Predicting Changes in Depression Levels in Person's with Parkinson's Disease using Machine Learning

Samantha J. Stoll[1]

[1]Jarvis College of Computing and Digital Media, DePaul University, Chicago, IL

Acknowledgements: The Fox Insight Study (FI) is funded by The Michael J. Fox Foundation for Parkinson's Research. We would like to thank the Parkinson's community for participating in this study to make this research possible.

Data Access: Data used in the preparation of this article were obtained from the Fox Insight database (https://foxinsight-info.michaeljfox.org/insight/explore/insight.jsp) on 01/10/2023. For up-to-date information on the study, visit https://foxinsight-info.michaeljfox.org/insight/explore/insight.jsp.

Table of Acronyms and Abbreviations:

Notation	Definition
CDS	Change in depression score
EQ-5D-5L	European Quality of Life 5 Dimensions 5-Level Version
FI	Fox Insight
GDS-15	Geriatric Depression Scale - Short Form
PASE	Physical Activity Scale for the Elderly
PD	Parkinson's Disease
PD NMS	Parkinson's Disease Non-Motor Symptoms
PDQ-8	Parkinson's Disease Questionnaire (8-items)

Abstract

Introduction: Depression is the most common non-motor symptom in Parkinson's Disease (PD) and can lead to worsening or longer duration of other PD symptoms. Further, depressive symptoms often precede motor symptoms and a diagnosis of PD. Therefore, the ability to detect factors leading to worsening depression levels in the context of PD can help guide interventions to improve persons with PD's quality of life. This study investigated the ability to predict longitudinal changes in depression levels in persons with PD using machine learning methodology.

Methods: This study included 2,464 participants with PD and at least three completed Geriatric Depressive Scale-Short Form (GDS-15) surveys prior to the COVID-19 pandemic. Change in depression was measured using binary categories of rate change in GDS-15 scores: Worse Depression and Stable Depression. Predictors for random forest models included time-varying and fixed variables spanning demographics, PD symptoms and treatment, and lifestyle. Ten weighted models were created to assess varying timepoint and predictor variable groupings and to prioritize balanced accuracy or recall.

Results: Overall, models had modest to low accuracy in predicting Worse and Stable Depression over time. Of models prioritizing balanced accuracy, the model with the top predictors from all three timepoints performed the best in overall accuracy and Worse Depression f1-score (69% and 42%). Models prioritizing recall achieved Worse Depression recall rates of over 90% but with the consequence of very low precision (~ 20%).

Conclusion: The leading performance of the parsimonious models suggests targeted interventions for improving depression levels in PD. The low accuracy of models underscores the complexity of predicting changes in depression levels in PD in comparison to predicting current state depression. Future research should utilize the temporal dynamics of longitudinal measures of depression and predictors and explore more complex machine learning methods (e.g., Long Short-Term Memory Networks).

Introduction

Parkinson’s disease (PD) is a neurodegenerative disease that affects over 8 million persons worldwide, with over 900,000 in the United States alone and prevalence doubling in the past 25 years as of 2023[1,2]. PD is most commonly characterized by motor symptoms, including rigidity, involuntary movements and tremors, and slow or imbalanced movement[1,2]. While the motor symptoms are the more easily recognizable, the onset of many non-motor symptoms, including depression, often occurs before motor symptoms and a formal PD diagnosis[1,3]. More worrisome, depression is the most common non-motor symptom of PD[3], can lead to worsening or longer duration of PD symptoms[3,4], and is one of the leading causes of disability worldwide[5]. Therefore, depressive symptoms, in conjunction with potentially debilitating motor symptoms, can greatly diminish quality of life in persons with PD.

While there is no cure for PD or depression, current research and technologies have focused on better predicting and managing PD and its symptoms. Specifically, there are several studies that have used predictive and machine learning models to predict depression in those with Parkinson’s disease[6-10]. However, there are few studies using machine learning to predict changes or trajectories of depression in persons with Parkinson’s disease. While it can be clinically useful to predict if persons with PD will develop depression, it can often be futile as many patients develop depressive symptoms before their motor symptoms and official PD diagnosis[1,3], decreasing chances for meaningful intervention.

Understanding what factors can predict an improvement or worsening in depression in those with PD, can help guide clinicians, providers, and caretakers hone in those areas to help improve their patients quality of life. Of studies that investigate longitudinal change in depression levels in those with PD, traditional statistical techniques such as linear mixed models are used[11]. With the rapidly evolving machine learning methods, it is essential that new and possibly more robust methods are applied to previously examined questions to draw new insights and discoveries. Therefore, given the harmful impacts of depression in the context of PD and the sparsity of new research investigating longitudinal change in depression within PD, this study examines the ability to predict longitudinal changes in depression in persons with PD using machine learning methodology.

Methods

Sample and Procedures

Data came from the publicly available Fox Insight (FI) study using data from January 2023 (https://foxden.michaeljfox.org/.). FI is an online longitudinal study beginning in 2017 that recruited adults with PD as well as their caretakers[12]. Questionnaires were sent to participants either once or at regular intervals to collect information on various aspects of their condition and wellbeing including current medical and health conditions, medications, PD symptoms, and quality of life metrics. Details on FI recruitment and study procedures have been documented elsewhere[12,13]. At the time of data analysis, there were 54,129 unique participants enrolled in the study.

This study excluded participants without a current PD diagnosis ($n_{excluded}$=15,868). Of those with a PD diagnosis, those without at least three fully complete depressive symptoms surveys were excluded ($n_{excluded}$ = 29,558). Lastly, we excluded data that occurred during or after the COVID-19 pandemic, using when participants completed the COVID- 19 Experience Survey to mark the onset of the pandemic, to ensure sure depressive symptoms were not influenced by the pandemic. The sample was then filtered again to ensure at three depressive symptom surveys were complete prior to the COVID-19 pandemic ($n_{excluded}$ = 6,239). In total, the analytic sample included 2,464 participants (see **Figure 1** for study inclusion flowchart).

Measures

Change in Depressive Symptoms

The Geriatric Depressive Scale–Short Form (GDS-15) was administered to participants at enrollment and annually thereafter. The GDS-15 measure depression symptomology, specifically in geriatric populations, where higher scores represent more depressive symptoms[14]. Change in depression score (CDS) was calculated by taking the overall change in participants' GDS-15 scores from their first and third timepoints and dividing by months elapsed between participants' timepoints to account for the varying time between survey completions amongst participants. Next, tertiles were used to categorize CDS into three groups: Improved Depression, No Change in Depression, and Worse Depression. To focus on predicting those who have an adverse outcome, Worse Depression, the Improved and No Change in Depression groups were collapsed into a Stable Depression group. Therefore, the Worse Depression group had GDS-15 scores at or above the 67th percentile of participants, and the Stable Depression group had scores below the 67th percentile.

