Smart Health 36 (2025) 100561

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

An adaptive multimodal fusion framework for smartphone-based medication adherence monitoring of Parkinson's disease

Chongxin Zhong [a], Jinyuan Jia [b], Huining Li [a,*]

[a] Department of Computer Science, North Carolina State University, United States
[b] Information Sciences and Technology, Pennsylvania State University, United States

ARTICLE INFO

Keywords:
Parkinson's disease
Smartphone
Multimodal fusion

ABSTRACT

Ensuring medication adherence for Parkinson's disease (PD) patients is crucial to relieve patients' symptoms and better customizing regimens according to patient's clinical responses. However, traditional self-management approaches are often error-prone and have limited effectiveness in improving adherence. While smartphone-based solutions have been introduced to monitor various PD metrics, including medication adherence, these methods often rely on single-modality data or fail to fully leverage the advantages of multimodal integration. To address the issues, we present an adaptive multimodal fusion framework for monitoring medication adherence of PD based on a smartphone. Specifically, we segment and transform raw data from sensors to spectrograms. Then, we integrate multimodal data with quantification of their qualities and perform gradient modulation based on the contribution of each modality. Afterward, we monitor medication adherence in PD patients by detecting their medicine intake status. We evaluate the performance with the dataset from daily-life scenarios involving 455 patients. The results show that our work can achieve around 94% accuracy in medication adherence monitoring, indicating that our proposed framework is a promising tool to facilitate medication adherence monitoring in PD patients' daily lives.

1. Introduction

Approximately 6.1 million people have been affected by Parkinson's disease (PD) worldwide (GBD 2016 Neurology Collaborators, 2019). Medications, as one of the effective treatment options, can relieve symptoms related to movement, voice, and tremors. Consistent medication adherence enables physicians to adjust treatment based on the patient's clinical response effectively. In contrast, non-adherence—resulting from missed, mistimed, or extra doses—can worsen Parkinsonism symptoms (Malek & Grosset, 2015). Studies show that significant medication non-adherence remains prevalent in patients with PD today (Grosset, Bone, & Grosset, 2005; Malek & Grosset, 2015). Several factors, including complex regimens and medical complications with PD, cause medication non-adherence to PD patients. 74.1% of patients reported lower adherence to the prescribed therapy (Radojević et al., 2022), which ultimately increases the average annual rates of hospitalizations and ancillary care visits (Fleisher & Stern, 2013; Malek & Grosset, 2015).

Practical suggestions to improve medication adherence include communication, education, and engagement of other people in the therapeutic process, which can help patients understand the disease itself and the results of non-adherence (Straka, Minár, Gažová, Valkovič, & Kyselovič, 2018). Dosing devices (McDonald, Garg, & Haynes, 2002), such as reminders and pillboxes, and easier dosing regimens (Sesar, Arbelo, & del Val, 2011) are also helpful in improving adherence. In summary, the above methods mainly

* Corresponding author.
E-mail addresses: czhong4@ncsu.edu (C. Zhong), jinyuan@psu.edu (J. Jia), hli83@ncsu.edu (H. Li).

https://doi.org/10.1016/j.smhl.2025.100561
Received 10 March 2025; Accepted 15 March 2025
Available online 25 March 2025
2352-6483/Published by Elsevier Inc.

rely on self-management. However, self-management is prone to errors due to patients' forgetfulness and the difficulty in accurately assessing medication efficacy (Shin, Habermann, & Pretzer-Aboff, 2015). Furthermore, these methods cannot detect non-adherence promptly and alert patients to take medicines.

In recent years, mobile technologies have been increasingly used for symptom assessment (Baxter, Carroll, Keogh, & Vandelanotte, 2020; Omberg et al., 2022). For example, smartphone built-in sensors can be used to collect daily life activity data (e.g., gait, voice) for symptom assessment of PD in a passive and continuous way. Given that PD drugs can take effect within an hour and alleviate symptoms, it is feasible to use a smartphone-based symptom assessment approach to detect whether a PD patient has taken the drug and ultimately infer medication adherence.

