative Sciences Unit, Department of Medicine, Stanford University, Palo Alto, CA, USA. ³Division of Occupational, Environmental, and Climate Medicine, San Francisco Veterans Affairs Medical Center, University of California–San Francisco, 4150 Clement Street, Box 127, San Francisco, CA 94121, USA. ⁴Department of Biomedical Informatics, Emory University, Atlanta, GA, USA. ⁵Center for Biomedical Informatics, University of Tennessee Health Science Center, Memphis, USA. ⁶Department of Neurology, Edward Hines Jr. VA Hospital, Hines, IL, USA. ⁷Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, TN, USA. ⁸Parkinson School of Health Sciences and Public Health, Loyola University Chicago, Maywood, IL, USA. ⁹Veterans Affairs Pacific Islands Health Care Systems, Honolulu, HI, USA. ¹⁰Pacific Health Research and Education Institute, Honolulu, HI, USA. ¹¹Kuakini Medical Center, Honolulu, HI, USA. ¹²Department of Geriatric Medicine, John A. Burns School of Medicine, University of Hawaii, Honolulu, HI, USA. ¹³Department of Neurology, Weill Institute for Neurosciences, University of California San Francisco, San Francisco, CA, USA. ✉email: samuel.goldman@ucsf.edu; oakbilgi@wakehealth.edu

Parkinson's disease (PD) is a systemic disease that is currently diagnosed when classic motor symptoms including resting tremors and bradykinesia, become evident and generally manifest after 50% of substantia nigra dopaminergic neurons are already dead or dying and with extensive striatal dopaminergic deafferentation^{1–4}. However, pathologic changes in the lower brainstem and peripheral autonomic nervous system likely begin years or even decades before significant nigral damage occurs^{5,6}, and symptoms referable to these sites of early pathologic injury may manifest many years before motor PD is diagnosed³. Putative interventions to delay or slow the progression from prodromal disease to parkinsonism could be implemented during this pathologic evolution if PD patients could be identified early with confidence.

Lewy pathology is found throughout the autonomic nervous system in PD⁷. Pathology of the sympathetic and parasympathetic ganglia, cardiac nerves and cardiac deafferentation are consistently seen in early PD^{8–11} and post-mortem examinations associate these pathologies with *incidental* nigral Lewy bodies (ILB)¹⁰. For this reason, cardiac sympathetic deafferentation as measured by ¹²³I-Metaiodobenzylguanidine (¹²³I-MIBG) scintigraphy is designated as a supportive criterion for the clinical diagnosis of PD in the MDS-PD diagnostic criteria¹². However, MIBG scintigraphy is invasive and expensive, and is not a viable tool for population-level screening for PD risk.

Sympathetic and parasympathetic nerve inputs mediate adaptive responses to varying physiological conditions and cardiovascular demands, all of which can manifest as heart rate variability (HRV). HRV can be assessed by measuring the duration between consecutive R waves on an ECG¹³. HRV is reduced in PD, likely reflecting underlying cardiac autonomic Lewy pathology and deafferentation^{14–22}. In addition, HRV has recently been shown to correlate with striatal dopamine depletion in PD²³. Paralleling MIBG observations, evidence suggests that HRV may be reduced in prodromal PD. HRV is reduced in rapid eye movement (REM) sleep behavior disorder (RBD)^{24,25}, a condition that is thought to be a biomarker and strongest predictor for prodromal PD, due to its direct correlation to the progression to PD^{26,27}.

A previous study using data from Atherosclerosis Risk in Communities Study Description (ARIC), a large prospective study, found that low HRV was associated with 2–3 fold increased subsequent risk of PD over a mean of 18 years follow up period²⁸. In addition, there has been research into predicting PD or prodromal PD using numerous clinical risk factors in combination with biological components extracted from serum samples, such as cytokines and chemokines within a decision tree algorithm²⁹. Research by Akbilgic et al. (2022) used cardiac electrical activity signal information from 60 subjects from the Honolulu Asia Aging Study (HAAS) within a logistic regression model following a Probabilistic Symbolic Pattern Recognition (PSPR) method to identify patients at high risk of PD but without a pre-specified time window³⁰. While this research reported an average AUC of 0.835, the HAAS cohort was a small and very homogeneous data sample and not representative of the general population. While HRV metrics, typically measured using 5-min ECGs, have been used in multiple studies, there is some debate whether such classical metrics can indeed capture substantial information on the likelihood of development of PD³⁰. In previous research, models were developed to predict future PD using machine learning algorithms that incorporate multiple features derived from the full waveform of standard 10-s ECGs³⁰. Such research indeed provides substantial preliminary indication that ECG-based signals can be used as a predictor of PD, with the possibility of predicting PD onset at a timely manner before diagnosis.

This research builds on prior artificial intelligence (AI) approaches but uses raw waveform ECGs rather than features extracted from ECGs. There is substantial utility, generalizability and simplicity of raw 12-lead ECGs, which are easily and routinely collected, to predict prodromal (and risk of) PD within a specified time window. Importantly, this novel study uses two independent study populations with available electronic ECGs obtained prior to PD onset, with one dataset used for model building and the other used for external validation.

Results

Our initial EHR-based search identified 131 eligible PD cases and 1,058 controls from the MLH database, with a total of 428 and 3584 available ECGs, respectively. After validation of PD diagnosis by manual chart review, the LUC dataset included 29 PD cases and 165 matched controls with one ECG each. The baseline characteristics of these cohorts are summarized in Table 1.