Predictors

Fixed-Predictors. Participants are asked to update their demographic information (e.g., education, income, race) at various timepoints throughout their enrollment in the FI Study. However, the current study used self-reported sex (male vs. female) and race (white vs. non-white) as static predictors of depressive symptoms as these did not change over time. Other sociodemographics were used as time-dependent predictors.

Time-Varying Predictors. Repeated measurements from FI Study surveys were included as time-dependent predictors if they had adequate variation A full list of time-dependent predictor variables can be found in the **Supplemental Table 1.** All instances of predictors occurred before or during the participants third GDS completion.

Sociodemographics: Highest education level, household income, age, employment status, veteran status, and current weight were used a time-dependent predictors of CDS to capture how changes in lifestyle or demographics may influence depressive symptomology changes. Current weight and age were scaled down by a factor of 1000 and 100, respectively, to fit the range of other predictor variables more closely.

Family history of depression: A binary variable indicating if anyone in the participants family had, or currently has, a history of depression was included as a time-dependent predictor to model participants possibly changing perception and knowledge of if anyone in their family has depression.

PD medications: A binary predictor of whether participants were currently taking any prescription medications for symptoms related to their PD diagnosis was included in the model. Further, a sum score of total medications being taken for participants' PD diagnosis was included.

PD Health-Related Quality of Life: Items from the 8-item Parkinson's Disease Questionnaire (PDQ-8) were included to measure participants' quality of life specific to their PD diagnosis[15]. Items ask how often participants have difficulty completing tasks or have uncomfortable feelings (e.g., depressed, embarrassment) on a scale of 0 (Never) – 4 (Always or cannot do at all) related to their PD diagnosis.

General Health-Related Quality of Life: Non-disease-specific quality of life was assessed using the European Quality of Life 5 Dimensions 5 Level Version (EQ-5D-5L) survey[16]. Five dimensions are captured, including (1) Mobility, (2) Self-Care, (3) Usual Activities, (4) Pain and Discomfort, and (5) Anxiety and Depression. Items were rated on a scale of 0 (No problems) – 4 (Unable to do tasks/Extreme pain or anxiety/depression). A global rating of overall health was reported on a scale of 0 (Worst health imaginable) – 100 (Best health imaginable), which was scaled down by a factor of 100 to fit the range of other predictor variables more closely.

Living Situation: To assess participant's possibly changing living situation, six binary questions were included to determine participants' current living situation. Possible living situations included were (1) Living alone, (2) Living with spouse, (3) Living with adult child/children, (4) Living with minor child/children, (5) Living with other family, and (6) Living in an assisted living. Items were not mutually exclusive.

PD Treatment: Participants were asked if their PD was being treated by various professionals. Included treatment providers were (1) General Neurologist, (2) Doctor/Primary Care Doctor, (3) Nurse Practitioner/Physician Assistant, (4) Movement Disorder Specialist, (5) Other provider. Items were not mutually exclusive.

Physical Activity: Items from Physical Activity Scale for the Elderly (PASE) were included to measure the presence and frequency of participants' physical activity[17]. Aspects of physical activity included household activities, leisure activities, walking, light recreational activities, moderate recreational activities, strenuous recreational activities, and muscle strength exercises. Household activities were measured on a binary presence/absence scale. All other activities were measured on a scale of 0 (Never) – 3 (Often; 5–7 days) regarding the previous seven days.

Non-motor Symptoms: To measure participants non-motor symptoms in the last month related to PD (e.g., dizziness, constipation, difficulty swallowing) participants were asked to complete the PD NMS Questionnaire[18]. Each item asked if the non-motor symptom was present or absent and a sum score was calculated to represent total non-motor symptoms present.

Acute Surgery: Participants noted if they had at least one surgery that required anesthesia in the last year. This question was used to measure if any new surgeries occurred in the last year that may influence depressive symptoms.

Data Analyses

Time Collapsing. Given the remote nature of the study, the exact date and amount of time between survey completions varied among participants. The first three consecutive instances participants completed the GDS were included in analyses. To systematically compare and analyze survey completions within a consistent framework, individual timepoints were collapsed into timepoint 1, timepoint 2, and timepoint 3.

Model Training, Tuning, and Testing. The sample dataset was split into training, validation, and testing sets, stratified by CDS categories, and proportions of males and females were ensured to be relatively equal among all datasets, as depressive symptoms tend to present and manifest differently between males and females[19]. Given that the imbalance of CDS categories in the full sample (66.6% Stable, 33.4% Worse Depression), the training dataset oversampled for Worse Depression (59.4% Stable, 40.6% Worse Depression) to improve training performance. Further, class weights were used (Stable = 1, Worse Depression = 3) to help specificity given the unbalanced distribution of Stable and Worse Depression outcomes in sample[19].

To better understand the quantity and temporal nature of timepoints needed to best predict CDS, five separate models were trained using Random Forest using: (1) Predictors from all three timepoints simultaneously, (2) Predictors from the first timepoint only, (3) Predictors from the second timepoint only, (4) Predictors from the third timepoint only, and (5) The most parsimonious set of predictors using feature importance results from models 1 – 4, using the mean decrease in impurity to calculate feature importance, as most predictors are not continuous or have high cardinality if categorical[20]. Lastly, Gradient Boost and AdaBoost algorithms were employed using the same variables from Model 5 to explore further machine learning techniques to reduce bias and handle the unbalanced dataset.

First, all models were trained on the training dataset using grid search cross validation to find the best performing parameters. The balanced accuracy score was used to select the best model hyperparameters. The trained model was then evaluated using prediction results from the validation set to determine if further tuning was required. Lastly, predictions from the final models using the test dataset were used as final model results.

The training of models 1 – 5 using grid search cross validation was repeated with the recall score used to select the best model parameters to minimize false negatives, which may be particularly important in health care and depression intervention settings. Models 1–5 trained and selected using the recall score are denoted with a .5 (e.g., Model 1.5, Model 2.5, etc.) to differentiate from models trained and selected using balanced accuracy. A summary of all models can be found in **Table 1.**

Imputation. Missingness rates for predictor variables ranged from 0%–63% with the majority having a missingness rate of less than 47%. Given the modest amounts of missingness, missing values for training, validation, and training sets were imputed using a mode imputation model from the training dataset.

