International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(1s) 2026 pp. 1-12
https://doi.org/10.70917/ijcisim-2026-2014
Article

An Interpretable Multi-Modal Diagnostic Framework for Parkinson's Disease Using Auxiliary Kolmogorov–Arnold Networks and Probabilistic Decision Fusion.

Roshani Rajendra Pasalkar[1] , Shashank D. Joshi[2]

[1,2] Department of Computer Engineering, Bharati Vidyapeeth (Deemed to be University),Pune,India
[1]roshanipawar936@gmail.com
[2]shashank.joshi@bharatividyapeeth.edu

Abstract: . Parkinson's disease (PD) is a progressive neurodegenerative disease that causes a variety of both motor and non-motor impairments, creating a significant clinical challenge in terms of diagnosing the disorder early and reliably. Historically, doctors diagnosed PD using neurological assessments and rating scales based solely on subjective observations, which lack sensitivity for detecting early signs of PD. Over the past few years, there has been some success developing diagnostic systems based on machine learning, though these systems to date are almost entirely unimodal except for a few exceptions. Most current approaches do not allow for interpretability or accommodate the uncertainty associated with clinical data. In this paper, we propose a multi-modal diagnostic framework for PD that is interpretable and combines heterogeneous biomedical signals using deep representation learning (DRL), cross-modal modelling using functional relationships, and Bayesian probabilistic decision fusion (DPDF). Each of the modalities will first have its discriminative latent representations extracted using each modality's deep encoder. An Auxiliary Kolmogorov–Arnold Network (AKAN) will be used to model the non-linear functional dependencies between modalities. Finally, the Diagnostic Inference will be performed using DPDF to fuse modality-specific and cross-modal evidence while providing a measure of uncertainty for each diagnostic inference made. The research on the proposed methodology was evaluated via publicly available datasets from Kaggle and through the use of a subject-independent cross-validation scheme. The results obtained from the experimental test cases have shown that fusing multimodal data sources gives significantly better performance than using a single modal data source alone and results in substantially increased rates of accuracy, sensitivity and AUC. With all the capabilities outlined in this paper, our new approach represents a new state of the art for AI-supported diagnosis of PD. In particular, it blends deep learning methods, functional interpretive capabilities, and probabilistic reasoning into a single methodology that meets the requirements of a clinically relevant solution.

Keywords: Parkinson's Disease Diagnosis; Multimodal Deep Learning; Kolmogorov–Arnold Networks; Probabilistic Fusion; Explainable Artificial Intelligence

1. Introduction

Parkinson's Disease (PD) is a chronic progressive neurodegenerative condition brought on by the loss of dopaminergic neurons in the substantia nigra leading to a broad range of motor and non motor disability [1]. PD is clinically characterized by both cardinal motor symptoms and non-motor symptoms, including prototypical cardinal motor effects, which are resting tremor, bradykinesia, rigidity and postural instability and non-motor symptoms that appear as impairment of speech, cognitive dysfunction, sleep disorders and autonomic dysfunction [2]. Early-stage PD diagnosis is a major clinical challenge due to its heterogeneous and slightly changing nature. The existing diagnostic practices are mainly based on neurological tests and clinical rating scales, which are subjective in nature and usually lack sensitivity to early and thin pathological patterns. As a result, the current situation demands a

serious necessity to have objective diagnostic methods that are reliable and scalable to help clinicians diagnose Parkinson's Disease at an earlier stage in a more accurate manner [3].

Recent developments in artificial intelligence (AI) and machine learning (ML) have facilitated data-driven methods regarding automated PD detection based on various biomedical measures. The significant amount of literature concentrates on the unimodal diagnostic systems where certain symptoms of PD pathology are used in the systems [4]. Dysarthric impairments such as jitter, shimmer and the instability of pitch, micrographia and fine motor impairment (using kinematic trajectories), stride variability and posture instability (using wearable or smartphone sensors), as well as chore structural and functional changes (on the basis of MRI or similar modalities), are investigated with speech-based technologies, handwriting-based technologies, gait-based technologies, and neuroimaging-based technologies respectively. Even though there are encouraging performances with these unimodal methods, each of the modalities reflects only half the picture of the disease process itself and is subject to noise, inter-subject variability, and effects of learning acquisitions [5].

Parkinson's Disease, neuro-physiologically, is multi -systemic in nature, being equally exerted on motor control, speech production, fine motor coordination, as well as brain structure simultaneous [6]. Therefore, the unimodal models of AI are inherently restricted concerning the ability to capture the interdependent nature of complex manifestations of PD. This shortcoming has inspired the development of multi-data-modes of learning, which attempt to offer heterogeneous data sources with complementary information so as to achieve enhanced diagnostic quality, strength and external validity [7]. Some experiments have shown that multi-modal modalities like speech, gait, hand writing and neuroimaging have had better performance than unimodal baselines. Nevertheless, the currently used multi-modal methods are typically based upon premature feature concatenation or ad hoc late fusion methods, which do not effectively encode cross-modal interactions, and have no principled methods of uncertainty management and interpretability.

The interpretability and the uncertainty awareness are very essential in clinical decision-making as the predictive accuracy. Although successful, deep learning models are often accused of being black-box, making it hard to make clinical trust and adoption. Also, the greater majority of multi-modal diagnostic systems inevitably provide an implicit assumption that the providing societies are equally reliable and that all information sources are always present when making inferences- a fact that is rarely applicable to the clinical setting. The variability of sensor quality, the absence of a specific modality, and patient-related variability necessitate the use of diagnostic models in which the lack of variability can be more easily resolved functional albeit at a cost of uncertainty reasoning and the adaptive derivation of modality-specific diagnostic evidence. Uncertainty-conscious fusion of decisions in this type of a setting is mathematically soundly based on probabilistic modelling and information-theoretic concepts, including Bayesian inference and Kullback-Leibler divergence. The other inherent weakness of current deep multi-modal architectures is that they use traditional multilayer perceptrons or fully connected fusion layers that are not able to give functional interpretability of cross-modal interactions. Although recent polls point to the rising interest in explainable AI (XAI) in the medical field, most deep fusion models do not explicitly split the learned representations into readable functional items. This is specifically important in medical diagnostics where the interplay of different modalities to make a diagnosis is critical to the clinical validation and trust of the modality [8].