The MLH data was split into 90% training and 10% internal validation. We developed 1D-CNN model (Fig. 1) using 90% MLH training data up to 100 epochs. The model from the epoch which resulted in the highest AUC in the 10% internal validation data was selected as the final model. This model resulted in an internal validation AUC of 0.73 with 95% CI of 0.51–0.96. To convert predicted risk values into predicted classes (PD or not), we identified the decision threshold as 0.139, which maximized the F1 score. Using this threshold, we obtained a sensitivity of 0.43 and a specificity of 0.96 in predicting prodromal PD up to 5 years before disease onset in the internal validation dataset. We note that this threshold can be adjusted according to one's purposes, to increase sensitivity at a cost of reduced specificity. For purposes of developing a clinical trial for a putative disease-modifying drug, high specificity is the preferred metric, because it would substantially reduce the required sample size. Conversely, when one or more disease-modifying drugs become available, assuming the drug is safe, it could make sense to select a test positivity cut-point that emphasizes sensitivity over specificity, and thereby treat a larger proportion of at-risk individuals. The MLH data driven model was then externally validated on the LUC data and resulted in an AUC of 0.67 (0.54, 0.79), correctly classifying 12 of 29 PD cases 5 years prior to disease diagnosis, and 123 of 165 controls, with a specificity of 0.75 and sensitivity of 0.41. In LUC external validation data, the accuracy increased slightly when predicting future PD within 3 years of diagnosis, with AUC = 0.69 (0.57–0.82) and average precision (AP) of 0.21, with the highest accuracy achieved, as expected, when predicting PD within one year of diagnosis with an AUC of 0.74 (0.62–0.87) and AP of 0.11 (Table 2, Fig. 2).

In addition to the 1D-CNN, models using multiple feature-engineering methods, including descriptive statistics extracted from 12 leads of the ECGs, sample entropy, PSPR etc., were also developed to predict prodromal

	MLH			LUC		
	Total (n = 1189)	Case (n = 131)	Control (n = 1058)	Total (n = 194)	Case (n = 29)	Control (n = 165)
Mean age at ECG (SD)	70.1 (9.5)	72.6 (8.0)	69.8 (9.6)	71.2 (8.7)	69.8 (8.7)	71.4 (8.8)
Mean age at PD diagnosis (SD)	75.3 (7.7)	75.3 (7.9)	–	72.3 (8.0)	71.0 (8.8)	–
Mean years between ECG and PD diagnosis (SD)	2.7 (1.9)	2.7 (1.9)	–	1.7 (1.0)	1.65 (1.0)	–
Male n (%)	639 (53.7%)	77 (58.8%)	562 (53.1%)	119 (61.3%)	18 (62.1%)	101 (61.2%)
Race n (%)						
White	753 (63.3%)	86 (65.7%)	667 (63.0%)	149 (76.8%)	23 (79.3)	126 (76.4)
Black or African American	412 (34.7%)	41 (31.3%)	371 (35.1%)	22 (11.3%)	3 (10.4)	19 (11.5)
Multiple	10 (0.8%)	0 (0%)	10 (1.0%)	0 (0%)	0 (0%)	0 (0%)
Other/unknown	5 (0.4%)	2 (1.5%)	3 (0.3%)	19 (9.8%)	3 (10.4)	16 (9.7)
Asian	8 (0.7%)	2 (1.5%)	6 (0.6%)	4 (2.1%)	0 (0.0%)	4 (2.4%)
Amer. Ind/Alaska Native	1 (0.1%)	0 (0%)	1 (0.1%)	0 (0%)	0 (0%)	0 (0%)

Table 1. Baseline characteristics. MLH, University of Tennessee-Methodist Le Bonheur Healthcare; LUC, Loyola University, Chicago.

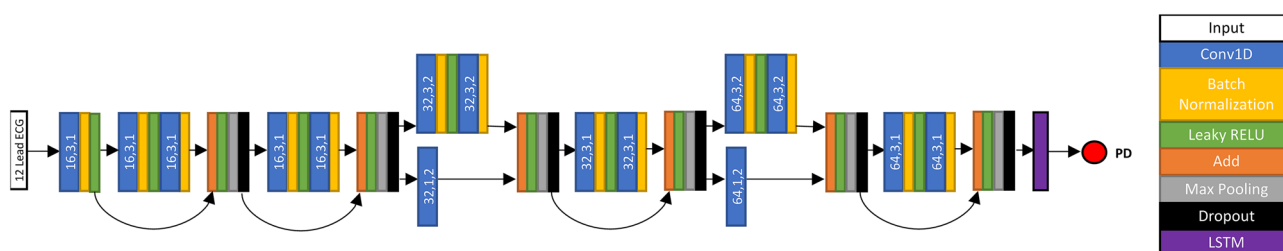


Figure 1. Proposed 1D-CNN architecture used to predict PD risk using 10 s 12-lead ECG data. Values in the Conv1D boxes refer to number of filters, kernel size and number of strides respectively.

Subgroup	Time from ECG until PD incident diagnosis date		
	6 months–1 year	6 months–3 years	All (6 months–5 years)
All subjects	0.74 [0.62–0.87]	0.69 [0.57–0.82]	0.67 [0.54–0.79]
Sex			
Male	0.73 [0.57–0.9]	0.71 [0.55–0.87]	0.69 [0.53–0.85]
Female	0.79 [0.6–0.98]	0.66 [0.45–0.86]	0.63 [0.43–0.84]
Race			
White	0.71 [0.56–0.86]	0.70 [0.56–0.85]	0.67 [0.52–0.81]
Black*	1.00 [1.00–1.00]	0.53 [0.13–0.91]	0.65 [0.27–1.03]
Age			
> =60	0.73 [0.6–0.87]	0.70 [0.57–0.83]	0.67 [0.53–0.80]
< 60*	0.78 [0.40–1.15]	0.66 [0.26–1.05]	0.66 [0.26–1.05]

Table 2. CNN model performance in the LUC external validation cohort. *The data size was very limited to reach any conclusion.