Results

Sample characteristics

Overall, the sample included a fairly equal proportion of males and females (53.1% male), with the large majority of the sample self-reporting as White and non-Hispanic (97.9% and 96.1%, respectively). Most participants had an initial household income of $75,000 or more (47.3%), at least a bachelor's degree (70.9%) and were of non-Veteran status (85.7%). The average age at enrollment was 66.2 years-old (SD= 8.12; min = 32.2; max= 91.5) A summary of sample descriptives can be found in **Table 2.** No predictor variables were highly correlated with GDS-15 or raw CDS scores (max r = 0.56 and 0.10, respectively) and no predictors variables were highly correlated with one another (max r = 0.72; max variance inflation factors = 2.22).

Change in Depressive Symptoms:

The average sample CDS score was 0.26 (SD=1.63; min=-18.0; max=11.15). The average time between participants first and third GDS-15 completion was 19.41 months (min=5; max = 39) with the overall average time between consecutive GDS-15 completions being 9.7 months (min=1; max=37). 33.4% (n=822) of the sample was in the Worse Depression CDS category.

Model Results

The training, validation, and testing datasets were 53.6%, 53.4%, and 50.7% male, respectively. 40.6%, 22.5%, and 22.5% of the samples were categorized as having Worse

Depression. All performance metrics for all models can be found in **Table 3** with results specific to Worse Depression in **Figure 2** and specific to Stable Depression in **Supplemental Figure 1.**

Model 1; balanced accuracy scoring (Model 1.5; recall scoring) – All predictors from all timepoints. Using all predictors from all three timepoints, Model 1 had an overall accuracy of 62% (32%). For Stable Depression and Worse Depression, Model 1 had a precision of 84% (91%) and 32% (24%) and recall of 64% (14%) and 58% (95%), respectively. The top five most importance predictors for Model 1, in decreasing order of importance were (1) Non-motor sum score at timepoint 3, (2) Global rating of health at timepoint 3, (3) Non-motor sum score at timepoint 1, (4) Age at timepoint 1, and (5) How anxious or depressed participants are at the time of the survey, from the EQ-5D-5L, at timepoint 3. Gini importance scores of the top five predictors ranged from 0.03 – 0.06.

Model 2; balanced accuracy scoring (Model 2.5; recall scoring) – All predictors from first timepoint. Using all predictors from just the first timepoint, Model 2 had an overall accuracy of 57% (26%). For Stable Depression and Worse Depression, Model 2 had a precision of 82% (81%) and 28% (23%) and recall of 57% (0.06%) and 56% (95%), respectively. The top five most importance predictors for Model 2, in decreasing order of importance were (1) Weight at timepoint 1, (2) Age at timepoint 1, (3) Non-motor sum score at timepoint 1, (4) Global rating of health at timepoint 1, and (5) Highest education level at timepoint 1. Gini importance scores of the top five predictors ranged from 0.04 – 0.12.

Model 3; balanced accuracy scoring (Model 3.5; recall scoring) – All predictors from second timepoint. Using all predictors from just the second timepoint, Model 3 had an overall accuracy of 44% (31%). For Stable Depression and Worse Depression, Model 3 had a precision of 80% (85%) and 24% (24%) and recall of 36% (14%) and 69% (92%), respectively. The top five most importance predictors for Model 3, in decreasing order of importance were (1) Age at timepoint 2, (2) Weight at timepoint 2, (3) Non-motor sum score at timepoint 2, (4) Global rating of health at timepoint 2, and (5) Sum of PD medications at timepoint 2. Gini importance scores of the top five predictors ranged from 0.04 – .09.

Model 4; balanced accuracy scoring (Model 4.5; recall scoring) – All predictors from third timepoint. Using all predictors from just the third timepoint, Model 4 had an overall accuracy of 54% (38%). For Stable Depression and Worse Depression, Model 4 had a precision of 84% (90%) and 28% (26%) and recall of 51% (23%) and 66% (91%), respectively. The top five most importance predictors for Model 4, in decreasing order of importance were (1) Age at timepoint 3, (2) Non-motor sum score at timepoint 3, (3) Weight at timepoint 3, (4) Global rating of health at timepoint 3, and (5) How anxious or depressed participants are at the time of the survey, from the EQ-5D-5L, at timepoint 3. Gini importance scores of the top five predictors ranged from 0.04 – 0.09.

Model 5; balanced accuracy scoring (Model 5.5; recall scoring) – Most important predictors from all timepoints. Using the predictors that appeared in the top 10 most important predictors in at least two models (see **Figure 3** and **Figure 4** for highest variable importance in all models), Model 5 had an overall accuracy of 66% (39%). For Stable Depression and Worse Depression, Model 5 had a precision of 84% (86%) and 34% (25%) and recall of 69% (26%) and 56% (86%), respectively. Among all models using balanced accuracy as a scoring metric, Model 5 had the highest overall accuracy and f1-scores among all models (**Table 2**). The top five most importance predictors for Model 5, in decreasing order of importance were (1) Non-motor sum score at timepoint 3, (2) Non-motor sum score at timepoint 1, (3) Global rating of health at timepoint 3, (4) Global rating of health at timepoint 2, and (5) Age at timepoint 1. Gini importance scores of the top five predictors ranged from 0.05 – 0.09.

No significant improvements for in detecting Worse Depression were made using Gradient Boos or AdaBoost algorithms for Model 5 (see **Table 3**).

Discussion

Results from this study suggest that when given longitudinal data for participants, information from the most recent observation is individually the most helpful in predicting an overall increase in depressive symptoms over time compared to individually examining other timepoints. This result highlights the importance of frequent monitoring and assessments of depressive symptoms in person with PD as well as focusing interventions and treatments on current experiences. Moreover, results revealed that a parsimonious model (see **Figure 3** and **Figure 4** for variable importance results) utilizing all three timepoints (Model 5) had comparable performance, and slightly advantageous metrics, to a more expansive model of variables at all available timepoints (Model 1). The similar performance of the parsimonious model suggests that crucial information for depression level trajectories could be honed in on to created targeted assessments to predict changes in depression levels. This is beneficial, as decreasing the battery of assessments required for evaluation is less time consuming and burdensome for both clinicians and patients, resulting in an efficient assessment process.