Recent smartphone-based symptom assessment mainly has two directions. One direction is to use one dedicated sensing modality. For example, PDVocal (Zhang, Song et al., 2019) detects PD using non-speech body sounds. PDMove (Zhang, Xu et al., 2019) monitors medication adherence of PD patients using gait assessment (Mirelman et al., 2019). PDAssess (Yang et al., 2023) assesses the stage of PD patients using free-speech voice (Rusz, Cmejla, Ruzickova, & Ruzicka, 2011). These works only use a single sensing modality for analysis. However, a single modality cannot always make accurate estimates in all scenarios due to the limited information it represents. For example, PDMove (Zhang, Xu et al., 2019) has to train a personalized model for each PD patient using transfer learning due to the unreliability of the gait data. Another direction is to fuse multiple sensing modalities. For example, PDLens (Zhang et al., 2020) detects drug effectiveness using gait, voice, and balance data. However, they most rely on basic feature concatenation. In practice, the efficacy of a sensor may change due to unknown user and environment dynamics. Also, the symptoms of different patients may not exactly be the same for a single data source, such as gait. These factors make data from different modalities have different reliabilities. If sensor data fusion is performed without quantifying data quality, it may struggle to enhance performance when estimating fine-grained characteristics, particularly in clinical-level applications.

To address the above challenges, we propose an adaptive multimodal fusion framework for smartphone-based medication adherence monitoring of Parkinson's disease. The framework takes data from multiple sensing modalities, performs data fusion with quantification, and monitors medication adherence based on the learning results from integrated data representations. First, the framework takes raw data from sensors as inputs, applies necessary filters to segment data, and transforms data segments into spectrograms for feature extraction. Second, the framework estimates the uncertainty of each modality using energy score (Liu, Wang, Owens, & Li, 2020) as the quality of data from each modality, integrates multimodal data with dynamic weights based on the estimated uncertainties, and performs gradient modulation based on the contributions of each modality to the final output. Finally, the framework detects whether PD patients take medicine or not to monitor medication adherence. With such quantified multimodal data integration, the framework can monitor the medication adherence of PD patients accurately with one general model for all PD patients.

We evaluate the framework on a dataset collected from PD patients' daily life scenarios containing two data modalities: gait and voice. The dataset involves 455 patients. The framework achieves an average accuracy of 94.76% in medication adherence detection.

We conclude our contributions as follows:

- We design and implement an adaptive multimodal fusion framework for smartphone-based medication adherence monitoring of Parkinson's disease. With such a framework, we can monitor the medication adherence of PD patients more accurately and avoid the burden of training personalized models for all patients.
- Our work is the first to monitor medication adherence of PD with quantified multimodal fusion. It is a general framework that can be adapted to arbitrary sensing modalities.
- We evaluate the framework with gait and voice modalities on the dataset collected from daily-life scenarios. The results show that the data fusion mechanism used in the framework is effective for multiple sensing modalities.

2. Methodology

In this section, we present the overview of the framework, which is shown in Fig. 1. We first apply a series of signal processing technologies on raw sensor data to obtain feature representations from each modality. Each data modality is transformed into spectrograms to preserve both time-domain and frequency-domain information. Despite being converted into a uniform format, the modalities exhibit distinct representations for the same stage of PD. Brute-force data fusion may get corrupted results. Therefore, we perform data fusion of all sensing modalities with quantification, which consists of uncertainty quantification and data integration based on the uncertainty. Finally, we monitor the PD patients' medication adherence with the integrated multimodal classifier by detecting the medicine intake status of patients.

2.1. Data preprocessing

In this module, we collect data from different sensing modalities and preprocess data to extract high-fidelity feature representations of each modality. The data preprocessing basically consists of segmentation, transformation, and feature extraction.