PD within a 5-year time period. This was done to compare multiple models with a raw ECG-based model. From all the methods listed in Supplementary Table S1, the highest LUC validation accuracy achieved was AUC = 0.58 when using descriptive ECG features. When all the features extracted from all the different feature engineering methods were included within one main LightGBM model, the AUC increased to 0.61 (0.52–0.79). These results show that the AUC of the external validation accuracy obtained from deep learning on raw digital ECGs, which is 0.67 (0.54–0.79), was better than to those obtained from machine learning approaches using engineered ECG features (Supplementary Table S1). Since there was no major improvement in the LightGBM model utilizing all the extracted features, the 1D-CNN ECG model was the basis of the subgroup analyses.

Subgroup analysis. Although statistical power was limited for some analyses, we implemented a comprehensive subgroup analysis of the LUC external validation cohort, stratifying by sex, race, and age. We also explored how predictive accuracy varied by time from ECG until PD incidence date (Table 2).

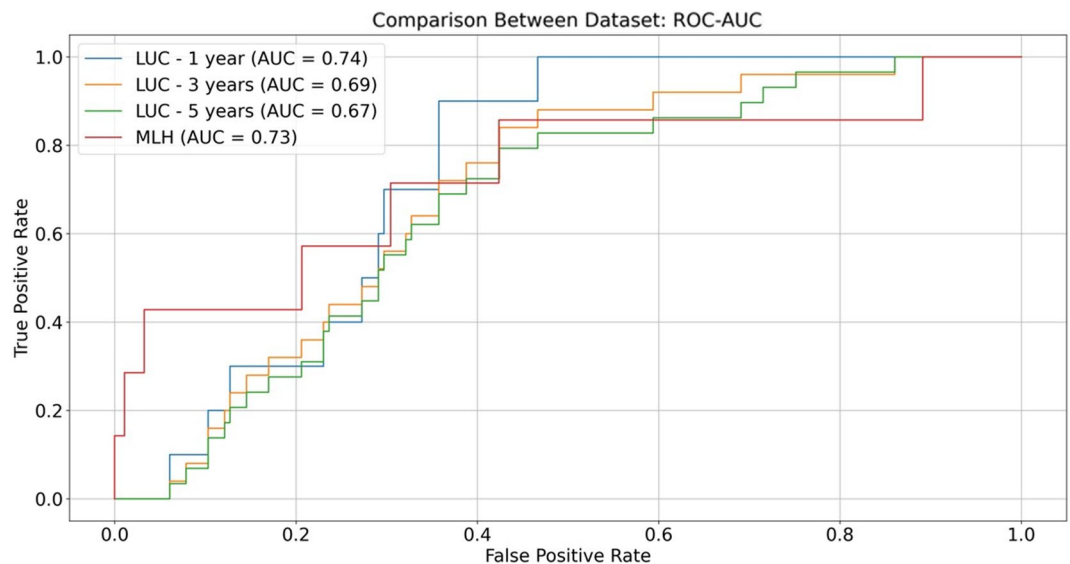


Figure 2. The area under the ROC Curves for LUC external validation and MLH internal validation sets.

As shown in Table 2, prediction accuracy of the model improved with a shorter time interval between the ECG and the PD diagnosis. The 1D-CNN model predicted prodromal PD in males at the highest overall accuracy, with AUC 0.69 (0.53–0.85), although accuracy was high in women within one year of PD diagnosis (AUC 0.79 [0.60–0.98]); sex-specific differences in classification accuracy were not statistically significantly different (Delong test $p = 0.62$). Overall accuracy was similar in those aged $< \text{or} > 60$, though was high in those < 60 within one year of PD diagnosis (AUC 0.78 [0.40–1.15]). We had limited statistical power for race-specific comparisons.

Discussion

There has been an increase in studies researching early identification of PD through biological models, but there still remains a major gap in early detection and, even more so, prediction of PD within the prodromal stage³¹. While multiple studies have assessed the use of various clinical features to predict prodromal PD, the implementation of such models on a population level may be limited by the lack of symptom reporting and/or diagnostic coding of these features in the medical record. Accurate identification of prodromal PD is essential for implementation of cost-effective clinical trials of putative neuroprotective agents, and ultimately for risk stratification and population-level disease prevention once therapeutic efficacy is established for one or more agents. In order to be effective on a population level, a predictive test of PD risk should be universally available, cost-effective and non-invasive. Studies found that heart rate variability (HRV) determined from 5-min ECGs is reduced in prevalent PD, and results from a single prospective study showed that lower HRV was associated with an increased risk of incident PD³². Prior work by our group utilized machine learning approaches to predict prodromal PD using clinical variables in one study³³ and a proof-of-concept study using standard 10-s printed ECGs in another study³⁰. Both studies resulted in moderate accuracy, however the latter study was conducted in a demographically homogeneous elderly population and lacked an external cohort for model validation³⁰.

This research builds on the above work, using a deep learning approach to develop a predictive model for prodromal PD using standard 10-s raw digitized ECGs stored in a large healthcare system EHR, and validates this model using an independent cohort from a different healthcare system EHR. The developed model had moderately good classification accuracy in the validation cohort, with an AUC of 0.74 between 6 and 12 months before PD diagnosis and an AUC of 0.67 when the ECG was obtained as long as 5 years before PD diagnosis. The data imbalance between cases and controls and the relatively small size of the validation cohort likely limited our classification accuracy, yet the model successfully distinguished between the majority of cases and controls in an independent population, and has important implications as an ancillary approach to detect early PD in population-level screening.

