JCN Open Access ORIGINAL ARTICLE

pISSN 1738-6586 / eISSN 2005-5013 / J Clin Neurol 2023;19(3):270-279 / https://doi.org/10.3988/jcn.2022.0160

# Long Short-Term Memory-Based Deep Learning Models for Screening Parkinson's Disease Using Sequential Diagnostic Codes

Seokjoon Yoon[a]
Minki Kim[a]
Woong-Woo Lee[b,c]

[a]College of Business, Korea Advanced Institute of Science and Technology, Seoul, Korea
[b]Department of Neurology, Nowon Eulji Medical Center, Eulji University, Seoul, Korea
[c]Department of Neurology, Eulji University College of Medicine, Daejeon, Korea

**Background and Purpose** It is challenging to detect Parkinson's disease (PD) in its early stages, which has prompted researchers to develop techniques based on machine learning methods for detecting PD. However, previous studies did not fully incorporate the slow progression of PD over a long period of time nor consider that its symptoms occur in a time-sequential manner. Contributing to the literature on PD, which has relied heavily on cross-sectional data, this study aimed to develop a method for detecting PD early that can process time-series information using the long short-term memory (LSTM) algorithm.

**Methods** We sampled 926 patients with PD and 9,260 subjects without PD using medical-claims data. The LSTM algorithm was tested using diagnostic histories, which contained the diagnostic codes and their respective time information. We compared the prediction power of the 12-month diagnostic codes under two different settings over the 4 years prior to the first PD diagnosis.

**Results** The model that was trained using the most-recent 12-month diagnostic codes had the best performance, with an accuracy of 94.25%, a sensitivity of 82.91%, and a specificity of 95.26%. The other three models (12-month codes from 2, 3, and 4 years prior) were found to have comparable performances, with accuracies of 92.27%, 91.86%, and 91.81%, respectively. The areas under the curve from our data settings ranged from 0.839 to 0.923.

**Conclusions** We explored the possibility that PD specialists could benefit from our proposed machine learning method as an early detection method for PD.

**Keywords** Parkinson's disease; deep learning; long short-term memory; medical-claims data; diagnostic code.

## INTRODUCTION

Parkinson's disease (PD) is a chronic neurodegenerative disorder that is clinically characterized by progressive motor dysfunctions including bradykinesia, resting tremor, rigidity, and postural instability.[1] Early detection of PD is challenging because many patients with PD—prior to showing such cardinal motor symptoms—could also have various prodromal nonmotor symptoms such as constipation, hyposmia, sleep disturbance (including REM sleep behavior disorder [RBD]), and depression. However, nonmotor features except for RBD are not so specific to PD.[2] Mild parkinsonism may also be missed in early PD because of its ambiguity. Hence, experts such as movement-disorder specialists find it difficult to detect PD in its early stage.[3,4]

Researchers have attempted to resolve this issue by making further developments in big-data processing and computer technology, with machine learning in particular being widely

**Received** April 19, 2022
**Revised** October 24, 2022
**Accepted** October 26, 2022

**Correspondence**
Woong-Woo Lee, MD, PhD
Department of Neurology, Nowon Eulji Medical Center, Eulji University, 68 Hangeulbiseok-ro, Nowon-gu, Seoul 01830, Korea
**Tel** +82-2-970-8312
**Fax** +82-504-267-0859
**E-mail** w2pooh@hanmail.net



 

used. While researchers have adopted machine learning methods for PD detection using various data sources, most studies focused on a specific measure related to motor symptoms, nonmotor symptoms, or other biomarkers, such as voice signals,[5-8] cardiovascular oscillations,[9] smell identification,[10] and motion signals.[11-14] Moreover, studies have used multiple measures of nonmotor symptoms (e.g., RBD and olfactory dysfunction) and biomarkers (e.g., cerebrospinal fluid measurements and SPECT imaging markers) in one model.[15,16] Deep learning with a convolutional neural network (CNN) structure has recently been widely adopted for detecting PD by processing image data, such as those from MRI, PET, and SPECT.[17-20]

Nevertheless, previous studies on PD detection using motor or nonmotor symptoms have barely accounted for the time-series information associated with PD progression in patients. PD symptoms occur in a time-sequential manner over a long period of time.[1,2,4] Generally, constipation appears first in the body-first type, and RBD, depression, dopaminergic PET/SPECT findings, and subtle parkinsonism tend sequentially follow.[2,21] Although the symptoms occur in a different sequence in the brain-first type, both subtypes have a common feature in that their symptoms progress sequentially. It can therefore be valuable to utilize the time-series information in the early detection of PD. However, most studies that we were aware of that explored the methods for PD detection typically relied on cross-sectional data without sufficiently reflecting time-dependent information. For comparison purposes, we have included Table 1, which summarizes the previous literature that adopted machine learning methods for PD detection.

To explore the extent to which a machine learning approach can help PD detection by processing time-series information, we investigated the performance of the long short-term memory (LSTM) model with a recurrent neural network (RNN) architecture for incorporating PD symptoms that occurred in a time-sequential manner over a long period of time. More specifically, we used sequential information from medical-claims data, which contained diagnostic codes from the visit of each patient.

## METHODS

This study was conducted in accordance with the Declaration of Helsinki. The analysis was performed retrospectively, and all subjects were anonymized and deidentified. Written informed consent for accessing the clinical records was not required because of the guaranteed anonymity and deidentification. The study protocol was approved by the Institutional Review Board of Korea Advanced Institute of Science and Technology (No.18-056).