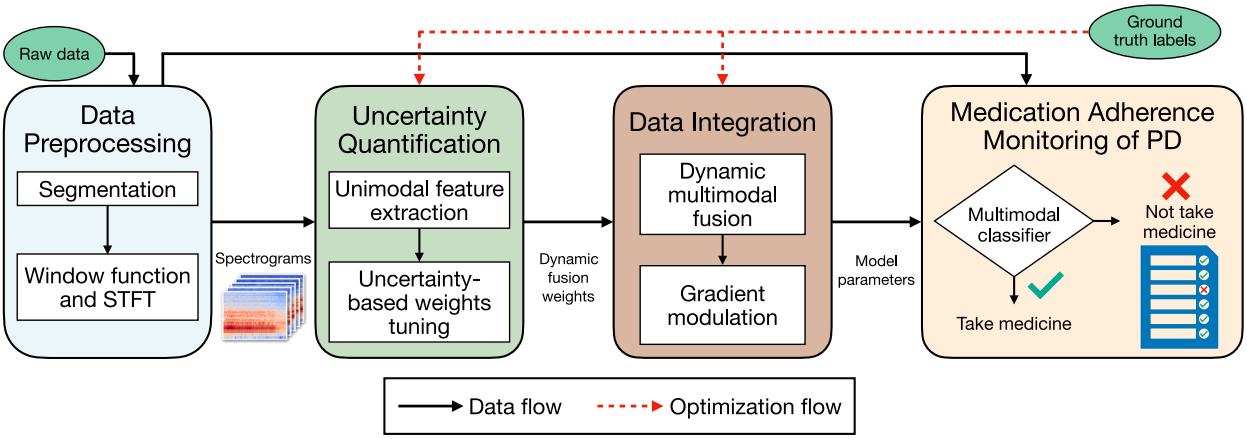

Fig. 1. The overview of our proposed framework.

Segmentation. The raw data collected directly from smartphone built-in sensors contains valuable information but is not immediately suitable for analysis or learning. In this module, we segment data based on their properties for further processing. In this work, we use gait and voice data as examples.

Gait data is analyzed based on cycles because of the walking pattern of a human (Zhang, Xu et al., 2019). Before segmentation, we first use a zero-phase filter, a type of finite impulse response (FIR), to remove the high-frequency components from the acceleration data without phase distortion. The length of the filter is set to 219. The passband is set to the range of [0.75 Hz, 2.25 Hz] to eliminate irrelevant components to gait cycles given the average step frequency and speed of walking according to the previous studies (Bohannon, 1997; Grieve & Gear, 1966). Then, we identify each local prominence from the filtered raw data to extract gait cycles. The normalized amplitude of the acceleration data is applied due to the fact that each person's amplitude value is unique.

The voice data is first partitioned into chunks with a duration of one second and zero overlapping rate. Then, we apply the Hanning window to each segment to reduce sidelobes.

Transformation. Both gait and voice data are transformed into spectrograms due to the ability to preserve features in both the time and frequency domains and their fitness to neural network input. Another FIR filter is applied to each gait segment to filter out components unrelated to PD patients' walking activity and tremors. The passband is set to under 12 Hz, which is approximately the highest frequency of PD tremor (Deuschl et al., 2000). The short-time Fourier transform (STFT) is applied to each segment to generate spectrograms.

2.2. Data fusion with quantification

As aforementioned, basic feature concatenation may lead to unsatisfied results. The issue is more common for gait data as patients in the same stage of PD may have different behaviors and symptoms. Previous work shows that accuracy can drop to as low as 70.0% when the uncertainty in gait data is not accounted for Zhang et al. (2020). Multimodal fusion with such low-quality data can adversely affect the final prediction results. To integrate data effectively, we need to quantify the uncertainty of each data modality first. After that, we perform the late fusion by assigning dynamic weights, which are determined based on the assessed uncertainty of each modality's input. Gradient modulation is then applied after each modality's inputs are adjusted by the dynamic weights to further optimize the learning process.

Uncertainty Quantification: The uncertainties in the learning process of deep neural networks mainly fall into two categories: aleatoric uncertainty and epistemic uncertainty (Kendall & Gal, 2017). Aleatoric uncertainty represents noises that lie within the data, while epistemic uncertainty refers to the uncertainty in the model. In this paper, we focus on aleatoric uncertainty quantification, as epistemic uncertainty can be mitigated with sufficient data from PD patients.