While the AUCs derived here can be deemed as somewhat low-moderate at 5 years before diagnosis, it should be noted that the main aim of this model is to serve as a tool to help clinicians assess probable risk of people developing PD in the future and not for diagnosis per se. Other researchers have reported higher AUCs for predictive models, however most of those models were built using large numbers of extracted features including, but not limited to, clinical variables, neuropsychological test scores, domain composite scores, brain section measurements and magnetic resonance imaging (AUC = 0.84)^{34,35}, while other models relied on physical movement tests (e.g. finger tapping) to extract bradykinesia features (AUC = 0.79–0.85) or PD detection based on language and/or speech data (AUC > 0.85)^{36–38}. Despite some of these biomarkers are also low cost and easy to collect, they are typically collected only from already symptomatic patients, therefore, not suitable for EMR based prodromal PD screening purposes. In contrast to a standard 10-s ECG, many of these features would not be available in the EMR, limiting their usefulness for population-level screening. The use of ECGs in combination with other

prodromal PD features available in the EMR will likely further improve our ability to screen the general population to identify persons at high risk for PD.

A strength of the current study is the validation of PD diagnosis and incidence dates in the external LUC validation population, thus ensuring accurate and generalizable performance metrics. Given that the model was developed in the MLH cohort without chart-based validation, a proportion of whom were likely misclassified, the predictive performance of the model in the replication cohort is all the more remarkable, and provides support for the broad generalizability of the use of ECGs for prodromal PD prediction: almost half of the future PD cases were correctly identified as such, while most of the controls were classified properly as not being at risk. We acknowledge that the deep learning model developed here needs improvement to maximize case identification, however, in addition to likely disease misclassification, it should also be noted that the MLH dataset used for training was largely imbalanced between cases and controls, making the model slightly more effective in identifying controls. The model's performance can be improved following re-training using a larger cohort or the implementation of transfer learning in combination with a larger cohort.

In comparison to the machine learning models utilizing ECG feature-engineering as inputs (Supplementary Table S1), the CNN model predicted prodromal PD at higher accuracy. These results are comparable to published research that included multiple different biological and demographic information, some of which can be quite costly to obtain (e.g. serum analysis), within a machine learning system. Furthermore, while the prediction accuracy of the CNN model within 5 years before PD onset is moderate, this research adds to our understanding that PD cardiac markers, which are found before the motor changes of PD³⁹, can act as signals within an artificial intelligence framework without the need of full ECG feature extraction. The inclusion of demographic and clinical variables as additional inputs with the predicted outcomes from the ECG CNN model will likely increase the accuracy of predicting prodromal PD within an extended time window. Karabayir et al.³³ showed that there is added benefit in including demographic and clinical variables in the prediction of prodromal PD, allowing for opportunity for exploration of improving this work. Such variables can also extend to the inclusion of data from the Cognitive Abilities Screening Instrument (CASI), olfactory tests and the simple choice reaction times test. At the time of development of the models in the current study, such clinical and demographic variables were not available from the MLH and LUC data sources.

It is also worth noting that recent literature has showcased the potential of leveraging fractional dynamics to capture long-range memory in ECG data analysis, offering valuable insights for addressing temporal dependencies in the field^{40,41}. In this study, we implemented various models including pure RNN, LSTM, and Transformers^{42–44}. While RNN and LSTM models struggled to capture long-term dependencies (AUC of 0.6 and 0.61, respectively), Transformers demonstrated promise with an AUC of 0.69, effectively modeling complex temporal relationships. Nevertheless, our model, combining 1D convolution layers and LSTM, slightly outperformed Transformers by effectively capturing both short and long-term dependencies, resulting in improved accuracy. While Transformers offer comparable accuracy, their computational complexity and memory requirements may present challenges. Our model strikes a balance between efficiency and performance. In summary, our study validates the effectiveness of our hybrid model and acknowledges the promising performance of Transformers.

Overall, the CNN model developed in this research classified individuals based solely on a 10-s ECG input to an artificial intelligence framework. The easy accessibility and routine collection of ECGs within any clinical setting increases their generalizability to the wider population since it substantially reduces bias towards any sex, race or age subgroup. This makes such models useable in the wider, possibly global, population, especially with modern technological advancement which allows AI models to be incorporated within smart technologies. It is highly likely that incorporation of our ECG-based predictive model with other known prodromal markers of PD would result in improved risk stratification for PD. This highlights the importance of routine collection of ECGs within cohort studies aiming at understanding the pathophysiology of PD and other neurodegenerative diseases.

This study has some limitations. We were able to perform chart reviews on LUC data, but not on MLH data. Based on the high false positive rate of PD diagnosis based exclusively on EHR ICD codes at LUC (only 29/47 were determined to have PD), a substantial proportion of MLH PD cases were likely to have been misclassified. Therefore, implementing transfer learning on our final deep learning model and retraining it on large well-annotated datasets will result in a more accurate PD prediction tool. This also highlights the challenges around PD case ascertainment using EHR-based queries of diagnostic codes. The inclusion of AI models in clinical practice and future clinical decision support could help to reduce diagnostic misclassification. The five-year PD risk prediction model we developed correctly classified only about half of future PD cases but 96% of controls on MLH validation while only 41% of future PD cases and 75% of controls on LUC validation data. However, the test positivity cut-point can be adjusted by an investigator or clinician to emphasize high specificity (e.g., for clinical trials of putative disease-modifying drugs) or high sensitivity (e.g., for selection of high-risk individuals to receive preventive treatment), depending on the particular application.

We also note that the raw ECG data used in this study was gathered from two different cardiology information systems: GE MUSE at LUC and Epiphany at MLH. Despite this limitation, the concordant results between internal testing and the validation process provides evidence that raw ECGs exported from different platforms can be used reliably within AI frameworks, with minimal pre-processing.

