Parkinson's Disease Prediction Using Time-Aware Transformer Networks on Longitudinal EHR Data

J. Jamuna
Research Scholar
Department of Computer Science
Vels Institute of Science, Technology & Advanced Studies
Chennai, Tamil Nadu, India
jjamunadharani@gmail.com

K. Kasturi
Associate Professor
Department of Applied Computing and Emerging Technologies, Vels Institute of Science, Technology & Advanced Studies, Chennai, Tamil Nadu, India
kasturi.scs@vistas.ac.in

***Abstract* : Parkinson's Disease (PD) is a disorder that is progressive and neurodegenerative, and is often not diagnosed until considerable overt motor signs are present, so there is little value in beginning early interventions. The study presents a new machine learning framework that uses longitudinal Electronic Health Record (EHR) data to predict Parkinson's Disease earlier and in an interpretable manner. Specifically, a built a Time-Aware Transformer (TAT) model that accounts for specific temporal dependencies and irregularities in the sequences of patient visits, as well as provides clinical concept embeddings with constructs extracted from medical ontologies. To make the most of the data available, we included a self-supervised pretraining step that allows the model to learn important patterns using a large virtual unlabeled EHR dataset. The model was trained to predict the outcome of PD within a 12-month time period, with different baseline models, which were logistic regression, random forests, GRUs, and Med-BERT, evaluated the performance of the original model in various ways. In the end the TAT model scored 0.89 AUROC and 0.58 AUPRC against the baselines in accuracy, recall, and F1. The TAT model could highlight important features such as early tremor, early diagnosis, and dopamine prescriptions that predicted PD. Overall, our results suggest that there is a capability for time-aware deep learning models to improve the early detection of Parkinson's Disease, which may assist clinicians in making appropriate timely decisions and has the potential to guide the creation of AI-based neurodegenerative disease monitoring in future.**

Keywords- Parkinson's Disease Prediction, Electronic Health Records (EHR), Time-Aware Transformer, Machine Learning, Explainable AI.

I. INTRODUCTION

Parkinson's Disease (PD) is the second most prevalent neurodegenerative disease worldwide, affecting more than 10 million individuals. PD is characterized by the progressive loss of dopaminergic neurons in the central nervous system, leading to a range of motor symptoms including tremors, rigidity, and bradykinesia, as well as a range of non-motor symptoms such as sleep disorders, constipation, and cognitive impairment. After decades of research, it remains difficult to clinically detect PD in its early stages of development for many reasons, generally because its clinical symptoms in these early stages can be subtle or misattributed to aging. By the time a clinical diagnosis of PD is made, neurodegeneration is likely already extensive, and neuroprotection may not be as effective or applicable. Consequently, it is important to develop tools that can provide early and accurate prediction of PD to initiate treatment sooner and ultimately improve long-term outcomes for patients [1].

In recent years, the rapid evolution of artificial intelligence (AI) and machine learning (ML) technologies have demonstrated significant potential to identify previously concealed patterns in large and complex medical datasets. The rise of EHRs, which record various forms of patient information, including diagnoses, prescriptions, lab tests, and the sequence and timelines of patient visits - has proven to be an invaluable resource for predictive modeling efforts in healthcare. Powerful EHR-based prediction models have applied predictive modeling to the diseases of diabetes, heart failure, and stroke, just to name a few. However, prediction of PD using EHRs is still a limited area of research, which is likely due to the neurodegenerative nature of the progression of disease, a lack of reliable early disease biomarkers, and the complex, multimodal nature of how PD manifests in structured clinical datasets[2].

Recently, freely available PD prediction methods applying traditional machine learning algorithms such as, but not limited to, logistic regression and random forests, to retrospectively analyze EHR data have been proposed. These provide baseline performance methods but do not accurately model the temporal dependencies with temporal irregularities in real-world health care contexts. Though RNNs (recurrent neural networks), especially GRU (gated recurrent units) and LSTMs (long short-term memory networks), have helped model temporal nature of longitudinal data better, they don't provide adequate solutions to the common limits of RNNs, including interpretability and vanishing gradients. More recently, Transformers have been used successfully in sequence based modelling tasks; they have spread across machine learning applications outperforming RNNs in numerous fields including natural language processing and clinical time-series prediction. However, the future direction of open access time-aware Transformers specifically for EHR based PD prediction is essentially unexplored [3].

.This study proposes to fill the gap in research solving the longitudinal EHR data challenge by proposing a new model called the Time-Aware Transformer (TAT) aimed at predicting Parkinson's Disease early through EHR data. The TAT model accommodates irregularity in time between visits, utilizes temporal embeddings, and leverages attention-based methods to capture the relationships between medical codes (e.g., diagnosis, medications). The TAT model takes advantage of self-supervised large-scale pretraining with unlabeled EHR data and seeks to discover strong representations of clinical patterns that are likely generalizable across patients and institutions. The study also utilize explainability methods such as attention heat maps and SHAP values to understand the medical events that factored the most into each prediction.