We denote the training dataset as $D = \{x_i, yi\}_{i=1}^{N}$, where N represents the number of samples in the dataset. Each x_i represents inputs from different modalities as $x_i = \{x_i^1, x_i^2, \ldots, x_i^M\}_{i=1}^{N}$ for M modalities. In our proposed application, we use gait and voice as input modalities, thus making $x_i = \{x_i^g, x_i^v\}_{i=1}^{N}$. For each modality, f^m represents the unimodal learning model for the mth modality. $y_i \in \{1, 2, \ldots, H\}$, where H is defined as the number of categories.

We use energy score (Liu et al., 2020; Zhang et al., 2023), which is a widely used metric for uncertainty quantification. The energy score represents the statistical distance between probability distributions and can be used to calculate the data density. The energy score can be calculated as:

$$E(x_i^m) = -T^m \cdot log \sum_{h}^{H} e^{f_h^m(x_i^m)/T^m}, \tag{1}$$

where $f_h^m(x_i^m), m \in \{g, v\}$ is defined as the output of the unimodal classifier f^m for the category h and T^m that originates from the energy-based model (Ranzato, Boureau, Chopra, & LeCun, 2007) is a parameter for each modality controlling how distinguishable the energy scores are between in- and out-of-distribution examples. A larger energy score means a more uniform distribution of the modality, indicating higher estimated uncertainty.

Uncertainty-driven Dynamic Data Fusion: With quantified uncertainty of each modality, we can integrate inputs from different modalities with confidence. Since the gait and voice impairment behaviors of PD patients vary, it is clear that the noise varies with different inputs. Therefore, the uncertainty of each input is different from that of the others, thus making dynamic weighting for each modality essential. The output after integration is calculated using uncertainty-driven fusion weighting:

$$f(x) = \sum_{m=1}^{M} \mathcal{W}^m(x_i^m)\mathcal{F}^m(x_i^m), \tag{2}$$

where W^m represents the function that calculates dynamic weights for modality m and $\mathcal{F}^m$ represents the output of each modality. The dynamic weights can be calculated as:

$$\mathcal{W}^m(x_i^m) = \alpha^m u^m(x_i^m) + c^m, \tag{3}$$

where α^m and c^m are the hyper-parameters that can be tuned for better fusion results, and $u^m(x)$ is the estimated uncertainty of the m modality. As mentioned before, we use energy score to estimate uncertainty. The dynamic weights are formulated as:

$$\mathcal{W}^m(x_i^m) = -\alpha^m T^m \cdot log \sum_{h}^{H} e^{f_h^m(x_i^m)/T^m} + c^m. \tag{4}$$

Since a higher energy score indicates a higher estimated uncertainty, α^m is a negative number.

Additionally, the estimated uncertainty based on energy score is regularized using the sample-wise loss to enhance its quality (Zhang et al., 2023). First, the average training loss τ_i^m for T epochs for mth modality is calculated after training for enough epochs. The estimated uncertainty is then regularized by learning the relationship between the average loss and the dynamic weight, ensuring that a higher average loss corresponds to a smaller dynamic weight. The regularization loss can be calculated as:

$$\mathcal{L}_{reg} = max(k(w_i^m, w_j^m)(\tau_i^m - \tau_j^m) + |w_i^m - w_j^m|, 0), \tag{5}$$

$$k(w_i^m, w_j^m) = \begin{cases} 1 & w_i^m > w_j^m \\ 0 & w_i^m = w_j^m \\ -1 & \text{others.} \end{cases} \tag{6}$$

The total loss of the multimodal fusion model is calculated as:

$$\mathcal{L} = \mathcal{L}_{fusion} + \sum_{m=1}^{M} \mathcal{L}^m + \lambda \mathcal{L}_{reg}, \tag{7}$$

where $\mathcal{L}_{fusion}$ is the loss after dynamic fusion, $\mathcal{L}^m$ is the loss of each modality before fusion, and λ is a hyper-parameter that controls the strength of the regularization. For spectrograms in our work, the losses are calculated using cross-entropy.

