IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 April 2024, accepted 21 May 2024, date of publication 24 May 2024, date of current version 3 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3405009

RESEARCH ARTICLE

Multimodal Reinforcement Learning for Embedding Networks and Medication Recommendation in Parkinson's Disease

HONGBUM KIM, CHANYOUNG PARK, JONG HOON KIM, SEJEONG JANG, AND HYO KYUNG LEE, (Member, IEEE)
School of Industrial and Management Engineering, Korea University, Seoul 02841, Republic of Korea

Corresponding author: Hyo Kyung Lee (hyokyunglee@korea.ac.kr)

This work was supported in part by the National Research Foundation (NRF) funded by Korean Government (MSIT) under Grant RS-2023-00212713, in part by Korea Institute for Advancement of Technology (KIAT) funded by Korean Government (MOTIE) through the Competency Development Program for Industry Specialist under Grant P0020649, and in part by the BK21 FOUR funded by the Ministry of Education of Korea and National Research Foundation of Korea.

ABSTRACT This study presents a comprehensive framework for analyzing Parkinson's disease (PD) progression and optimizing medication recommendations based on multimodal data integration and reinforcement learning (RL) techniques. The research framework integrates diverse biomarkers from structured clinical assessments, imaging data, and biospecimen information, creating various single, bimodal, and trimodal datasets. Deep learning-based embedding networks are employed to transform high-dimensional data into lower-dimensional representations, capturing essential patterns related to PD progression. The learned representations are evaluated using visualization methods and predictive modeling to differentiate clinically relevant outcomes. Subsequently, medication recommendation models are developed based on the learned representations, assisting clinicians in making informed decisions to manage PD progression effectively. Optimal medication policies are derived considering different modality settings, revealing insights into medication prescribing tendencies across modalities. The findings highlight the significance of integrating diverse data modalities in understanding and managing PD progression, showcasing the potential of RL-based medication recommendation systems in clinical decision-making. This research contributes to the advancement of personalized medicine strategies for PD patients, emphasizing the importance of tailored medication regimes based on multimodal data insights.

INDEX TERMS Reinforcement learning, Parkinson's disease, representation learning, multimodal data, medication recommendation, personalized medicine.

I. INTRODUCTION

Parkinson's disease (PD) is a chronic neurological disorder characterized by impaired function of dopamine-producing cells in the central nervous system, leading to a gradual decline in the ability to control movement [1]. The primary symptoms of PD include progressive bradykinesia, rigidity, rest tremor, and postural instability, which significantly hinder activities of daily living such as walking and communication. Due to the absence of specific diagnostic tests for PD, diagnosis typically relies on the comprehensive assessment of various clinical criteria, encompassing motor symptoms, response to medication, speech, imaging, and others [2]. PD may present with a diverse array of symptoms, typically becoming apparent upon the loss of 60 to 80% of dopamine-producing cells. Currently, there exists no definitive cure for PD, with treatment primarily focusing on symptom alleviation. The severity and progression

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino.

of symptoms vary among individuals, emphasizing the importance of early detection and personalized therapeutic approaches for effective management.

Previous studies have explored machine learning and deep learning approaches to enhance PD diagnosis and prediction. Given the diverse clinical criteria in PD diagnosis, diagnostic methodologies leveraging various data modalities have been suggested [3], [4], [5], [6], [7], [8], [9], [10]. Diminished motor abilities being a primary PD symptom, motor assessment scores and movement data are crucial for diagnosis. Shahid and Singh [6] predicted PD progression using Motor-Unified Parkinson's Disease Rating Scale (UPDRS) [11] and Total-UPDRS scores with a principal component analysis-based deep neural network model, outperforming previous studies by accurately capturing feature space nonlinearity, and El Maachi et al. [8] predicted PD diagnosis and severity using a deep learning-based gait analysis algorithm. Speech impairment is also a prevalent symptom in PD, and utilizing speech signals has enabled early diagnosis and prediction of disease severity. Grover et al. [3] proposed a deep neural network for predicting PD severity using the Parkinson's Telemonitoring Voice Data Set. Sadek et al. [4] suggested a method using artificial neural network models with voice recording data to aid PD identification. Senturk [5] proposed a feature selection method combining machine learning algorithms with voice data for PD identification. Furthermore, image data such as Neuromelanin-Sensitive Magnetic Resonance Imaging Diffusion Tensor Imaging (NMS-MRI) provides insights into dopamine cell loss and is utilized in PD identification where Shinde et al. [9] proposed a convolutional neural network-based PD prediction framework using NMS-MRI.

Recently, a novel approach utilizing machine learning and deep learning methods to integrate previously individually analyzed data modalities has been proposed. By combining diverse data types including non-motor features, cerebrospinal fluid measurements, and dopamine imaging markers, methodology has shown promise in facilitating early diagnosis of PD. Makarious et al. [10] proposed a multimodal prediction model utilizing genetics, clinico-demographic, and transcriptomics features. This model demonstrated superiority over single-data-focused approaches by leveraging the fusion of machine learning with multimodal data.

Given that PD is a complex disorder characterized by heterogeneity in clinical presentation, progression rate, and risk of complications, data-driven approaches have explored the existence of different PD subtypes, which could have implications for understanding disease etiology [12]. Defining and categorizing subtypes through diverse cluster analyses enabled early prediction and the provision of interventions specifically tailored to each subtype [13], [14], [15], [16], [17]. Previous research has delineated PD subtypes by employing data-driven cluster analyses encompassing both motor and non-motor characteristics. These methodologies provide a comprehensive understanding of the clinical and pathophysiological clustering within PD, thereby facilitating the development of personalized therapeutic strategies tailored to the specific conditions of individual patients.

The predominant pharmacological interventions in PD management involve the administration of medications, notably levodopa [18] and carbidopa, aimed at modulating dopamine levels to alleviate symptoms. Additionally, adjunctive pharmacotherapies such as anticholinergics and monoamine oxidase-B (MAO-B) inhibitors are utilized for symptom management [19]. However, it should be noted that prolonged medication use may lead to adverse effects such as dyskinesia and the wearing-off phenomenon. Given that medication administration may not consistently yield favorable outcomes, it becomes imperative to devise an optimal medication delivery strategy that carefully takes into consideration a spectrum of patient conditions.

In recent years, reinforcement learning (RL) has gained prominence in the medical field as a method for discovering optimal treatment policies [20], [21], [22], [23], [24], [25]. RL is a subfield of machine learning, where an agent interacts with an environment to learn the optimal actions to take, making it particularly suitable for problems where sequential decision-making is involved. A notable trend in RL research is the representation of states as latent vectors [23], [24], [25]. This approach involves transforming high-dimensional input data into lower-dimensional representations, often referred to as latent spaces. By doing so, the complexity of the data is reduced, making it more amenable to learning by RL algorithms. This strategy has been increasingly favored over directly using high-dimensional data due to its ability to capture complex patterns and relationships more efficiently. However, previous studies have often underutilized the potential of high-dimensional data and have typically focused on a subset of variables or single modality, limiting the comprehensiveness of their analyses.

Prior research in RL for optimizing PD drug management has been proposed, albeit not as extensively pursued as studies focused on early prediction. Of the few studies, Watts et al. [26] introduced a framework aimed at optimizing personalized medication regimens for PD, with a focus on two primary symptoms continuously monitored using data from wearable sensors: bradykinesia and dyskinesia. Their approach enhanced clinical practice by leveraging deep reinforcement learning (DRL) algorithms to determine optimal drug dosages and administration timings. Baucum et al. [27] developed a data-driven RL framework for optimizing PD medication therapy using wearable sensors, constructing simulation models to understand how individual patient's motor symptoms respond to medication administration. However, there were limitations, including a relatively small sample size and a lack of consideration for patient subtypes. The primary focus of these studies revolved around the analysis of data obtained from wearable sensors, thereby constraining their capacity to incorporate extensive patient

state information. Despite the success in early prediction tasks within the PD domain, RL's potential in optimizing medication delivery strategies remains underexplored, highlighting the need for more comprehensive methodologies.

In parallel, efforts have been made to predict optimal medication dosages, as demonstrated by Riasi et al. [28] and Gutowski et al. [29]. Riasi et al. [28] conducted a study using Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures to optimize the dosages of key treatments, including levodopa. Gutowski et al. [29] optimized drug scheduling using biomedical sensor data, which increased treatment precision by about 7%. While these studies have contributed valuable insights, they are often limited by their reliance on single-modality data, which may not capture the full complexity of PD progression. Moreover, Kim et al. [30] leveraged single modal data from the Parkinson's Progression Markers Initiative database to determine the effective medication combinations for managing the disease state. They employed policy iteration of the Markov decision process (MDP) to accomplish this task. However, patient state was defined according to simple clinical guidelines and thus was not able to capture the complex relationships among the visit or the temporal dependencies and sequential patterns in the data. Moreover, potential adverse effects of medication were not considered as well.

To address these limitations and advance the field of PD medication management, our study takes a comprehensive and multimodal approach. We analyze a wide range of data modalities, including structured clinical assessments, imaging data, and biospecimen information, to identify combinations that best reflect disease progression and treatment response. Specifically, given that longitudinal health records of PD patients are high-dimensional with numerous variables collected across diverse modalities, we explore various embedding methods that best represent complex patient health states. The learned latent representations of the patient health state are further utilized to develop a medication recommendation model where RL is used to derive optimal medication policies. Our goal is to compare the optimal medication policies across multiple modalities and provide insights into modality differences. Consequently, we aim to contribute to the advancement of personalized medicine strategies for PD patients, emphasizing the importance of tailored medication regimes based on multimodal data insights.

Of particular relevance to our study, Bhattarai et al. explored optimal treatment regimens for Alzheimer's disease (AD), a neurodegenerative disorder sharing similarities with PD [31]. Their study involved gathering diverse structured indicators and employing RL to propose a framework for automated decision support systems in AD treatment. This research resonates with our own investigation, as both studies aim to leverage advanced computational techniques, such as reinforcement learning, to optimize treatment strategies for neurodegenerative diseases. However, a key difference lies in the RL model state representation, where Bhattarai et al. derived the states using a decision tree algorithm solely based on data from the patient's latest clinic visit. This resulted in 13 discrete states which is not capable in capturing temporal dependencies among the visits and lacks in providing a diverse and nuanced understanding of patient conditions.

Our work contributes to the field in several key ways. First, we developed a comprehensive framework integrating diverse biomarkers from structured clinical assessments, imaging data, and biospecimen information, providing a comprehensive view of PD progression and treatment response. Second, we employed advanced embedding techniques to transform high-dimensional data into lower-dimensional representations, facilitating the identification of clinically relevant patterns and trends. Third, we developed medication recommendation models based on the learned representations. These models assisted clinicians in making informed decisions to manage PD progression effectively, thus contributing to the advancement of personalized medicine strategies for PD patients. By comparing optimal medication policies across multiple modalities, our study offers insights into the relative effectiveness of different treatment strategies and highlights the importance of personalized medicine in PD management. Ultimately, our research aims to improve the quality of care for PD patients by leveraging multimodal data insights to optimize medication regimens and enhance treatment outcomes.