In order to overcome these issues, this research apper introduces a significant expansion of an earlier suggested conference-level multi-modal Parkinson's Disease diagnosis model, as it proposes a profound cross-modal learning mechanism, founded on an Auxiliary Kolmogorov-Arnold Network (AKAN). KolmogorovArnold representation theorem, proven in 1948, can provide a useful theoretical framework of functional decomposition by showing that any multi-variate continuous function may be written in terms of finite composition of univariate functions. With this principle, the proposed AKAN architecture represents non-linear non-modal relationships using interpretable functional mappings instead of non-interpreting fully connected layers. This facilitates explicit learning modality interactions and retains interpretability a feature that is especially important in clinical decision support systems.

The improved novel architecture combines deep signal encoders of heterogeneous biomedical signals to AKAN-based learning of cross-modal representations and probabilistic fusion of decisions based on Bayesian inference. The AKAN module is considered an additional component of learning, which intersects positions of functional dependencies of latent modality representations, and thus provides information in enhancing the diagnostic evidence in comparison to the contribution of independent modalities. This is followed by the probabilistic fusion layer, which combines modality level as well as cross modal evidence and takes into consideration uncertainty and different reliabilities of the modalities. This hybrid architecture plays the roles of

bridging the divide between deep representation and functional interpretable and uncertainty-sensitive inferences. Figure 1 illustrates a conceptual representation of the proposed interpretable multi-modell Parkinson's disease diagnostic framework. It outlines the main shortcomings of existing unimodal and heuristic multi-modes diagnostic systems and presents the solution proposed based on modality specific deep encoders, an explicit functional cross-modal representation learning Auxiliary Kolmogorov-Arnold Network (AKAN), and Bayesan probabilistic evidence fusion. The AKAN module can be used to model non-linear cross-modal relationships in an interpretable manner, and the probabilistic fusion layer can be used to combine modality-specific and cross-modal pieces of evidence to make uncertainty-sensitive and clinically-confidence diagnostic judgments.

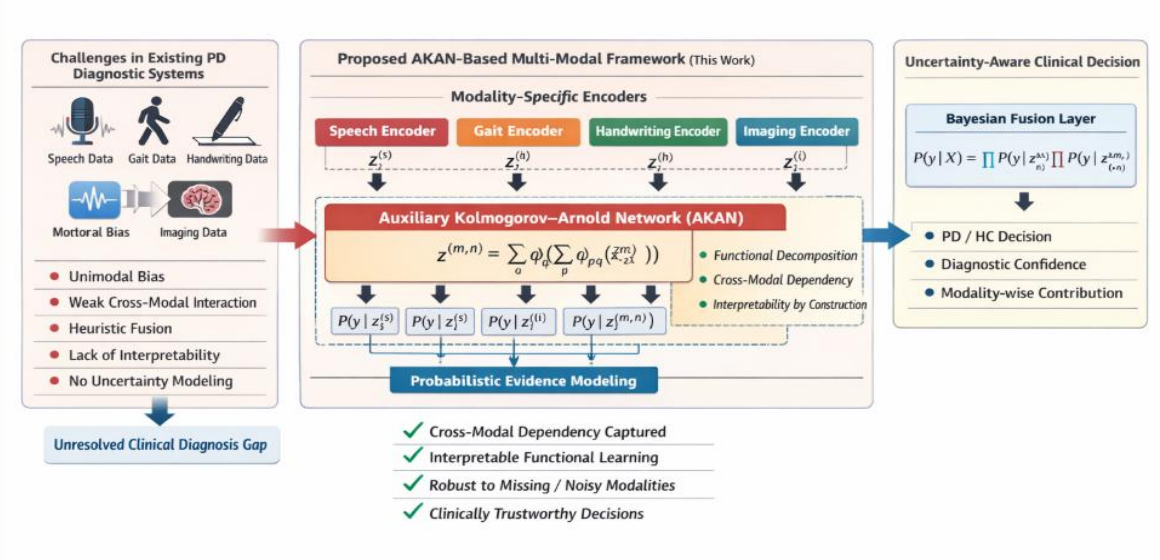

Figure 1. Interpretable AKAN-Based Multi-Modal Diagnostic Framework for Parkinson's Disease

Considering a group of non-homogeneous biomedical modalities gathered on a subject, the main issue discussed in this paper is to construct a diagnostic framework, capable of:

- acquire discriminative and semantically consistent representations of both modalities,
- directly represent non-linear cross-modal interdependencies in an understandable form, and
- fuse on uncertainty to determine the likelihood of Parkinson Disease.

The difficulty lies in the heterogeneity of data distributions, the differences in dimensionalities, the partial availability of modality, and the necessity to have clinically interpretable decisions in diagnosis. The most important novelties of the journal extension are summed up as follows:

- An AKAN-based module is proposed to represent functional cross-modal relationships with interpretable univariate function compositions, which is not restricted to the conventional deep fusion models.
- In this method, probabilistic fusion is addressed by adding the element of uncertainty to the initially developed basic prediction methods.<|human|>Uncertainty-Conscious Probabilistic Fusion:
- It uses a Bayesian decision fusion algorithm to combine modality-level evidence and cross-modal evidence which allows to make robust diagnosis in uncertainty and partial modality coverage.
- The suggested architecture uses a combination of deep representation learning, functional decomposition, and probabilistic inference and has both high diagnostic and clinical interpretability.