Gradient Modulation after Dynamic Fusion: The uncertainty-driven data fusion integrates data with consideration of their uncertainty. However, it does not measure the contribution of each modality after dynamic weighting. Inspired by recent works (Peng, Wei, Deng, Wang, & Hu, 2022), incorporating gradient modulation for each modality after dynamic weighting offers a practical solution. This is achieved by monitoring the discrepancy between their contributions to the final output. For gait and voice modality $m \in \{g, v\}$, the discrepancy ratio (Peng et al., 2022) ρ_t^m for the tth epoch is calculated as:

$$\rho_t^g = \frac{\sum_{i \in N} s_i^g}{\sum_{i \in N} s_i^v}, \tag{8}$$

where s_i^g and s_i^v are scores that measure the contribution of gait and voice modality to the output of the fusion model. ρ_t^v is defined as the reciprocal of ρ_t^g. The scores are calculated as:

$$s_i^m = \sum_{h=1}^{H} 1_{h=y_i} \cdot S(\mathcal{W}_t^m(x_i^m)\mathcal{F}_t^m(x_i^m))_h, \tag{9}$$

where $m \in \{g, v\}$ and S is the softmax function. Given the contribution discrepancy between the gait and voice modalities, the gradient of each modality m can be modulated as:

$$g_t^m = \begin{cases} 1 - \tanh(\beta \cdot \rho_t^m) & \rho_t^m > 1 \\ 1 & \text{others,} \end{cases} \tag{10}$$

where β is a hyper-parameter controlling the strength of modulation. With the gradient modulation, the modality contributing more to the output ($\rho_t^m > 1$) can be prioritized for further training, while modalities with higher noise levels remain unaffected.

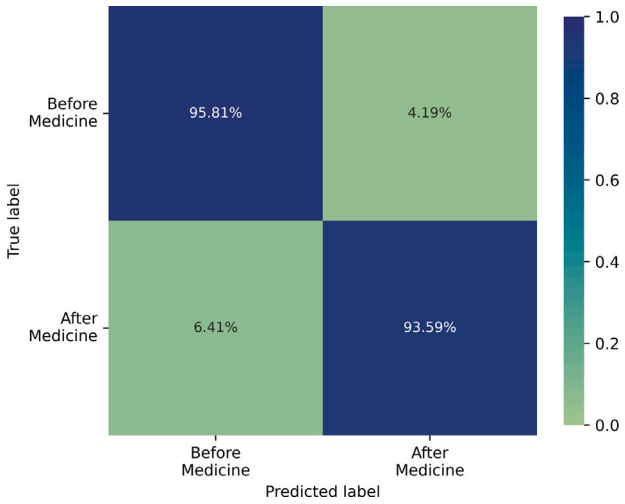

Fig. 2. The normalized confusion matrix of medicine-taking detection.

2.3. Medication adherence monitoring of PD

We implement the PD medication adherence monitoring application using the trained multimodal classifier by detecting whether the patient has taken medicine or not. It is a passive monitoring approach that only uses built-in sensors in the smartphone without active cooperation from patients. The patients need to register their medication schedule before monitoring starts. Then, our system passively collects the patient's gait and voice data and detects medicine intake status using the multimodal classifier. Afterward, we can examine missed or extra doses by comparing our medicine intake detecting results with the patient's medication schedule and ultimately infer medication adherence.

3. Evaluation

3.1. Experimental setup

3.1.1. Dataset

We use a PD dataset involving 455 PD patients (Bot et al., 2016). The gait and voice data are collected by smartphone either before or after patients take medicine. The labels belong to two categories: "before medicine" and "after medicine". We have 155 289 gait samples and 186 095 voice samples. Gait and voice data with the same label within a day are paired as inputs of the multimodal learning process. After pairing, we have 97 733 samples in total. 80% of the samples are used as training data and the rest 20% are used for testing.

3.1.2. Learning model implementation

The experiments are performed on an NVIDIA RTX A6000 GPU. Before data fusion, a ResNet-18 classifier is used to extract feature representations from spectrograms for each modality. To train a multimodal fusion model, we choose the SGD optimizer with 5e−4 initial learning rate, 1e−3 weight decay, and 0.9 momentum factor. The learning rate is reduced with 70 step and 0.3 multiplicative factor of learning rate decay. We conduct the five-fold cross-validation to evaluate the performance of the framework.