II. MATERIALS AND METHODS

Figure 1 illustrates the overall research framework where various biomarkers collected from diverse modalities including structured clinical assessments (STR), imaging data (IMG), and biospecimen (BIO) information were utilized. Upon necessary data preprocessing, different types of datasets were created based on the varying combination of the modalities, consisting of 3 single modal (BIO, IMG, STR), 3 bimodal (BIO+IMG, BIO+STR, IMG+STR), and 1 trimodal (ALL; BIO+IMG+STR) datasets. To effectively integrate heterogeneous data modalities, embedding networks were utilized to transform high-dimensional data into lower-dimensional representations that capture important patterns and relationships. The learned latent representations in the lower-dimensional space can serve as state representations of the observed data as they encode relevant information about PD progression, such as disease severity and symptom manifestations. The embedded states of various modality datasets were evaluated using visualization methods and comparing the predictive ability of the latent states in differentiating a clinically relevant outcome metric of PD. Once the suitable latent representations for each data modality have been selected, the state representations were subsequently used to develop medication recommendation models. By predicting future disease states based on the learned representations, our framework can assist clinicians in making informed decisions about medication regimes, such as adjusting medication dosage or recommending different types of drugs to slow

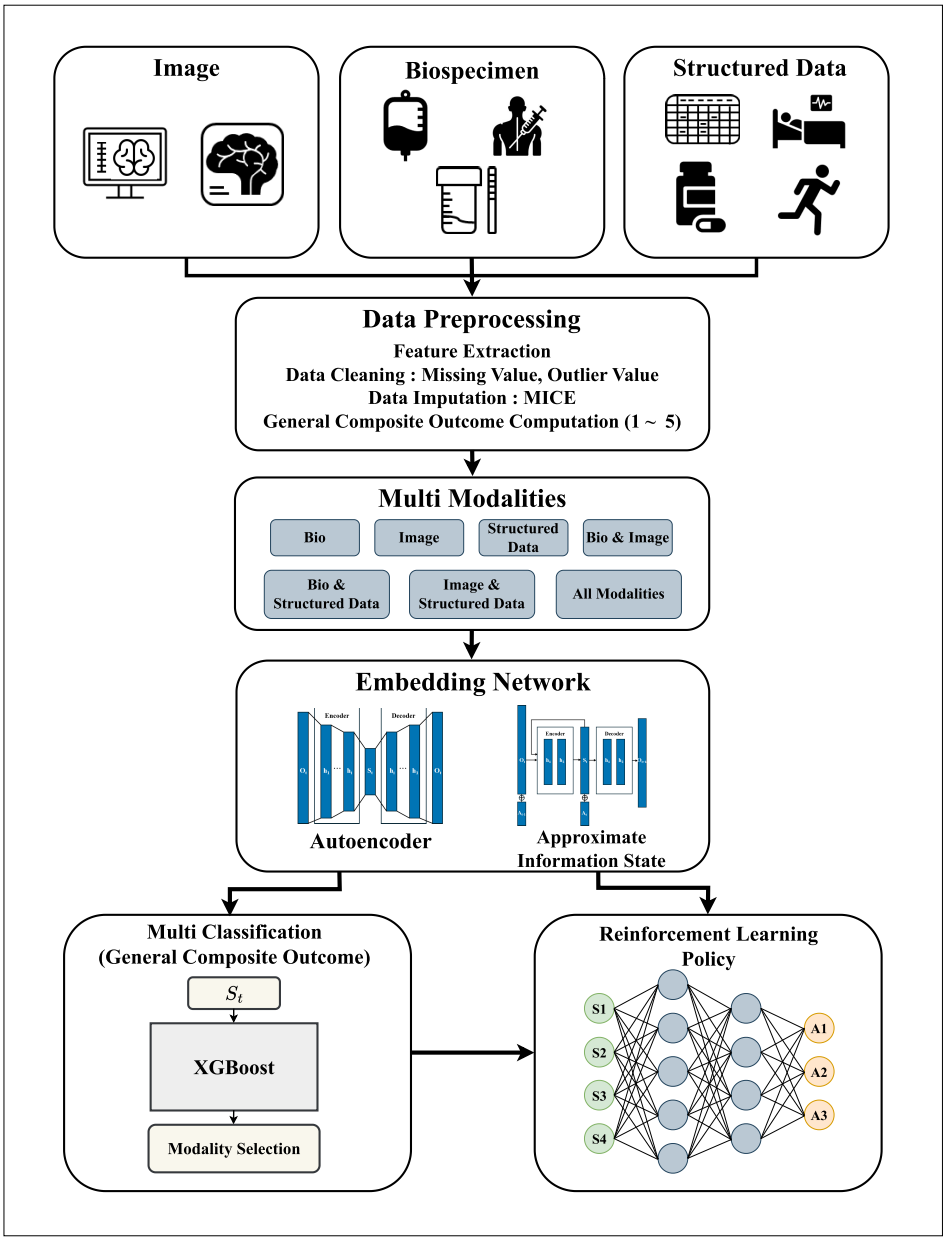


FIGURE 1. Multimodal data integration and embedding framework for Parkinson's disease progression analysis and medication recommendations.

disease progression or minimize the occurrence of adverse events. Furthermore, the optimal medication policies derived under varying modality datasets can provide insights into modality differences.

A. DATA DESCRIPTION

This study utilized data from the Parkinson's Progression Markers Initiative (PPMI) [32], an international multicenter observational study aimed at enhancing the understanding of PD by collecting a diverse array of potential biomarkers. The collected biomarkers encompass patient demographics, clinical assessments, imaging data, and biospecimens indicative of PD progression. The study collects and utilizes information from these diverse modalities to improve understanding and prediction of PD progression (see Table 1). To investigate the relationship between medication usage and PD progression, our analysis concentrated on individuals with early-stage PD who possessed extensive visit records suitable for longitudinal examination. The inclusion criteria for participants consisted of: (1) an initial Hoehn and Yahr stage of 2 or lower, (2) a history of more than 5 years of follow-up visits, and (3) complete documentation for all relevant features. As a result, a cohort of 594 patients, comprising a total of 8,610 visit records spanning from 2010 to 2023, was chosen for this study.

- Structural modalities (STR) consist of motor and non-motor assessments [33], including medical history evaluations designed to capture various manifestations

TABLE 1. Detailed feature information for each modality where the numbers in parentheses refer to the number of included features within the category. UPDRS(Unified Parkinson's Disease Rating Scale), DTI MRI(Diffusion Tensor Imaging Magnetic Resonance Imaging), HVLT(Hopkins Verbal Learning Test), BJLO(Benton Judgment of Line Orientation), SDMT(Symbol Digit Modalities Test), MSF(Modified Semantic Fluency), MOCA(Montreal Cognitive Assessment), LNS(Letter Number Sequencing test), MCI(Mild Cognitive Impairment), GDS(Geriatric Depression Scale), STAI(State-Trait Anxiety Inventory), QUIP(Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease), HnY(Hoehn & Yahr (HnY) Stage), MSEADLG(Schwab and England ADL scale), ESS(Epworth Sleepiness Scale), SCOPA-AUT(The Scale for Outcomes in Parkinson's disease for Au-tonomic symtoms), RBD(Rapid Eye Movement Sleep Behavior Disorder), CAUD(Caudate nucleus test), PUT(Putamen Test).

Modality	Category	Description	Details
Image	Imaging	Magnetic Resonance Imaging	DTI MRI, Resting state sequences, MRI results
		DatScan	Striatal Binding Ratio of the Caudate/Putamen/Anterior brain region
Biospecimen	Biospecimen	Biospecimen	Blood test(27), Cerebrospinal Fluid(16), Etc(10)
Structure	Demographics	Base	Age, Sex, Education
	Motor Assessments	UPDRS	UPDRS II – IV
	Non-motor Assessments	Neuropsychology	HVLT, BJLO, SDMT, MSF, MOCA, LNS, MCI
		Neuroethology	GDS, STAI, QUIP, HnY, MSEADLG
		etc.	ESS, SCOPA-AUT, RBD, UPDRS I, CAUD, PUT
	Medical History	Medication Information	Medication, Adverse Event

of PD in patients. Changes in drug history are considered biomarkers that can detect changes in patients' conditions. Particularly, the AESEVER (Adverse Event Severity) serves as an indicator for the severity of side effects experienced by patients. A score of 1 signifies mild severity, with symptoms including rash and lightheadedness, while a score of 2 indicates moderate severity, accompanied by symptoms such as back pain and migraine. Lastly, a score of 3 denotes severe cases, encompassing symptoms like syncope, convulsions, and sciatica.

- Imaging modalities (IMG), such as Single Photon Emission Computed Tomography (SPECT) with DatScan, were employed to assess dopamine transporter (DAT) functionality in the brain, which is crucial for the diagnosis and study of PD [34]. Additionally, Magnetic Resonance Imaging Diffusion Tensor Imaging (MRI DTI) was used to evaluate cerebral activity, further enhancing the diagnostic capabilities in clinical trials. SPECT imaging, as per the PPMI imaging protocol, was acquired at designated PPMI imaging centers and then forwarded to the imaging core lab at the Institute for Neurodegenerative Disorders (IND) for visual interpretation. The procedure for reviewing all SPECT images for dopamine transporter deficit is meticulously outlined in [35], ensuring standardized and accurate assessments.
- Biospecimen modalities (BIO) utilized in PD research include blood, urine, and cerebrospinal fluid (CSF) samples [36]. These samples are particularly crucial for collecting longitudinal data on PD, with CSF and urine being vital for tracking the progression of the disease. Blood samples were further categorized into plasma and serum components, each offering unique insights into the physiological status and biomarker profiles indicative of PD progression. In the analysis of proteomics data, CSF from both PD patients and healthy individuals was examined using the SOMAscan platform [37]. This involves comprehensive quality control steps to exclude outliers, calibrators, buffers, and non-human SOMAmers. The processing steps include hybridization normalization, scaling by plate, median normalization within each plate, followed by calibration at SomaLogic. Additionally, the data undergoes log2 transformation, inter-plate median normalization, and batch correction at the plate level, ensuring the accuracy and reliability of the data for further study.

Given the diverse range of biomarkers used in diagnosing PD, there is a need for a single measure to represent the overall severity of the condition. In addition to the features obtained from the three modalities, General Composite Outcome (GCO) was computed to assess the overall severity of PD and categorize disease severity into five levels based on quintile cutoff values. By utilizing the most critical manifestations, the GCO allows for a broad assessment of clinically relevant outcomes rather than relying on a potentially biased single outcome. Specifically, the derivation of GCO utilized Motor Symptoms (UPDRS-II, IV), Motor Signs (UPDRS-III), Cognition (MCI), and Non-motor Manifestations (GDS). GCO metric allowed for a more accurate assessment of the patient's condition in subsequent analyses, where a higher value indicates greater severity.

B. STATE EMBEDDING

Learning the latent representations of health state via embedding would be particularly useful in understanding PD for several reasons. The PPMI dataset is high-dimensional, with numerous variables collected from each patient across diverse modalities. By transforming the high-dimensional data into lower-dimensional embeddings, latent variables driving PD progression can be revealed. Furthermore, embedding methods provide a unified framework for integrating heterogeneous data modalities into a common representation space, allowing for joint analysis and interpretation of different types of data. For example, embedding could

capture relationships between clinical assessments, imaging findings, and biomarker levels, facilitating a comprehensive understanding of PD progression. In particular, as PD is a progressive neurodegenerative disorder characterized by changes in symptoms and disease severity over time, embedding methods based on deep learning models capable of capturing temporal dynamics could be helpful in modeling the sequential nature of PD progression. In this study, two types of embedding networks were introduced, the Autoencoder (AE) and the Approximate Information State (AIS), where the embedding results on different modalities were compared.