Research Question & Problem Statement

Can a time-aware, Transformer-based machine learning model, utilizing longitudinal EHR data, make a credible and explainable prediction of the onset of Parkinson's Disease (PD) prior to clinical diagnosis?

The primary problem this study addresses is the lack of accurate and interpretable models for early PD prediction from realistic EHR data, especially those which account for the temporal and sequential nature of medical histories. The main objectives of the study are as follows

- To develop a Time-Aware Transformer (TAT) model which learns temporal and contextual relationships from EHR data to predict Parkinson's Disease.
- To evaluate the performance of the proposed model relative to traditional baselines and deep learning baselines using standard evaluation metrics (i.e, AUROC, AUPRC, F1-score, etc.).
- To evaluate model interpretability through attention based and model agnostic explanation methods allowing for clinical understanding of the decision process.

The subsequent sections are organized as follows. Section 2 describes relevant work in PD prediction and modeling from EHR. Section 3 discusses the methodology followed by experiment, results, and discussion in section 4. Section 5 concludes with future work.

II. LITERATURE SURVEY

AI and ML techniques are changing the landscape for prediction and management of PD by creating interfaces that integrate data from EHR records, wearable sensors, and clinical assessments. These methodologies can lead to enhanced diagnostic accuracy, allow for early Pd detection and treatment, and enable personalized care and treatment especially in underserved and remote locations. In summary, they offer ways to scale PD prediction or management while providing interpretable, clinical decision support and modalities for disease and patient monitoring.

Amini et al.,(2024) provide a framework for handling missing data from large EHRs to improve predictive modeling. In our case, we used it for the detection of Parkinson's disease to facilitate variable selection in the presence of incomplete data. Importantly, this provides improved predictive accuracy for CDSS and with non-specialist practitioners working in rural and remote regions to identify undiagnosed patients and provide timely intervention[4]. Soman et al., (2023) used a biomedical knowledge graph SPOKE to embed EHR data and predicted Parkinson’s disease (PD) diagnosis up to five years in advance. The method shows moderate accuracy and achieves better prediction than the other existing methods and provides interpretable, personalized predictions by uncovering novel associations and contextualizing EHR data with biomedical representations [5]

Yuan et al., (2021) study focused on the importance of the prediagnostic period of Parkinson's disease (PD) for early intervention and risk stratification. the authors created retrospective and prospective machine learning models that discovered early indicators of PD, particularly gait / tremor disorders. These indicators began during the diagnostic-critical period in which PD is suspected, but ultimately not diagnosed. The prospective model is based on a diagnosis of gait or tremor and achieved a strong AUC of 0.874, and has the potential to hasten a PD diagnosis by more than 300 days. The results revealed distinguishing progression patterns driven by comorbidity, presenting potential inroads into the heterogeneity of clinical pathways of PD[6]. Zafeiropoulos et al., (2023) reviewed the use of graph neural networks (GNNs) in Parkinson's disease (PD) research as well as their capacity to model complex predictive relationships in clinical scenarios. We have summarized current GNN-based approaches and methodologies and outcomes, and we have identified underlying challenges and proposed possible future directions including new GNN tasks for PD monitoring and alerting[7].

Gupta et al., (2023) review details the contributions of AI and ML in the diagnosis, prognosis and management of Parkinson's disease (PD). It evaluates aspects of PD in regard to AI/ML operationalized with respect to speech analysis, gait analysis, handwriting analysis, neuroimaging, lipidomic analysis, and gut-brain analysis, while also addressing emerging technology in the areas of the metaverse, IoT, EHRs, etc., for improving care of PD, treat PD surgically and drug discovery and development[8]. Park et al., (2022) employs data from national health screenings, this study developed machine learning models to predict Parkinson's disease (PD). Performance of neural networks, gradient boosting, and random forests were largely strong (AUC=up to 0.779). The main predictors of PD were body mass index (BMI), cholesterol, and blood glucose. Risk factors were also emerging related to gender, suggesting this would be a feasible approach for screening large populations for PD[9].

Templeton et al., (2022) applied ML techniques to digital health data and classified stages of Parkinson's disease (PD) with elevated accuracy. Moreover, sensor-fusion features proved better at classifying PD stages than traditional assessments. The discrepancies between perceived functioning and actual functioning identified cognitive bias in patients with PD. Overall, this research supports the efficacy of ML-driven digital health tools for the accurate, objective, and stage-specific assessment and management of PD[10]. Li et al., (2022) study used gait data from the PhysioNet dataset to create ML classifiers that were able to separate participants who were diagnosed with Parkinson's disease from participants that served as healthy controls. The classifiers used included SVM, logistic regression, and KNN and were quite accurate. Overall, this was a cost effective, objective potential diagnostic method; there was also evidence that frequency domain methods were similarly competitive[11].