### Data and subjects

We used the National Sample Cohort data from the National Health Insurance Service of Korea (NHIS-NSC) for the time period of 2002–2013. These cohort data include information on demographics, medical histories, medications, medical claims, and regular health checkup data.[22] The NHIS-NSC is

**Table 1.** Comparison of previous machine learning approaches for PD detection

| Study | Methods of PD detection | Data type | Time-series information |
|---|---|---|---|
| Karapinar Senturk (2020)[6] | CART, ANN, and SVM | Voice signals | X |
| Almeida et al. (2019)[5] | KNN, MLP, optimum-path forest, and SVM | Voice signals | X |
| Lahmiri and Shmuel (2019)[7] | SVM and Bayesian optimization technique | Voice signals | X |
| Valenza et al. (2016)[9] | SVM | Spontaneous cardiovascular oscillations | X |
| Silveira-Moriyama et al. (2009)[10] | Logistic regression | Smell identification data | X |
| El Maachi et al. (2020)[13] | CNN | Gait signals | X |
| Abdulhay et al. (2018)[11] | Medium tree and medium Gaussian SVM | Gait impairments and tremor occurrences | X |
| Zeng et al. (2016)[14] | Radial basis function neural networks | Gait patterns | X |
| Prashanth et al. (2016)[15] | SVM, naïve Bayes, boosted trees, and RF | Nonmotor symptoms and biomarkers (PPMI) | X |
| Wang et al. (2020)[16] | Three deep learning models that were all feed-forward neural networks with two hidden layers | Nonmotor symptoms and biomarkers (PPMI) | X |
| Chakraborty et al. (2020)[17] | CNN | 3-T T1-weighted MRI scans (PPMI) | X |
| Shinde et al. (2019)[19] | CNN | NMS-MRI | X |
| Searles Nielsen et al. (2017)[35] | Logistic regression and elastic net algorithm | Medical-claims data | X |
| Current study | LSTM | Medical-claims data (diagnostic codes) | O |

ANN, artificial neural network; CART, classification and regression trees; CNN, convolutional neural network; KNN, K-nearest neighbors; LSTM, long short-term memory; MLP, multilayer perceptron; NMS, neuromelanin sensitive; PD, Parkinson's disease; PPMI, Parkinson's Progression Markers Initiative; RF, random forest; SVM, support vector machine.

a population-based sample cohort that contains information on 1,025,340 participants, which is equivalent to 2.2% of the total population of Korea, and the sample was randomly selected from the Korean National Health Insurance database. Among the cohort data, we used diagnosis database, constructed based on the medical claims. The data include information on patient ID, diagnosis date, diagnosis code, and code of the department where the diagnosis was performed. The diagnostic codes in the NHIS-NSC data followed the criteria from the International Classification of Diseases-10 (ICD-10) provided by the World Health Organization. The diagnostic codes contain information about disease categories and subcategories.

Among the sample cohort data, we first selected patients with PD and their counterparts (subjects without PD) based on the age and sex distribution of the patients. We defined patients with PD as those diagnosed using the diagnostic code G20 by the Neurology Department, and who had never been diagnosed with another parkinsonism-related disease (code G11.0–G11.9, G21.0–G21.9, G23.0–G23.9, or G31.8) before or after the PD diagnosis. We monitored at least 4 years of diagnostic history after the first code-G20 diagnosis for every patient with PD to determine whether they were misdiagnosed. We included patients with PD older than 50 years who have visited outpatient clinics more than once and were diagnosed with PD for the first time during 2006–2009. We chose age- and sex-matched subjects without PD, including those who had not been diagnosed with PD or other parkinsonism-related diseases during the data period (PD:non-PD=1:10).

## Model

We considered LSTM with an RNN architecture so as to incorporate PD symptoms that occur in a time-sequential manner over a long period of time.[23] More specifically, we employed one hidden-layer bidirectional LSTM (bi-LSTM) trained on two-class (i.e., PD or non-PD) classifications. An RNN-based model processes entire data sequences (e.g., sentences, speeches, or videos) instead of single data points (e.g., images) and returns its sequential output. Consequently, researchers have applied RNN-based models to areas where time-series information is more important than cross-sectional information, such as natural language processing, speech recognition, or anomaly detection in network traffic.[24-26] Details on the mechanism and logic of LSTM are presented in the Supplementary Material (in the online-only Data Supplement).

The input for our proposed model was a sequence of diagnostic codes for each patient. For example, the input vector of patient i, who was diagnosed T times at their outpatient clinic during the data period, was presented as $X_i=[x_{i1}, x_{i2}, \cdots, x_{iT}]$. Each diagnosis was converted to a diagnosis embedding of dimension 128. The interactions of input $x_{it}$ in both the previous and next input period ($x_{i(t-1)}$ and $x_{i(t+1)}$) were included by using bi-LSTM, which comprised forward and backward processing. The size of the hidden layer in bi-LSTM is 128. After the LSTM layer, a fully connected layer specified the two classes (i.e., PD and non-PD) and the softmax layer calculated the probability of each class as an output. The focal loss was calculated in the output layer. This loss function is often used in a classification problem with imbalanced classes.[27] A flowchart of our model and detailed examples are shown in Fig. 1A.

## Data setup

We tested the model performance using 4 different data-period settings. Clinical diagnoses that were strongly related to PD tended to occur in the periods closer to the time of the PD diagnosis. First, we used sequential 12-month data on diagnostic codes from 1 year prior to the first PD diagnosis (year -1 to year 0; diagnosis of PD in year 0). To determine the performance of the model in early PD detection, we used settings for 12-month data from the 2, 3, and 4 years prior to the first PD diagnosis (year -2 to year -1, year -3 to year -2, and year -4 to year -3, respectively). We defined these data periods as Periods 1, 2, 3, and 4, as presented in Fig. 1B.

Our diagnostic codes comprised three to five letters. The first three letters contained relatively broad diagnostic information, and the fourth and fifth letters indicated more specific subdivisions of the disease. In our study, we only used the first three letters of diagnostic codes to prevent an excessive amount of inputs, thereby avoiding overfitting and achieving greater applicability for the model. We trained our model using two different diagnostic-code settings by modifying the input set of diagnostic codes. Using three-letter diagnostic codes, the number of unique diagnostic codes were 1,032, 1,014, 987, and 1,014 for Periods 1, 2, 3, and 4, respectively. Setting 1, as a baseline model, covered every three-letter diagnostic code as inputs. Because Setting 1 used every diagnostic code, the data inputs involved relatively rare symptoms and the amount of data required for model training was quite large. We therefore only used diagnostic codes that corresponded to the top 100 frequencies in the model for Setting 2.