Correspondingly, several factors were deemed most important in predicting CDS categories across multiple models. One such factor was the total number of participants' non-motor PD symptoms. As non-motor symptoms are often uncomfortable and can interfere with daily activities (e.g., drowsiness, urinary dysfunction, constipation), there is an increased risk for diminished quality of life as more non-motor symptoms are present[21], which can lead to an increased risk for depression or worsening of symptoms. This relationship between non-motor symptoms and depression can be seen in this study sample, where total non-motor symptoms had a moderately positive correlation with GDS-5 scores ($r=0.56$; $p < 0.001$). Therefore, increased efforts should be placed on maximizing the quality of life in those with increased PD-related non-motor symptoms to decrease the risk for depression. Clinicians should ask patients early and directly if they are experiencing non-motor symptoms as they often precede a PD diagnosis[1], and patients may be embarrassed to report symptoms freely or unaware that they could be related to PD[22]. Relatedly, overall perception of health was an important predictor for CDS categories. Self-reported perception of health had a moderately negative correlation with GDS-15 scores ($r=-0.053$; $p = <0.001$), showing that those who have a more positive perception of their health tend to have fewer depressive symptoms. However, it is unclear if depressive symptoms predate or influence health perceptions or if health perception influences depressive symptoms. Research in positive psychology has shown that increasing positive emotions and resources (e.g., developing personal strengths, building positive engagements) can aid in treating depression[23], therefore giving an opportunity to focus PD interventions on building positive perspectives on health to help decrease risk for increased depressive symptoms.

Overall, despite oversampling for Worse Depression and additional class weights included in models to aid prediction, all tested models' precision and recall for detecting Worse Depression were only slightly better than chance probability. This outcome underscores the complexity of predicting changes in depression levels rather than current state symptomology. However, if detecting Worse Depression is of utmost importance, even at the risk of false positives, Models 1.5 – 5.5 were able to detect Worse Depression with recall scores raging from 0.86 – 0.95 which may benefit early intervention and prevention efforts.

This study remains to have several limitations and suggested directions for future research. First, the handling of the temporal dynamics of the data was a major limitation of the current study. Although utilizing the longitudinal nature of observations in prediction models, this study did not explicitly account for the varying time between observations for each participant. Further, each timepoint for each predictor was handled as a separate predictor rather than taking into account the within-person variation for each variable at the various timepoints, as is done in traditional longitudinal data analysis. Addressing these limitations could increase model performances. For example, two recently published papers provide new techniques that marry longitudinal data analysis methods and random forests. These methods have been compiled into R packages, DynForest[24] and LongituRF[25], and have been utilized in low-dimensional (e.g., < 25 predictors) and high-dimensional (omics-data) datasets, respectively.

Further, upon noting that most recent observations may be particularly beneficial in predicting changes in depression levels over time, researchers weight more recent timepoints in predictive models or use more complex methods, such as Long Short-Term Memory Networks which focuses on more recent, or short-term, “memories”[26].

In summary, while the current study was able to detect Worse Depression in a large cohort study of persons with PD with high recall rates using longitudinal data, further advancements and methodologies may be necessary to capture to complexity of individuals’ depression symptoms over time in the context of PD.

References

1. Armstrong MJ, Okun MS. Diagnosis and treatment of Parkinson disease: a review. Jama 2020;323(6):548-560.
2. World Health Organization. Parkinson disease. (https://www.who.int/news-room/fact-sheets/detail/parkinson-disease).
3. Ahmad MH, Rizvi MA, Ali M, Mondal AC. Neurobiology of depression in Parkinson's disease: Insights into epidemiology, molecular mechanisms and treatment strategies. Ageing research reviews 2023:101840.
4. Cong S, Xiang C, Zhang S, Zhang T, Wang H, Cong S. Prevalence and clinical aspects of depression in Parkinson's disease: A systematic review and meta-analysis of 129 studies. Neuroscience & Biobehavioral Reviews 2022;141:104749.
5. Friedrich MJ. Depression is the leading cause of disability around the world. Jama 2017;317(15):1517-1517.
6. Gu S-C, Zhou J, Yuan C-X, Ye Q. Personalized prediction of depression in patients with newly diagnosed Parkinson's disease: A prospective cohort study. Journal of affective disorders 2020;268:118-126.
7. Farabaugh AH, Locascio JJ, Yap L, et al. Assessing depression and factors possibly associated with depression during the course of Parkinson's disease. Annals of clinical psychiatry: official journal of the American Academy of Clinical Psychiatrists 2011;23(3):171.
8. Byeon H. Development of a depression in Parkinson's disease prediction model using machine learning. World Journal of Psychiatry 2020;10(10):234.
9. Yang Y, Yang Y, Pan A, et al. Identifying Depression in Parkinson's Disease by Using Combined Diffusion Tensor Imaging and Support Vector Machine. Frontiers in Neurology 2022;13:878691.
10. Nguyen HV, Byeon H. Prediction of Parkinson's Disease Depression Using LIME-Based Stacking Ensemble Model. Mathematics 2023;11(3):708.
11. Zhu K, van Hilten JJ, Marinus J. Associated and predictive factors of depressive symptoms in patients with Parkinson's disease. Journal of neurology 2016;263:1215-1225.
12. Dobkin RD, Amondikar N, Kopil C, et al. Innovative recruitment strategies to increase diversity of participation in Parkinson's disease research: the Fox Insight cohort experience. Journal of Parkinson's disease 2020;10(2):665-675.
13. Smolensky L, Amondikar N, Crawford K, et al. Fox Insight collects online, longitudinal patient-reported outcomes and genetic data on Parkinson's disease. Scientific data 2020;7(1):67.
14. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. Journal of psychiatric research 1982;17(1):37-49.