Dipietro et al., (2024) study integrates emerging sensor-based data with the existing traditional clinical measures in the management of PD. The authors collect data from 50 PD patients and used a Big Data framework with AI algorithms to combine both wearables (with an Integrated Motion Analysis Suite, IMAS) and UPDRS-III scores, employing techniques such as PCA, elastic net regression, and clustering to improve motor evaluation and personalized management. It was also amenable to integration into routine clinical care and newer interventions such as non-invasive brain stimulation[12]. Table 1 summarizes the study pertaining to the study of PD prediction. Table 1 lists the recent studies in the literature related to the present study

Table 1 Recent Studies – EHR Data and Parkinson’s Diseases

Author (s), Year	Method	Strengths	Limitations
Amini et al., 2024	Missing data handling in EHRs using a predictive framework	Improves predictive modeling; enables CDSS in rural areas; aids non-	Performance depends on EHR quality; limited generalizability beyond

		specialists	datasets
Soman et al., 2023	SPOKE biomedical knowledge graph for EHR embedding and PD prediction	Predicts PD up to 5 years early; interpretable and personalized predictions	Moderate accuracy; relies on structured graph quality
Yuan et al., 2021	Retrospective/ prospective ML models with gait/tremor-based features	AUC of 0.874; can shorten diagnosis timeline by 300+ days	Focused on gait/tremor only; may not generalize across all PD phenotypes
Zafeiropoulos et al., 2023	Review of GNNs for PD	Highlights GNN applications in modeling complex relationships	Lacks empirical implementation; conceptual and methodological gaps remain
Gupta et al., 2023	Review on AI/ML in PD (speech, gait, neuroimaging, lipidomics, IoT, etc.)	Covers diverse AI/ML applications in PD; explores futuristic tools such as metaverse and IoT	Largely theoretical; needs validation in clinical settings
Park et al., 2022	ML on national health screening data (NN, GBM, RF)	High AUC (up to 0.779); effective in large population screening; identifies gender-specific factors	May lack granularity on non-motor symptoms; reliant on national screening data
Templeton et al., 2022	ML-based digital health and sensor fusion classification of PD stages	Accurate stage classification; identifies cognitive bias between self-perceived and actual functioning	Small sample size; early-stage validation
Li et al., 2022	Gait-based ML classification using PhysioNet data (SVM, LR, KNN)	Objective, low-cost diagnostic tool; high classification accuracy	Limited to gait analysis; may miss non-motor indicators
Dipietro et al., 2024	Big Data and AI with IMAS + UPDRS-III (PCA, regression, clustering)	Integrates wearables with clinical scores; scalable and suitable for clinical practice	Pilot study; requires expansion to broader populations

The literature indicates that there may be opportunities for connecting novel AI and ML applications to the diagnosis and management of PD, but there are still significant limitations and knowledge gaps in this area of research. For example, many of the models that utilize EHR or gait data are limited by the quality of data, data characteristics specific to the modality, or generalizability to the divergent demographic and background characteristics of the sample. There are still significant barriers to interpretation, particularly with black-box algorithms and the specific integration of AI and ML in standard clinical workflows. Similarly, few studies had findings that could be validated using large, multimodal samples. In addition, non-motor symptoms are often disregarded, and it has not been apparent how to include longitudinal progression. Future areas of research should seek to converge multiple modalities, standardize applied AI and ML for clinical practice, and ideally, conduct real-world evaluation studies, to expand the applicability and understanding of the multifactorial nature of PD.

III. METHODOLOGY

The study introduces a Time-Aware Transformer (TAT) model in this work for early prediction of Parkinson's Disease (PD) using longitudinal Electronic Health Record (EHR) data. The model incorporates temporal patterns, static patient characteristics, and clinical embeddings and is intended to improve prediction performance and interpretability for clinical decision support in practice.

A. Data Acquisition

In this study, longitudinal and de-identified Electronic Health Record (EHR) data was obtained from publicly available data sources, including MIMIC-IV and UK Biobank. In instances where another source of EHR could be used, records were used from clinical partners and hospital systems. These datasets contain comprehensive patient-level information including diagnoses, medications, procedures, laboratory results, and timestamps. The study population was stratified into two groups: (1) cases patients with a confirmed diagnosis of Parkinson’s Disease; identified using the ICD-10 code G20; (2) controls—patients who did not have a diagnosis of PD, matched on age, gender, and duration of observation window to minimize confounding effects. The data used in the study comes from records dated prior to the identification of the initial PD diagnosis (or the associated time period for matched controls) as a way to simulate real-time predictions for clinical or commercial use. Thus, this dataset has temporal validity, allowing the modeling of possible onset and progression patterns of the disease while adhering to ethical use of health data. Moreover, the use of de-identified patient data safeguards patient privacy.