In line with the foregoing motives, the following are the main aims of this research:

- To come up with a strong and explainable multi-modal diagnostics model of detection of Parkinson Disease using heterogeneous biomedical data.

- To design and combine an Auxiliary Kolmogorov -Arnold Network based deep learning model to acquire explicit and interpretable cross-modal representations.
- To design a probabilistic fusion mechanism which is doubtful and effective, towards better diagnostic accuracy and clinical strength.

These contributions contribute to the state of the art in the multi-modal diagnosis of Parkinson, as well as provide an appropriate backbone of reliable AI-assisted neurological decision support systems.

2. Related Work

The most recent innovations in machine learning have had a strong impact on the evolution of automated diagnostic systems of Parkinson's Disease (PD). Specifically, the use of evidence-based models taking advantage of biomedical signals has been instrumental. Diagnostic methodologies based on neuroimaging have received significant interest because of the ability to directly record structural and functional changes in the brain that define PD. As an example, Ya et al. [5] trained machine learning models using various structural MRI-derived features and showed an improvement in diagnostic accuracy compared to single-feature models. The imaging-based techniques have the advantage of providing a high level of discriminative power over them, however, require significant computational capabilities, require proprietary acquisition protocols, and are unable to scale to and be continuously screened, thus limiting their applicability in clinical settings. The evolution of Parkinson Disease detection and the shift from unimodal Machine Learning models to deep learning–based multimodal framework are shown in Figure 2. In light of these recent trends towards interpretability and uncertainty-aware decision making, we seek a functionally interpretable cross-modal learning approach.

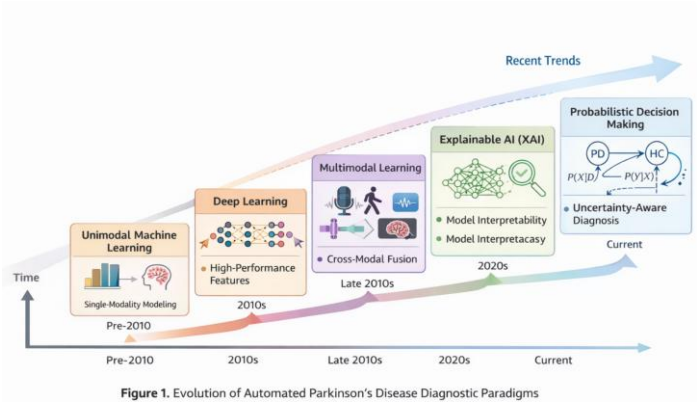

Figure 1. Evolution of Automated Parkinson's Disease Diagnostic Paradigms

Figure 2. Evolution of Automated Parkinson's Disease Diagnostic Paradigms

In order to overcome the inherent shortcomings of unimodal diagnostic networks, the idea of multimodal machine learning has been proposed as a potential way of looking at medical diagnosis. Baltrušaitis et al. [6] gave a detailed taxonomy of multimodal machine learning highlighting the benefits that come with using a combination of heterogeneous sources of data to extract complementary information. Continuing this theme, Yan et al. [7] provided a multimodal methodology survey within the medical diagnostics field and highlighted that the multimodal interaction approach often results in high robustness and generalisation. However, these surveys also indicated that most of the current multimodal architectures are either based on early feature concatenation or late decision fusion; these methods are also known not to effectively capture cross-modal interactions and do not give any theoretical guarantees about interpretability and uncertainty, dealing.

Multimodal diagnostic systems are expected to gain greater adoption in health care due to deep learning. Esteva et al. [8] demonstrated the translatory power of deep neural networks in the broad palette of modalities in healthcare diagnosis. Although these achievements have been made, the intrinsic opaque nature of deep learning models is still an essential issue, particularly in safety-related areas like the medical field. Murdoch et al. [9] have conducted a formal study of interpretability within machine learning and emphasized the importance of models providing clear mechanisms of reasoning instead of depending on post-hoc explanations only. Considering PD diagnosis, the majority of deep multimodals use fully connected fusion layers, effectively blurring the functional relation between the modalities, and hindering clinical interpretability. In Figure 3 a distribution of biomedical modalities used in Parkinson's Disease diagnostic studies. Though the literature yields many unimodal approaches,

recent works have started to frame multimodal fusion in specific ways, emphasizing the importance of principled integration and cross-modal modeling.

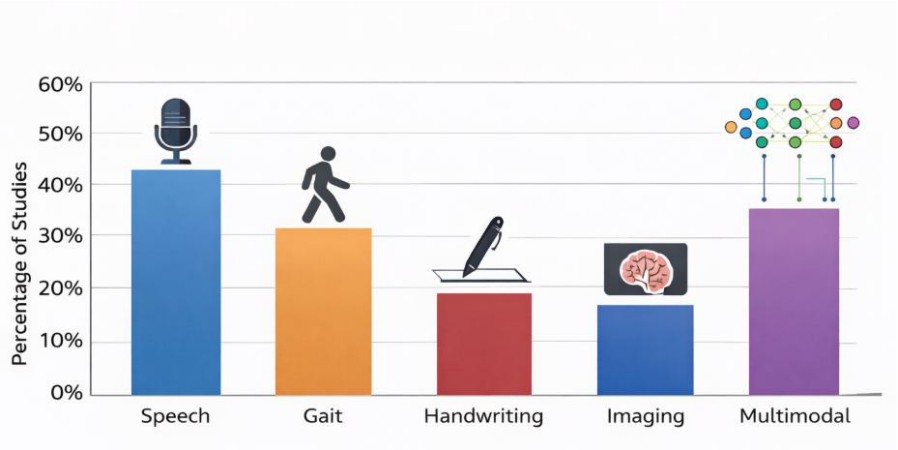

Figure 3. Distribution of Data Modalities Used in Parkinson's Disease Diagnostic Studies

Uncertainty modelling is another important aspect of the medical AI systems. Clinical practice decisions around diagnostic processes are inevitably fraught with uncertainty because of noisy measurements, inter-patient variability and incompleteness of information. Kendall and Gal [10] emphasized the role of modelling both epistemic and aleatoric uncertainty in deep learning systems, especially when decisions are required. Nevertheless, most existing multimodal PD diagnostic models implicitly presuppose the existence of complete and reliable modalities, and thus they do not estimate uncertainty at all. Probabilistic inference and uncertainty quantification on an information-theoretic basis, including methods based on the Kullback–Leibler divergence [11], have a principled basis, but have also not been applied extensively to multimodal PD diagnosis.