The prediction accuracy of our model is highest using ECGs 6-months-1-year prior to PD diagnosis, which may pose a limitation for use in recruitment for some clinical trials. However, it still has moderate accuracy using ECGs up to 3-years or even 5-years prior to PD diagnosis. When specificity was maximized in our example (to reduce false positives), sensitivity for identifying persons with future PD was less than 50%. We acknowledge this relatively low sensitivity as a major limitation. However, for clinical trials of future disease modifying therapies, and in contrast with some diagnostic tests, high specificity is far more important than high sensitivity. We anticipate that future work that additionally incorporates demographic, clinical variables, and genetic information within the current deep learning model will offer greater sensitivity without sacrificing the high specificity of

our ECG-only model. We anticipate that continuous improvement of model prediction accuracy, for example, through the use of larger and more accurate cohort data, and inclusion of other ‘simple’ variables will lead to the integration of these models within ECG-capable smart wearables. The development of smart applications within wearables will allow for cost-effective, non-invasive and non-burdensome screening of people for PD risk, allowing for early detection, quicker follow-up times and timely intervention strategies.

This work provides proof-of-principle that a deep learning predictive model using only simple 10-s ECGs correctly classifies individuals with prodromal PD with modest accuracy. This model was effective in distinguishing between future PD cases and controls in an independent cohort, increasing in accuracy closer to disease diagnosis, although still notably within the prodromal stage of the disease. The use of standard ECGs, which are easily and routinely collected by healthcare providers, may help identify individuals at high-risk of PD, allowing for timely inclusion in disease-modifying therapeutic trials and possible intervention strategies to delay or slow disease progression.

Methods

Standard protocol approvals and patient consents. This retrospective case–control study was conducted using data from two hospital centers in the United States (Loyola University Chicago (LUC), Maywood, IL and University of Tennessee Health Science Center (UTHSC), Memphis, TN). The study was reviewed and approved by the institutional review boards of the participating centers (Loyola University Chicago and University of Tennessee Health Science Center), and all methods were carried out in accordance with the relevant guidelines and regulations. Due to its retrospective nature and exempt status, written informed consent was waived by the institutional review boards of the participating centers.

Study cohorts. This study used electronic health record (EHR)-based datasets from Loyola University Chicago (LUC), Chicago, IL and University of Tennessee-Methodist Le Bonheur Healthcare (MLH), Memphis, TN. The study design is summarized in Fig. 3. The participants in the analytic cohorts were aged 26–89 years at the time that the ECG was recorded. While 26 to 40 year-old participants fall within the ‘early-onset’ criteria of PD^{45,46} and is indeed rare, we utilized this information to build a more inclusive model, unbiased towards age, with the main objective of determining whether the developed models can incorporate a wide range of prodromal-to-PD scenarios.

Case identification. In both datasets, we identified patients aged 26–89 years who had received at least two ICD diagnostic codes for PD (ICD9 332.0, ICD10 G20) as the primary reason for the visit at least 30 days apart, between May 1, 2014 and January, 2020 at LUC and January 1, 2015 and January, 2020 at MLH. PD incidence date was defined as the earliest of the first PD diagnostic code or PD-related medication prescription. For the external validation phase, we included individuals from LUC who had a standard 12-lead ECG within 5 years before the incidence date. In order to have the largest possible training sample, we included individuals from MLH within 10-years prior to the PD incidence date.

Case validation. We reviewed the medical records of individuals from the LUC dataset to confirm the PD diagnosis and date of disease incident diagnosis. To carry out this evaluation on the LUC data, the patients’ charts were accessed using the *Epic* software, and the related clinical notes were reviewed by a movement disorders specialist (co-author KC). ‘Care Everywhere’ notes in *Epic* from other healthcare systems were also examined when available. We considered diagnostic consistency, clinical features, response to medication, and possible alternative diagnoses. All patient charts were also searched for the presence of medical diagnoses or medications potentially inconsistent with a diagnosis of PD. Based on these descriptions, we identified 131 cases at MLH and 29 eligible chart-reviewed cases (47 cases before manual chart review) at LUC and extracted their corresponding ECGs from cardio servers.

Control identification. Randomly selected control subjects from each institution were matched to cases by age at ECG, date of ECG recording, race and sex, and were required to have been in the system for at least

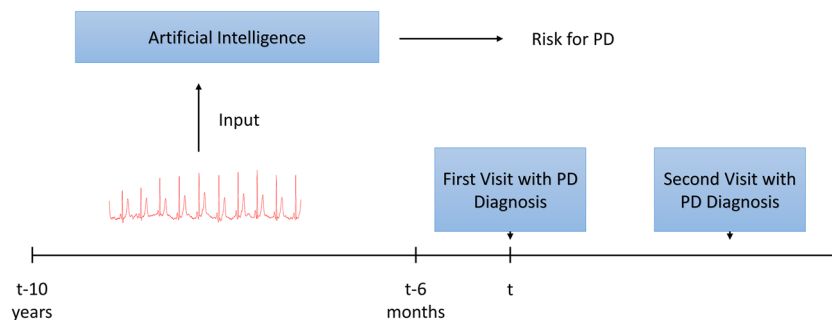


Figure 3. Study design.

10 years (for MLH) or 5 years (for LUC) with no diagnostic codes for PD or another form of parkinsonism. We identified 1058 control subjects at MLH and 65 at LUC, whose ECGs were retrieved for analysis.

Covariate data. Covariates were also obtained from the EHR including age, sex, and race.