1) AUTOENCODER

An Autoencoder (AE) [38] is an unsupervised learning model that compresses high-dimensional data features into a lower-dimensional latent representation. This model consists of two parts: an encoder and a decoder. The encoder transforms the input data O_t into the latent space, producing a latent representation S_t. The decoder then reconstructs the data back into its original form from this space. Through this structure, the AE aims to effectively preserve the important information of the original data in the latent space while minimizing reconstruction loss. In this process, the AE can learn non-linear relationships and complex patterns, allowing a deeper understanding of the intrinsic characteristics of the data. An illustration of the AE structure is shown in Figure 2-(a).

Compared to conventional representation learning methods such as Principal Component Analysis (PCA) and Multi-Dimensional Scaling (MDS), which primarily rely on linear techniques, the AE operates differently. PCA and MDS identify directions or maximum variance in the data, with PCA aiming to preserve data variance through principal components and MDS preserving object distances in lower-dimensional space. While effective for linear relationships, these methods are limited in capturing non-linear data structures. In constrast, the AE employs non-linear activation functions, enabling it to capture non-linear characteristics present in data. This flexibility enables the AE to effectively represent complex and abstract data structures that PCA and MDS may miss. Consequently, the AE can adapt more readily across various modalities of data, making it a versatile tool for representation learning.

2) APPROXIMATE INFORMATION STATE

The Approximate Information State (AIS) [39] is a state estimation technique that enhances effective decision-making in complex reinforcement learning environments. This concept plays a crucial role particularly in Partially Observable Markov Decision Processes (POMDPs), where not all state information is directly accessible to the observer. In POMDPs, agents must infer the current state based solely on the history of observations and actions.

AIS provides a structure that helps agents make more accurate state estimations in such environments. This system

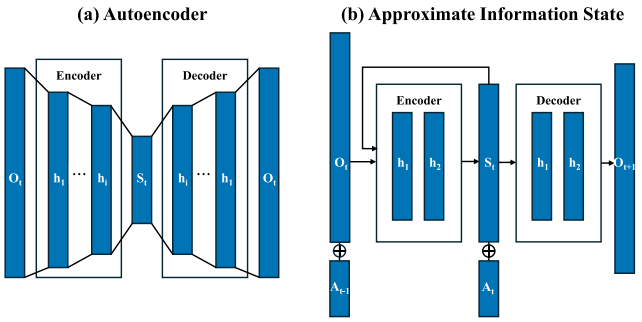


FIGURE 2. Structural architecture of Autoencoder (left) and Approximate Information State (right). O_t, A_t, S_t notations represent observation, action, state at time t, respectively, and h_i represent the i-th hidden layers.

uses a neural network comprising an encoder and a decoder to approximate the state. The encoder receives current observations O_t and previous actions A_{t-1} as input and estimates the current state, typically utilizing recurrent neural network structures like GRU or LSTM. These structures capture the dependencies in data over time, effectively incorporating past information into current decisions. Also, decoder uses the state estimated by the encoder S_t along with the current action A_t to predict observations at the next time point O_{t+1}. Through this process, AIS reduces the uncertainty of the current state and enables the agent to more accurately predict future conditions. An illustration of the AIS structure is shown in Figure 2-(b).

The operational efficiency of the AIS in POMDP environments is further underscored by its adherence to the Markov property and its function as a sufficient statistic, integral aspects for the RL models it supports. The Markov property is exemplified in the AIS model through the encoder's capacity to absorb and integrate current observations O_t and previous actions A_{t-1} into a comprehensive state representation S_t. This encapsulation ensures that all necessary information for making future state predictions and decisions is self-contained, eliminating the need for external historical data references. Simultaneously, the AIS's ability to serve as a sufficient statistic is realized through its compression of all pertinent environmental and historical data into a concise state estimate that fully represents the decision-relevant information. This efficient data synthesis allows the neural network to base decisions on a distilled and actionable dataset, thus enhancing the overall strategic efficacy of the agent operating within a POMDP framework. Through these properties, the AIS not only streamlines complex decision-making processes but also optimizes the action selection in environments characterized by uncertainty and partial observability. Thus, AIS not only captures the sequential connections between observations within a dynamic programming context but also incorporates the actions executed, positioning it as a more apt approach for RL settings than conventional representation techniques.

3) EVALUATION OF EMBEDDING METHODS

AE and AIS were employed to select an appropriate embedding method for each multimodal dataset. The search for optimal hyperparameters was conducted using Optuna [40], which efficiently finds the optimal settings. To compare the embedding methods, performance metrics specific to the method were utilized to compare AE and AIS embeddings based on their ability to predict GCO labels using a machine learning model. The detailed hyperparameter settings for each modality are shown in Table 7. In addition, the learned representations were visualized to gain insights into how well each embedding model has learned meaningful representations of the data. Effective embeddings will result in data points with similar characteristics (i.e., same GCO level) being clustered together in the visualization, indicating that the model has captured important features of the data in the embedding space.

C. MULTIMODAL REINFORCEMENT LEARNING

Once the appropriate latent representations for each data modality have been determined, these state representations were utilized to construct medication recommendation models. Leveraging predictions of future disease states derived from the learned representations, our multimodal RL model aimed to support clinicians in making well-informed decisions regarding medication regimens.

1) REINFORCEMENT LEARNING

RL is a machine learning technique focused on teaching an agent how to interact with its environment and make sequential decisions. The components of RL consist of states s, actions a, rewards r, transition probabilities $P(s, a, s')$, and discount factors γ. In each step, the agent observes the current state, selects an action, and the environment transitions to the next state based on the chosen action and provides a reward. The primary objective of RL is for the agent to learn an optimal policy that maximizes its expected cumulative reward R over time. The set of actions taken by the agent is referred to as a policy π, and the set of optimal actions for all states, which leads to the maximum return, is known as the optimal policy π^*. To learn optimal policies, RL algorithms use various techniques such as value iteration, policy iteration, Q-learning, and deep Q-learning. Of particular, deep Q learning [41] uses deep neural networks to approximate action value functions Q, allowing agents to process continuous, high-dimensional state spaces. Hence, deep Q-learning was utilized in this study to represent the state using embedding methods.

For multimodal RL, effective representation learning is particularly crucial as it facilitates better decision-making by enabling the agent to understand complex relationships and dependencies present in multimodal datasets. This understanding allows the agent to make more informed and context-aware decisions, leading to improved performance in various tasks. The primary objective of this study was to develop medication recommendation models using various modalities of PD biomarkers and compare the optimal policies to provide insights into modality differences.

2) COMPONENTS OF REINFORCEMENT LEARNING

The key components of the RL-based medication recommendation model were defined as follows:

State: To effectively represent the PD health state, we employed deep learning-based embedding networks (AE and AIS) to extract latent representations from the diverse biomarkers obtained from structured clinical assessments, imaging data, and biospecimen information. These networks utilized neural architectures to learn hierarchical features from the input data, effectively capturing complex relationships and patterns associated with PD progression. The learned representations served as compact encodings of the multimodal data, and were used to represent the state of the RL model. From both AE and AIS embedding networks, the latent vector S_t was extracted and used to represent the states upon model convergence, which were represented as a set of continuous vector.

Action: The primary treatment for PD is centered around medication. Medications play a crucial role in alleviating the primary symptoms of PD, with various medication options offering different short- and long-term effects on patients. In the early stages of the disease, patients typically start with lower potency medications that offer a milder adverse effect profile and are administered at lower dosages. As PD advances, medication regimens are adjusted to achieve optimal symptom management or mitigate medication-related adverse effects. Among the most commonly prescribed medications are levodopa, dopamine agonists, and MAO-B inhibitors. To standardize the comparison of dosage levels across different types of medications, we employed a metric called Levodopa Equivalent Daily Dose (LEDD) [42]. LEDD was computed based on the concept of the levodopa equivalent dose (LED), which represents the dosage of a medication that produces a similar effect to levodopa. By summing all the LEDs administered throughout the day, we derived the total levodopa equivalent dose, denoted as LEDD. This metric holds particular significance in monitoring changes in medication policy by observing alterations in LEDD. In instances where LEDD values were unavailable, they were estimated using a formula: LEDD = (Levodopa Equivalent Factor) $\times$ (Daily Dose) $\times$ (Frequency). If any component of this calculation was absent, LEDD was approximated by taking the median of LEDDs associated with medications sharing similar labels.

As depicted in Table 2, a total of 11 actions have been devised, taking into account both the medication type and the LEDD. The actions predominantly consist of prescriptions for three types of drugs: levodopa, dopamine agonists [43], and MAO-B inhibitors [44]. All other single drug prescriptions are considered as 'Other Medication', and a combination of multiple drugs is referred to as 'Combination'

TABLE 2. Action categorization based on medication type and dosage binarization where the cutoff for dosage binarization is set as the third quartile (Q3) of LEDD distribution for each medication type. *L, D, M, O, C, N* refer to levodopa, dopamine agonists, MAO-B inhibitors, other medication, and no drug, respectively, and subscript *L* and *H* each refer to low and high dosage.

	Action		Cutoff
Medication	**Low**	**High**	**LEDD (Q3)**
Levodopa	L_L	L_H	563.75
Dopamine - agonists	D_L	D_H	200
MAO-B	M_L	M_H	100
Other Medication	O_L	O_H	300
Combination	C_L	C_H	305
No Drug	N		-

action. Considering the intensity of drug prescription, each drug type was categorized into binary actions; 'Low' representing low dosage and 'High' referring to high dosage. To determine the cutoff for binarization, we used the third quartile (Q3) value of the dosage distribution for each specific drug type. This approach accounts for the diverse effects of each medication and allows for a clear distinction between low and high dosages. Lastly, the action of prescribing no drug was represented as a single action 'No Drug'.

Reward: In formulating the reward function R, we focused on the key metric GCO by penalizing high GCO values of 4 or 5 which indicates severe patient conditions. The occurrence of adverse events was also penalized by considering varying AESEVER severity criteria of 1, 2, or 3. Thus, indicator functions for GCO and AESEVER were introduced in the reward function. $\mathbb{1}_{\{GCO \geq \lambda_{GCO}\}}$ equals 1 if the GCO value is greater than or equal to a predefined threshold λ_{GCO}. Similarly, $\mathbb{1}_{\{AESEVER \geq \lambda_{AESEVER}\}}$ equals 1 if the AESEVER value is greater than or equal to a predefined threshold $\lambda_{AESEVER}$. By combining these components, the reward function penalizes high GCO values (indicating severe patient conditions) and adverse events, where the thresholds λ_{GCO} and $\lambda_{AESEVER}$ determine the severity levels at which penalties are applied. Additionally, we introduced parameter α to assess the impact of AESEVER relative to GCO in reward formulation, assuming $0 \leq \alpha \leq 1$ since GCO was our main patient outcome target. As the choice of α is arbitrary, we sought to derive the best α setting by experimenting with varying α values within the range of 0 to 1 in increments of 0.1.

$$R = -\mathbb{1}_{\{GCO \geq \lambda_{GCO}\}} - \alpha \cdot \mathbb{1}_{\{AESEVER \geq \lambda_{AESEVER}\}},$$
$$\lambda_{GCO} \in \{4, 5\}, \quad \lambda_{AESEVER} \in \{1, 2, 3\}.$$

3) EVALUATION OF REINFORCEMENT LEARNING MODEL

The optimal hyperparameter settings for RL training were selected using Optuna where we conducted a comprehensive search for six key hyperparameters of the DQN: Batch Size, Hidden Layer Size, Loss Function, Learning Rate, Learning Decay, and Learning Step. Additionally, the optimal reward setting was determined through hyperparameter searching as well, including the parameters α, λ_{GCO}, and $\lambda_{AESEVER}$ (see Table 7).