With 4 different data periods (i.e., Periods 1, 2, 3, and 4) and 2 different diagnostic-code settings (i.e., settings with all and with the top 100 diagnostic codes), 8 data settings were trained and tested by the model. We divided each setting of the data set into the training (70%) and testing (30%) sets. We partitioned the patients using stratified sampling; that is, both sets had the same proportion of patients with PD and subjects without PD in each input. For each setting, we repeated the process 100 times by randomly splitting the sample data into training and testing data. The model was trained us-

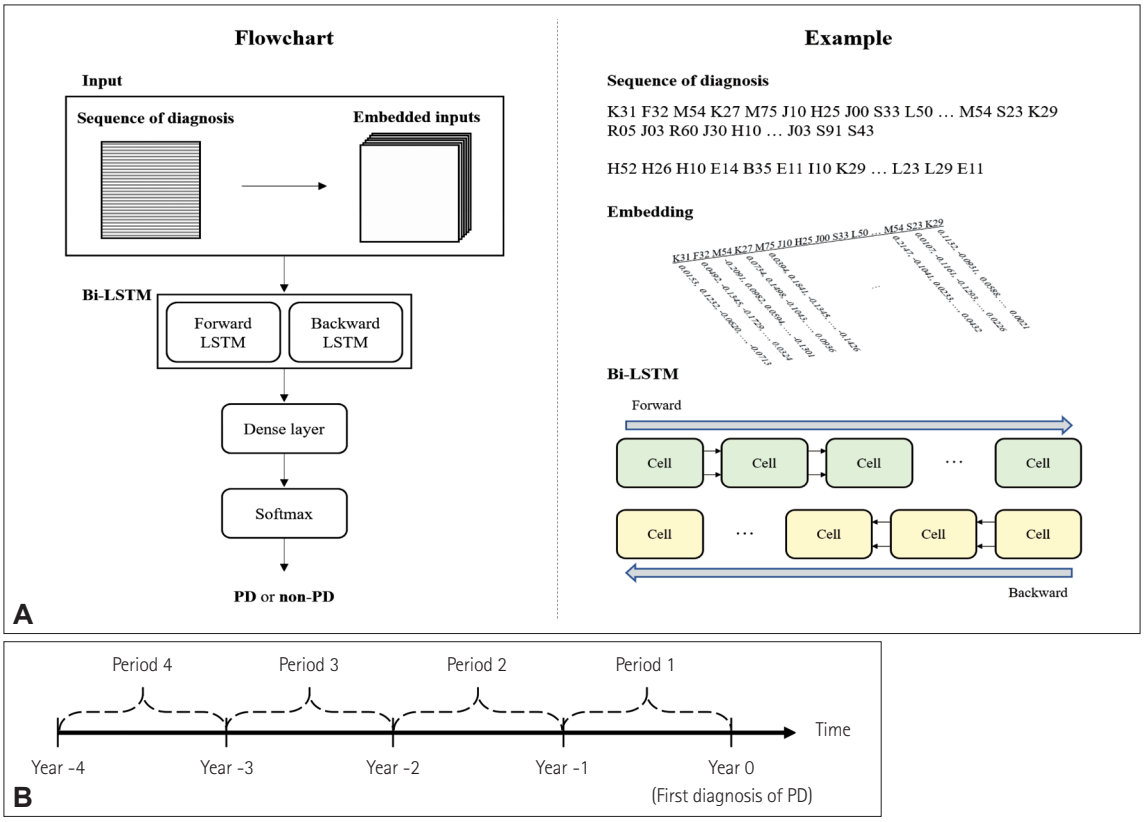


**Fig. 1.** Summary of model and data setup. A: Flowchart of the model. B: Data setup with four different periods. LSTM, long short-term memory; PD, Parkinson's disease.

ing the training data, and patients in the testing data were predicted to be PD or non-PD using the trained model. We computed the accuracy, sensitivity, specificity, and areas under the receiver operating characteristic (ROC) curves (AUCs) for the testing data.

We compared the diagnostic-code distributions of the PD and non-PD groups to determine which diagnostic codes appeared frequently in the PD group. The diagnostic-code distribution was summarized for each data-period setting. We first summarized the top 20 diagnostic codes, which frequently occurred in patients with PD during the period, and we presented the corresponding ranks of diagnostic codes for subjects without PD.

## RESULTS

We obtained summary statistics for 926 patients with PD and 9,260 subjects without PD sampled in this study. The female-to-male ratio of our sample was 0.625, and the age at diagnosis was 69.05±8.41 years (mean±standard deviation) in both the PD and non-PD groups. The total numbers of diagnoses in our sample over the course of 4 years from the first PD diagnosis were 238,735 and 1,733,398 cases in the PD and non-PD groups, respectively. The patients with PD and subjects without PD in our sample visited clinics 85.9 and 62.4 times on average per year, respectively. Also, the patients with PD (28.9 times/year) visited clinics more frequently than the subjects without PD (23.4 times/year). The prevalence and incidence rates of PD in our study are listed in Table 2.

**Table 2.** Prevalence and incidence of PD in this study

| | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|
| Prevalence | 545 | 670 | 806 | 932 |
| Incidence | 188 | 223 | 258 | 257 |

Data are presented as *n*.

### Model performance

Table 3 lists the model performance measured on the testing data for each setting. The performance—quantified as the accuracy, sensitivity, and specificity—for Setting 1 according to period were 94.25%, 82.91%, and 95.26%, respectively, for Period 1; 92.27%, 77.68%, and 93.61% for Period 2; 91.86%, 75.38%, and 93.42% for Period 3; and 91.81%, 75.65%, and 93.28% for Period 4. The settings that covered Period 1 demonstrated better performance than those that covered Periods 2, 3, and 4. Among the settings for each data period, Setting 1 (all 3-letter diagnostic codes) demonstrated better performance than Setting 2 (only the top 100 diagnostic codes).

The AUCs of Setting 1 were 0.923, 0.890, 0.870, and 0.870 for Periods 1, 2, 3, and 4, respectively. Fig. 2 presents the ROC plot for one testing data set in each setting.

**Table 3.** Model performance comparisons across data settings

| Data period | Data setting | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|---|---|---|---|---|
| Period 1 | Setting 1 | 94.25±0.53 | 82.91±2.53 | 95.26±0.61 |
| | Setting 2 | 93.36±0.52 | 79.35±2.77 | 94.62±0.57 |
| Period 2 | Setting 1 | 92.27±0.52 | 77.68±2.92 | 93.61±0.62 |
| | Setting 2 | 91.73±0.76 | 75.09±2.92 | 93.26±0.85 |
| Period 3 | Setting 1 | 91.86±0.56 | 75.38±8.28 | 93.42±0.93 |
| | Setting 2 | 91.59±0.67 | 73.93±8.04 | 93.25±1.03 |
| Period 4 | Setting 1 | 91.81±0.68 | 75.65±2.90 | 93.28±0.77 |
| | Setting 2 | 91.49±0.72 | 70.91±7.82 | 93.36±1.04 |

Data are presented as mean±standard deviation.

