Genetic and brain imaging phenotype joint prediction of longitudinal Parkinson's Disease subtypes

L.M. Polfus[1*], H. Masnadi Shirazi[1*], A. Reardon[2], M. Varga[1], A. Tchourbanov[1], T. Gosselin[2], A. Dadu[3], F. Faghri[3], M. Nalls[3], J. Hesterman[2], R. Gunn[2], A. Chamberlin[1], B. Avants[2]

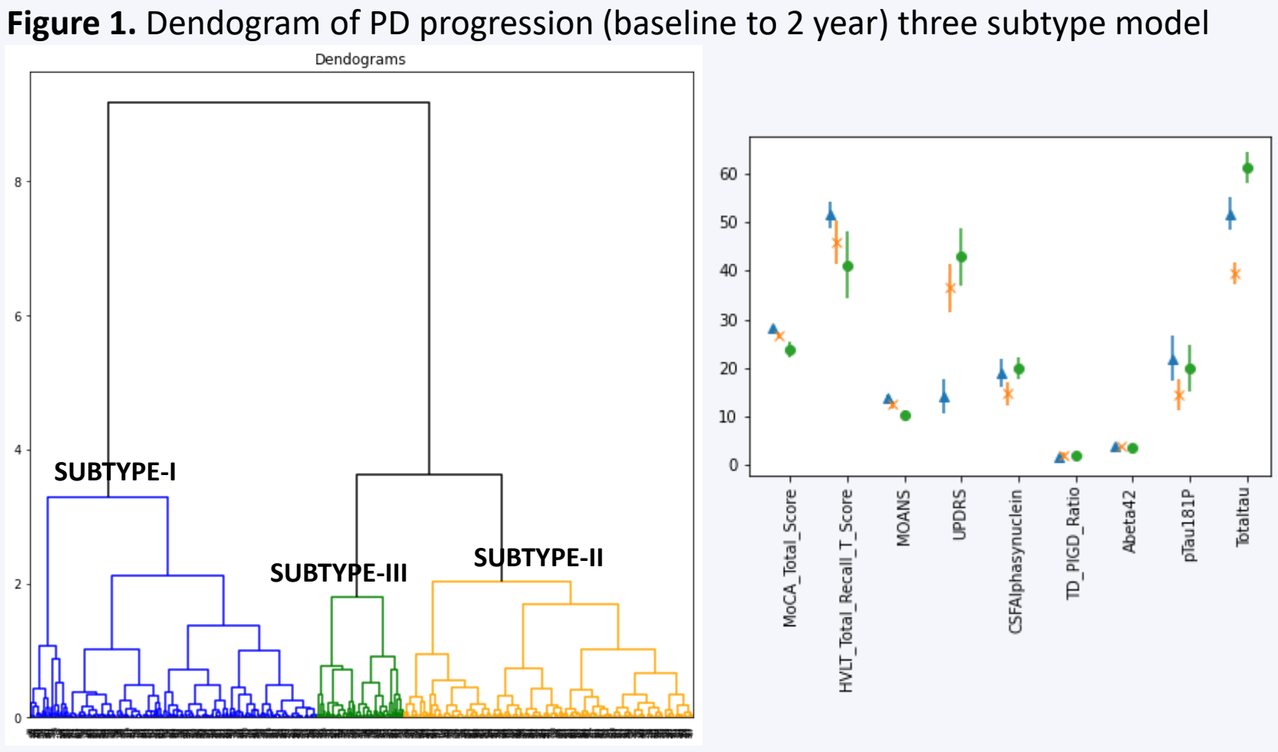
[1]REALM IDx/Ambry Genetics, Medical Sciences and Advanced Analytics, Aliso Viejo, CA, United States of America [2]REALM IDx/Invicro, Image Analysis, Needham, MA, United States of America, [3]Data Tecnica International, Data Science, Washington DC, United States of America

Poster #: P18.015.C

Background: Among neurodegenerative diseases, the prevalence of Parkinson's Disease (PD) remains as the fastest rising in the US. The distribution of an aging population impacts PD incidence (90,000 individuals newly diagnosed annually) with an estimated 1.2 million individuals living with PD by 2030 (https://www.parkinson.org/understanding-parkinsons/statistics).