Assessment of RL model results involves evaluating the performance of the learned policy or Q-values in achieving the RL model's objective. In this section, we first sought to identify the best hyperparameter settings by evaluating the performance of the learned Q-values in achieving the RL model's objective. We can use the learned Q-values to make predictions related to the RL task and then assess the quality of these predictions using appropriate metrics such as AUROC (Area Under the Receiver Operating Characteristic curve) for binary classification tasks. As the classification task should benefit from the knowledge captured in the learned Q-values, we utilized the same task as in the RL model reward function formulation, i.e., classification of $GCO \geq \lambda_{GCO}$ and $AESEVER \geq \lambda_{AESEVER}$. As $\lambda_{GCO} \in \{4, 5\}$ and $\lambda_{AESEVER} \in \{1, 2, 3\}$, we computed the average AUROC across six combinations of λ_{GCO} and $\lambda_{AESEVER}$. This approach allowed us to objectively evaluate the RL model's performance based on reward settings.

To further evaluate RL model results, the optimal policies were evaluated by examining the Q-values to assess how well the RL model's optimal decisions align with the expected reward as indicated by the learned Q-values. Specifically, to further support the hypothesis that a serious condition exists among PD patients and that mitigation or prevention through medication may be possible, we sort patients according to their visit sequence when their GCO level first worsened. Here, the criterion of worsened GCO was set as level 5. Thus, we aimed to examine the trends in Q-values before reaching GCO level 5 and after reaching it, and compared the clinicians' actual prescription policy to the recommended policy of the RL model. This allows the evaluation of the adequacy of administered and recommended medications in different modality settings. To examine both the trends of deterioration and recovery, we selected two visit records prior to reaching level 5, and two visit records after reaching level 5. Hence, patients who never experienced GCO level 5 or patients with insufficient prior or post data were excluded from the analysis.

III. RESULTS

A. DATA PREPROCESSING

For each modality, missing values occurred depending on when the test was performed, resulting in varying missing value ratios: BIO: 54.59%, STR: 14.23%, IMG: 56.31%, BIO+IMG: 59.55%, BIO+STR: 41.92%, IMG+STR: 28.79%, ALL: 43.46%. Variables with more than 95% missing values were dropped. To handle missing data, we employed the Multiple Imputation by Chained Equations (MICE) [45] method for imputation. MICE effectively replaces missing data over time, minimizing information loss between visits, which is a limitation associated with commonly utilized zero or forward imputation methods. In this study, a total of 89 variables were utilized, including

TABLE 3. Optimal hyperparameter settings for embedding methods (AE and AIS), reinforcement learning (DQN), and reward function components.

Model	Parameters	ALL	BIO + STR	IMG + STR	STR
AE	Batch Size	64	64	64	32
	Hidden Layer Size	[83, 41]	[76, 38]	[4]	[7]
	Learning Rate	0.0003	0.0003	0.0003	0.0003
AIS	Batch Size	32	64	8	128
	Hidden Layer Size	[1006, 893]	[1010, 359]	[857, 696]	[534, 553]
	Learning Rate	0.0003	0.0003	0.0003	0.0003
DQN	Batch Size	64	64	32	64
	Hidden Layer Size	[64, 32]	[64, 32]	[32]	[32, 32]
	Loss function	Mean Squared Error	Mean Squared Error	Smooth L1 Loss	Smooth L1 Loss
	Learning rate	0.005	0.00005	0.001	0.00001
	Learning decay	0.85	0.75	0.75	0.9
	Learning step	100	500	2	200
Reward Setting	α	0.5	0.2	0.4	0.7
	λ_{GCO}	3	3	5	4
	$\lambda_{AESEVER}$	2	1	1	1

26 structured variables, 53 biological variables, and 10 imaging variables, which are summarized in Table 1.

B. OPTIMAL HYPERPARAMETER SETTING

Table 3 displays the optimal hyperparameter settings used for state embedding and RL model training, all of which were optimized using Optuna. For the embedding process, optimal values were derived for three key hyperparameters: Batch Size, Hidden Layer Size, and Learning Rate, with training conducted to minimize the loss. For the DQN reinforcement learning, six hyperparameters were considered: Batch Size, Hidden Layer Size, Loss Function, Learning Rate, Learning Decay, and Learning Step. These parameters are crucial as they significantly influence the outcomes, and the training was oriented towards maximizing AUROC that reflects both GCO and AESEVER. The selection of the optimal embedding method for each modality was also determined through this process where the selections were AIS for ALL and IMG+STR and AE for BIO+STR and STR. Finally, the optimal settings for the reward function components α, λ_{GCO}, and $\lambda_{AESEVER}$ were also derived based on the best AUROC performance setting. Notably, for the BIO+STR modality, an α value of 0.2 suggests less consideration for AESEVER in the reward setting, whereas the STR modality with an α value of 0.7 indicates a more balanced consideration of AESEVER and GCO.

C. COMPARISON OF EMBEDDING METHODS

1) EFFECTIVENESS OF EMBEDDINGS IN GCO PREDICTION

The embedding method plays a crucial role in reducing the dimensionality of high-dimensional data while preserving important patterns. In our study, we employed two approaches, AE and AIS, to effectively capture the unique characteristics of each modality. To appropriately evaluate the learned embeddings, we assessed their capability of predicting the overall severity level represented as the GCO metric. Specifically, we assessed whether the embeddings contained relevant patient information by developing the XGboost [46] model to predict GCO label as a multi-class classification task. To do so, GCO features were excluded from the input dataset and used solely as target labels.

TABLE 4. GCO level multi-classification results of the two embedding methods for each modality. The numbers in parentheses correspond to the standard deviation.

	Embedding	AUROC	F1-score
BIO	AE	0.544(0.009)	0.228(0.007)
	AIS	0.540(0.008)	0.227(0.011)
IMG	AE	0.525(0.008)	0.169(0.009)
	AIS	0.527(0.009)	0.164(0.013)
STR	AE	0.710(0.005)	0.372(0.013)
	AIS	0.703(0.005)	0.375(0.008)
BIO + IMG	AE	0.552(0.006)	0.237(0.011)
	AIS	0.551(0.005)	0.236(0.013)
BIO + STR	AE	0.708(0.007)	0.373(0.014)
	AIS	0.686(0.007)	0.358(0.007)
IMG + STR	AE	0.600(0.009)	0.270(0.019)
	AIS	0.726(0.004)	0.400(0.006)
All	AE	0.702(0.005)	0.382(0.003)
	AIS	0.675(0.009)	0.346(0.015)

Table 4 displays the AUROC and F1-score performance of the embeddings across various modalities using AE and AIS methods. For BIO, IMG, and BIO+IMG modalities, both AE and AIS embeddings failed to achieve an AUROC exceeding 0.6, suggesting inadequate representation of GCO-relevant information. While for STR and BIO+STR modalities, AE and AIS demonstrated similar performance with an AUROC 0.7, AIS significantly outperformed AE in the IMG+STR modality with an AUROC of 0.726, and AE showed better performance for the ALL modality with an AUROC of 0.702. These findings highlight the impact of embedding method selection on modality representation in capturing GCO information.

2) VISUALIZING THE LEARNED EMBEDDINGS

To further examine the outcomes of the embedding process, we employed t-SNE (t-distributed Stochastic Neighbor Embedding) [47], a dimensionality reduction technique for

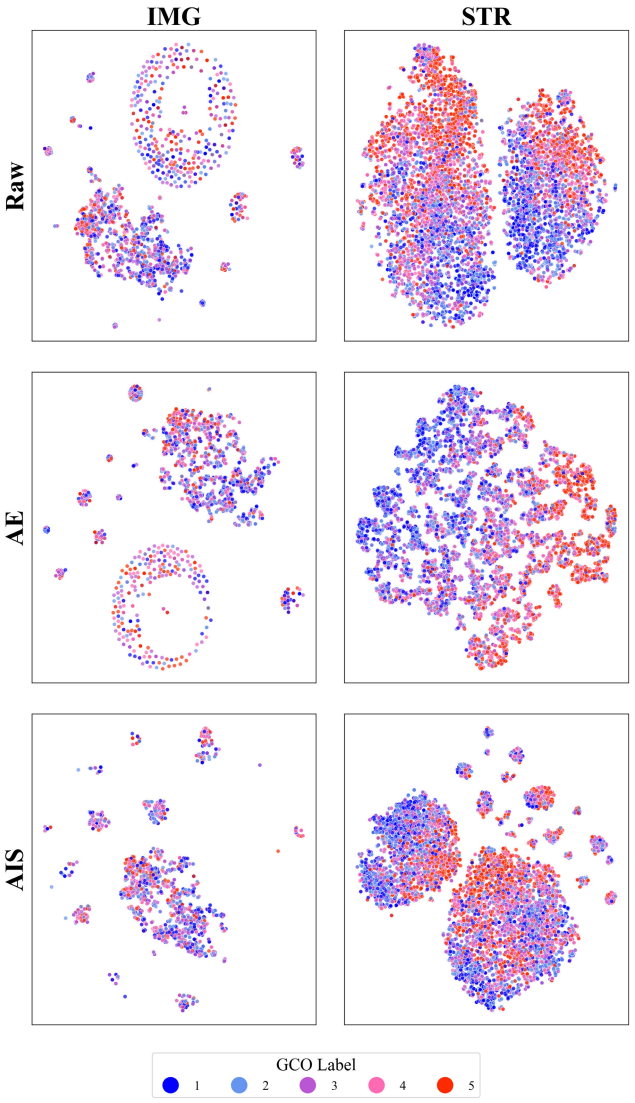


FIGURE 3. Visual comparison of embedding results for STR and IMG modalities with raw dataset.

visualizing high-dimensional data, renowned for preserving intricate data patterns effectively. Embedding methods often produce high-dimensional representations of data, making it challenging to visualize and interpret directly. t-SNE is a popular dimensionality reduction technique that projects data points from a high-dimensional space to a lower-dimensional space while preserving local structure and capturing non-linear relationships between data points. Figure 3 illustrates the t-SNE visualization outcomes of the embedding results along with the raw dataset, comparing the STR modality, which exhibited the highest machine learning (ML) prediction performance, with the IMG modality, characterized by the lowest performance. Effective embeddings should exhibit distinct clusters corresponding to different patient health states. By coloring the points based on the corresponding GCO labels (representing varying patient health states), we can visually inspect whether the embedded representations exhibit distinct clusters corresponding to different health states. The visualization showcases that the embedding technique had adeptly organized the STR modality data with respect to the GCO label, facilitating a clear differentiation between patient statuses, particularly using the AE method. Note that the states with same GCO label tend to be clustered together where the clusters show clear stratification under the AE method. Conversely, the IMG modality, displaying inferior ML performance, demonstrated a poor discernible organization post-embedding, indicating a deficiency in capturing and representing its intrinsic characteristics effectively. Varying GCO labels were mixed within the same cluster, not effectively representing patient health states. Moreover, the widely dispersed points within clusters may represent diverse or ambiguous states. Both embedding methods were not able to differentiate severity levels, aligning with the low ML performance results.