## Distribution of diagnostic codes

Comparisons of the diagnostic-code distributions of the PD and non-PD groups are listed in Table 4. The top 20 diagnostic codes that frequently occurred in patients with PD during the period were summarized in descending order in the Table 4. We also report the corresponding ranks for subjects without PD. Fig. 3 presents the diagnostic codes with ranks that were at least ten times higher in the PD group than in the non-PD group.

In Period 1, depressive episode had the largest rank difference, sequentially followed by cerebral infarction, vestibular function disorder, other anxiety disorders, light headedness, angina pectoris, gastric ulcer, and shoulder lesions. In Period 2, depressive episode and cerebral infarction presented the largest rank difference. While vestibular function disor-

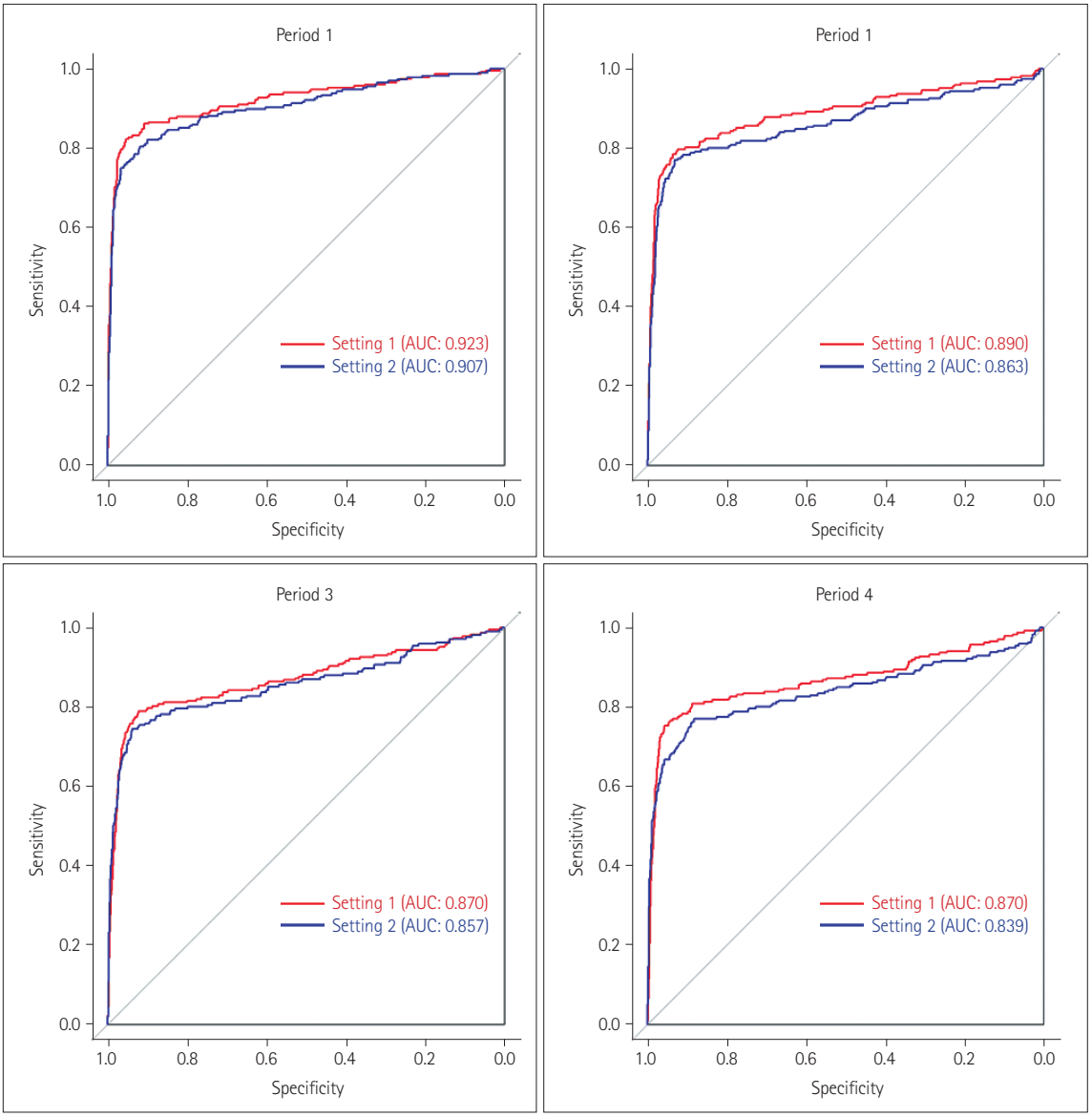


**Fig. 2.** Receiver operating characteristic (ROC) plots for all settings. AUC, area under the ROC curve.

**Table 4.** Frequency ranking of diagnostic codes for each data period

| Period 1 | | | | Period 2 | | | | Period 3 | | | | Period 4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Diagnostic code | Frequency ranking | | Rank difference | Diagnostic code | Frequency ranking | | Rank difference | Diagnostic code | Frequency ranking | | Rank difference | Diagnostic code | Frequency ranking | | Rank difference |
| | PD | non-PD | | | PD | non-PD | | | PD | non-PD | | | PD | non-PD | |
| I10 | 1st | 1st | 0 | I10 | 1st | 1st | 0 | I10 | 1st | 1st | 0 | I10 | 1st | 1st | 0 |
| K29 | 2nd | 2nd | 0 | K29 | 2nd | 2nd | 0 | K29 | 2nd | 2nd | 0 | K29 | 2nd | 2nd | 0 |
| E11 | 3rd | 4th | -1 | E11 | 3rd | 3rd | 0 | E11 | 3rd | 3rd | 0 | E11 | 3rd | 3rd | 0 |
| E78 | 4th | 6th | -2 | M54 | 4th | 5th | -1 | M17 | 4th | 4th | 0 | M54 | 4th | 5th | -1 |
| M17 | 5th | 3rd | 2 | E78 | 5th | 6th | -1 | M54 | 5th | 5th | 0 | M17 | 5th | 4th | 1 |
| M54 | 6th | 5th | 1 | M17 | 6th | 4th | 2 | E78 | 6th | 6th | 0 | E78 | 6th | 6th | 0 |
| I63* | 7th | 37th | -30 | K59 | 7th | 14th | -7 | J20 | 7th | 7th | 0 | I63* | 7th | 40th | -33 |
| F32* | 8th | 47th | -39 | I63* | 8th | 38th | -30 | K59* | 8th | 19th | -11 | K59 | 8th | 16th | -8 |
| K59 | 9th | 12th | -3 | J20 | 9th | 7th | 2 | I63* | 9th | 40th | -31 | J20 | 9th | 7th | 2 |
| F41* | 10th | 34th | -24 | J30 | 10th | 8th | 2 | J30 | 10th | 8th | 2 | K25 | 10th | 15th | -5 |
| K25* | 11th | 23rd | -12 | M81 | 11th | 9th | 2 | M81 | 11th | 9th | 2 | J30 | 11th | 8th | 3 |
| K21 | 12th | 20th | -8 | F41* | 12th | 35th | -23 | K25 | 12th | 14th | -2 | H10 | 12th | 14th | -2 |
| J30 | 13th | 7th | 6 | H04 | 13th | 12th | 1 | H10 | 13th | 12th | 1 | F41* | 13th | 34th | -21 |
| M75* | 14th | 25th | -11 | K25 | 14th | 18th | -4 | H25 | 14th | 17th | -3 | I20* | 14th | 24th | -10 |
| J20 | 15th | 8th | 7 | F32* | 15th | 46th | -31 | J45 | 15th | 18th | -3 | K58* | 15th | 25th | -10 |
| I20* | 16th | 30th | -14 | I20* | 16th | 28th | -12 | N40 | 16th | 23th | -7 | K30 | 16th | 10th | 6 |
| R42* | 17th | 38th | -21 | H25 | 17th | 15th | 2 | M13 | 17th | 10th | 7 | M13 | 17th | 9th | 8 |