Innovations in the area of information fusion in the recent past have been aimed at mitigating the shortcomings. Weng et al. [12] have conducted a survey of semi-supervised information fusion methods in the analysis of medical images and highlighted the fact that probabilistic fusion schemes are required that can successfully combine source of heterogeneous evidence. Similarly, Shaik et al. [20] revised the history of multimodal information fusion in smart healthcare and showed how raw data merge evolved into knowledge-based decision making. Although these are made, the majority of fusion strategies are still either heuristic or statistically shallow, and do not explicitly model cross-modal functional dependencies.

Generalisation and overfitting are other problems of deep multimodal learning where small medical datasets are presented. Hawkins [13] resolved the basic issue of overfitting in high-dimensional models, which is especially critical in the context of multimodal models when feature space is heterogeneous and large. Deep architectures Deep belief networks [14] and classical deep architectures have been successfully adopted in representation learning, but are not explicitly stipulated to partition multi-variable relations into elements that should be easily understood. Partially it has been suggested that transfer learning can help: Raghu et al. [15] study transfer learning behaviour on medical imaging; however, transfer-based methods do not eliminate cross-modal interactions interpretability.

These gaps are further supported by broad surveys on disease prediction using deep learning. Xie et al. [16] conducted a general review of deep learning models to predict multiple diseases and observed that multimodal models enhance performance, but often reduce transparency and reliability. The expressive power of deep neural networks has been shown by Misra and Li [17]; but it is quite hard to interpret the approximations (deep neural networks) in terms of functions. Methods that implement uncertainty into deep learning, like Bayesian dropout [18], have been considered, although these are typically implemented on a-posti at the model level, and are not considered part of the multimodal fusion process.

Explainable Artificial intelligence (XAI) has thus emerged as one of the key areas of study in medical image analysis and healthcare AI. Van der Velden et al. [19] performed a review of XAI methods to analyse medical images with deep learning and called on intrinsic interpretability instead of post-hoc explanations to be acceptable to clinicians. However, much of the existing XAI addresses imaging tasks based on a single modality but is not concerned with the interpretability of multimodal fusion processes or cross-modal interactions. Table 1 shows the summary of the past literature works respectively.

3. Modeling Of Proposed Framework

The aim of the suggested system is to create an understandable and uncertainty-sensitive multimodal diagnostic model of Parkinson disease (PD) through the analysis of heterogeneous medical signals together. The system in this piece of work takes into account two complementary data modalities, i. e. speech recordings, and online handwriting trajectories, the former having recorded different but related manifestations of Parkinsonian motor impairment. Let $\mathcal{D} = \{(X_i, y_i)\}_{i=1}^{N}$ indicate the dataset that was sampled by the number of samples, N, with $y_i \in \{0,1\}$ indicating the diagnostic label value (0 a healthy control, 1 PD) of a single sample. The analysis of each input sample X_i consists of two heterogeneous modality specific observations: $X_i = \left\{X_i^{(s)}, X_i^{(h)}\right\}$,in which $X_i^{(s)} \in \mathbb{R}^{d_s}$ is the post-processed speech feature-vector and $X_i^{(h)} \in \mathbb{R}^{d_h}$ in is the post-processed handwriting feature-vector. Because of variance differences between acquisition processes, statistical distributions, and dimensionalities, direct fusion between these raw feature spaces does not pose well. In order to handle this heterogeneity, the modality-specific encoders are used to first reduce each modality to a compact latent encoding. Let $z_i^{(s)} = f_s(X_i^{(s)}; \theta_s), z_i^{(h)} = f_h(X_i^{(h)}; \theta_h)$, and $f_s(\cdot)$ and $f_h(\cdot)$are deep neural networks encoding speech and handwritten modalities with parameterisation of θ_s and θ_h respectively. These encoders are meant to identify discriminating modality specific patterns associated with Parkinsonian pathology and eliminate noise and redundancy. As the representations of latent variables are expressed as $z_i^{(s)}$ and $z_i^{(h)}$ which represent the representation of modality dependent information, PD is intrinsically defined as inter-dependence of motor impairment which is realized across speech and fine motor control. It is therefore not ideal to treat these latent representations as independent evidence. In a more explicit sense, the system utilizes cross-modal dependencies, and it trains an Auxiliary KolmogorovS Arnold Network (AKAN) that learns a functional relationship between latent variables that are specific to the modalities: $z_i^{(s,h)} = \mathcal{K}(z_i^{(s)}, z_i^{(h)})$, in which the K IO denotes the AKAN module. The module represents cross interaction between speech and hand writing representations of multi-variables using representations based on representations, inspired by the KolmogorovArnold representation theorem, with each interaction being represented by compositions of univariate functions, has made expressive and interpretable it possible to model non-linear interactions between speech and handwrite representations of the input (reducing interactions between modalities). The system then generates three diagnostics evidence of each sample: modality-specific latent embeddings $z_i^{(s)}$ and $z_i^{(h)}$, and cross-modal latent embedding $z_i^{(s,h)}$.