Overall, since modalities incorporating structure information exhibited relatively higher GCO classification performance, we selected the following four modalities consisting of STR for further analysis: STR, BIO+STR, IMG+STR, and ALL. With the appropriate embedding method selected, we subsequently re-trained the embeddings using GCO information explicitly for complete state representation. Figure 4 visualizes the embedding results post-inclusion of the GCO variable, revealing its distinct differentiation across all modalities. By examining the proximity of points within and between clusters, the discriminative power of the embedded representations in capturing relevant features of patient states can be seen. Furthermore, evaluating the GCO classification task based on the re-trained embeddings resulted in all four modalities exhibiting excellent performance with AUROC scores of 0.97 or higher. From such results, it can be concluded that the re-trained embeddings adequately captured PD severity, and thus were subsequently used as the state vector for RL model development.

D. EVALUATION OF MULTIMODAL REINFORCEMENT MODEL

1) USING LEARNED Q-VALUES TO PREDICT SEVERE PD STATUS

To assess the RL model results, the learned Q-values were used to perform classification tasks relevant to the RL reward function, and the quality of these classifications was evaluated using AUROC. The tasks of classifying GCO $\geq \lambda_{GCO}$ and AESEVER $\geq \lambda_{AESEVER}$ for $\lambda_{GCO} \in \{4, 5\}$ and $\lambda_{AESEVER} \in \{1, 2, 3\}$ were evaluated. Table 5 displays the average AUROC values, ranked in descending order of STR, ALL, BIO+STR, and IMG+STR, with values of 0.950, 0.849, 0.806, and 0.718, respectively. We noted that across all modalities, AUROC values were higher for severe patient conditions (GCO 5) and lower for relatively better conditions (GCO 4). It is also of note that AUROC performance does not show monotonicity with respect to the number of modalities used, and utilization of all modalities (ALL) does not necessarily lead to the best performance. Conversely,

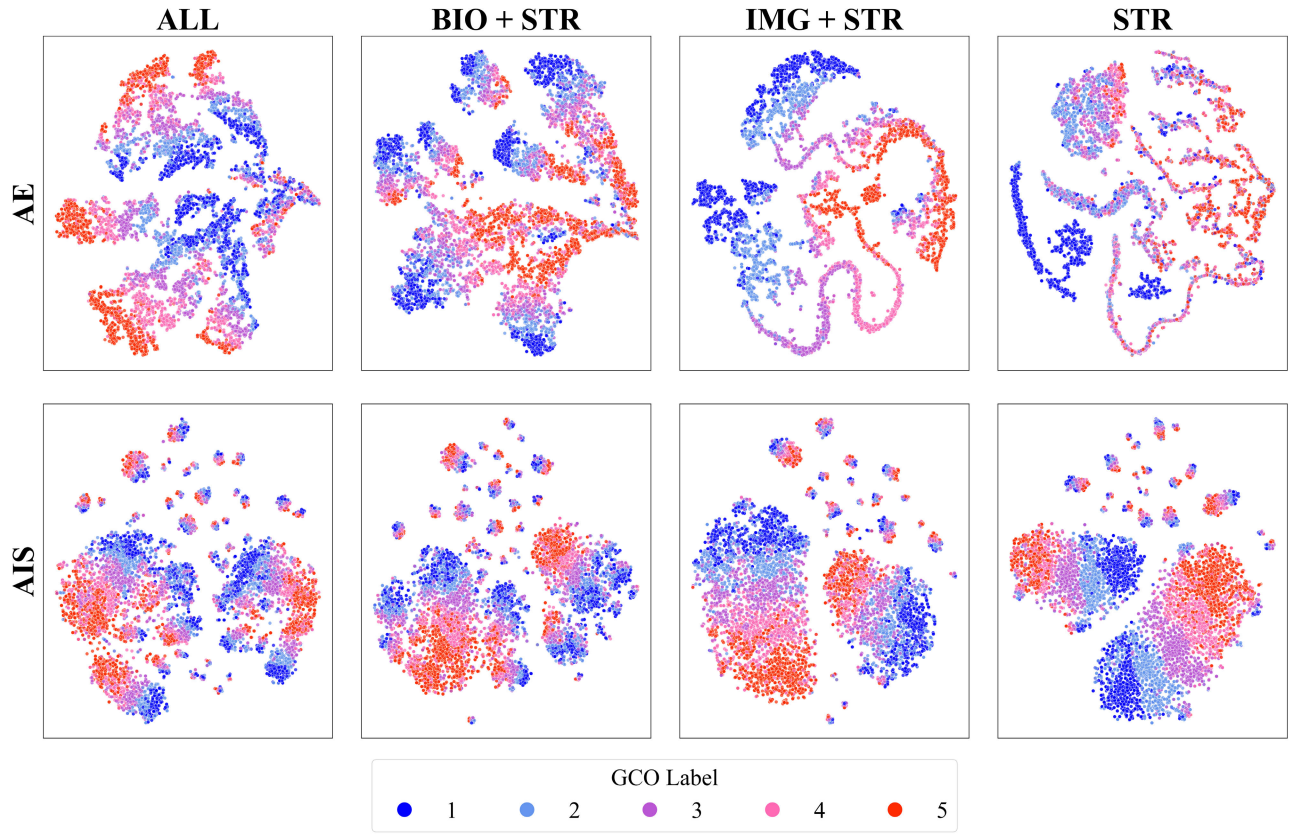


FIGURE 4. Visualization of the embedding results post-inclusion of the GCO feature for All, BIO+STR, IMG+STR, and STR modalities.

TABLE 5. AUROC results of predicting severe PD status using learned Q-values for each modality. Varying settings of severity are represented as a tuple of (λ_{GCO}, λ_{AESEVR}), and AVG refers to the average of the AUROC scores.

Modality	λ_{GCO}, $\lambda_{AESEVER}$	Score	Modality	λ_{GCO}, $\lambda_{AESEVER}$	Score
ALL	(4,1)	0.820	BIO + STR	(4,1)	0.750
	(4,2)	0.803		(4,2)	0.760
	(4,3)	0.787		(4,3)	0.801
	(5,1)	0.896		(5,1)	0.853
	(5,2)	0.892		(5,2)	0.836
	(5,3)	0.897		(5,3)	0.836
	AVG	0.849		AVG	0.806
Modality	λ_{GCO}, $\lambda_{AESEVER}$	**Score**	**Modality**	λ_{GCO}, $\lambda_{AESEVER}$	**Score**
IMG + STR	(4,1)	0.698	STR	(4,1)	0.860
	(4,2)	0.641		(4,2)	0.973
	(4,3)	0.585		(4,3)	0.990
	(5,1)	0.798		(5,1)	0.893
	(5,2)	0.773		(5,2)	0.986
	(5,3)	0.810		(5,3)	0.996
	AVG	0.718		AVG	0.950

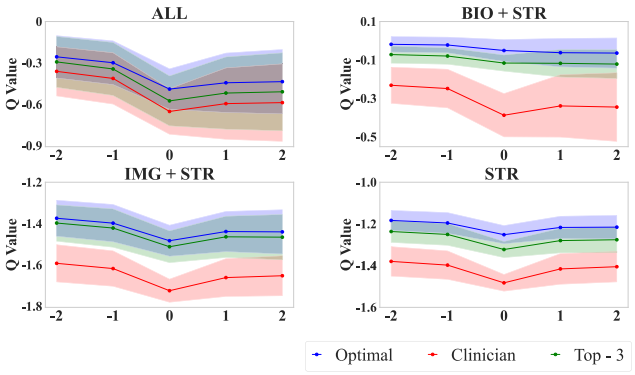


FIGURE 5. Trend of Q-values around the first entrance into severe PD deterioration of GCO level 5 for each modality.

IMG+STR modality showed the worst performance with AUROC showing very poor performance for classifying GCO ≥ 4 and AESEVER ≥ 3. This could potentially imply these two modalities may not integrate seamlessly, thus leading to poor PD state representation.

2) TREND OF Q-VALUES AROUND THE FIRST SEVERE DETERIORATION

To examine the trend of Q-values before reaching GCO level 5 and after reaching it, we aligned patients according to the point in their care when the patient first entered GCO level 5. The results of this analysis suggest two main points. First, Figure 5 reveals that across all modalities, the optimal prescription policies learned by the RL (denoted as Optimal) exhibited better outcomes compared to those prescribed by clinicians (denoted as Clinician). We further selected the top 3 recommended medications to account for variety in medication selection, where the average Q-values of these medications (denoted as Top - 3) also consistently showed better outcomes than that of clinicians' policy. The gap between the value of recommended medications and

actual prescribed medications was significant for BIO+STR, IMG+STR, and STR modalities, indicating a possible mitigation of GCO worsening by adequate medication prescription. Conversely, the ALL modality showed a close overlap between the optimal and clinician policy, indicating the actual prescribed medications were similar to those recommended by the RL model. Second, a gradual decline in the patient's condition was observed before reaching GCO level 5, with a distinct lowest point aligning with the time of entering the worsened GCO status. Thus, the learned Q-values adequately represent PD severity and the RL model's decisions align with expected rewards as indicated by the learned Q-values. In addition, recovery or maintenance of the patient's condition was observed after reaching GCO level 5. The extent of recovery for clinician policy varied depending on the modality, implying PD severity can be perceived differently for the same patient health condition depending on the choice of modality.

Since the reward function formulation setting α varies between modalities, a direct comparison of Q-values was not possible. Nevertheless, the trend of optimal policy Q-values could be compared where BIO+STR results showed the least change in values. This observation suggests a potential benefit in terms of preventing worsening conditions through the appropriate prescription of medications.

3) COMPARISON OF RECOMMENDED MEDICATIONS

With the RL model results validated, we compared the prescription patterns by investigating the recommended policy from the RL model using different modality datasets. Table 6 illustrates the distribution of drugs recommended by the RL model. In the ALL modality, drugs M_L, O_L, and O_H are prescribed at a notably high rate, consisting of 90% of prescriptions in total. Conversely, for the IMG+STR modality, three drugs, L_H, M_L, and O_L, are prescribed at an identical rate of 33.3%. These three drugs also feature prominently in the STR modality, with rates of 33.3%, 30.8%, and 28.6%, respectively. It is worth noting that approximately 7-8% of cases in the ALL and STR modalities do not receive any drug prescription (N). In contrast, the BIO+STR modality exhibits a diverse range of drug prescriptions despite a relatively high rate of 31% for O_L, as observed in other modalities. This suggests a propensity to prescribe a variety of drugs within the BIO+STR modality.