ECG data. Digital 12-lead, 10-s ECG, referring to time–voltage data, was obtained for each case and control. The ECGs at LUC were exported from MUSE Cardiology Information System in XML format at 500 Hz. Base 64 encoded waveform data was further decoded using the Python programming language. ECGs at MLH were exported in DICOM format and the associated waveforms were parsed out. Some MLH ECGs were recorded at 500 Hz and some at 250 Hz and therefore, to ensure comparability and based on our previous successful application on utilization of ECG via deep learning^{47,48}, all ECGs at 500 Hz were down-sampled to 250 Hz by removing one of every two amplitudes from each lead. In addition, we excluded the first second of each ECG before inputting them into the model. This was done to avoid any possible noise due to the patient's movement at the beginning of the ECG recording. It should be noted that while ECGs were obtained from different service providers, i.e. GE MUSE and Epiphany, there is no reported difference in the ECG data exported from either service apart from the storage method (DICOM: binary vs XML: human-readable)⁴⁹.

Method development and external validation. Because the MLH study cohort was significantly larger than the LUC cohort, we used MLH data for training and LUC data for external validation. We note that we had access to one ECG per patient from LUC. However, we could retrieve multiple ECGs per patient at MLH, and we utilized all ECGs during training. We split the MLH dataset into 90% training and 10% internal validation sets based on patients rather than ECGs, since some patients have multiple ECGs. All available ECGs were used for the 90% training dataset. However, while multiple ECGs were available, in the 10% MLH internal validation dataset only one randomly selected ECG per patient was used. This was done with the aim of achieving representative samples to that within the LUC external validation dataset. The model's hyperparameters were tuned based on the highest accuracy when using the internal validation set. This model, i.e. the one providing the highest AUC, was used for external validation on LUC data. In addition to the AUC, the best performing model was assessed using sensitivity, specificity, positive predictive value, negative predictive value, and average precision score (AP). Although AUC is a widely used metric to evaluate the performance of classifiers, it should be carefully utilized in evaluation of severely imbalanced classification problems since it only measures how well predictions are ranked. This makes AUC optimistic for the evaluation of classification models that include imbalanced and/or small datasets. Therefore, we have also provided the area under AP and Precision-Recall (PR) curve (AUPRC) (Figure S1) as an alternative metric to effectively evaluate the proposed model by focusing on the minority class (PD).

Deep learning method. We used a one-dimensional convolutional neural network (1D-CNN) with raw digital ECGs as inputs with the risk for PD being the determined output. The 1D-CNN, inspired from ResNet⁵⁰, takes advantage of skipping connections between layers, which eases the optimization of the kernels in the network and therefore obtains robust and more generalizable models than those yielded by traditional or shallow networks. CNN architectures include several hyperparameters including the number of convolutional filters, the kernel dimension, and the stride parameter of convolution layers. The parameters and the overall architecture of the CNN are depicted in Fig. 1. In our architecture, we employed 15 1D convolution layers and one LSTM layer to extract abstract spatial representations and capture both long and short-term dependencies from ECG data, enabling us to effectively capture complex patterns for accurate prediction of PD risk. To optimize the model's performance, we employed the categorical cross entropy loss function, which effectively minimized the difference between predicted and true class labels. This choice significantly improved accuracy and enabled reliable predictions, especially for multi-class classification tasks like the one presented in our work. Furthermore, we utilized the leaky rectified linear units (Leaky-RELU) activation function⁵¹ to obtain a better representation of the ECG by activating the intermediate neurons and allowing negative inputs. After the hyperparameter tuning process, dropout layers with a rate of 0.1 were included within the CNN architecture to penalize the kernels and help prevent model overfitting⁵². The Adam optimization algorithm⁵³ (beta₁ = 0.9, beta₂ = 0.999 and learning rate = 0.001) was used to optimize the kernels in the model. The batch size for data input was set to 128. We have used the grid search approach to tune architecture based on several hyperparameters including the number of convolutional filters, the kernel dimension, and the stride parameter of convolution layers. The model's training process was stopped once reaching consecutive decrease (five times in a row) on the AUC of the internal validation dataset.

Conventional machine learning method. To compare the proposed deep learning algorithm, we also utilized Light Gradient Boosting Machine (LightGBM) by inputting ECG features, which were extracted from a variety of signal processing algorithms⁵⁴. In the development of the LightGBM model, the same five-fold cross validation strategy used in the development of the CNN model was used to ensure comparability. Details of the feature engineering-based machine learning model can be found in Supplementary Material.

Data availability

Requests for patient-related data not included in the article will not be considered. The code for our proposed model can be accessed at https://github.com/ikarabayir/PD_ECG_CNN_Risk.