3.2. Performance metrics

We use accuracy, precision, recall, and F1 score to evaluate the performance of medicine intake status detection. The label "after PD medication" is defined as the positive class.

3.3. Overall performance

Fig. 2 shows the normalized confusion matrix of medicine intake status detection. The X-axis is the ground truth, and the Y-axis is the prediction results of our model. Overall, the framework achieves 94.76% accuracy, 95.26% precision, 93.59% recall, and 94.42% F1 score, which shows high medicine intake detection performance. The precision is higher than recall, showing that our model is more likely to predict a sample as "before medicine". One reason is that the efficacy of medicine may not be evident for some patients who have developed drug resistance.

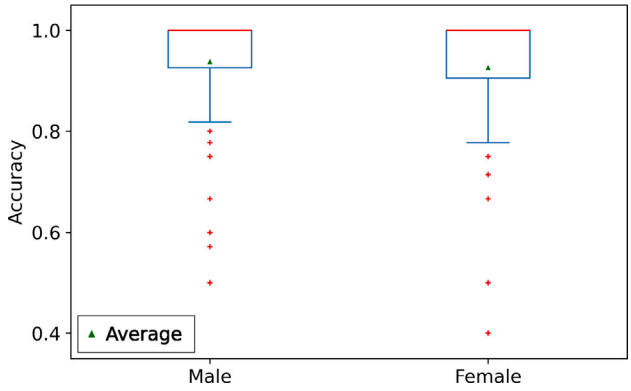

Fig. 3. The comparison of the accuracy of medicine intake detection between different genders.

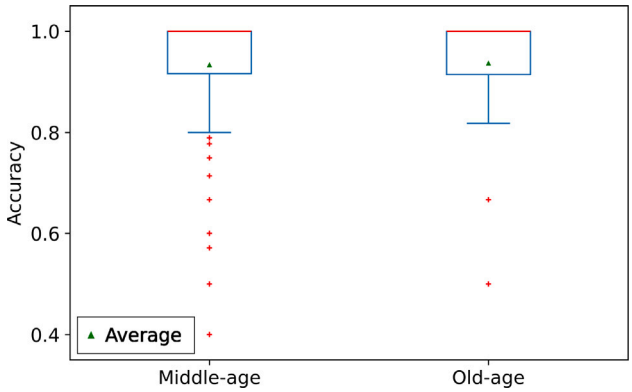

Fig. 4. The comparison of the accuracy of medicine intake detection between different age groups.

Table 1
The comparison of the accuracy between different learning setups.

Method	Gait	Voice	Concat	GM	DW	Our work
Accuracy (%)	59.21	89.04	88.94	92.00	93.96	**94.76**

3.4. Performance in different demographic groups

We measure the model's performance against different demographic groups, including gender, age, and onset year. There are 65% male patients and 35% female patients. 87% of patients are under 75 years old, which is considered middle age. The rest are 75 years old or over, considered old age. Onset year refers to the time between the first appearance of symptoms and the experiment. 36% of patients' onset years are less than or equal to three years, and 64% of patients' onset years are greater than three.

Fig. 3 shows the comparison of the accuracy of medicine intake detection between different genders. The red line at the top of the box indicates the median. It shows that the model achieves 100% accuracy for more than half of the patients. The average accuracy for males and females is 93.74% and 92.56%, respectively. Our results implicate that gender is an independent biological factor.

Fig. 4 shows the comparison of accuracy between different age groups. The mean accuracy for middle-age and old-age patients is 93.31% and 94.37%, respectively. The result shows that our model can detect medicine intake status accurately for both middle-age and old-age patients. The middle-age patients have more outliers. This may be because some middle-age patients have not unnoticeable symptoms that are hard to detected by the model.

Fig. 5 shows the accuracy comparison between different onset years of patients. The average accuracy of the two groups is 93.78% and 93.14%, respectively. The result shows that the progress of the disease will not affect the accuracy of the medicine intake detection.