Overall, the distribution of recommended medications varied substantially depending on the chosen modality. Levodopa was frequently prescribed when considering a single modality of STR or the bimodal combination of IMG+STR. MAO-B and Other medication were frequently recommended regardless of the modality used, with notable differences in prescription dosage. High dosages of Other medication were only prescribed under the ALL or BIO+STR modalities. The modality resulting in the least diverse set of medications was IMG+STR, which also exhibited the lowest embedding performance as shown in Table 4. Conversely, BIO+STR recommended the most diverse array of medications, with its embedding performance being the second best in terms of both AUROC and F1-score. Therefore, it can be inferred that the effectiveness of state embeddings does impact the resulting RL model's optimal policy.

TABLE 6. The distribution of recommended medications from the RL model across the ALL, BIO+STR, IMG+STR, and STR modalities.

Action	ALL	BIO + STR	IMG + STR	STR
L_L	0.0%	2.6%	0.0%	0.0%
L_H	0.2%	0.2%	33.3%	33.3%
D_L	0.0%	20.6%	0.0%	0.0%
D_H	0.0%	18.1%	0.0%	0.0%
M_L	33.3%	8.4%	33.3%	30.8%
M_H	2.2%	1.5%	0.0%	0.1%
O_L	26.8%	31.0%	33.3%	28.6%
O_H	28.7%	5.7%	0.0%	0.0%
C_L	0.0%	6.2%	0.0%	0.0%
C_H	0.0%	0.0%	0.0%	0.3%
N	8.7%	5.7%	0.0%	7.0%

IV. DISCUSSION

We further compared modality differences by analyzing the characteristics of the recommended policy from the RL model. The objective of this analysis was twofold: 1) for each modality, examine whether PD severity was considered when prescribing medications, and 2) compare medication prescription pattern differences across modalities. The medications with the top 3 highest Q-values were considered as the recommended policy where the policies are visualized for each modality in Figure 6. To examine the difference in prescription policies according to patient condition, GCO was set as the PD severity metric and the recommended medications were plotted for each condition. In addition, using AESEVER as the PD severity metric, a similar recommended medication distribution plot is presented in Figure 8.

A. PRESCRIPTION PATTERN COMPARISON

Analyzing the characteristics of each modality reveals distinct trends in prescription patterns. First, both IMG+STR and STR modalities tended to prescribe strong doses of levodopa (L_U) and weaker doses of MAO-B inhibitors (M_D) and other medication (O_D). Notably, for the IMG+STR modality, variations in prescriptions based on the patient's GCO level were minimal which could be a sign of insufficient consideration of PD severity when prescribing medications. The STR modality exhibited a slight increase in no medication (N) and strong doses of medication combinations (C_U) as the patient's condition worsened. Interestingly, the selection of non-prescription was more frequently selected as the GCO level worsened, which may be due to drug

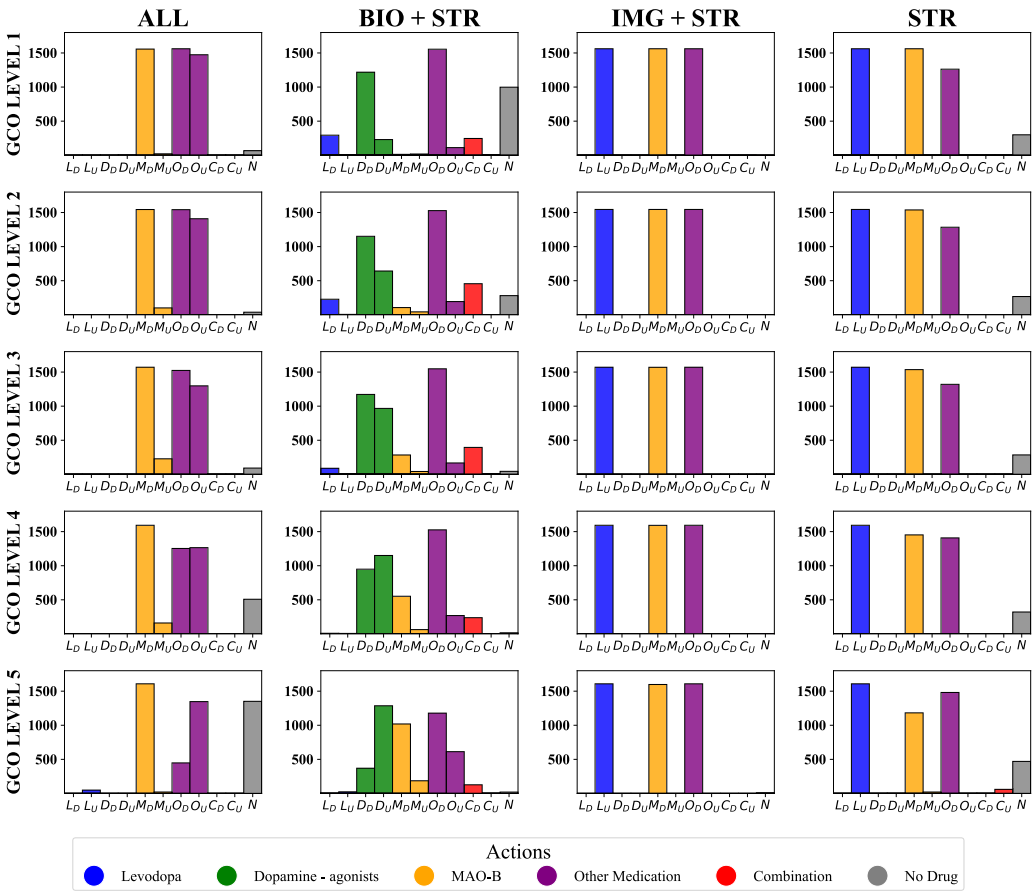


FIGURE 6. Distribution of the recommended medications stratified by GCO level. The x-axis refers to the actions of the RL model, the y-axis represents the frequency of the recommended action, and the colored bars refer to the different types of medications.

side effects. However, the clinical applicability of such a prescription pattern needs further justification.

In ALL modality, there was a predominant prescription of MAO-B inhibitors and other drugs, maintaining consistent drug strength during stable patient conditions. However, as the patient's condition worsened, there was a decrease in prescribing other medication (O_U) and a substantial increase in not prescribing medications (N). Notably, the recommendation of not prescribing medications was most frequent under the ALL modality. Again, such prescription patterns could be attributed to a significant shift in medication for identifying patients with deteriorating conditions and consideration of side effects, thus requiring further investigation.

On the other hand, the BIO+STR modality encompassed the most wide range of medications. As the condition deteriorated, there was a decrease in prescribing strong levodopa drugs (L_D), combination drugs (C_D), and non-prescriptions (N). Additionally, for dopamine agonists, strong dosage (D_U) was increasingly prescribed as the condition worsened. These distinct prescription patterns indicate that the BIO+STR modality incorporated the patient's condition when recommending medications. Furthermore, compared to other modalities, rather than completely avoiding drug prescriptions and selecting non-prescriptions, alternative prescriptions were recommended based on the potency of the drugs.

A similar analysis was conducted using AESEVER as PD severity metric where the distribution of medication prescriptions based on AESEVER severity is shown in Figure 8. Note that given the imbalanced AESEVER distribution, cases with AESEVER of 2 or higher were grouped and visualized together. The overall distribution of prescriptions was similar to those of GCO. Specifically, within the ALL modality, there was no significant change in medication trends with increasing AESEVER severity. Also, MAO-B inhibitors and other drugs were prescribed in large quantities similar to GCO-based prescriptions. For the IMG+STR and STR modalities, a consistent pattern of high reliance on specific medication types (L_U, M_D, O_D) was observed, mirroring previous results. However, this pattern may not represent appropriate prescriptions due to the lack of change in drug prescription despite the consideration of adverse events. In contrast, the BIO+STR modality showed a decrease in drug dosage as AESEVER severity increased, suggesting consideration for potential drug side effects. These characteristics were particularly evident in dopamine agonists.

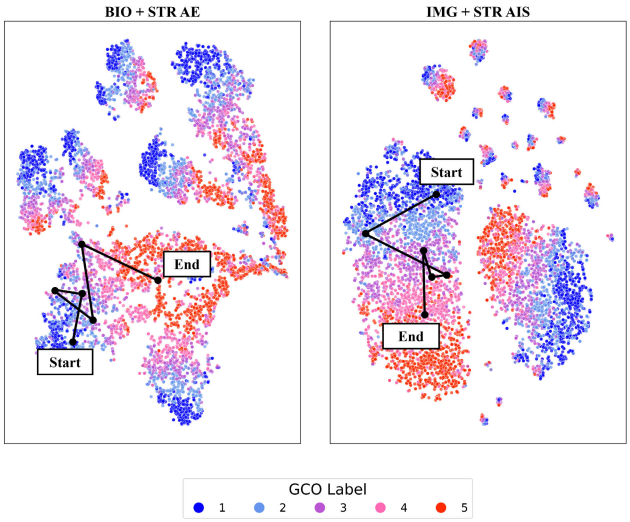


FIGURE 7. A patient trajectory visualized in the embedded space for the BIO+STR and IMG+STR modalities. The visit sequence begins from "Start" and progresses through the ordered numbers until reaching the final visit marked as "End."

Overall, these findings allowed us to understand the distinct characteristics of each modality. Specifically, a single modality alone could not provide sufficient information regarding PD severity and the recommended prescriptions mainly relied on three medication types with the same dosage. Regarding multimodal analysis, the IMG modality did not provide additional information to the STR modality, as the resulting prescriptions still did not accurately represent PD severity and exhibited a consistent pattern. In contrast, integrating BIO modality information into the STR modality led to prescriptions that more closely reflected the patient's state. The recommended prescriptions utilized a wide range of medications similar to what was observed in the dataset, and alternative medications were prescribed for severe PD status rather than avoiding drug prescriptions. However, utilizing all three modalities did not necessarily lead to the most appropriate prescription decisions as non-prescription was most frequently recommended for severe GCO levels and prescriptions relied on a limited set of medications.

B. ILLUSTRATIVE CASE STUDY OF A SINGLE PATIENT TRAJECTORY

We further provide a comparison of modalities through the example of a single patient case. Figure 7 shows trajectory data for a specific patient record using two different modality datasets. This patient had a total of six visit sequences with an adverse event (AESEVER=2) reported on the third visit, starting with a benign GCO level 1 and ending with a relatively severe GCO level 4. Both modalities show a clear trend of the patient's condition worsening over time, transitioning from a predominantly blue-dominated area to a prominently red-marked area. The results are based on actual prescriptions and do not indicate a clear trend of recovery due to the absence of medication administration.

However, the prescriptions generated by the RL model varied across modalities. In the IMG+STR modality, three medications (L_U, M_L, O_L) were consistently prescribed regardless of GCO severity or AESEVER occurrence. Conversely, in the BIO+STR modality, the prescription strategy changed around the third visit when adverse event occurred. Levodopa (L_D) was initially prescribed until the second visit. However, following the onset of AESEVER, MAO-B inhibitor (M_L) was chosen, succeeded by a dopamine agonist (D_U), reflecting a dynamic drug selection based on patient response and side effect management. These observations highlight that despite the same PD severity and adverse event status, the characteristics of each modality lead to distinct prescriptions and potential for recovery.