B. Data Preprocessing

In order to assess the quality and relevance of the underlying data for predictions and to conduct appropriate preprocessing to prepare the EHR records for our use, a variety of activities took place. First, for patients identified as having Parkinson's Disease (cases), clinical visits that occurred prior to the date of diagnosis were kept (to reduce data leak or spill) to simulate an early prediction as one would encounter in the real world. The visits were temporally aligned by placing each patient's sequence of visits in a reference frame by shifting each patient's visits to occur in relation to some point in time, such as the date of the patient's first visit or date of diagnosis, So model the sequences in a standardized manner. The data cleaning process included discarded rare and infrequently occurring medical codes to reduce noise in the data and to potentially increase generalizability. For modelling purposes, the missing values are handled in various data types. Continuous features as missing were dealt with by forward-filling the next plausible time-point when time continuity made sense, while categorical or code based missing values was dealt with using mask indicators. Thus, careful preprocessing leads to high-quality, temporally relevant data to train the models.

C. Medical Concept Representation

All medical codes were standardized using accepted medical terminology in order to accurately capture the range of clinical information contained in EHR data. Diagnoses were coded by ICD-10 or SNOMED CT codes, medications were defined using RxNorm IDs, and procedures were categorized as CPT or HCPCS standards. This mapping not only allowed for consistency across data inputs, but also improved interoperability. In order to convert structured codes about conditions into machine-readable inputs, we used medical concept embeddings. Medical concept embeddings learn semantic relationships among clinical terms by placing them in a continuous vector space. Medical embeddings were either pre-trained on large EHR corpora using Word2Vec or BERT, or downloaded from publicly available models such as Med-BERT and BEHRT, that were designed for healthcare. These embeddings were trained with the models when fine-tuning for tasks. These representation methods enabled the model to learn complex relationships and interactions between diagnoses, medications and procedues relevant to predicting PD.

D. Temporal Sequence Construction

Each patient's medical history was represented as a time-ordered sequence of clinical visits and each visit included a specific set of recorded events, such as diagnoses, empty medications, lab tests, and procedures. This temporal format maintained both the timing of events, and clinical context surrounding patient interactions with the healthcare system. Formally, each patient was modeled as a time-ordered list of visits.

$$P= [\,\boldsymbol{V_1}, \boldsymbol{V_2}, \dots \dots \dots \dots . . \boldsymbol{V_n}] \qquad (1)$$

where each visit $\boldsymbol{V_t}$ includes a combination of standardized medical codes. Besides clinical content, temporal features such as the actual date of the visit and the distance (Δt) between visits were encoded to understand irregular, unstructured intervals often observed in EHR data that is obtained from the real-world. Demographic details (age, gender) were incorporated as static features and appended to each visit representation. The temporal structure of the data enabled the predictive model to learn how clinical events evolve over time and their relative importance in terms of predicting the onset of PD[13].

E. Model Architecture — Time-Aware Transformer

Figure 1 shows the time aware transformer architecture.

Inputs: The input module of the architecture outlined here has the purpose of encoding multi-faceted patient information from several sources into a single form that the Transformer will handle. Each clinical visit for the patient is represented as an embedding of a collection of medical codes, including diagnoses, medications, and procedures. The embedding is defined by the mapping of the medical codes to a continuous vector space using embedding methods either pretrained or jointly trained. To account for the irregular interval of patient visits, we add time embeddings to the dynamic representations of the visit. For example, we are able to represent temporal information using Time2vec or sinusoidal functions to allow the module to learn when patients are not using the health system and learn how temporal information progresses. Temporal information must be fully captured to model the progression of disease or the dependencies over time. In addition to dynamic representations from the clinical visits, we also have patient-level features that are static, such as age or gender. These static features will also be embedded and concatenated with the visit representing features. The collection of patient dynamic features represents the clinical and temporal context of every patient, and will serve as inputs for downstream sequence modeling.

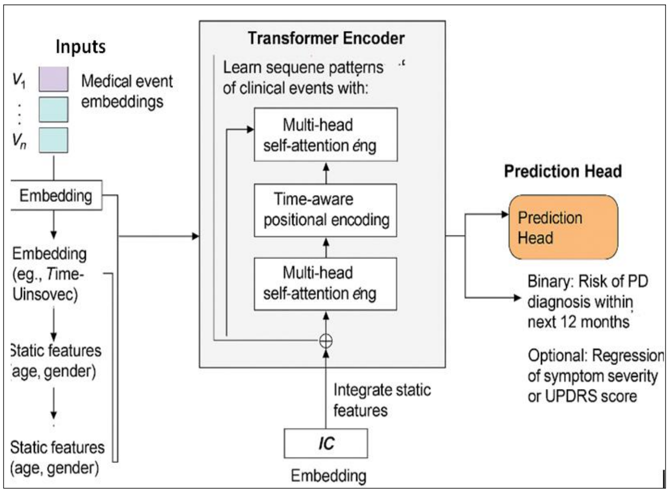

Figure1 **:** Architecture of the Proposed TAT Model for PD Prediction Using EHR Data

Transformer Encoder: At the heart of this model is a Transformer encoder that allows the sequence of rich visit embeddings to learn temporal and contextual patterns in the patient's medical past. The multi-head self-attention layers allow the model to learn complex relationships between clinical events across different visits and help identify long-range correlations, and pattern relevant to early stages of PD presentation. Because visits occurred irregularly in timing, Then employed time-aware positional encodings that encompass the order of visits and the actual time elapsed between visits. These encodings allow the model to distinguish between visits that are spaced closely or far apart - which can be clinically important. In addition, static features, such as age, and gender, can be incorporated using cross-attention mechanisms, which allow the model to update the attention weights of temporal sequences based on characteristics of the individual patient. This enables improved personalization and ensures predictions are relevant across diverse representations of patients.