| H04 | 18th | 13th | 5 | K58* | 18th | 30th | -12 | F41* | 18th | 36th | -18 | M81 | 18th | 11th | 7 |
| N40 | 19th | 24th | -5 | R42* | 19th | 36th | -17 | H04 | 19th | 13th | 6 | M75 | 19th | 26th | -7 |
| H81* | 20th | 50th | -30 | N40 | 20th | 25th | -5 | I20 | 20th | 27th | -7 | J45 | 20th | 23rd | -3 |

*Denotes diagnostic code (disease) with a ranking difference between the PD and non-PD groups of at least.
E11, type 2 diabetes mellitus; E78, disorders of lipoprotein metabolism and other lipidemias; F32, depressive episode; F41, other anxiety disorders; H04, lacrimal system disorders; H10, conjunctivitis; H25, senile cataract; H81, vestibular function disorders; I10, high blood pressure; I20, angina pectoris; I63, cerebral infarction; J20, acute bronchitis; J30, spasmodic rhinorrhea; J45, asthma; K21, gastro-oesophageal reflux disease; K25, gastric ulcer; K29, gastritis and duodenitis; K30, indigestion; K58, irritable colon; K59, other functional intestinal disorders; M13, other arthritis; M17, gonarthrosis (arthrosis of the knee); M54, dorsalgia; M75, shoulder lesions; M81, osteoporosis without pathological fracture; N40, prostate hyperplasia; R42, light headedness; PD, Parkinson's disease.

der, gastric ulcer, and shoulder lesions disappeared, irritable colon first presented in this period. In Period 3, at least ten rank differences were observed in cerebral infarction, other anxiety disorders, and other functional intestinal disorders. In Period 4, cerebral infarction, other anxiety disorders, angina pectoris, and irritable colon were observed.

## DISCUSSION

Every model setting for each data period demonstrated good prediction performance, with AUCs value higher than 0.839 (0.870 when only considering Setting 1), and the model with Period 1 presented the best performance. This was a predictable result, because more PD-specific symptoms are likely to develop around the time of a PD diagnosis, and those symptoms provide key information when detecting PD. Meanwhile, it is worth noting that the models set up with Periods 2, 3, and 4 showed comparable prediction performances. Given that Periods 2, 3, and 4 covered the time farther past from the diagnosis of PD, we can claim that our proposed model detects PD with symptoms less specific to PD.

To provide supplementary information for clinicians, we compared the annual number of visits and diagnoses of both groups. We found that patients with PD visited clinics more frequently and had more diagnoses, meaning that people with PD have more complaints from the prodromal phase. To identify which diagnoses were linked to prodromal PD, we investigated the diagnostic codes that were prevalent in the PD group for each data period. The diagnostic codes, which were only highlighted in Period 1, were H81 (vestibular function disorder), K25 (gastric ulcer), and M75 (shoulder lesions). Codes H81 and M75 may reflect subtle parkinsonism, including postural instability and rigidity-related shoulder problems in the prediagnostic period. Code K25 may be related to decreased gastrointestinal motility, which is a nonmotor symptom of PD. Code F32 (depressive episode), which was the most