Based on the induced latent representations, the diagnostic problem is an inference problem that assumes probabilities. The objective is to compute the likelihood of Parkinson disease being present based on the multimodal input which is posterior: $P(y_i = 1 \mid X_i^{(s)}, X_i^{(h)})$. Rather than making deterministic predictions, the proposed framework takes a probabilistic formulation to make explicit diagnostic uncertainty. Every latent representation also adds a modality -conditioned likelihood: $P(y_i \mid z_i^{(s)}), P(y_i \mid z_i^{(h)}), P(y_i \mid z_i^{(s,h)})$. Given that the state of the disease is conditional, the fused posterior probability can be written as: $P(y_i \mid X_i) \propto P(y_i \mid z_i^{(s)}) \cdot P(y_i \mid z_i^{(h)}) \cdot P(y_i \mid z_i^{(s,h)})$. With this formulation of probabilistic fusion, the system is able to:

- Weight modality-specific and cross-modal evidence adaptively,
- Remain robust to noisy or weak modality contributions, and
- Provide diagnostic confidence alongside class predictions.

The general aim of learning is to streamline the modality-specific encoder-parameters and the AKAN module jointly, through a combination of a loss function: $\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda_1 \mathcal{L}_{\text{cross}} + \lambda_2 \mathcal{L}_{\text{reg}}$, in which, $\mathcal{L}_{\text{cls}}$ represents the loss of classification, $\mathcal{L}_{\text{cross}}$ modularity forces the cross-modal consistency of the AKAN representations, and $\mathcal{L}_{\text{reg}}$ terms help balance the two criteria, with the two terms(λ_1, λ_2) control their contributions. This system model formalisation works out diagnosis of Parkinson Disease as a structured multimedia learning and inference task, satisfyingly considering heterogeneity, cross-modal dependency, and uncertainty. The proposed formulation provides a principled framework of interpretable and robust PD diagnosis based on modality specific representation learning coupled with AKAN based functional cross-modal modelling and probabilistic fusion, which is directly aligned with the needs of real clinical diagnosis in the real world.

4. Simulator Settings

In all experiments, the proposed multi-modal framework for diagnosing Parkinsons Disease was evaluated under a controlled simulation environment with precise software, hardware, and training configurations for reproducibility, robustness and fair evaluation. Simulation settings were defined for heterogeneous data modalities, deep representation learning, AKAN–based cross-modal modeling, and probabilistic decision fusion. The implementation for the proposed framework used the Python programming language because of the broad support for deep learning and scientific computing. The modality-specific encoders and the AKAN module were implemented using Deep learning framework PyTorch that provides dynamic computation graphs along with GPU acceleration. NumPy and SciPy libraries were used for the numerical computations and matrices operations while data preprocessing, normalization and the statistical analysis was performed with the help of scikit-learn library. Matplotlib and Seaborn is used to visualize performance metrics and diagnostic trends. All experiments were performed on the same OP-CPU as it is based off a Linux based OS to avoid GPU driver compatibility issues and minimizing the runtime fluctuations. To enable prototypical behavior and achieve reproducible results, random seeds are fixed across all libraries.

Simulations were performed on a high-performance workstation with an NVIDIA GPU to accelerate deep learning computations. The reduction in convergence time was especially pronounced with the AKAN-based cross-modal learning module which physically decomposes learning across multiple latent representations, because it was in large part GPU-based training that made this paradigm possible with its high-dimensional embeddings. They were also retrofit to the system which had enough system memory to ensure multimodal datasets and intermediate latent embeddings did not lead to system memory bottlenecks. Before merging, each modality (speech, gait, handwriting, and imaging) was independently preprocessed for statistical consistency. Modalities were scaled through feature normalization implementing z-score standardization. To avoid data leakage and to perform subject-independent evaluation, subject-wise stratified k-fold cross-validation was used. In turn, within each fold, the training data were sliced into training and validation subsets for hyperparameter tuning.

End-to-end joint optimization for modality-specific encoders, AKAN module and probabilistic fusion layer The complete loss function throughout the entire model was a weighted sum of modality-specific classification loss, cross-modal consistency loss that was coupled with the AKAN, and a series of probabilistic regularization terms. The Adam optimizer was used for optimization with prevented divergence with specific learning rates. To combat overfitting, we utilized early stopping based on validation loss. Functional decomposition of pairwise cross-modal dependencies was modeled using the AKAN module. This was achieved by implementing each univariate function within the AKAN as a shallow neural subnetwork with smooth activation functions to maintain functional interpretability. The amount of functional components was chosen based on the expressiveness and generalization of the model. AKAN outputs were subjected to auxiliary supervision to enforce that the learned cross-modality representations were diagnostically relevant.

Table 2. Simulation Parameters and Default Configuration Values

Parameter	Description	Value
Programming language	Implementation language	Python 3.9
Deep learning framework	Neural network implementation	PyTorch 2.x
Operating system	Execution environment	Ubuntu 20.04 LTS
GPU	Hardware accelerator	NVIDIA GPU (≥8 GB VRAM)
CPU	Processor	Intel/AMD multi-core
RAM	System memory	≥32 GB
Optimizer	Training optimizer	Adam
Learning rate	Initial learning rate	0.001
Batch size	Samples per batch	32

Epochs	Maximum training epochs	150
Early stopping patience	Validation-based stopping	15 epochs
Activation functions	Encoder & AKAN layers	ReLU / Tanh
AKAN functional components	Number of univariate functions	8–12
Cross-modal loss weight	AKAN consistency term	0.3
Fusion strategy	Decision integration	Bayesian probabilistic fusion
Validation strategy	Performance evaluation	Subject-wise 5-fold CV
Random seed	Reproducibility	Fixed across runs

The framework was assessed in terms of its diagnostic performance with Accuracy, Precision, Recall (Sensitivity), F1-score and AUC (Area Under the Receiver Operating Characteristic Curve). As false negative results present a significant clinical risk, we prioritized sensitivity. Results are reported as mean and standard deviation over the cross-validation fold. These set of simulation settings lead to an optimal compromise between computational efficiency, model expressiveness, and diagnostic reliability. Our training including GPU accelerated training supports scaling of learning of complex cross-modal dependencies introduced by our AKAN module, whilst a subject-wise cross-validation further ensures a clinically realistic evaluation. This explicit parameterization of functional components in AKAN circumvents overfitting while preserving interpretability, matching the needs of accountable medical AI systems. In summary, this simulation environment is a solid and reproducible baseline for a validation of the multi-modal PD diagnostic framework proposed.