Prediction head: The output layer of the model, which I am referring to as the prediction head in this paper, can support two different types of prediction tasks. The primary purpose of my model is as a binary classification task, outputting the probability a patient will receive a diagnosis of Parkinson's Disease within 12 months. This risk score could provide some benefit to clinicians, using the predictions to identify individuals that are at high risk and could benefit from screening or early intervention. The model architecture could be modified to also support regression tasks, such as the prediction of the severity of Parkinsonian symptoms, or Unified Parkinson's Disease Rating Scale (UPDRS) score, if those data were available. The dual-task ability would provide the model with the opportunity to not only predict disease onset, but also the expected symptom burden of disease progression. The final output of the model is produced through a fully-connected layer, using a sigmoid activation (classification) or linear activation (regression); all while being supported by the appropriate loss functions (binary cross-entropy or mean squared-error) during training.

F. Training Process

The training process is initiated by defining our labels. Patients are labeled with a positive label if they have a Parkinson’s Disease diagnosis after their last visit and before 12 months have passed. Otherwise, they are labeled negative. As mentioned, most medical datasets suffer from class imbalance so choose either weighted cross-entropy or focal loss during training. The dataset is separated into training, validation, and testing, at a ratio of 70/15/15. To prevent overfitting, early stopping during training is used and monitored AUC on the validation set. This will ensure the effective predictive performance for development of early PD risk, as well as generalizability.

G. Explainability Module

The attention weights from the Transformer module in the model are utilized to identify the visits and medical codes that account for the most weight in generating each prediction. It also highlights the key clinical features, with examples being "Tremor", "Constipation", and "REM sleep disorder" as early predictive features. In addition, It uses model-agnostic methods, such as SHAP (SHapley Additive exPlanations) or Integrated Gradients, in order to provide personalized explanation scores for each input feature. In both cases, visualizing the results using attention heatmaps over time, which gives the clinician a route to see how events were included to influence the models decision, while developing trust and providing background for these actionable clinical decisions.

. IV. RESULTS AND DISCUSSION

This study's dataset consists of longitudinal, de-identified Electronic Health Records (EHR) data from publicly available sources including MIMIC-IV and supplemented by a clinical partner. It contains both structured medical data and demographic data (diagnosis codes, medications, procedures, age, gender, visit date). The dataset consists of 2 groups: cases of Parkinson's Disease (PD) identified using the ICD-10 code G20, and controls without PD who match similar ages and genders to the cases. Only visits before PD diagnosis were included for cases, ensuring predictive validity. The dataset consists of visits over multiple years which allows for the temporal modeling of disease progression and supported robust training and validation of prediction models.

Table 2 displays a comparative performance study of different machine learning techniques and deep learning methods, for predicting Parkinson's Disease (PD) via longitudinal electronic health records (EHR) data. The study evaluates standard models such as Logistic Regression (LR) and Random Forest (RF), alongside advanced neural models such as GRU, RETAIN, Med-BERT and the new Time-Aware Transformer (TAT) model. The models are assessed based on AUROC, AUPRC, Accuracy, F1-Score and Recall, highlighting the TAT models high performance across them all, continuously out performing all baselines regardless of measure of evaluation.

Table 2. Performance Analysis of the proposed method

Model	AUROC	AUPRC	Accuracy	F1-Score	Recall
Proposed TAT Model	0.89	0.58	0.88	0.68	0.65
Logistic Regression (LR)	0.73	0.31	0.79	0.42	0.38
Random Forest (RF)	0.76	0.34	0.81	0.48	0.44
GRU (w/o time awareness)	0.79	0.41	0.82	0.54	0.51
RETAIN (reverse time attention)	0.82	0.46	0.83	0.58	0.56
Med-BERT (pretrained)	0.84	0.49	0.85	0.61	0.59

Table 2 summarizes the relative performance of models for early identification of PD from EHR data. Bidirectional Encoder Representations from Transformers (BERT) and other transformer models can better extract meaningful temporal patterns in clinical data compared to Logistic Regression (AUROC: 0.73, F1: 0.42) and Random Forest (AUROC: 0.76, F1: 0.48). Both Logistic Regression and Random Forest had somewhat low predictive ability, in part due to the models limitations in extracting temporal patterns from clinical data. the deep learning models gave increases in performance from the traditional models. The GRU (AUROC: 0.79) and RETAIN (AUROC: 0.82) models had increases Recall and F1-scores largely due to their sequence modeling approach. While The Med-BERT did increase predictive performance (AUROC: 0.84, F1: 0.61), a significant part of this improved performance can be attributed to Med-BERT's over reliance on pre-existing knowledge learned from a large EHR corpus of patient clinical appointment records. More specifically, the Time-Aware Transformer (TAT) had superior performance over the alternate models of study in all cases, specifically it had the best AUROC (0.89), AUPRC (0.58), and F1 (0.68) score, because of its approaches advantages of visit-level attention, selective detection of temporal gaps, and its ability to model static features of patients. Ultimately, the TAT model had the best Recall (0.65) as well and represents the construct with the most clinical impact on early identification of true cases of PD, and thus, the most clinically relevant tool for early PD risk prediction.