15. Jenkinson C, Fitzpatrick R, Peto V, Greenhall R, Hyman N. The PDQ-8: development and validation of a short-form Parkinson's disease questionnaire. Psychology and Health 1997;12(6):805-814.
16. Alvarado-Bolaños A, Cervantes-Arriaga A, Rodríguez-Violante M, et al. Convergent validation of EQ-5D-5L in patients with Parkinson's disease. J Neurol Sci 2015;358(1-2):53-7. (In eng). DOI: 10.1016/j.jns.2015.08.010.
17. Washburn RA, Smith KW, Jette AM, Janney CA. The Physical Activity Scale for the Elderly (PASE): development and evaluation. Journal of clinical epidemiology 1993;46(2):153-162.
18. Chaudhuri KR, Martinez-Martin P, Schapira AH, et al. International multicenter pilot study of the first comprehensive self-completed nonmotor symptoms questionnaire for Parkinson's disease: the NMSQuest study. Movement disorders: official journal of the Movement Disorder Society 2006;21(7):916-923.
19. Mohammadi S, Seyedmirzaei H, Salehi MA, et al. Brain-based Sex Differences in Depression: A Systematic Review of Neuroimaging Studies. Brain Imaging and Behavior 2023:1-29.
20. Loecher M. Unbiased variable importance for random forests. Communications in Statistics-Theory and Methods 2022;51(5):1413-1425.
21. Hinnell C, Chaudhuri KR. The effect of non-motor symptoms on quality of life in Parkinson’s disease. Eur Neurol Rev 2009;4(2):29-33.
22. Chaudhuri KR, Prieto-Jurcynska C, Naidu Y, et al. The nondeclaration of nonmotor symptoms of Parkinson's disease to health care professionals: an international study using the nonmotor symptoms questionnaire. Movement Disorders 2010;25(6):704-709.
23. Lim WL, Tierney S. The effectiveness of positive psychology interventions for promoting well-being of adults experiencing depression compared to other active psychological treatments: a systematic review and meta-analysis. Journal of happiness studies 2023;24(1):249-273.
24. Devaux A, Proust-Lima C, Genuer R. Random Forests for time-fixed and time-dependent predictors: The DynForest R package. arXiv preprint arXiv:230202670 2023.
25. Capitaine L, Genuer R, Thiébaut R. Random forests for high-dimensional longitudinal data. Statistical methods in medical research 2021;30(1):166-184.
26. Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation 1997;9(8):1735-1780.

Table 1. Description of Study Models

Model	Timepoints Included	Predictors Included
Model 1* (Model 1.5**)	All timepoints (1 – 3)	All predictors
Model 2* (Model 2.5**)	Timepoint 1	All predictors
Model 3* (Model 3.5**)	Timepoint 2	All predictors
Model 4* (Model 4.5**)	Timepoint 3	All predictors
Model 5* (Model 5.5**)	All timepoints (1 – 3)	Predictors that were in top 10 of importance for at least two other models

*Models 1–5 trained using balanced accuracy; **Models 1.5–5.5 trained using recall

Table 2. Sample Descriptives at Enrollment (N=2,464)

	N	%
Sex		
Female	1156	46.9%
Male	1308	53.1%
Race		
White	2412	97.9%
Non-White	52	2.1%
Ethnicity		
Non-Hispanic	2369	96.1%
Hispanic	95	3.9%
Household Income		
< $50k	478	19.4%
$50k - $75k	432	17.5%
$75k - $100k	374	15.2%
> $100k	791	32.1%
Missing	389	15.8%
Employment		
Full-time	422	17.1%
Part-time	209	8.5%
Retired	1732	70.3%
Unemployed	95	3.9%
Missing	6	20.0%
Education		
< HS	16	0.6%
HS or GED	149	6.0%
Some college	355	14.4%
Associates	191	7.8%
Bachelors	761	30.9%
Masters	670	27.2%
Professional school	170	6.9%
Doctorate	145	5.9%
Missing	7	0.3%
Veteran Status		
Non-Veteran	2111	85.7%
Veteran	352	14.3%
Missing	1	0.0%

Table 3. Model Results Predicting Change in Depression Score

	F1-Score		**Precision**		**Recall**		**Accuracy**
Model	**Stable**	**Worse Depression**	**Stable**	**Worse Depression**	**Stable**	**Worse Depression**	
Model 1 (balanced accuracy) *All predictors from all timepoints*	0.72	0.41	0.84	0.32	0.64	0.58	0.62
Model 1.5 (recall) *All predictors from all timepoints*	0.24	0.39	0.91	0.24	0.14	0.95	0.32
Model 2 (balanced accruacy) *All predictors from first timepoint*	0.67	0.37	0.82	0.28	0.57	0.56	0.57
Model 2.5 (recall) *All predictors from first timepoint*	0.11	0.37	0.81	0.23	0.06	0.95	0.26
Model 3 (balanced accuracy) *All predictors from second timepoint*	0.50	0.36	0.80	0.24	0.36	0.69	0.44
Model 3.5 (recall) *All predictors from second timepoint*	0.23	0.38	0.85	0.24	0.14	0.92	0.31
Model 4 (balanced accuracy) *All predictors from third timepoint*	0.63	0.39	0.84	0.28	0.51	0.66	0.54
Model 4.5 (recall) *All predictors from third timepoint*	0.37	0.40	0.90	0.26	0.23	0.91	0.38
Model 5 (balanced accuracy) *Top predictors from Models 1–4*	0.76	0.42	0.85	0.34	0.69	0.56	0.66

Model 5.5 (recall) *Top predictors from Models 1–4*	0.40	0.39	0.86	0.25	0.26	0.86	0.39
Model 5 - Gradient Boost	0.77	0.38	0.82	0.33	0.72	0.47	0.66
Model 5 - AdaBoost	0.82	0.41	0.83	0.4	0.82	0.41	0.73

All Random Forest models had class weights of 1 for Stable and 3 for Worse Depression

Figure 1. Study Inclusion Criteria

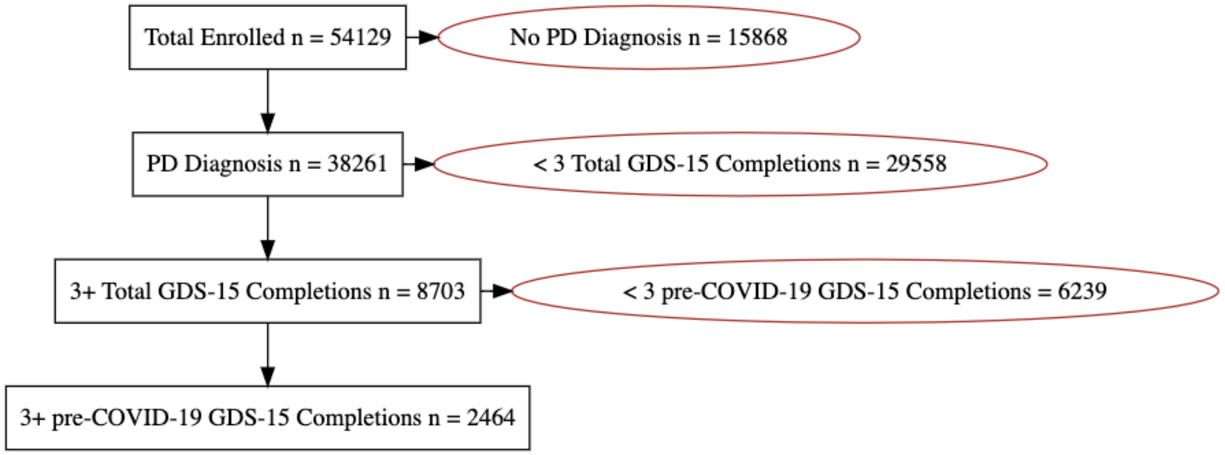

Total Enrolled n = 54129
No PD Diagnosis n = 15868
PD Diagnosis n = 38261
< 3 Total GDS-15 Completions n = 29558
3+ Total GDS-15 Completions n = 8703
< 3 pre-COVID-19 GDS-15 Completions = 6239
3+ pre-COVID-19 GDS-15 Completions n = 2464