3.5. Ablation study

In this section, we evaluate the improvement contributed by each factor. Table 1 shows the average accuracy under different training setups. The details of different setups are listed as follows: **Gait**: Learning only with gait data; **Voice**: Learning only with

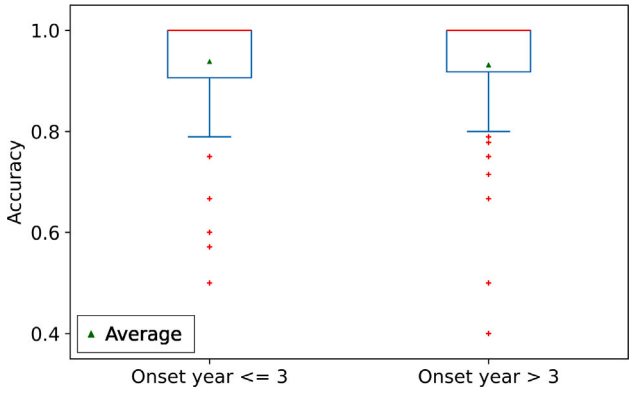

Fig. 5. The comparison of the accuracy of medicine intake detection between different onset years.

voice data; **Concat**: Learning with simple concatenation of gait and voice data; **GM**: Multimodal fusion with gradient modulation; **DW**: Multimodal fusion using dynamic weights; **Our work**: Multimodal fusion with both gradient modulation and dynamic weights.

Because of its large uncertainty, the accuracy of learning only with gait data is under 60%. The accuracy of learning with simple concatenation is lower than that of learning with voice data, showing the influence of low-quality data in multimodal fusion without quantification. The results show that our work can better harness the potential of each data modality and gain more accurate results on medicine intake detection.

4. Conclusion

In this paper, we presented a framework for medication adherence monitoring of Parkinson's disease with quantified multimodal fusion to better leverage the valuable information in each data modality. It works by preprocessing raw data from multiple sensing modalities, calculating dynamic weights based on uncertainty estimation, and applying gradient modulation to the dynamically integrated data. This approach enables accurate detection of medication intake status, which can help monitor the medication adherence of PD patients. The evaluation shows that our work can achieve better accuracy than existing methods.

CRediT authorship contribution statement

Chongxin Zhong: Methodology, Software, Writing – original draft, Writing – review & editing. **Jinyuan Jia:** Writing – review & editing, Investigation. **Huining Li:** Conceptualization, Project administration, Resources, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We thank all anonymous reviewers for their insightful comments on this paper. This work was supported by the mPower dataset.

Data availability

The authors do not have permission to share data.

References

Baxter, C., Carroll, J.-A., Keogh, B., & Vandelanotte, C. (2020). Assessment of mobile health apps using built-in smartphone sensors for diagnosis and treatment: Systematic survey of apps listed in international curated health app libraries. *JMIR MHealth UHealth*, *8*(2), Article e16741.

Bohannon, R. W. (1997). Comfortable and maximum walking speed of adults aged 20-79 years: reference values and determinants. *Age and Ageing*, *26*(1), 15–19.

Bot, B. M., Suver, C., Neto, E. C., Kellen, M., Klein, A., Bare, C., et al. (2016). The mpower study, parkinson disease mobile data collected using ResearchKit. *Scientific Data*, *3*(1), 1–9.

Deuschl, G., Raethjen, J., Baron, R., Lindemann, M., Wilms, H., & Krack, P. (2000). The pathophysiology of parkinsonian tremor: a review. *Journal of Neurology*, *247*(S5), V33–V48.

Fleisher, J. E., & Stern, M. B. (2013). Medication nonadherence in parkinson's disease. *Current Neurology and Neuroscience Reports*, *13*, 1–4.

GBD 2016 Neurology Collaborators (2019). Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the global burden of disease study 2016. *Lancet Neurology*, *18*(5), 459–480.

Grieve, D. W., & Gear, R. J. (1966). The relationships between length of stride, step frequency, time of swing and speed of walking for children and adults. *Ergonomics*, *9*(5), 379–399.

Grosset, K. A., Bone, I., & Grosset, D. G. (2005). Suboptimal medication adherence in parkinson's disease. *Movement Disorders*, *20*(11), 1502–1507.

Kendall, A., & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems*, *30*.