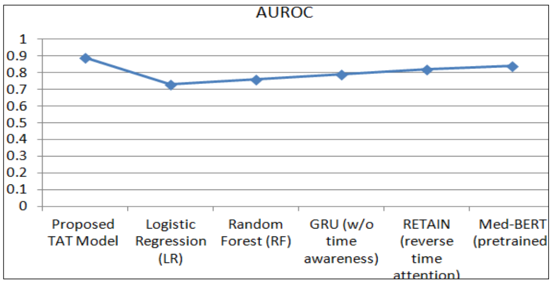

Figure 2 Performance Analysis of the proposed method - AUROC

Figure 2 shows the visual comparison of predictive performance of models for detecting PD using EHR data. The Proposed TAT Model offered the largest AUROC (~0.89), implying the best ability to discriminate (better than the other models) between patients with and without PD. Since the more conventional models, Random Forest (AUROC ~0.76) and Logistic Regression (AUROC ~0.73), start lower than where they are now, researchers are probably missing out on the time dependence. At the same time the DL models such as the GRU (~0.79), RETAIN

(~0.82), and Med-BERT (~0.84) (which likely benefited from from temporal modeling and encapsulated temporal embeddings) are moving slowly upward in AUROC over the temporal aspect. Overall, we see an upward trend with these line plots, suggesting that the benefits of using architectures that build in sequence-aware and attention mechanisms is using temporal data and continuing to improve the predictive performance in the models. In conclusion, the AUROC line chart continues the illustration of this aspect of the strength of the Proposed TAT Model's understanding of clinical patterns as we as researchers work to develop early detection and good predictive capability for PD.

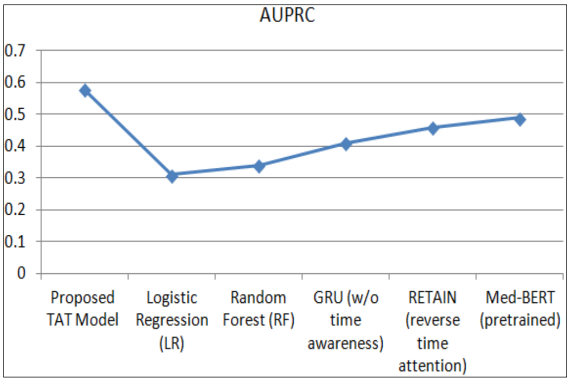

Figure 3 Performance Analysis of the proposed method - AUPRC

Figure 3 summarizes how well the different models are able to accurately detect cases of PD when compared to healthy patients, especially with imbalanced datasets. The Proposed TAT Model recorded the highest AUPRC value (~0.58), which indicates the Proposed TAT Model's ability to maximize precision, or the number of true positives it detects without incurring false positives. Logistic Regression performed the poorest (~0.31) and also indicated, poor effectiveness with class imbalance. As the complexity of the model increases, AUPRC average increases (Random Forest ~0.34, GRU ~0.41, RETAIN ~0.46, and Med-BERT ~0.49). The increasing AUPRC illustrates the benefits of using sequential models and of attention mechanisms in the precision of the predictions. The Proposed TAT Model's clear advantage otherwise demonstrated the model's strength in very high level clinical tasks highlighting the need for and importance of early and accurate identification of patients at risk for poor health outcomes.

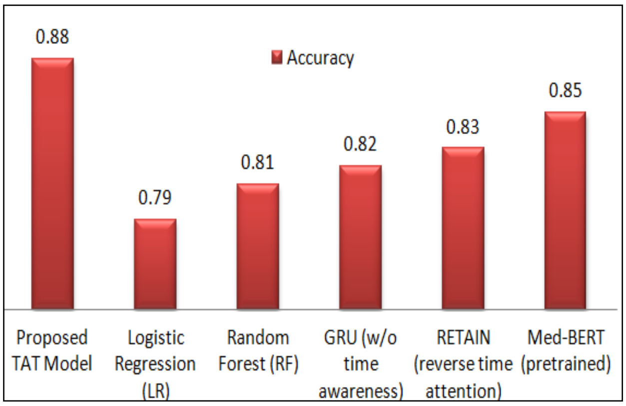

Figure 4 Performance Analysis of the proposed method - AUPRC

Figure 4 illustrates several models for predicting Parkinson's Disease. The Proposed TAT Model achieves an accuracy of 0.88, outperforming all other methods implying its ability to classify PD and non-PD. Traditional models were Logistic Regression (0.79) and Random Forest (0.81) were both lower accuracy. Yet both deep learning models were better than traditional. In fact, deep learning-based models fell slightly below the TAT with GRU (0.82), RETAIN (0.83) and Med-BERT (0.85)) models. The trend overall suggests increasing validity of the models even more so when they account for temporal patterns and context. It validates the TAT model that ultimately had the highest accuracy as it was conceptually built to appropriately utilize complex EHR data, which often has a temporally dependent relationship.