Both genetic and non-genetic factors contribute to PD with level of involvement dependent on familial or sporadic diagnosis. Both motor and non-motor features define clinical presentation with subtle interventional prodromal disease symptoms occurring prior to formal PD diagnosis. Neuropathological indicators include the loss of dopaminergic neurons located in the midbrain of the substantia pars compacta and insoluble alpha-synuclein aggregates, Lewy bodies, Lewy neutrites spreading to the cortex. The primary mechanism of motor function disruption is caused by an intracellular aggregation of alpha-synuclein which leads to dysfunction in basal ganglia circuitry. PD categorization by established Hoehn and Yahr disease staging calibrates by onset of clinical manifestations, however, machine learning with additional inputs such as brain health captured by imaging or genetic predisposition captured by sequencing uncovers deeper understanding to define progression subtypes.

Figure 1. Dendogram of PD progression (baseline to 2 year) three subtype model

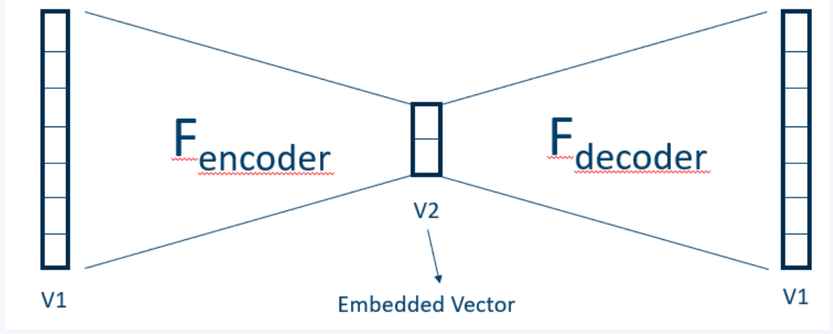


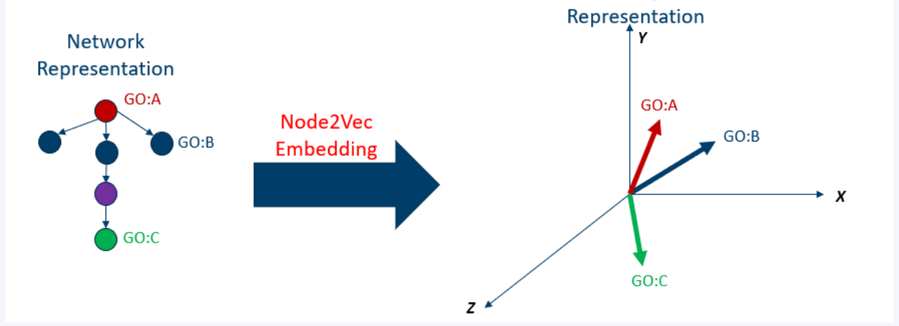
Methods: We implemented an embedded node2vec approach within Gene Ontology (GO) and Human Phenotype Ontology (HPO) terms within a longitudinal machine learning framework to predict PD subtypes. Our study utilized -OMICs data (genomic and imaging) along with HPO/GO term embedding, CSF biomarkers, and detailed longitudinal clinical measurements to predict PD progression subtypes (N=606 individuals).

We employed a Neural Net based Long Short-Term Memory (LSTM) ML model utilizing Parkinson's Progression Marker Initiative (PPMI) (**https://www.ppmi-info.org/**) clinical features at 2 time points (baseline and 2 years), motor changes, cognitive impairment and memory scores accounting for ON/OFF medication use, genetic variation, HPO/GO terms, and imaging (T1-weighted neuroimaging for volume and thickness). These inputs were trained in the LSTM model to create 2-D embedded vectors. Each of the 606 patients and their sequence of clinical visits over time is represented by a 2-D vector resulting in 606 2-D vectors.