V. CONCLUSION

Our study employed a multimodal dataset to comprehensively compare and analyze prescriptions for PD patients. Specifically, we utilized various embedding techniques to extract latent vectors from diverse modalities as state information, which was used in training an RL model considering GCO and AESEVER. Our findings confirm that prescription characteristics vary across modalities. Notably, while the ALL, IMG+STR, and STR modalities exhibited consistent medication prescribing tendencies, the BIO+STR modalities considered a variety of drug strengths in their prescriptions, taking into account the patient's condition (GCO) and side effect (AESEVER). This underscores the importance of considering which modality information is used in making prescription decisions when patients seek medical treatment.

However, our study does come with several limitations. First, the lack of data on prescriptions involving the combination of two or more drugs prevented us from fully considering the complex interactions between multiple prescribed medications. Second, the study overlooked the diversity of information included under "Other Medications" and disregarded potential drug interactions and side effects. Lastly, the results were derived using a limited size of dataset and need further validation on an external dataset. However, we provided insights into how multimodal data integration could play a key role in capturing essential patterns related to PD progression, and affect medication prescription policies accordingly. The findings highlight the significance of integrating diverse data modalities in understanding and managing PD progression, emphasizing the importance of tailored medication regimes based on multimodal data insights.

APPENDIX.

The complete set of hyperparameter setting values shown in Table 7 is used as input for Optuna for hyperparameter tuning. Note that the resultant optimal setting is depicted using an asterisk. Furthermore, the distribution of the recommended medications from the RL model is depicted in Figure 8 using AESEVER as the PD severity metric.

TABLE 7. The complete set of hyperparameter setting values where the asterisk refers to the optimized setting identified via Optuna.

Model	Parameters	ALL	BIO + STR	IMG + STR	STR
AE	Batch Size	[8, 16, 32, 64*, 128]	[8, 16, 32, 64*, 128]	[8, 16, 32, 64*, 128]	[8, 16, 32*, 64, 128]
	Hidden Layer Size	[83, 41]	[76, 38]	[4]	[7]
	Learning Rate	[1e-4, 3e-4*, 1e-5]	[1e-4, 3e-4*, 1e-5]	[1e-4, 3e-4*, 1e-5]	[1e-4, 3e-4*, 1e-5]
AIS	Batch Size	[8, 16, 32*, 64, 128]	[8, 16, 32, 64*, 128]	[8*, 16, 32, 64, 128]	[8, 16, 32, 64, 128*]
	Hidden Layer Size	[1006, 893]	[1010, 359]	[857, 696]	[534, 553]
	Learning Rate	[1e-4, 3e-4*, 1e-5]	[1e-4, 3e-4*, 1e-5]	[1e-4, 3e-4*, 1e-5]	[1e-4, 3e-4*, 1e-5]
DQN	Batch Size	[16, 32, 64*, 128, 256]	[16, 32, 64*, 128, 256]	[16, 32*, 64, 128, 256]	[16, 32, 64*, 128, 256]
	Hidden Layer Size	[[32], [64], [32,32], [32,64], [64, 32]*, [64,64]]	[[32], [64], [32,32], [32,64], [64, 32]*, [64,64]]	[[32]*, [64], [32,32], [32,64], [64,32], [64,64]]	[[32], [64], [32, 32]*, [32,64], [64,32], [64,64]]
	Loss function	[Mean Squared Error*, Smooth L1 Loss]	[Mean Squared Error*, Smooth L1 Loss]	[Mean Squared Error, Smooth L1 Loss*]	[Mean Squared Error, Smooth L1 Loss*]
	Learning Rate	[1e-6, 5e-6, 1e-5, 5e-5, 1e-4, 5e-4, 1e-3, 5e-3*, 1e-2]	[1e-6, 5e-6, 1e-5, 5e-5*, 1e-4, 5e-4, 1e-3, 5e-3, 1e-2]	[1e-6, 5e-6, 1e-5, 5e-5, 1e-4, 5e-4, 1e-3*, 5e-3, 1e-2]	[1e-6, 5e-6, 1e-5*, 5e-5, 1e-4, 5e-4, 1e-3, 5e-3, 1e-2]
	Learning decay	[0.75, 0.8, 0.85*, 0.9, 0.95, 1.0]	[0.75*, 0.8, 0.85, 0.9, 0.95, 1.0]	[0.75*, 0.8, 0.85, 0.9, 0.95, 1.0]	[0.75, 0.8, 0.85, 0.9*, 0.95, 1.0]
	Learning step	[2, 5, 10, 20, 30, 50, 100* ,200 ,250 ,500 ,1000]	[2, 5, 10, 20, 30, 50, 100 ,200 ,250 ,500* ,1000]	[2*, 5, 10, 20, 30, 50, 100 ,200 ,250 ,500 ,1000]	[2, 5, 10, 20, 30, 50, 100 ,200* ,250 ,500 ,1000]
Reward setting	Alpha	[0.1, 0.2, 0.3, 0.4, 0.5*, 0.6, 0.7, 0.8, 0.9, 1.0]	[0.1, 0.2*, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]	[0.1, 0.2, 0.3, 0.4*, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7*, 0.8, 0.9, 1.0]
	GCO	[2, 3*, 4, 5]	[2, 3*, 4, 5]	[2, 3, 4, 5*]	[2, 3, 4*, 5]
	AESEVER	[1, 2*, 3]	[1*, 2, 3]	[1*, 2, 3]	[1*, 2, 3]
Model	Embedding Method	[AE, AIS*]	[AE*, AIS]	[AE, AIS*]	[AE*, AIS]

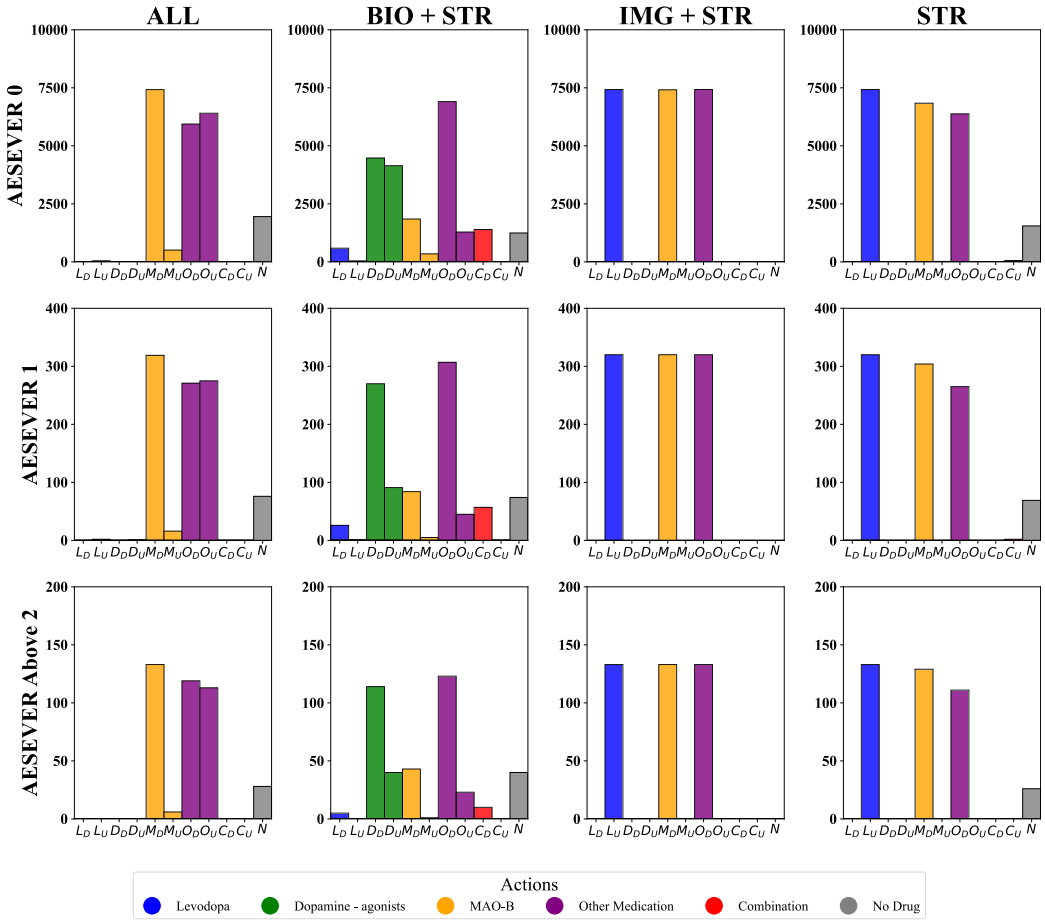


FIGURE 8. Distribution of the recommended medications stratified by AESEVER. The x-axis refers to the actions of the RL model, the y-axis represents the frequency of the recommended action, and the colored bars refer to the different types of medications.

REFERENCES

[1] B. R. Bloem, M. S. Okun, and C. Klein, "Parkinson's disease," *Lancet*, vol. 397, no. 10291, pp. 2284–2303, Jun. 2021.

[2] J. Jankovic, "Parkinson's disease: Clinical features and diagnosis," *J. Neurol. Neurosurg. Psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[3] S. Grover, S. Bhartia, A. Yadav, and K. R. Seeja, "Predicting severity of Parkinson's disease using deep learning," *Proc. Comput. Sci.*, vol. 132, pp. 1788–1794, Jan. 2018.

[4] R.M. Sadek, S. A.Mohammed, A. R. K. Abunbehan, A. K. H. A. Ghattas, M. R. Badawi, M. N. Mortaja, B. S. Abu-Nasser, and S. S. Abu-Naser, "Parkinson's disease prediction using artificial neural network," *Int. J. Academic Health Med. Res.*, vol. 3, no. 1, pp. 1–8, 2019.

[5] Z. K. Senturk, "Early diagnosis of Parkinson's disease using machine learning algorithms," *Med. Hypotheses*, vol. 138, May 2020, Art. no. 109603.

[6] A. H. Shahid and M. P. Singh, "A deep learning approach for prediction of Parkinson's disease progression," *Biomed. Eng. Lett.*, vol. 10, pp. 227–239, May 2020.

[7] S. Bind, A. K. Tiwari, A. K. Sahani, P. Koulibaly, F. Nobili, M. Pagani, O. Sabri, T. Borght, K. Laere, and K. Tatsch, "A survey of machine learning based approaches for Parkinson disease prediction," *Int. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 2, pp. 1648–1655, 2015.

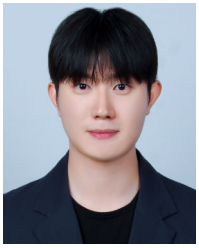
[8] I. El Maachi, G.-A. Bilodeau, and W. Bouachir, "Deep 1D-convnet for accurate Parkinson disease detection and severity prediction from gait," *Exp. Syst. Appl.*, vol. 143, Apr. 2020, Art. no. 113075.

[9] S. Shinde, S. Prasad, Y. Saboo, R. Kaushick, J. Saini, P. K. Pal, and M. Ingalhalikar, "Predictive markers for Parkinson's disease using deep neural nets on neuromelanin sensitive MRI," *NeuroImage, Clin.*, vol. 22, 2019, Art. no. 101748.