Figure 2. All Model Results for Predicting Worse Depression

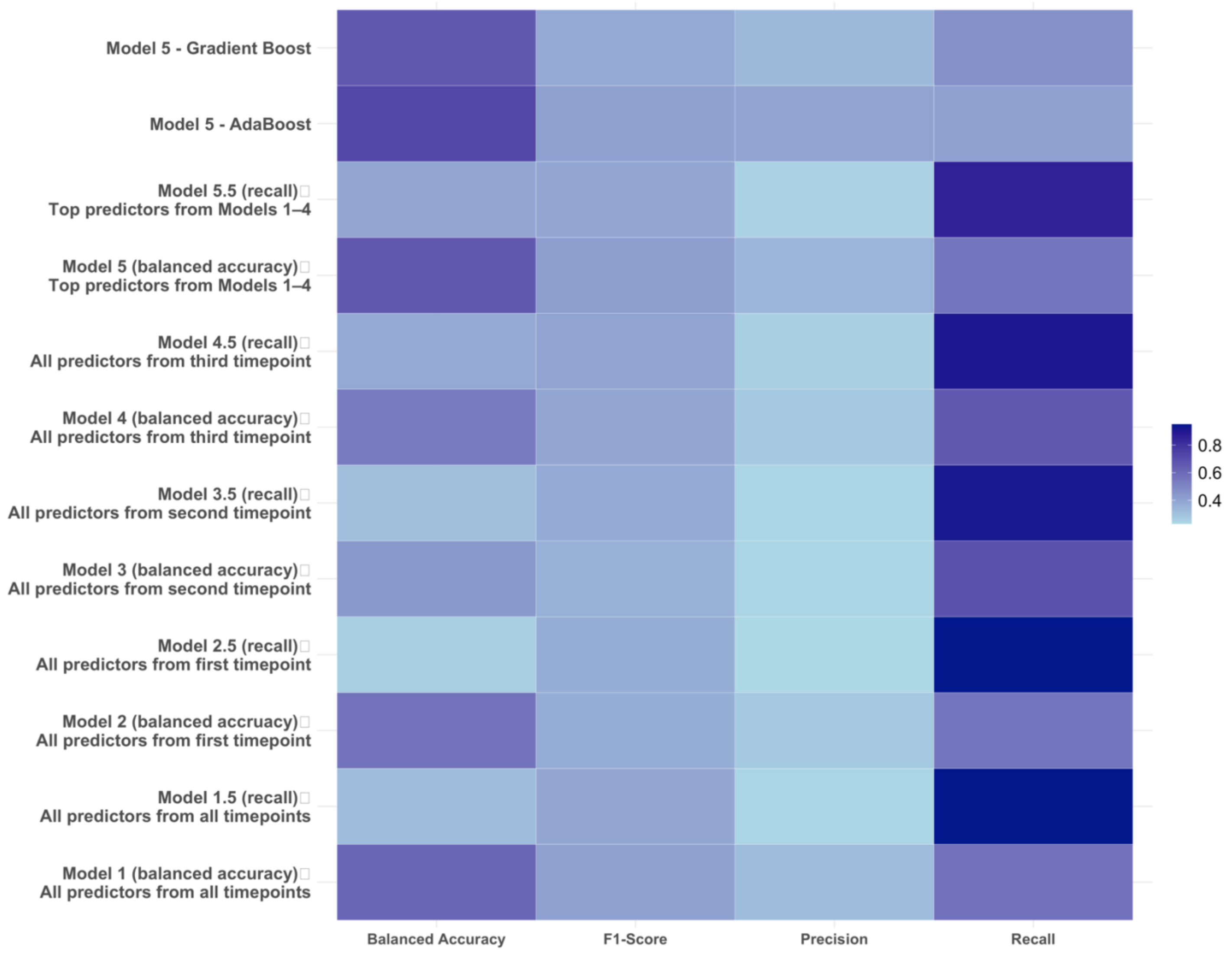