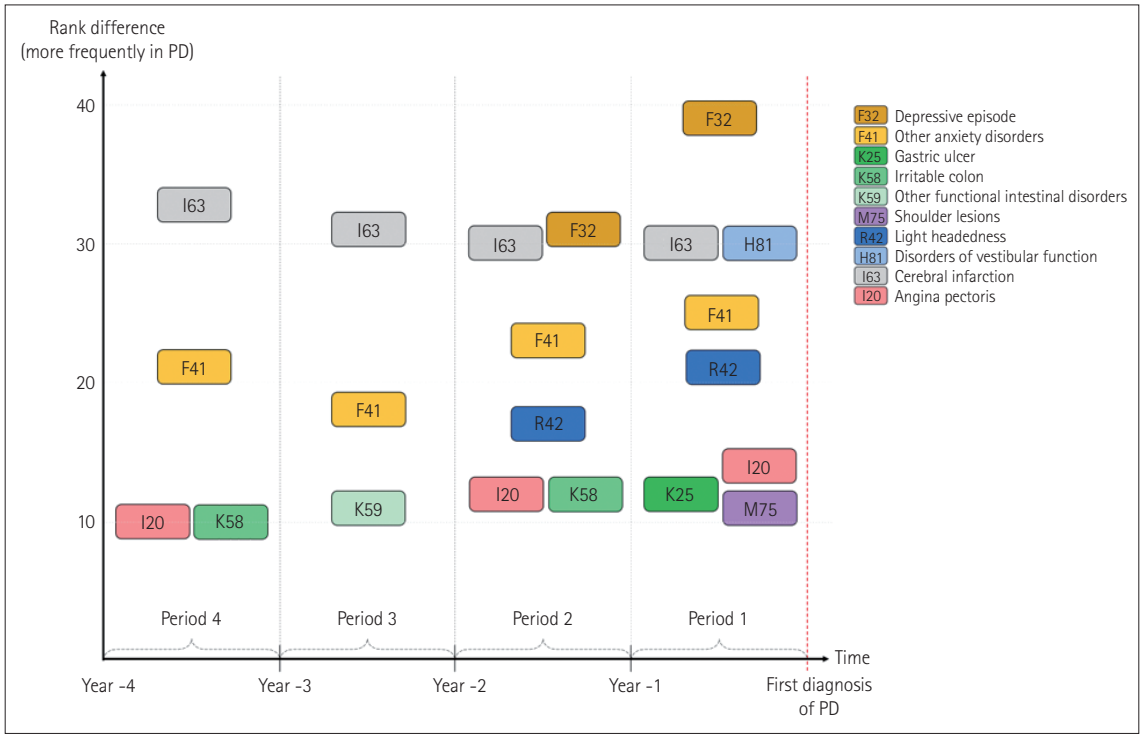


**Fig. 3.** Diagnostic codes that presented more frequently in the Parkinson's disease (PD) group than in the non-PD group.

common in Periods 1 and 2, is a common prodromal nonmotor symptom in PD.[1,2] In Periods 3 and 4, codes I63 (cerebral infarction), F41 (other anxiety disorders), K58 (irritable colon), K59 (other functional intestinal disorders), and I20 (angina pectoris) occurred the most frequently in patients with PD. Constipation and anxiety are common symptoms in the prodromal stage of PD,[1,2] which could explain the high frequency of intestinal symptoms and anxiety disorders during this stage. Codes reflecting motor symptoms (i.e., R42, H81, and M75) did not appear in Periods 3 and 4. This was consistent with the previous studies finding that nonmotor symptoms such as anxiety and intestinal problems preceded motor symptoms.[1,2]

Diagnostic code I63 (cerebral infarction) was observed frequently regardless of the data period. There have been several reports of an association between cerebral infarction and PD.[28-30] Although the pathogenesis on how ischemic stroke causes PD is still unclear, the inflammatory process after neuronal ischemia may play a role in developing a PD-like pathology.[31] It should also be considered that a minor or incidental silent infarction could mimic the motor symptoms of PD.

Among the significant prodromal symptoms for PD, RBD and hyposmia were not detected in the current analysis. In particular, RBD is the strongest diagnostic marker for prodromal PD.[2] RBD was not captured because the ICD-10 system does not contain an appropriate code representing RBD. Hyposmia seems to have been ignored by patients without olfactory function tests, perhaps because they perceived the symptoms as being less serious than other prodromal symptoms.[32] Although RBD and hyposmia did not contribute directly, the current LSTM model based on prediagnostic medical-claims data performed well, which could be attributable to its ability to incorporate sequential information for these diagnostic codes.

The amount of data used for the model training was large in Setting 1 (all 3-letter diagnostic codes), in which there were approximately 1,000 unique diagnostic codes. Concerning the data size required for model training, Setting 2 reduced the number of unique diagnostic codes to 100 by only using the top 100 codes in terms of frequency. The results confirmed that the model demonstrated decent performance, even under the condition in which the data input only covered 100 major diagnostic codes.

Table 5 summarizes the prediction performances from the previous literature on PD detection. Previous studies that used gait or voice signals showed relatively high accuracy compared with our study.[5,7,11,13,14] The studies that included nonmotor symptoms and biomarkers of PPMI also performed well in detecting PD.[15,16] The studies that used MRI, PET/CT, or SPECT images in a CNN model demonstrated good performances.[17-19] However, these studies used fine-motor features or well-refined data that are relatively inaccessible. Most of the studies generated their own customized data with spe-

cific study ideas. To use the prediction method as an early detection tool for patients with PD, the data must be easy to access, and hence the primary data used in previous studies is of little use in the development of PD detection tools.

In contrast, considering that our model used medical-claims data, the performance of the model indicated that it would work reasonably well as a preliminary screening tool, such as helping to detect at-risk patients in the early stage and developing health policies for them. Medical-claims data are secondary data related to the health information of patients that is communicated with insurance providers. Although records in medical-claims data may not always represent the exact state of a patient, the cost of accessing the data is low, and the data cover a wide range of patients. Moreover, more countries (e.g., the United States, Finland, and the United Kingdom) are now providing open access to such data (e.g., the United States National Institutes of Health, Findata of Finland, and the National Health Service of the United Kingdom). Researchers have used medical-claims data to predict the onset of diseases such as heart failure, diabetes, and cancer.[33,34] Searles Nielsen et al.[35] demonstrated that analyzing medical-claims data from Medicare using logistic regression and an elastic net algorithm was effective at detecting PD (Table 5).

This study may have had limitations. First, we cannot fully exclude the possibility that other factors affected the development of PD. Although we revealed the distribution of diagnostic codes for PD and non-PD groups in each data period (Table 4), this could only be a descriptive explanation for the results of the model. The factors that affected the model—to predict a patient as PD or non-PD—could be precisely investigated using artificial intelligence methods as an option in the deep learning area, which we left to future research. Second, not all patients with PD followed specific time-sequential progression, such as the Braak stages or body-first subtype.[4,36] There were many cases of clinical heterogeneity that did not show prodromal nonmotor symptoms before motor