- A total of 1457 HPO terms and 970 GO terms were used for node2vec embedding.
- Embedding added 300 dimensional "Word2Vec" representation for the HPO terms and 100 dimensional "Node2Vec" representation for the GO terms.
- Ninety known PD associated genome wide association study variants were used as genetic inputs.

Prior to training, the features were standardized by subtracting the mean and dividing by the standard deviation. A leave-one-out cross-validation was also performed.





Results: Utilizing embedding algorithms along with HPO/GO terms increased the Area Under the Curve (AUC) by ~5% compared to using genetic variants directly. Clinical + imaging inputs provided the highest AUC for two (93.2%) or three (77.4%) subtype models (**Tables 1 and 2; Figure 1 and 2**).

The key drivers of clinical measures/biomarkers for subtypes were UPDRS and Total Tau. PPMI study results identified specific genes (*CTNNB1*, *SHH*, and *GLI2*), HPO embeddings/biologic pathways via GO terms combined with brain regions characterizing fast/slow PD progressors independent of Hoehn and Yahr disease stage.

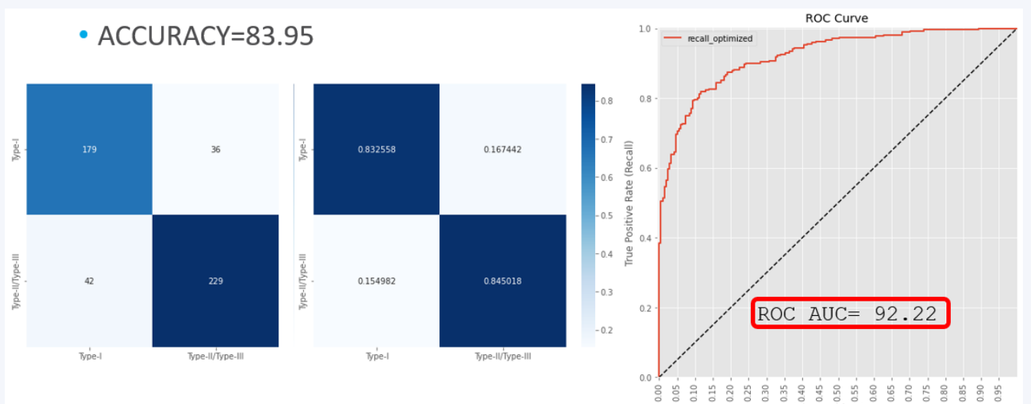
Table 1. Two class method: Type I versus Type II/III by Input Type

Two Class Method:	Area Under Curve %
Clinical + Images	93.21%
Clinical	92.67%
Images	62.11%
90 SNP + GO + HPO	52.16%

Table 2: Three class method: Type I versus Type II versus Type III by Input Type

Three Class Method:	Area Under Curve %
Clinical (Baseline + First Visit)	78.05%
Clinical (Baseline)	77.72%
Clinical + Images	77.39%
Clinical + 90 SNP	76.73%
Clinical + GO + HPO	75.57%
90 SNP	43.23%
90 SNP + GO + HPO	42.40%
GO + HPO	39.93%

Figure 2. Leave One Out Cross-validation Clinical + Imaging



Genes indicated from embedding methods found in **Tables 3 and 4** may point to biologic mechanism progression of PD. *CTNNB1* may rejuvenate the microenvironment, and promote neurorescue and regeneration, part of the pathway for *(Wnt)/ß-catenin* (WβC) signaling is a vital pathway for dopaminergic (DAergic) neurogenesis (1). Another target potentially for neurprotectivity involves the inhibition of Sonic Hedgehog (SHH) signal transduction signaling (2). Also part of the SHH pathway, the *GLI2* gene encodes a transcription factor that participates in the development of the dopaminergic system during embryogenesis. The sex-determining region Y box 9 (SOX9) gene is a known initiator of gliogenesis, during early astrocyte differentiation.