[10] M. B. Makarious, H. L. Leonard, D. Vitale, H. Iwaki, L. Sargent, A. Dadu, I. Violich, E. Hutchins, D. Saffo, and S. Bandres-Ciga, "Multi-modality machine learning predicting Parkinson's disease," *Npj Parkinson's Disease*, vol. 8, no. 1, p. 35, 2022.
[11] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, and R. Dodel, "Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Scale presentation and clinimetric testing results," *Movement Disorders, Off. J. Movement Disorder Soc.*, vol. 23, no. 15, pp. 2129–2170, 2008.
[12] M. Lawton, Y. Ben-Shlomo, M. T. May, F. Baig, T. R. Barber, J. C. Klein, D. M. A. Swallow, N. Malek, K. A. Grosset, and N. Bajaj, "Developing and validating Parkinson's disease subtypes and their motor and cognitive progression," *J. Neurol., Neurosurg. Psychiatry*, vol. 89, no. 12, pp. 1279–1287, 2018.
[13] S. M. van Rooden, F. Colas, P. Martinez-Martin, M. Visser, D. Verbaan, J. Marinus, R. K. Chaudhuri, J. N. Kok, and J. J. van Hilten, "Clinical subtypes of Parkinson's disease," *Movement Disorders*, vol. 26, no. 1, pp. 51–58, 2011.
[14] R. Erro, M. Picillo, C. Vitale, R. Palladino, M. Amboni, M. Moccia, M. T. Pellecchia, and P. Barone, "Clinical clusters and dopaminergic dysfunction in de-novo Parkinson disease," *Parkinsonism Rel. Disorders*, vol. 28, pp. 137–140, Jul. 2016.
[15] S.-M. Fereshtehnejad, S. R. Romenets, J. B. M. Anang, V. Latreille, J.-F. Gagnon, and R. B. Postuma, "New clinical subtypes of Parkinson disease and their longitudinal progression: A prospective cohort comparison with other phenotypes," *JAMA Neurol.*, vol. 72, no. 8, pp. 863–873, 2015.
[16] S.-M. Fereshtehnejad, Y. Zeighami, A. Dagher, and R. B. Postuma, "Clinical criteria for subtyping Parkinson's disease: Biomarkers and longitudinal progression," *Brain*, vol. 140, no. 7, pp. 1959–1976, Jul. 2017.
[17] E. D. Pablo-Fernandez, A. J. Lees, J. L. Holton, and T. T. Warner, "Prognosis and neuropathologic correlation of clinical subtypes of Parkinson disease," *JAMA Neurol.*, vol. 76, no. 4, pp. 470–479, 2019.
[18] R. Katzenschlager and A. J. Lees, "Treatment of Parkinson's disease: Levodopa as the first choice," *J. Neurol.*, vol. 249, pp. 19–24, Sep. 2002.
[19] M. J. Armstrong and M. S. Okun, "Diagnosis and treatment of Parkinson disease: A review," *Jama*, vol. 323, no. 6, pp. 548–560, 2020.
[20] M. Fatemi, T. W. Killian, J. Subramanian, and M. Ghassemi, "Medical dead-ends and learning to identify high-risk states and treatments," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 4856–4870.
[21] S. A. Zadeh, W. N. Street, and B. W. Thomas, "Optimizing warfarin dosing using deep reinforcement learning," *J. Biomed. Informat.*, vol. 137, Jan. 2023, Art. no. 104267.
[22] K. A. Yau, Y.-W. Chong, X. Fan, C. Wu, Y. Saleem, and P.-C. Lim, "Reinforcement learning models and algorithms for diabetes management," *IEEE Access*, vol. 11, pp. 28391–28415, 2023.
[23] S. Job, X. Tao, L. Li, H. Xie, T. Cai, J. Yong, and Q. Li, "Optimal treatment strategies for critical patients with deep reinforcement learning," *ACM Trans. Intell. Syst. Technol.*, vol. 15, no. 2, pp. 1–22, Apr. 2024.
[24] X. Wu, R. Li, Z. He, T. Yu, and C. Cheng, "A value-based deep reinforcement learning model with human expertise in optimal treatment of sepsis," *Npj Digit. Med.*, vol. 6, no. 1, p. 15, Feb. 2023.
[25] T. Nanayakkara, G. Clermont, C. J. Langmead, and D. Swigon, "Unifying cardiovascular modelling with deep reinforcement learning for uncertainty aware control of sepsis treatment," *PLOS Digit. Health*, vol. 1, no. 2, Feb. 2022, Art. no. e0000012.
[26] J. Watts, A. Khojandi, R. Vasudevan, and R. Ramdhani, "Optimizing individualized treatment planning for Parkinson's disease using deep reinforcement learning," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 5406–5409.
[27] M. Baucum, A. Khojandi, R. Vasudevan, and R. Ramdhani, "Optimizing patient-specific medication regimen policies using wearable sensors in Parkinson's disease," *Manag. Sci.*, vol. 69, no. 10, pp. 5964–5982, Oct. 2023.
[28] A. Riasi, M. Delrobaei, and M. Salari, "A decision support system based on recurrent neural networks to predict medication dosage for patients with Parkinson's disease," *Sci. Rep.*, vol. 14, no. 1, p. 8424, Apr. 2024.
[29] T. Gutowski, R. Antkiewicz, and S. Szlufik, "Machine learning with optimization to create medicine intake schedules for Parkinson's disease patients," *PLoS ONE*, vol. 18, no. 10, Oct. 2023, Art. no. e0293123.
[30] Y. Kim, J. Suescun, M. C. Schiess, and X. Jiang, "Computational medication regimen for Parkinson's disease using reinforcement learning," *Sci. Rep.*, vol. 11, no. 1, p. 9313, Apr. 2021.
[31] K. Bhattarai, S. Rajaganapathy, T. Das, Y. Kim, Y. Chen, Q. Dai, X. Li, X. Jiang, and N. Zong, "Using artificial intelligence to learn optimal regimen plan for Alzheimer's disease," *J. Amer. Med. Inform. Assoc.*, vol. 30, no. 10, pp. 1645–1656, Sep. 2023.
[32] K. Marek, D. Jennings, S. Lasch, A. Siderowf, C. Tanner, T. Simuni, C. Coffey, K. Kieburtz, E. Flagg, and S. Chowdhury, "The Parkinson progression marker initiative (ppmi)," *Prog. Neurobiol.*, vol. 95, no. 4, pp. 629–635, 2011.
[33] F. Ba, M. Obaid, M. Wieler, R. Camicioli, and W. R. W. Martin, "Parkinson disease: The relationship between non-motor symptoms and motor phenotype," *Can. J. Neurological Sci.*, vol. 43, no. 2, pp. 261–267, Mar. 2016.
[34] C. Scherfler, J. Schwarz, A. Antonini, D. Grosset, F. Valldeoriola, K. Marek, W. Oertel, E. Tolosa, A. J. Lees, and W. Poewe, "Role of DAT-SPECT in the diagnostic work up of parkinsonism," *Movement Disorders, Off. J. Movement Disorder Soc.*, vol. 22, no. 9, pp. 1229–1238, 2007.
[35] J. L. Cummings, M. J. Fine, I. D. Grachev, C. R. Jarecke, M. K. Johnson, P. H. Kuo, K. L. Schaecher, J. A. Oberdorf, M. Rezak, and D. E. Riley, "Effective and efficient diagnosis of parkinsonism: The role of dopamine transporter SPECT imaging with ioflupane I-123 injection (DaTscanTM)," *Amer. J. Managed Care*, vol. 20, no. 5, pp. S97–S109, 2014.
[36] S. Hall, Y. Surova, A. Öhrfelt, H. Zetterberg, D. Lindqvist, and O. Hansson, "CSF biomarkers and clinical progression of Parkinson disease," *Neurology*, vol. 84, no. 1, pp. 57–63, Jan. 2015.
[37] H. Kobayashi, H. C. Looker, E. Satake, P. J. Saulnier, Z. I. M. Dom, K. O'Neil, K. Ihara, B. Krolewski, A. T. Galecki, M. A. Niewczas, J. M. Wilson, A. Doria, K. L. Duffin, R. G. Nelson, and A. S. Krolewski, "Results of untargeted analysis using the SOMAscan proteomics platform indicates novel associations of circulating proteins with risk of progression to kidney failure in diabetes," *Kidney Int.*, vol. 102, no. 2, pp. 370–381, Aug. 2022.
[38] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.
[39] J. Subramanian and A. Mahajan, "Approximate information state for partially observed systems," in *Proc. IEEE 58th Conf. Decis. Control (CDC)*, Dec. 2019, pp. 1629–1636.
[40] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 2623–2631.
[41] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with deep reinforcement learning," 2013, *arXiv:1312.5602*.
[42] C. L. Tomlinson, R. Stowe, S. Patel, C. Rick, R. Gray, and C. E. Clarke, "Systematic review of levodopa dose equivalency reporting in Parkinson's disease," *Movement Disorders*, vol. 25, no. 15, pp. 2649–2653, Nov. 2010.
[43] D. J. Brooks, "Dopamine agonists: Their role in the treatment of Parkinson's disease," *J. Neurol., Neurosurg. Psychiatry*, vol. 68, no. 6, pp. 685–689, Jun. 2000.
[44] L. Dezsi and L. Vecsei, "Monoamine oxidase B inhibitors in Parkinson's disease," *CNS Neurological Disorders-Drug Targets*, vol. 16, no. 4, pp. 425–439, 2017.
[45] M. J. Azur, E. A. Stuart, C. Frangakis, and P. J. Leaf, "Multiple imputation by chained equations: What is it and how does it work?" *Int. J. Methods Psychiatric Res.*, vol. 20, no. 1, pp. 40–49, Mar. 2011.
[46] B. Dou, Z. Zhu, E. Merkurjev, L. Ke, L. Chen, J. Jiang, Y. Zhu, J. Liu, B. Zhang, and G.-W. Wei, "Machine learning methods for small data challenges in molecular science," *Chem. Rev.*, vol. 123, no. 13, pp. 8736–8780, Jul. 2023.
[47] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.

HONGBUM KIM received the B.S. degree from the Department of Industrial Management Engineering, Hankuk University of Foreign Studies, Seoul, South Korea, in 2021. He is currently pursuing the M.S. degree with the School of Industrial and Management Engineering, Korea University, Seoul. His current research interest includes clinical decision support systems using artificial intelligence.

CHANYOUNG PARK received the B.S. degree in electronic and electrical engineering from Hongik University, Seoul, South Korea, in 2022. He is currently pursuing the M.S. degree with the School of Industrial and Management Engineering, Korea University, Seoul.



JONG HOON KIM received the B.S. degree from the Department of Information and Communication Engineering, Hankuk University of Foreign Studies, Seoul, South Korea, in 2024. He is currently pursuing the M.S. degree with the School of Industrial and Management Engineering, Korea University, Seoul.

SEJEONG JANG received the B.S. degree in industrial and information systems engineering from Seoul National University of Science and Technology, Seoul, South Korea, in 2024. He is currently pursuing the M.S. degree with the School of Industrial and Management Engineering, Korea University, Seoul.

HYO KYUNG LEE (Member, IEEE) received the B.S. degree from the Department of Information and Industrial Engineering, Yonsei University, Seoul, South Korea, in 2012, the M.S. degree in industrial and systems engineering from Georgia Institute of Technology, Atlanta, GA, USA, in 2013, and the Ph.D. degree from the Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, WI, USA, in 2019. She was an Assistant Professor with the University of Pittsburgh, from 2019 to 2021. She is currently an Assistant Professor with the School of Industrial and Management Engineering, Korea University, Seoul. Her current research interests include healthcare decision making and prediction models for precision medicine.