Liu, W., Wang, X., Owens, J., & Li, Y. (2020). Energy-based out-of-distribution detection. *Advances in Neural Information Processing Systems*, *33*, 21464–21475.

Malek, N., & Grosset, D. G. (2015). Medication adherence in patients with parkinson's disease. *CNS Drugs*, *29*(1), 47–53.

McDonald, H. P., Garg, A. X., & Haynes, R. B. (2002). Interventions to enhance patient adherence to medication prescriptions: scientific review. *JAMA*, *288*(22), 2868–2879.

Mirelman, A., Bonato, P., Camicioli, R., Ellis, T. D., Giladi, N., Hamilton, J. L., et al. (2019). Gait impairments in parkinson's disease. *Lancet Neurology*, *18*(7), 697–708.

Omberg, L., Chaibub Neto, E., Perumal, T. M., Pratap, A., Tediarjo, A., Adams, J., et al. (2022). Remote smartphone monitoring of parkinson's disease and individual response to therapy. *Nature Biotechnology*, *40*(4), 480–487.

Peng, X., Wei, Y., Deng, A., Wang, D., & Hu, D. (2022). Balanced multimodal learning via on-the-fly gradient modulation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8238–8247).

Radojević, B., Dragašević-Mišković, N. T., Milovanović, A., Svetel, M., Petrović, I., Pešić, M., et al. (2022). Adherence to medication among parkinson's disease patients using the adherence to refills and medications scale. *International Journal of Clinical Practice*, *2022*, Article 6741280.

Ranzato, M., Boureau, Y.-L., Chopra, S., & LeCun, Y. (2007). A unified energy-based framework for unsupervised learning. In *Artificial intelligence and statistics* (pp. 371–379). PMLR.

Rusz, J., Cmejla, R., Ruzickova, H., & Ruzicka, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinson's disease. *Journal of the Acoustical Society of America*, *129*(1), 350–367.

Sesar, A., Arbelo, J. M., & del Val, J. L. (2011). Treatment of parkinson disease, time and dosage: "does simple dosage facilitate compliance and therapeutic goals?". *Neurologist*, *17*(6 Suppl 1), S43–6.

Shin, J. Y., Habermann, B., & Pretzer-Aboff, I. (2015). Challenges and strategies of medication adherence in parkinson's disease: A qualitative study. *Geriatric Nursing*, *36*(3), 192–196. http://dx.doi.org/10.1016/j.gerinurse.2015.01.003, URL https://www.sciencedirect.com/science/article/pii/S019745721500049X.

Straka, I., Minár, M., Gažová, A., Valkovič, P., & Kyselovič, J. (2018). Clinical aspects of adherence to pharmacotherapy in parkinson disease. *Medicine (Baltimore)*, *97*(23), Article e10962.

Yang, B., Hu, Q., Xie, W., Wang, X., Luo, W., & Zhang, Q. (2023). PDAssess: A privacy-preserving free-speech based parkinson's disease daily assessment system. In *Proceedings of the 21st ACM conference on embedded networked sensor systems* (pp. 251–264).

Zhang, H., Guo, G., Song, C., Xu, C., Cheung, K., Alexis, J., et al. (2020). Pdlens: smartphone knows drug effectiveness among parkinson's via daily-life activity fusion. In *Proceedings of the 26th annual international conference on mobile computing and networking* (pp. 1–14).

Zhang, H., Song, C., Wang, A., Xu, C., Li, D., & Xu, W. (2019). Pdvocal: Towards privacy-preserving parkinson's disease detection using non-speech body sounds. In *The 25th annual international conference on mobile computing and networking* (pp. 1–16).

Zhang, Q., Wu, H., Zhang, C., Hu, Q., Fu, H., Zhou, J. T., et al. (2023). Provable dynamic fusion for low-quality multimodal data. In *Proceedings of the 40th international conference on machine learning*. JMLR.org.

Zhang, H., Xu, C., Li, H., Rathore, A. S., Song, C., Yan, Z., et al. (2019). Pdmove: Towards passive medication adherence monitoring of parkinson's disease using smartphone-based gait assessment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *3*(3), 1–23.