Figure 3 shows the SHAP diagram of GO terms for SMAD phosphorylation and axon near this axis with HPO terms point to balance, coordination, muscle weakness, abnormal insulin and cholesterol levels. Structural measurements of the medial visual cortex, medial temporal and mid-brain regions contribute to subtype differentiation; these regions are reinforced by the biological pathway analysis.

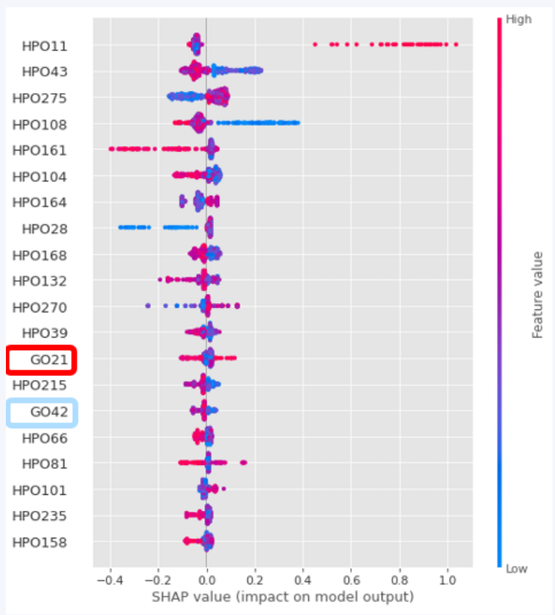
Table 3. Associated gene counts of top 10 HPO embeddings and 50 codes matching across embeddings

Gene:	CTNNB1	SHH	GLI2	GLI3	SOX9	PTCH1	NOTCH1	CREBBP	TP63
Count:	710	630	590	568	562	561	561	555	553

Table 4. Associated gene counts of top 10 GO embeddings and 50 codes matching across embeddings

Gene:	CTNNB1	SHH	GLI2	CREBBP	NOTCH1	SOX9	GLI3	PTCH1	CAV1
Count:	751	637	601	590	586	579	575	573	573

Figure 3 .SHAP Diagram of features near axis



Conclusions: PD subtyping to identify fast or slow disease progressors or stratifying by genetic carrier status remains especially challenging due to small sample size. Our PPMI study results identified specific genes (*CTNNB1*, *SHH*, and *GLI2*) from HPO embeddings/biologic pathways (using gene to GO term mappings as an intermediary) combined with brain regions characterizing fast/slow PD progressors independent of Hoehn and Yahr disease stage. Utilizing embedding algorithms along with HPO/GO terms increased the Area Under the Curve (AUC) by ~5% compared to using genetic variants directly. Clinical + imaging inputs provided the highest AUC for two (93.2%) or three (77.4%) subtype models.

- Findings will provide mapping between genetic risk variants, symptom progression along with brain network health useful for clinical trial inclusion.
- Future personalized medicine applications include specific drug targets by brain region and PD-related biologic mechanism.

Next Steps: Currently, PD progression milestone inputs to enhance clinical outcome interpretation are being incorporated.

Plans to replicate our approach in additional Accelerating Medicines Partnership (AMP) datasets, such as PD Biomarkers Program (PDBP), LRRK2 Cohort Consortium, Harvard Biomarker Study,

- especially for longitudinal clinical measurements, imaging (where available), and genomic data for HPO/GO term (node2vec) embedding.

Disclosures: M.A.N.'s, F.F.'s and A.D.'s participation in this project was part of a competitive contract awarded to Data Tecnica International LLC by the National Institutes of Health to support open science research, part of an open science collaboration between CARD and Invicro. M.A.N. is also advisor for Clover Therapeutics and Neuron23 Inc. L.M.P., H.M.S., M.V., A.T., T.G., A.C., R.G., J.H., and B.A. are employees of REALM IDx.

References:

1. Marchetti B, et al. Aging Cell. 2020;19:e13101.
2. Schimdt S., et al. Nat Commun. 2022 Aug 16;13(1):4819