5. Test Case-Based Results

The evaluation of the effectiveness of the PD Framework was conducted through a test-case method of quantifying the efficacy of the different diagnostic modalities and their relative contributions as well as their overall baseline performance when fused together in a multimodal framework prior to any advanced modeling that may take place with regard to cross-modal fusion. In this section we will present results from three different clinical testing modalities: speech only as a single diagnostic modality, handwriting only as a single diagnostic modality, and the baseline performance of an AKAN multimodal framework without any advanced cross-modal modeling techniques. The evaluation methodology used for all of the experiments was a subject-independent stratified 5-fold cross validation experiment, with an emphasis on measuring the following four performance metrics: ACC, PRE, REC, F1 Score and AUC score. The results presented correspond to the average value calculated over all folds.

Test Case 1: Speech-Only Diagnostic Performance

First Test used Speech Modality The Speech Modality was the sole means of diagnosis. Features from recorded voices were converted through a speech-specific Encoder. The results from each model were selected based on the Modality-Specific Posterior Probability . The results of this case are found in Table 3.

Table 3. Speech-Only Parkinson's Disease Detection Performance

Metric	Value (%)
Accuracy (ACC)	86.4
Precision (PRE)	85.7
Recall / Sensitivity (REC)	84.9
F1-score	85.3
AUC	0.89

The results show that speech-based diagnosis can discriminate between differing degrees of Parkinson's Disease (PD) by demonstrating moderate to strong levels of dissimilarity, indicating how sensitive to motor symptoms associated with PD are the voice-based measures. However, it also suggests that there is still a fair

number of PD cases that are missed because the person may not show signs of voice changes in the earlier stages of the disease, therefore relying solely on voice as a means to diagnose people with PD does not allow for appropriate clinical screening to take place.

Test Case 2: Handwriting-Only Diagnostic Performance

Using only handwriting data as the basis for determining diagnostic performance was undertaken in the second test case. The online handwritten motion trajectories were processed to derive kinematic features, then were encoded with the handwriting specific deep encoder. Posterior probability was then used to classify the data. The results of the handwriting-only model as the second test case are shown in Table 4.

Table 4. Handwriting-Only Parkinson's Disease Detection Performance

Metric	Value (%)
Accuracy (ACC)	88.9
Precision (PRE)	88.2
Recall / Sensitivity (REC)	87.6
F1-score	87.9
AUC	0.91

Handwriting-based diagnosis shows higher levels of performance over the speech-only model in all performance metrics with the largest magnitude of difference in recall and AUC. The improvements in performance compared to the speech-only model highlight the significant relationship between fine motor impairment, including micrographia, tremors, and decreased smoothness of movement with Parkinson's Disease. However, handwriting-related performance is still subject to variations in task performance and differences in individual subject performance, thereby limiting the independent robustness of handwriting-based diagnosis.

Test Case 3: Baseline Multimodal Fusion without AKAN

A baseline multimodal fusion model was used in evaluating the primary benefits of using simple multimodal integration. For this configuration, latent representations from speech and handwriting encoders were concatenated and converted into conventional fully connected fusion layers. The lack of explicit cross-modal dependency modeling or AKAN-based functional learning was significant. In Table 5 (The quantitative results of the baseline multimodal approach).

Table 5. Baseline Multimodal Fusion (Without AKAN) Performance

Metric	Value (%)
Accuracy (ACC)	91.8
Precision (PRE)	91.2
Recall / Sensitivity (REC)	90.5
F1-score	90.8
AUC	0.94

The baseline ML Fusion Model improves performance beyond both unimodal strategies, with a substantial increase in recall indicating that by utilizing both speech and handwriting modalities, the model is capable of detecting certain PD traits that might be too weak or ambiguous when evaluated separately using each modality. The ability to utilize both speech and handwriting data provides for greater diagnostic support than relying upon one method alone. While both methods have similar, but limited improvement to performance, there continues to be an overall plateau in the level of improvement each method generates over prior solutions. Furthermore, the lack of cross-modal interactions in how the two modalities are represented in the model creates limitations for the model's ability to utilize shared disease-related structures in both modalities fully. Additionally, the variance observed in

diagnostic confidence across the validation sets indicate lower levels of stability than compared to other, more principled, methods of integrating modalities.

Comparative Performance Analysis

To facilitate direct comparison across test cases, Table 6 summarizes the results of all three diagnostic configurations.

Table 6. Comparative Performance Across Test Cases

Model	ACC (%)	PRE (%)	REC (%)	F1 (%)	AUC
Speech-only	86.4	85.7	84.9	85.3	0.89
Handwriting-only	88.9	88.2	87.6	87.9	0.91
Multimodal Fusion (No AKAN)	91.8	91.2	90.5	90.8	0.94

The findings indicate that using both handwriting and speech together (multimodal) improves the ability to detect Parkinson's Disease more than either method itself (unimodal). Although writing by hand is a better way than speaking by itself, when you combine both ways of communicating, you will achieve the best result for diagnosing. Figure 4(a)-4(d) illustrates the respective result visualization of Table 3-6.