**Table 5.** Performance predictions of the previous literature on PD detection

| Study | Methods of PD detection | Data type | | Performance (%) | | |
|---|---|---|---|---|---|---|
| | | | | Accuracy | Sensitivity | Specificity |
| Karapinar Senturk (2020)[6] | CART, ANN, and SVM | Voice signals | | 93.84 | - | - |
| Almeida et al. (2019)[5] | KNN, MLP, optimum-path forest, and SVM | Voice signals | | 94.55 | - | - |
| Lahmiri and Shmuel (2019)[7] | SVM and Bayesian optimization technique | Voice signals | | 92.13 | 82.79 | 95.27 |
| Valenza et al. (2016)[9] | SVM | Spontaneous cardiovascular oscillations | | 73.47 | 80.83 | 76.00 |
| Silveira-Moriyama et al. (2009)[10] | Logistic regression | Smell identification data | | - | 90.40 | 85.50 |
| El Maachi et al. (2020)[13] | CNN | Gait signals | | 98.70 | - | - |
| Abdulhay et al. (2018)[11] | Medium tree and medium Gaussian SVM | Gait impairments and tremor occurrences | | 92.70 | - | - |
| Zeng et al. (2016)[14] | Radial basis function neural networks | Gait patterns | | 96.39 | 96.77 | 95.89 |
| Prashanth et al. (2016)[15] | SVM, naïve Bayes, boosted trees, and RF | Nonmotor symptoms and biomarkers (PPMI) | | 96.40 | 97.03 | 95.01 |
| Wang et al. (2020)[16] | Three deep learning models that were all feed-forward neural networks with two hidden layers | Nonmotor symptoms and biomarkers (PPMI) | | 96.45 | - | - |
| Kotsavasiloglou et al. (2017)[37] | Naïve Bayes, AdaBoost, logistic regression, SVM | Handwriting | | 90.90 | 88.00 | 95.00 |
| Gupta et al. (2020)[38] | SVM ranking method | Handwriting | | 83.75 | - | - |
| Arroyo-Gallego et al. (2017)[39] | Logistic regression, SVM, AdaBoost | Mobile touchscreen typing | | - | 81.00 | 81.00 |
| Chakraborty et al. (2020)[17] | CNN | 3-T T1-weighted MRI scans (PPMI) | | 95.29 | 92.70 | 94.30 |
| Shinde et al. (2019)[19] | CNN | NMS-MRI | | 80.00 | - | - |
| Searles Nielsen et al. (2017)[35] | Logistic regression and elastic net algorithm | Medical-claims data | | - | 73.50 | 83.20 |
| Current study | LSTM | Medical-claims data (diagnostic codes) | Period 1 | 94.25 | 82.91 | 95.26 |
| | | | Period 2 | 92.27 | 77.68 | 93.61 |
| | | | Period 3 | 91.86 | 75.38 | 93.42 |
| | | | Period 4 | 91.81 | 75.65 | 93.28 |

ANN, artificial neural network; CART, classification and regression trees; CNN, convolutional neural network; KNN, K-nearest neighbors; LSTM, long short-term memory; MLP, multilayer perceptron; NMS, neuromelanin sensitive; PD, Parkinson's disease; PPMI, Parkinson's Progression Markers Initiative; RF, random forest; SVM, support vector machine.

dysfunction onset. While our model could capture the prevalent sequential patterns of the diagnoses of patients with PD, it might fail to capture abnormal patterns. Third, the prevalence and incidence of PD in our study were differed slightly from those found by Lee et al.,[22] who analyzed the same NHIS-NSC cohort data. During the diagnostic period (2006–2009), we enrolled 926 cases with PD, whereas that previous study enrolled 1,106. This discrepancy in incidence was probably caused by the application of different diagnostic criteria. We narrowed down our sample by guaranteeing at least 4 years of monitoring to exclude other parkinsonian disorders, because our study had a case–control design. Finally, there may have been inaccuracies in the disease diagnoses in the medical-claims data. The diagnostic codes were recorded based on the decisions of physicians, and the codes in medical-claims data may therefore not represent the exact states of patients. Though we selected patients with PD using solid criteria to overcome such issues, there may still have been contaminated samples.

In conclusion, this study found that PD detection with time-series information is feasible by using patient-level diagnostic codes. Given the possibility of early PD detection using prodromal data obtained 2–4 years before the first diagnosis, we believe that our proposed method would be useful as a preliminary screening program for people at risk of PD.

## Supplementary Materials

The online-only Data Supplement is available with this article at https://doi.org/10.3988/jcn.2022.0160.

### Availability of Data and Material

The datasets generated or analyzed during the study are available from the corresponding author on reasonable request.

### ORCID iDs

Seokjoon Yoon https://orcid.org/0000-0002-3033-4680
Minki Kim https://orcid.org/0000-0001-8696-2494
Woong-Woo Lee https://orcid.org/0000-0002-8767-7967

### Author Contributions

Conceptualization: all authors. Data curation: Seokjoon Yoon. Formal analysis: Seokjoon Yoon. Funding acquisition: Minki Kim. Investigation: Seokjoon Yoon. Methodology: Seokjoon Yoon. Project administration: Minki Kim, Woong-Woo Lee. Resources: Minki Kim. Software: Seokjoon Yoon. Supervision: Minki Kim, Woong-Woo Lee. Validation: Seokjoon Yoon. Visualization: Seokjoon Yoon. Writing—original draft: Seokjoon Yoon, Woong-Woo Lee. Writing—review & editing: all authors.

### Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

### Funding Statement

This paper is based on the first author's doctoral dissertation. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A1A14045741).