Received: 15 November 2022; Accepted: 14 July 2023

Published online: 29 July 2023

References

- Scherman, D. *et al.* Striatal dopamine deficiency in Parkinson's disease: Role of aging. Research support. *Non-U.S. Gov't. Ann. Neurol.* **26**(4), 551–557. <https://doi.org/10.1002/ana.410260409> (1989).
- Bernheimer, H., Birkmayer, W., Hornykiewicz, O., Jellinger, K. & Seitelberger, F. Brain dopamine and the syndromes of Parkinson and Huntington. Clinical, morphological and neurochemical correlations. *J. Neurol. Sci.* **20**(4), 415–455 (1973).
- Poewe, W. *et al.* Parkinson disease. *Nat. Rev. Dis. Primers* **3**, 17013. <https://doi.org/10.1038/nrdp.2017.13> (2017).
- Obeso, J. A. *et al.* Past, present, and future of Parkinson's disease: A special essay on the 200th anniversary of the Shaking Palsy. *Mov. Disord.* **32**(9), 1264–1310. <https://doi.org/10.1002/mds.27115> (2017).
- Braak, H. *et al.* Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging* **24**(2), 197–211. (2003).
- Braak, H., de Vos, R. A., Bohl, J. & Del Tredici, K. Gastric alpha-synuclein immunoreactive inclusions in Meissner's and Auerbach's plexuses in cases staged for Parkinson's disease-related brain pathology. *Neurosci. Lett.* **396**(1), 67–72 (2006).
- Dickson, D. W. *et al.* Neuropathology of non-motor features of Parkinson disease. *Parkinsonism Relat. Disord.* **15**(Suppl 3), S1–5. [https://doi.org/10.1016/S1353-8020\(09\)70769-2](https://doi.org/10.1016/S1353-8020(09)70769-2) (2009).
- Iwanaga, K. *et al.* Lewy body-type degeneration in cardiac plexus in Parkinson's and incidental Lewy body diseases. *Neurology* **52**(6), 1269–1271 (1999).
- Orimo, S. *et al.* Degeneration of cardiac sympathetic nerve begins in the early disease process of Parkinson's disease. *Brain Pathol.* **17**(1), 24–30. <https://doi.org/10.1111/j.1750-3639.2006.00032.x> (2007).
- Orimo, S. *et al.* Axonal alpha-synuclein aggregates herald centripetal degeneration of cardiac sympathetic nerve in Parkinson's disease. *Brain* **131**(Pt 3), 642–650 (2008).
- Fujishiro, H. *et al.* Cardiac sympathetic denervation correlates with clinical and pathologic stages of Parkinson's disease. *Mov. Disord.* **23**(8), 1085–1092 (2008).
- Postuma, R. B. *et al.* MDS clinical diagnostic criteria for Parkinson's disease. *Mov. Disord.* **30**(12), 1591–1601. <https://doi.org/10.1002/mds.26424> (2015).
- ESC/NASPE. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Eur. Heart J.* **17**(3), 354–381 (1996).
- Haapaniemi, T. H. *et al.* Ambulatory ECG and analysis of heart rate variability in Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **70**(3), 305–310 (2001).
- Kallio, M. *et al.* Comparison of heart rate variability analysis methods in patients with Parkinson's disease. *Med. Biol. Eng. Comput.* **40**(4), 408–414 (2002).
- Maetzler, W. *et al.* Time- and frequency-domain parameters of heart rate variability and sympathetic skin response in Parkinson's disease. *J. Neural Transm.* **122**(3), 419–425. <https://doi.org/10.1007/s00702-014-1276-1> (2015).
- Arnao, V. *et al.* Impaired circadian heart rate variability in Parkinson's disease: A time-domain analysis in ambulatory setting. *BMC Neurol.* **20**(1), 152. <https://doi.org/10.1186/s12883-020-01722-3> (2020).
- Bugalho, P. *et al.* Heart rate variability in Parkinson disease and idiopathic REM sleep behavior disorder. *Clin. Auton. Res.* **28**(6), 557–564. <https://doi.org/10.1007/s10286-018-0557-4> (2018).
- Cygankiewicz, I. & Zareba, W. Heart rate variability. *Handb. Clin. Neurol.* **117**, 379–393. <https://doi.org/10.1016/b978-0-444-53491-0.00031-6> (2013).
- Heimrich, K. G., Lehmann, T., Schlattmann, P. & Prell, T. Heart rate variability analyses in Parkinson's disease: A systematic review and meta-analysis. *Brain Sci.* <https://doi.org/10.3390/brainsci11080959> (2021).
- Li, Y. *et al.* Association between heart rate variability and Parkinson's disease: A meta-analysis. *Curr. Pharm. Des.* **27**(17), 2056–2067. <https://doi.org/10.2174/187152731966620090512222> (2021).
- Terroba-Chambi, C., Abulafia, C., Vigo, D. E. & Merello, M. Heart rate variability and mild cognitive impairment in Parkinson's disease. *Mov. Disord.* **35**(12), 2354–2355. <https://doi.org/10.1002/mds.28234> (2020).
- Kitagawa, T. *et al.* Association between heart rate variability and striatal dopamine depletion in Parkinson's disease. *J. Neural. Transm. (Vienna)* **128**(12), 1835–1840. <https://doi.org/10.1007/s00702-021-02418-9> (2021).
- Valappil, R. A. J. *et al.* Assessment of heart rate variability during wakefulness in patients with RBD. Abstract. *Mov. Disord.* **24**(Suppl. 1), S321–S322 (2009).
- Postuma, R. B., Lanfranchi, P. A., Blais, H., Gagnon, J. F. & Montplaisir, J. Y. Cardiac autonomic dysfunction in idiopathic REM sleep behavior disorder. *Mov. Disord.* **25**(14), 2304–2310. <https://doi.org/10.1002/mds.23347> (2010).
- Postuma, R. B., Gagnon, J. F., Bertrand, J. A., Genier Marchand, D. & Montplaisir, J. Y. Parkinson risk in idiopathic REM sleep behavior disorder: Preparing for neuroprotective trials. *Neurology* **84**(11), 1104–1113. <https://doi.org/10.1212/WNL.0000000000001364> (2015).
- Schenck, C. H., Boeve, B. F. & Mahowald, M. W. Delayed emergence of a parkinsonian disorder or dementia in 81% of older men initially diagnosed with idiopathic rapid eye movement sleep behavior disorder: A 16-year update on a previously reported series. *Sleep Med.* **14**(8), 744–748. <https://doi.org/10.1016/j.sleep.2012.10.009> (2013).
- Alonso, A., Huang, X., Mosley, T. H., Heiss, G. & Chen, H. Heart rate variability and the risk of Parkinson's disease: The Atherosclerosis Risk in Communities (ARIC) Study. *Ann. Neurol.* <https://doi.org/10.1002/ana.24393> (2015).
- Ahmadi Rastegar, D., Ho, N., Halliday, G. M. & Dzakmo, N. Parkinson's progression prediction using machine learning and serum cytokines. *Npj Parkinson's Dis.* **5**(1), 14. <https://doi.org/10.1038/s41531-019-0086-4> (2019).
- Akbalgic, O. *et al.* Electrocardiographic changes predate Parkinson's disease onset. *Sci. Rep.* **10**(1), 11319. <https://doi.org/10.1038/s41598-020-68241-6> (2020).
- Yuan, W. *et al.* Accelerating diagnosis of Parkinson's disease through risk prediction. *BMC Neurol.* **21**(1), 201. <https://doi.org/10.1186/s12883-021-02226-4> (2021).
- Alonso, A., Huang, X., Mosley, T. H., Heiss, G. & Chen, H. Heart rate variability and the risk of Parkinson disease: The Atherosclerosis Risk in Communities study. *Ann. Neurol.* **77**(5), 877–883. <https://doi.org/10.1002/ana.24393> (2015).
- Karabayir, I. *et al.* Predicting Parkinson's disease and its pathology via simple clinical variables. *J. Parkinsons Dis.* <https://doi.org/10.3233/JPD-212876> (2021).
- Shin, N.-Y. *et al.* Cortical thickness from MRI to predict conversion from mild cognitive impairment to dementia in Parkinson disease: A machine learning-based model. *Radiology* **300**(2), 390–399 (2021).
- Nishat, M. M., Hasan, T., Nasrullah, S. M., Faisal, F., Asif, M. A. A. R. & Hoque, M. A. Detection of Parkinson's disease by employing boosting algorithms 1–7. (IEEE, 2021).
- Laganas, C. *et al.* Parkinson's disease detection based on running speech data from phone calls. *IEEE Trans. Biomed. Eng.* **69**(5), 1573–1584 (2021).
- Lacy, S. E., Smith, S. L. & Lones, M. A. Using echo state networks for classification: A case study in Parkinson's disease diagnosis. *Artif. Intell. Med.* **86**, 53–59 (2018).