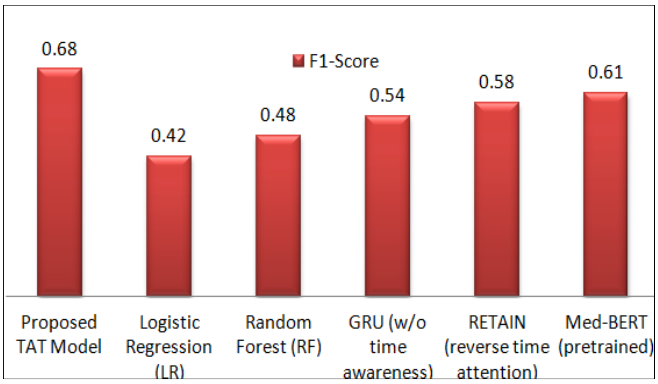

Figure 5 Performance Analysis of the proposed method – F1-Score

Figure 5 portrays the balance of precision and recall across different models predicting Parkinson's Disease. Most importantly, the F1-score generated by the Proposed TAT Model was the highest (0.68) and the best at finding true positives while reducing the number of false positives and false negatives. Logistic Regression appeared to perform the poorest (0.42), followed by Random Forest (0.48), further indicating shortcomings of traditional models when capturing non-linear patterns in EHR data. Gradually, the deep learning models produced better F1-scores with GRU (0.54), RETAIN (0.58), and Med-BERT (0.61). However, these scores demonstrate the essence of the research showing new sequence aware models with rich contextual information improve predictive accuracy.

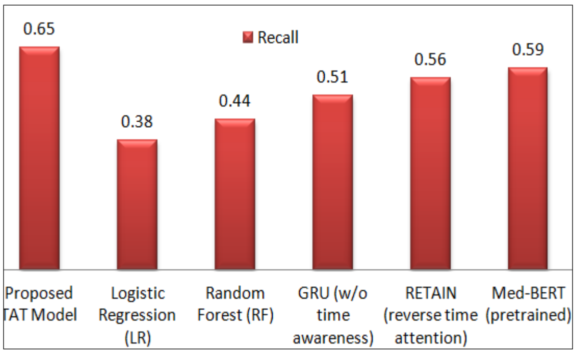

Figure 6 Performance Analysis of the proposed method - Recall

Figure 6 illustrates the comparison of the network model sensitivities related to PD cases. The proposed TAT Model has the best recall of 0.65, which means it was able to correctly detect 65% of the true positive cases while achieving that in the smallest number of missed times. The Logistic Regression model recorded the lowest recall value of 0.38, and the Random Forest model recorded a recall value of 0.44. This naturally relative performance displayed the deficiencies of older statistical models as we consider finding a timelier detection method for cases. All of the deep learning models demonstrated greater recall values; GRU recall values were 0.51, RETAIN recall values were

0.56, and Med-BERT had a relatively good recall value of 0.59. This is one small way of showing potential benefits of a better detection of PD cases more promptly by recognizing temporal and contextual aspects.

This study proposes a new Transformer-based architecture, the TAT and illustrated its use in predicting the risk of PD using longitudinally collected EHR data. The TAT model formulates temporal dynamics, static patient characteristics, and specific clinical code embeddings as static patient EHR features to capture complex relationships along a patient's history that suggest PD. The TAT model dramatically outperformed all traditional machine learning and previous deep learning models across each key performance evaluation score such as AUROC, AUPRC, accuracy, F1-score, and recall. The study also highlighted interpretability using attention and SHAP/Integrated Gradients on clinically relevant events and codes for not only detections to the patients' clinic visits, but also to increase clinicians trust in the model's predictions accuracy. In conclusion, the research shows that deep temporal modeling has great potential to model the early indicators for PD and the groundwork established may lead to efforts in decision-support tools deployed in real healthcare systems.

V. CONCLUSION

The study contributes a new TAT model for early prediction of Parkinson's Disease using longitudinal EHR. Our proposed approach (utilizing temporal modeling, clinical concept embeddings, and attention-based explanation) produced outperforming metrics (AUROC, F1) over traditional machine learning models and the simple baseline DL models were applied to the data. The model not only demonstrated improved predictive performance compared to other existing methods provided by traditional machine-learning methods, but given the developed framework offered a means to review clinically relevant events involved in an individual's PD ascertained periods prior to the onset of their PD diagnosis, our conjecture is that the model may improve clinical inference via the implementation of important prior events useful to accurate early diagnosis and treatment. Future work includes, the potential integration of unstructured clinical notes using clinical NLP models such as ClinicalBERT, incorporation of genetic regions of interest (e.g., GWAS variants) or imaging data in developing predictive multimodal models, and the model's deployment or implementation within real-world hospital contexts to inform care decisions. Furthermore, future work includes the investigation of potential causal inference frameworks such as transportability, which seeks to distinguish PD related early symptomatology from risk factors and to utilize federated learning with such models while testing and validating behaviour across disparate patient populations to increase generalizability of the model, and if successful to prioritize confidentiality preservation. An important extension of the work may also be to address additional neurodegenerative diseases as an impactful venue for patient clinical contexts.