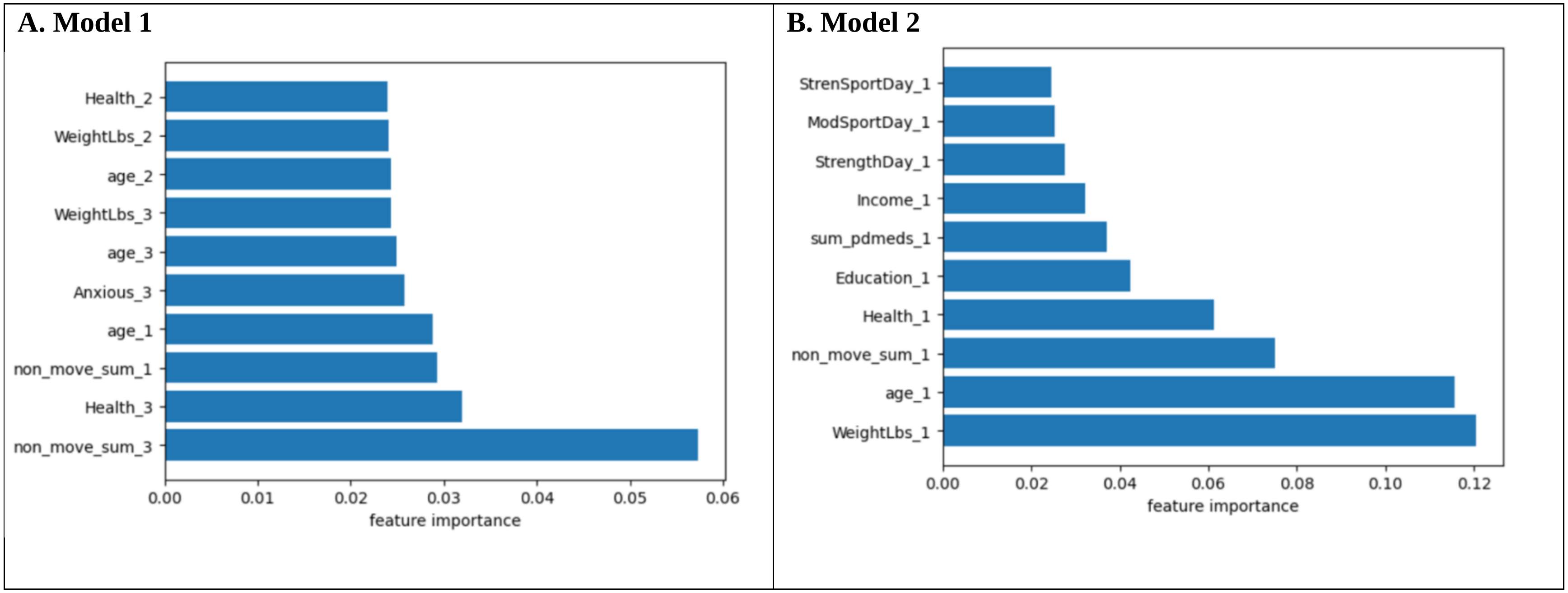
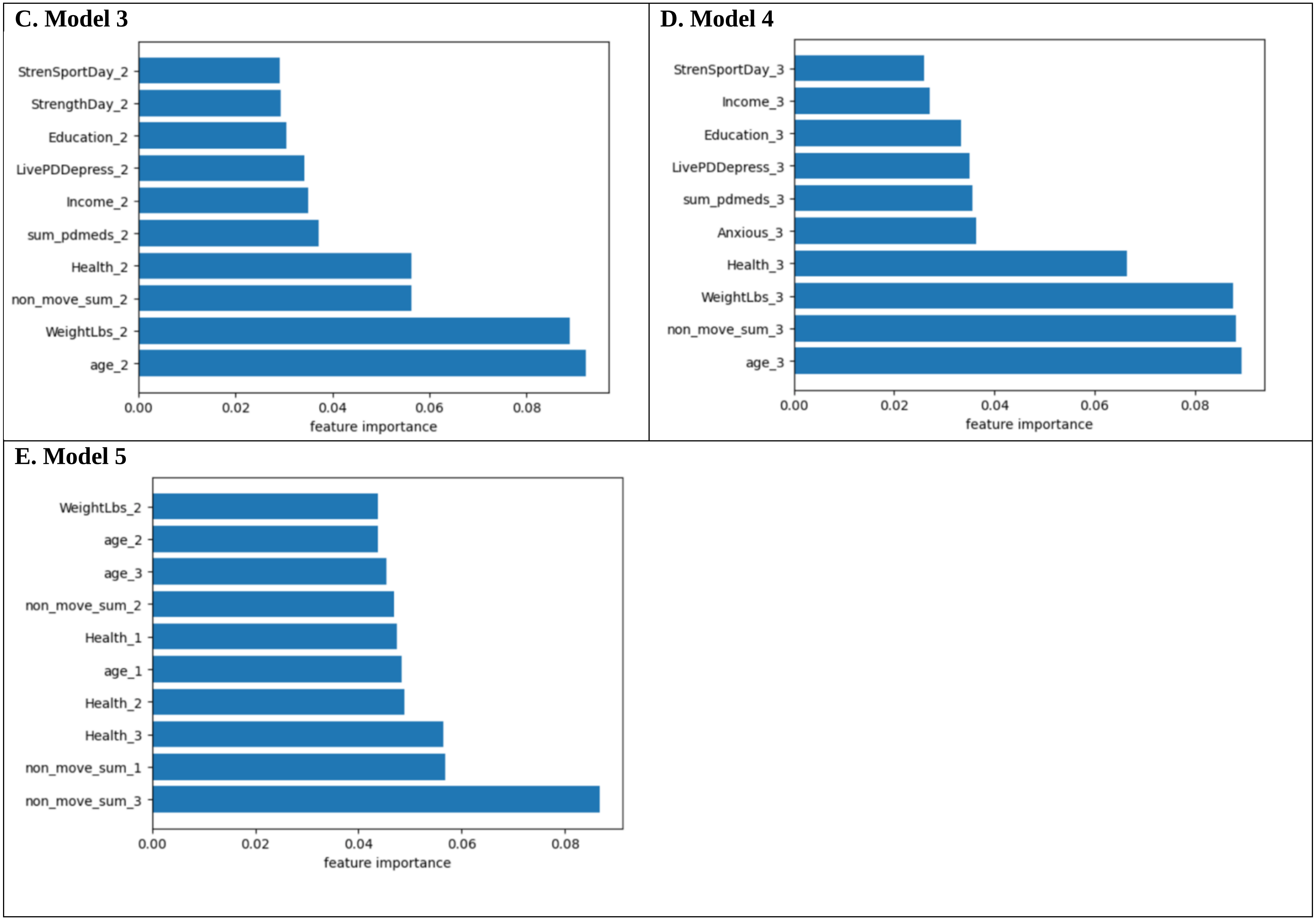