38. Hoq, M., Uddin, M. N. & Park, S.-B. Vocal feature extraction-based artificial intelligent model for Parkinson's disease detection. *Diagnostics* **11**(6), 1076 (2021).
39. Chahine, L. M. & Stern, M. B. Diagnostic markers for Parkinson's disease. *Curr. Opin. Neurol.* **24**(4), 309–317. <https://doi.org/10.1097/WCO.0b013e3283461723> (2011).
40. Yin, C. *et al.* Fractional dynamics foster deep learning of COPD stage prediction. *Adv. Sci.* **10**(12), 2203485. <https://doi.org/10.1002/advs.202203485> (2023).
41. Gupta, G., Yin, C., Deshmukh, J. V. & Bogdan P. Non-Markovian reinforcement learning using fractional dynamics 1542–1547 (2021).
42. Graves, A., Mohamed, A. R. & Hinton, G. Speech recognition with deep recurrent neural networks 6645–6649 (2013).
43. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735> (1997).
44. Vaswani, A., Shazeer, N., Parmar, N., *et al.* Attention is all you need. Presented at: *Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA* (2017).
45. Marttila, R., Rinne, U. & Marttila, R. Progression and survival in Parkinson's disease. *Acta Neurol. Scand.* **84**(S136), 24–28 (1991).
46. Ferguson, L. W., Rajput, A. H. & Rajput, A. Early-onset versus late-onset Parkinson's disease: A clinical-pathological study. *Can. J. Neurol. Sci.* **43**(1), 113–119 (2016).
47. Akbilgic, O. *et al.* ECG-AI: Electrocardiographic artificial intelligence model for prediction of heart failure. *Eur. Heart J. Digit. Health* <https://doi.org/10.1093/ehjdh/ztab080> (2021).
48. Akbilgic, O. *et al.* Artificial intelligence applied to ECG improves heart failure prediction accuracy. *J. Am. Coll. Cardiol.* **77**(18), 1 (2021).
49. Bond, R. R., Finlay, D. D., Nugent, C. D. & Moore, G. A review of ECG storage formats. *Int. J. Med. Inf.* **80**(10), 681–697 (2011).
50. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition 770–778 (2016).
51. Maas, A. L., Hannun, A. Y. & Ng, A. Y. Rectifier nonlinearities improve neural network acoustic models. *CiteSeer* **3**, 1 (2013).
52. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014).
53. Kingma, D. P., Ba, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
54. Ke, G., Meng, Q., Finley, T., *et al.* LightGBM: A highly efficient gradient boosting decision tree. Presented at: *Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017; Long Beach, California, USA* (2017).

Acknowledgements

We thank Michael J Fox Foundation for supporting this research (MJFF Grant ID 17267, PI Akbilgic). The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Author contributions

Types of authors' roles: 1. Research project: A. Conception, B. Organization, C. Execution; 2. Data A. Preparing and/or handling, B. Chart Reviews 3. Statistical and Machine Learning Analyses: A. Design, B. Execution, C. Review and Critique; 4. Manuscript: A. Writing of the First draft, B. Review and Critique. I.K. contributed to 2A, 3A, 3B, 3C, 4A, and 4B. F.G. contributed to 3A, 3B, 3C, 4A, and 4B. S.M.G. contributed to 1A, 1B, 1C, 1D, 4A, and 4B. R.K. and R.L.D. contributed to 1A, 1B, 1C, 1D, and 4B. K.C. contributed to 2B and 4B. L.C. contributed to 2A and 4B. J.L.J., K.B., W.R., H.P., K.M., L.B. and C.T. contributed to 4B. O.A. contributed to 1A, 1B, 1C, 1D, 3A, 3C, 4A, and 4B.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-38782-7>.

Correspondence and requests for materials should be addressed to S.M.G. or O.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023