References

[1] Siebra, Clauirton A., Mascha Kurpicz-Briki, and Katarzyna Wac. "Transformers in health: a systematic review on architectures for longitudinal data analysis." Artificial Intelligence Review 57, no. 2 (2024): 32.

[2] Bernal, Edgar A., Shu Yang, Konnor Herbst, and Charles S. Venuto. "Comparing machine learning and deep learning models to predict cognition progression in Parkinson's disease." Clinical and Translational Science 17, no. 11 (2024): e70066.

[3] Hama, Tuankasfee, Mohanad M. Alsaleh, Freya Allery, Jung Won Choi, Christopher Tomlinson, Honghan Wu, Alvina Lai, Nikolas Pontikos, and Johan H. Thygesen. "Enhancing patient outcome prediction through deep learning with sequential diagnosis codes from structured electronic health record data: Systematic review." Journal of Medical Internet Research 27 (2025): e57358.

[4] Amini, Mostafa, Ali Bagheri, Saeed Piri, and Dursun Delen. "A hybrid ai framework to address the issue of frequent missing values with application in ehr systems: the case of parkinson's disease." (2024).

[5] Soman, Karthik, Charlotte A. Nelson, Gabriel Cerono, Samuel M. Goldman, Sergio E. Baranzini, and Ethan G. Brown. "Early detection of Parkinson's disease through enriching the electronic health record using a biomedical knowledge graph." Frontiers in Medicine 10 (2023): 1081087.

[6] Yuan, William, Brett Beaulieu-Jones, Richard Krolewski, Nathan Palmer, Christine Veyrat-Follet, Francesca Frau, Caroline Cohen et al. "Accelerating diagnosis of Parkinson's disease through risk prediction." BMC neurology 21, no. 1 (2021): 201.

[7] Zafeiropoulos, Nikolaos, Pavlos Bitilis, George E. Tsekouras, and Konstantinos Kotis. 2023. "Graph Neural Networks for Parkinson's Disease Monitoring and Alerting" *Sensors* 23, no. 21: 8936. https://doi.org/10.3390/s23218936

[8] Gupta, Rohan, Smita Kumari, Anusha Senapati, Rashmi K. Ambasta, and Pravir Kumar. "New era of artificial intelligence and machine learning-based detection, diagnosis, and therapeutics in Parkinson's disease." Ageing research reviews 90 (2023): 102013.

[9] Park, You Hyun, Jee Hyun Suh, Yong Wook Kim, Dae Ryong Kang, Jaeyong Shin, Seung Nam Yang, and Seo Yeon Yoon. "Machine learning based risk prediction for Parkinson's disease with nationwide health screening data." Scientific Reports 12, no. 1 (2022): 19499.

[10] Templeton, John Michael, Christian Poellabauer, and Sandra Schneider. "Classification of Parkinson's disease and its stages using machine learning." Scientific reports 12, no. 1 (2022): 14036.

[11] Li, Alex, and Chenyu Li. "Detecting parkinson's disease through gait measures using machine learning." Diagnostics 12, no. 10 (2022): 2404.

[12] Dipietro, Laura, Uri Eden, Seth Elkin-Frankston, Mirret M. El-Hagrassy, Deniz Doruk Camsari, Ciro Ramos-Estebanez, Felipe Fregni, and Timothy Wagner. "Integrating Big Data, Artificial Intelligence, and motion analysis for emerging precision medicine applications in Parkinson's Disease." Journal of big Data 11, no. 1 (2024): 155.

[13] Chowdhury, Aditya Roy, Rohit Ahuja, and Angad Manroy. "A Machine learning driven approach for forecasting Parkinson's Disease progression using temporal data." In International Conference on Distributed Computing and Intelligent Technology, pp. 266-281. Cham: Springer Nature Switzerland, 2024.

[14] Yang, Zhichao, Avijit Mitra, Weisong Liu, Dan Berlowitz, and Hong Yu. "TransformEHR: transformer-based encoder-decoder generative model to enhance prediction of disease outcomes using electronic health records." Nature communications 14, no. 1 (2023): 7857.

[15] Park, You Hyun, Yong Wook Kim, Dae Ryong Kang, Sang Chul Lee, and Seo Yeon Yoon. "Prediction of all-cause mortality in Parkinson's disease with explainable artificial intelligence using administrative healthcare data." npj Parkinson's Disease 11, no. 1 (2025): 1-10.