Figure 3. Variable Importance for Models 1–5



A. Model 1

B. Model 2

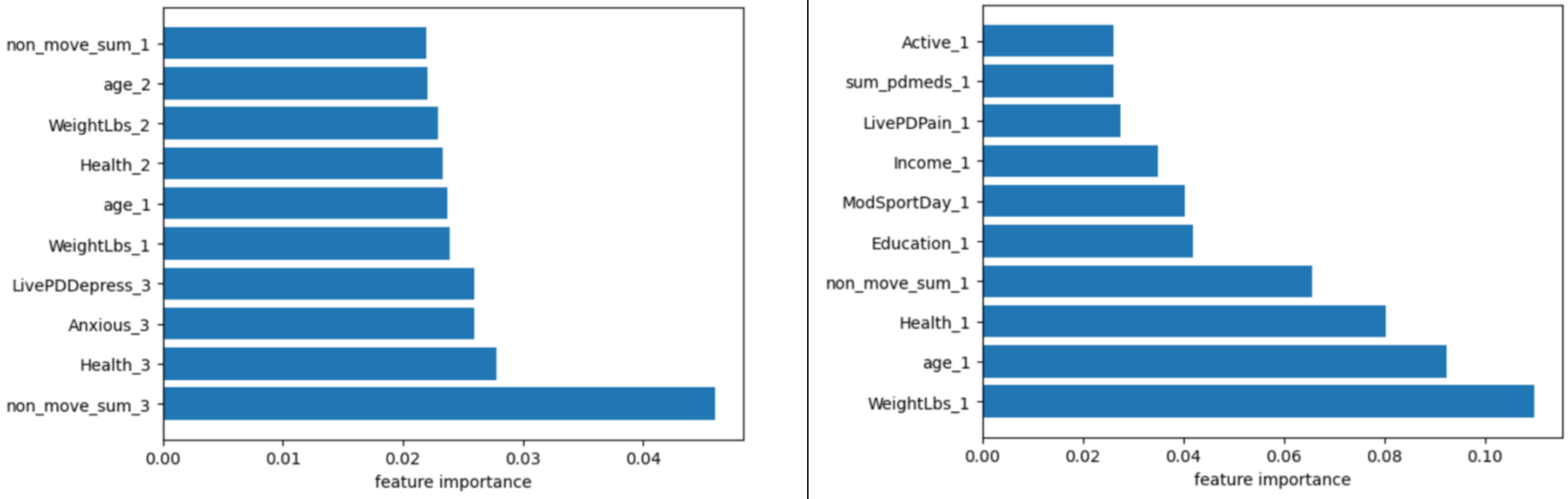


C. Model 3

D. Model 4

E. Model 5

Figure 4. Variable Importance for Models 1.5–5.5

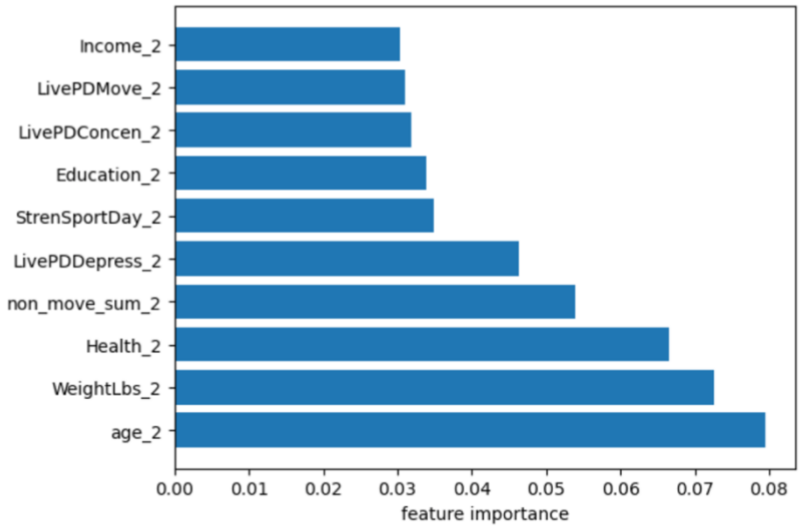


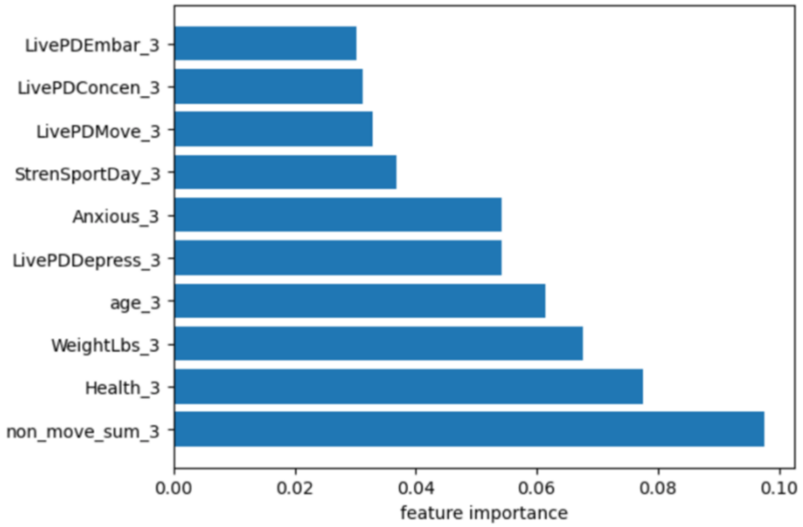
A. Model 1.5

B. Model 2.5

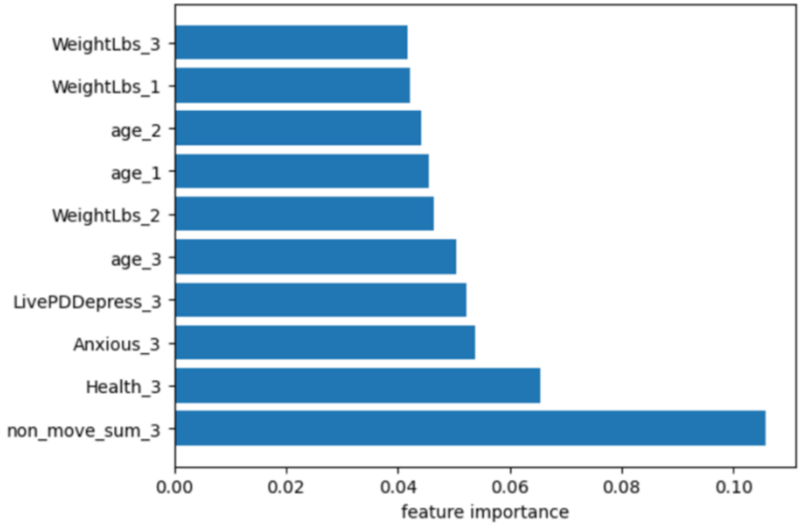
C. Model 3.5

D. Model 4.5


Income_2
LivePDMove_2
LivePDConcen_2
Education_2
StrenSportDay_2
LivePDDepress_2
non_move_sum_2
Health_2
WeightLbs_2
age_2
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
feature importance


LivePDEmbar_3
LivePDConcen_3
LivePDMove_3
StrenSportDay_3
Anxious_3
LivePDDepress_3
age_3
WeightLbs_3
Health_3
non_move_sum_3
0.00
0.02
0.04
0.06
0.08
0.10
feature importance

E. Model 5.5


WeightLbs_3
WeightLbs_1
age_2
age_1
WeightLbs_2
age_3
LivePDDepress_3
Anxious_3
Health_3
non_move_sum_3
0.00
0.02
0.04
0.06
0.08
0.10
feature importance