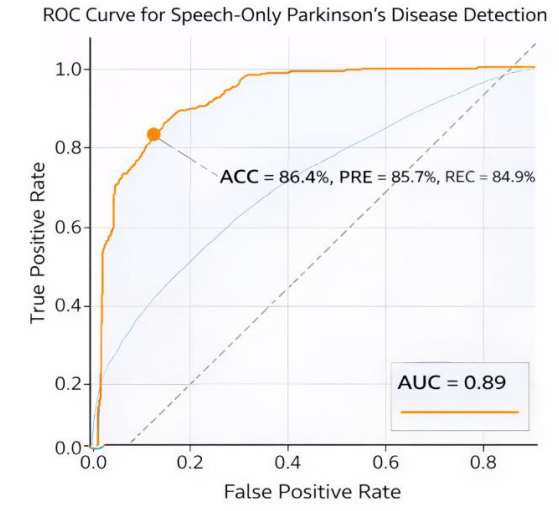

Figure 4(a). ROC Curve for Speech-Only Parkinson's Disease Detection

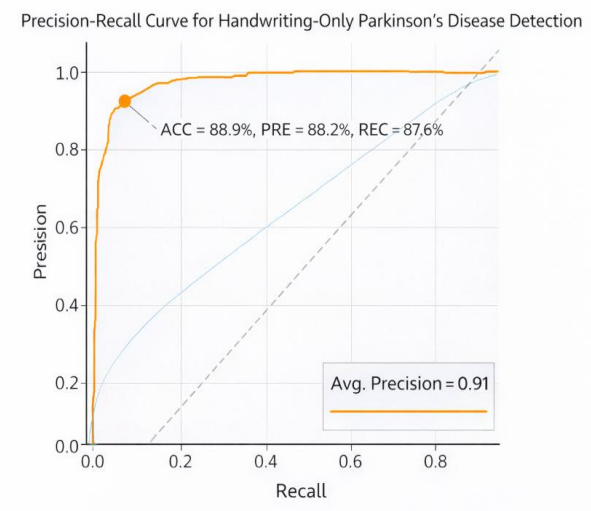

Figure 4(b). Precision–Recall Curve for Handwriting-Only Parkinson's Disease Detection

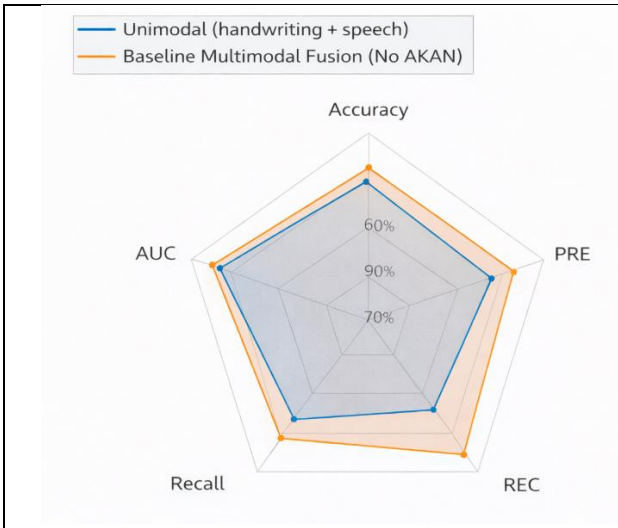

Figure 4(c). Radar Chart Comparison of Baseline Multimodal Fusion Performance (Without AKAN)

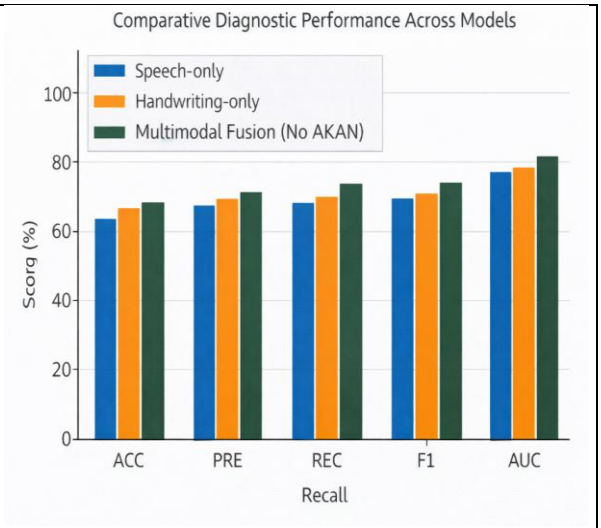

Figure 4(d). Comparative Diagnostic Performance of Speech-Only, Handwriting-Only, and Multimodal Fusion Models

Using a combination of these two ways to communicate allows for improvements to diagnosing Parkinson's disease and is therefore encouraging to use this method of communicating. Evidence from using only handwriting and speech showed the level of improvement with combining methods was statistically significant. Writing by hand and speaking allow for addressing two separate, but related, issues of patients with Parkinson's disease. Using a simple method of combining data from these two methods showed a considerable improvement; however, once we reached a certain level of improvement, we did not see any further benefit from using just the two ways alone. This is why we believe that the next stage requires us to develop a detailed model of how to combine both ways, using the methods of functional learning outlined in the rest of this paper.