## REFERENCES

1. Kalia LV, Lang AE. Parkinson's disease. *Lancet* 2015;386:896-912.
2. Postuma RB, Berg D. Advances in markers of prodromal Parkinson disease. *Nat Rev Neurol* 2016;12:622-634.
3. Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord* 2015;30:1591-1601.
4. Tolosa E, Garrido A, Scholz SW, Poewe W. Challenges in the diagnosis of Parkinson's disease. *Lancet Neurol* 2021;20:385-397.
5. Almeida JS, Rebouças Filho PP, Carneiro T, Wei W, Damaševičius R, Maskeliūnas R, et al. Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques. *Pattern Recognit Lett* 2019;125:55-62.
6. Karapinar Senturk Z. Early diagnosis of Parkinson's disease using machine learning algorithms. *Med Hypotheses* 2020;138:109603.
7. Lahmiri S, Shmuel A. Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine. *Biomed Signal Process Control* 2019;49:427-433.
8. Parisi L, RaviChandran N, Manaog ML. Feature-driven machine learning to improve early diagnosis of Parkinson's disease. *Expert Syst Appl* 2018;110:182-190.
9. Valenza G, Orsolini S, Diciotti S, Citi L, Scilingo EP, Guerrisi M, et al. Assessment of spontaneous cardiovascular oscillations in Parkinson's disease. *Biomed Signal Process Control* 2016;26:80-89.
10. Silveira-Moriyama L, Petrie A, Williams DR, Evans A, Katzenschlager R, Barbosa ER, et al. The use of a color coded probability scale to interpret smell tests in suspected parkinsonism. *Mov Disord* 2009;24:1144-1153.
11. Abdulhay E, Arunkumar N, Narasimhan K, Vellaiappan E, Venkatraman V. Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease. *Future Gener Comput Syst* 2018;83:366-373.
12. Cavallo F, Moschetti A, Esposito D, Maremmani C, Rovini E. Upper limb motor pre-clinical assessment in Parkinson's disease using machine learning. *Parkinsonism Relat Disord* 2019;63:111-116.
13. El Maachi I, Bilodeau GA, Bouachir W. Deep 1D-convnet for accurate Parkinson disease detection and severity prediction from gait. *Expert Syst Appl* 2020;143:113075.
14. Zeng W, Liu F, Wang Q, Wang Y, Ma L, Zhang Y. Parkinson's disease classification using gait analysis via deterministic learning. *Neurosci Lett* 2016;633:268-278.
15. Prashanth R, Dutta Roy S, Mandal PK, Ghosh S. High-accuracy detection of early Parkinson's disease through multimodal features and machine learning. *Int J Med Inform* 2016;90:13-21.
16. Wang W, Lee J, Harrou F, Sun Y. Early detection of Parkinson's disease using deep learning and machine learning. *IEEE Access* 2020;8:147635-147646.
17. Chakraborty S, Aich S, Kim HC. Detection of Parkinson's disease from 3T T1 weighted MRI scans using 3D convolutional neural network. *Diagnostics (Basel)* 2020;10:402.
18. Choi H, Kim YK, Yoon EJ, Lee JY, Lee DS; Alzheimer's Disease Neuroimaging Initiative. Cognitive signature of brain FDG PET based on deep learning: domain transfer from Alzheimer's disease to Parkinson's disease. *Eur J Nucl Med Mol Imaging* 2020;47:403-412.
19. Shinde S, Prasad S, Saboo Y, Kaushick R, Saini J, Pal PK, et al. Predictive markers for Parkinson's disease using deep neural nets on neuromelanin sensitive MRI. *Neuroimage Clin* 2019;22:101748.
20. Magesh PR, Myloth RD, Tom RJ. An explainable machine learning model for early detection of Parkinson's disease using LIME on DaTSCAN imagery. *Comput Biol Med* 2020;126:104041.
21. Horsager J, Andersen KB, Knudsen K, Skjærbæk C, Fedorova TD, Okkels N, et al. Brain-first versus body-first Parkinson's disease: a multimodal imaging case-control study. *Brain* 2020;143:3077-3088.
22. Lee J, Lee JS, Park SH, Shin SA, Kim K. Cohort profile: the national

health insurance service–national sample cohort (NHIS-NSC), South Korea. *Int J Epidemiol* 2017;46:e15.
23. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9:1735-1780.
24. Graves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing*; 2013 May 26-31; Vancouver, Canada: IEEE; 2013:6645-6649.
25. Malhotra P, Vig L, Shroff G, Agarwal P. Long short term memory networks for anomaly detection in time series. *Proceedings of the 23rd European Symposium on Artificial Neural Networks*; 2015 Apr 22-24; Bruges, Belgium: ESANN; 2015:89-94.
26. Tang D, Qin B, Liu T. Document modeling with gated recurrent neural network for sentiment classification. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*; 2015 Sep 17-21; Lisbon, Portugal: EMNLP; 2015:1422-1432.
27. Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell* 2020;42:318-327.
28. Kummer BR, Diaz I, Wu X, Aaroe AE, Chen ML, Iadecola C, et al. Associations between cerebrovascular risk factors and Parkinson disease. *Ann Neurol* 2019;86:572-581.
29. Huang YP, Chen LS, Yen MF, Fann CY, Chiu YH, Chen HH, et al. Parkinson's disease is related to an increased risk of ischemic stroke-a population-based propensity score-matched follow-up study. *PLoS One* 2013;8:e68314.
30. Song IU, Lee JE, Kwon DY, Park JH, Ma HI. Parkinson's disease might increase the risk of cerebral ischemic lesions. *Int J Med Sci* 2017;14:319-322.
31. Lohmann S, Grigoletto J, Bernis ME, Pesch V, Ma L, Reithofer S, et al. Ischemic stroke causes Parkinson's disease-like pathology and symptoms in transgenic mice overexpressing alpha-synuclein. *Acta Neuropathol Commun* 2022;10:26.
32. Postuma RB, Aarsland D, Barone P, Burn DJ, Hawkes CH, Oertel W, et al. Identifying prodromal Parkinson's disease: pre-motor disorders in Parkinson's disease. *Mov Disord* 2012;27:617-626.
33. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: predicting clinical events via recurrent neural networks. *JMLR Workshop Conf Proc* 2016;56:301-318.
34. Khalilia M, Chakraborty S, Popescu M. Predicting disease risks from highly imbalanced data using random forest. *BMC Med Inform Decis Mak* 2011;11:51.
35. Searles Nielsen S, Warden MN, Camacho-Soto A, Willis AW, Wright BA, Racette BA. A predictive model to identify Parkinson disease from administrative claims data. *Neurology* 2017;89:1448-1456.
36. Fearon C, Lang AE, Espay AJ. The logic and pitfalls of Parkinson's disease as "brain-first" versus "body-first" subtypes. *Mov Disord* 2021;36:594-598.
37. Kotsavasiloglou C, Kostikis N, Hristu-Varsakelis D, Arnaoutoglou M. Machine learning-based classification of simple drawing movements in Parkinson's disease. *Biomedical Signal Processing and Control* 2017;31:174-80.
38. Gupta U, Bansal H, Joshi D. An improved sex-specific and age-dependent classification model for Parkinson's diagnosis using handwriting measurement. *Computer Methods and Programs in Biomedicine* 2020;189:105305.
39. Arroyo-Gallego T, Ledesma-Carbayo MJ, Sánchez-Ferro A, Butterworth I, Mendoza CS, Matarazzo M, et al. Detection of motor impairment in Parkinson's disease via mobile touchscreen typing. *IEEE Trans Biomed Eng* 2017;64:1994-2002.