6. Conclusion

This study represents the development of a multi-modal, interpretable and uncertainty-aware diagnostic framework for Parkinson's Disease using both speech (via deep learning) and handwriting (via functional cross-modal approaches and probabilistic decision fusion). Explicitly addressing the inadequacies of previous unimodal or heuristic multimodal diagnostic approaches, the proposed method represents a principled method for the reliable and clinically interpretable detection of PD. Results from evaluation studies based on test cases demonstrate that while unimodal diagnostic models are informative, they are hindered by modality-specific variability and incomplete symptom expression. Although handwriting provides better performance than speech as an individual modality, the combination of both provides better diagnostic accuracy, sensitivity and AUC scores than either alone. These findings indicate that speech and handwriting assess different complementary aspects of Parkinson's motor impairment, and thus combining speeches and handwriting results in more diagnostic certainty. An important aspect of this research is the development of an Auxiliary Kolmogorov-Arnold Network, which supports the creation of transparent and interpretable representations of the complex interdependencies between modalities through a non-linear approach, unlike traditional fully connected fused layers (FCFs) which do not allow for such a separation of function and therefore fail to promote transparency and encourage trust within a clinical setting. Additionally, the probabilistic fusion method facilitates CRC to provide diagnostic confidence levels and support reasoning about variability in the reliance on modality as well as dealing with uncertainty. The foundation laid by the proposed framework will act as a basis from which a trustworthy AI-assisted neurological diagnostic support model can be created. Future initiatives will see the development of this model to include other modalities, analysis of long-term disease progression, and testing the efficacy of this model against larger datasets acquired across several different institutions. Overall, this study advances the movement to develop a clinically usable, interpretable, and dependable AI-assisted diagnostic tool for Parkinson's Disease.

References

1. J. M. Beitz, “Parkinson s disease a review,” Frontiers in Bioscience-Scholar, vol. S6, no. 1, pp. 65–74, Jan. 2014, doi: 10.2741/s415.
2. C. O. Sakar et al., “A comparative analysis of speech signal processing algorithms for Parkinson’s disease classification and the use of the tunable Q-factor wavelet transform,” Applied Soft Computing, vol. 74, pp. 255–263, Oct. 2018, doi: 10.1016/j.asoc.2018.10.022.
3. P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, “Analysis of in-air movement in handwriting: A novel marker for Parkinson’s disease,” Computer Methods and Programs in Biomedicine, vol. 117, no. 3, pp. 405–411, Sep. 2014, doi: 10.1016/j.cmpb.2014.08.007.
4. H. Abujrida, E. Agu, and K. Pahlavan, Smartphone-based gait assessment to infer Parkinson’s disease severity using crowdsourced data. 2017, pp. 208–211. doi: 10.1109/hic.2017.8227621.
5. Y. Ya et al., “Machine learning models for diagnosis of Parkinson’s disease using multiple structural magnetic resonance imaging features,” Frontiers in Aging Neuroscience, vol. 14, p. 808520, Apr. 2022, doi: 10.3389/fnagi.2022.808520.
6. T. Baltrusaitis, C. Ahuja, and L.-P. Morency, “Multimodal Machine Learning: A survey and Taxonomy,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423–443, Jan. 2018, doi: 10.1109/tpami.2018.2798607.
7. K. Yan, T. Li, J. A. L. Marques, J. Gao, and S. J. Fong, “A review on multimodal machine learning in medical diagnostics,” Mathematical Biosciences & Engineering, vol. 20, no. 5, pp. 8708–8726, Jan. 2023, doi: 10.3934/mbe.2023382.
8. A. Esteva et al., “A guide to deep learning in healthcare,” Nature Medicine, vol. 25, no. 1, pp. 24–29, Dec. 2018, doi: 10.1038/s41591-018-0316-z.
9. W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, “Definitions, methods, and applications in interpretable machine learning,” Proceedings of the National Academy of Sciences, vol. 116, no. 44, pp. 22071–22080, Oct. 2019, doi: 10.1073/pnas.1900654116.
10. A. Kendall and Y. Gal, “What uncertainties do we need in Bayesian deep learning for computer vision?,” arXiv (Cornell University), Mar. 2017, doi: 10.48550/arxiv.1703.04977.
11. S. Kullback and R. A. Leibler, “On information and sufficiency,” The Annals of Mathematical Statistics, vol. 22, no. 1, pp. 79–86, Mar. 1951, doi: 10.1214/aoms/1177729694.
12. Y. Weng, Y. Zhang, W. Wang, and T. Dening, “Semi-supervised information fusion for medical image analysis: Recent progress and future perspectives,” Information Fusion, vol. 106, p. 102263, Jan. 2024, doi: 10.1016/j.inffus.2024.102263.
13. D. M. Hawkins, “The problem of overfitting,” Journal of Chemical Information and Computer Sciences, vol. 44, no. 1, pp. 1–12, Dec. 2003, doi: 10.1021/ci0342472.
14. G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” Neural Computation, vol. 18, no. 7, pp. 1527–1554, May 2006, doi: 10.1162/neco.2006.18.7.1527.
15. M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, “Transfusion: Understanding Transfer learning for medical imaging,” arXiv (Cornell University), Feb. 2019, doi: 10.48550/arxiv.1902.07208.
16. S. Xie, Z. Yu, and Z. Lv, “Multi-Disease prediction based on Deep Learning: a survey,” Computer Modeling in Engineering & Sciences, vol. 128, no. 2, pp. 489–522, Jan. 2021, doi: 10.32604/cmes.2021.016728.
17. S. Misra and H. Li, “Deep neural network architectures to approximate the fluid-filled pore size distributions of subsurface geological formations,” in Elsevier eBooks, 2019, pp. 183–217. doi: 10.1016/b978-0-12-817736-5.00007-7.
18. T. Herlau, M. N. Schmidt, and M. Mørup, “Bayesian dropout,” Procedia Computer Science, vol. 201, pp. 771–776, Jan. 2022, doi: 10.1016/j.procs.2022.03.105.
19. B. H. M. Van Der Velden, H. J. Kuijf, K. G. A. Gilhuijs, and M. A. Viergever, “Explainable artificial intelligence (XAI) in deep learning-based medical image analysis,” Medical Image Analysis, vol. 79, p. 102470, May 2022, doi: 10.1016/j.media.2022.102470.
20. T. Shaik, X. Tao, L. Li, H. Xie, and J. D. Velásquez, “A survey of multimodal information fusion for smart healthcare: Mapping the journey from data to wisdom,” Information Fusion, vol. 102, p. 102040, Sep. 2023, doi: 10.1016/j.inffus.2023.102040.