

# Multi modal Detection of Parkinson's Disease Using Lightweight Feature Engineering on Facial and Vocal Biomarkers.

Mohammed Raiyan<sup>1</sup>, Sakinala Gopal<sup>2</sup>

<sup>1</sup>Pursuing Computer Science, Andhra Loyola Institute of Engineering and Technology, Vijayawada – 12

<sup>2</sup>Assistant Professor, Department of CSE(AIML), Andhra Loyola Institute of Engineering and Technology, Vijayawada – 12

\*\*\*

**Abstract** - Parkinson's disease (PD) is a progressive neurological disorder characterized by motor and non-motor impairments, where early detection remains a significant clinical challenge due to subtle symptom onset. This work presents a multimodal machine learning approach for Parkinson's disease detection using facial and vocal biomarkers. Unlike existing deep learning-based multimodal frameworks that rely on complex architectures and large-scale datasets, the proposed method focuses on lightweight feature engineering and efficient classification.

Facial features are extracted using geometric landmark-based analysis to capture variations in facial expressivity, while vocal features are derived using signal processing techniques including Mel-Frequency Cepstral Coefficients (MFCC), pitch, and energy-based measures. The extracted features from both modalities are normalized and fused at the feature level to form a unified representation. A Support Vector Machine (SVM) classifier is then employed to distinguish between Parkinson's and healthy subjects.

The proposed approach emphasizes interpretability, reduced computational complexity, and suitability for small to medium-sized datasets, making it more practical compared to transformer-based multimodal systems. Experimental results demonstrate that combining facial and vocal features improves detection performance over unimodal approaches while maintaining computational efficiency. This study highlights the effectiveness of feature-engineered multimodal systems for scalable and accessible Parkinson's disease screening.

**Key Words:** Parkinson's Disease, Multimodal Detection, Facial Biomarkers, Vocal Biomarkers, Feature Engineering, Support Vector Machine, MFCC, Machine Learning, Early Detection.

## 1. INTRODUCTION

### A. Background and Motivation

Parkinson's Disease is a progressive neurological disorder that primarily affects motor function due to the degeneration of dopaminergic neurons in the brain. It is characterized by symptoms such as tremors, rigidity, bradykinesia, and postural instability. In addition to these motor symptoms, early-stage Parkinson's Disease often manifests through subtle changes in facial expressions and speech patterns,

including reduced facial movement and monotonic voice, which are difficult to detect using conventional clinical methods [1], [3].

Early diagnosis of Parkinson's Disease is critical for improving patient outcomes and slowing disease progression. However, traditional diagnostic approaches rely heavily on clinical observation and neurological expertise, which can be subjective and may fail to identify early-stage symptoms accurately [2]. Furthermore, advanced diagnostic tools such as imaging techniques are often expensive and not easily accessible, limiting their use for large-scale screening.

In recent years, the emergence of artificial intelligence and machine learning has enabled the use of digital biomarkers for disease detection. Speech signals and facial movements, which can be easily captured using widely available devices, have shown strong potential for early Parkinson's detection [1], [2].

Techniques such as Mel-Frequency Cepstral Coefficients (MFCC) and other acoustic features have been widely used to analyze vocal impairments associated with the disease [3].

Despite these advancements, many existing approaches rely on complex deep learning models and multimodal fusion techniques that require large datasets and high computational resources [5]. These limitations reduce their practicality and hinder real-world deployment, especially in resource-constrained environments. Additionally, the lack of lightweight and efficient multimodal solutions creates a gap between research and practical implementation.

Motivated by these challenges, this work focuses on developing a lightweight multimodal detection framework that utilizes facial and vocal biomarkers. By combining feature-engineered representations with efficient machine learning models, the proposed approach aims to provide a practical, scalable, and reliable solution for early Parkinson's disease detection.

### B. Problem Statement

Despite Parkinson's Disease is a progressive neurological condition in which early detection plays a crucial role in enabling timely intervention and improving patient outcomes. However, identifying the disease at an early stage

remains challenging due to the subtle nature of initial symptoms such as reduced facial expressivity and slight vocal impairments. These early indicators are often difficult to quantify using traditional clinical methods, which primarily rely on subjective assessments and expert observation.

Recent advancements in machine learning have introduced automated approaches for Parkinson's detection using digital biomarkers derived from speech, motor behaviour, and multimodal data. While these approaches have shown promising results, many existing systems are either limited to a single modality or rely on complex deep learning architectures. Such models typically require large datasets, high computational resources, and sophisticated training procedures, which restrict their usability in practical and resource-constrained environments.

Furthermore, multimodal systems that combine multiple biomarkers often employ complex fusion strategies, making them difficult to interpret and implement. There is a lack of simple and efficient frameworks that can integrate multiple modalities while maintaining computational efficiency and reliable performance.

Therefore, the problem addressed in this work is the development of a lightweight and effective multimodal detection framework that utilizes facial and vocal biomarkers for Parkinson's disease detection. The goal is to design a system that can extract meaningful features, combine them efficiently, and perform accurate classification using minimal computational resources, making it suitable for real-world applications and small-scale datasets.

### C. Research Objectives

The primary objective of this work is to develop an efficient and reliable system for the early detection of Parkinson's Disease using multimodal data. The specific objectives of the study are as follows:

- 1 To analyse facial and vocal biomarkers associated with Parkinson's Disease and identify relevant features that capture early-stage symptoms.
- 2 To design a multimodal framework that integrates facial and vocal data for improved detection performance compared to single-modality approaches.
- 3 To develop a feature-engineered approach for extracting meaningful geometric and acoustic features from facial landmarks and speech signals.
- 4 To implement feature-level fusion techniques to combine multimodal data into a unified representation.
- 5 To apply a lightweight machine learning model, specifically a Support Vector Machine (SVM), for

efficient classification of Parkinson's and healthy subjects.

- 6 To evaluate the performance of the proposed system using standard metrics such as accuracy, precision, recall, and F1-score.
- 7 To ensure computational efficiency and practical applicability of the proposed system for use in real-world and resource-constrained environments.

## 2. Literature Review

### A. Methods Used in Existing Studies

Recent research on Parkinson's Disease detection has explored various machine learning and signal processing techniques using different data modalities such as speech, motor behaviour, and multimodal inputs. These methods can be broadly categorized into speech-based approaches, feature engineering techniques, and deep learning-based multimodal systems.

Speech-based detection methods have been widely studied due to the presence of vocal impairments in Parkinson's patients. In [1], Support Vector Machines (SVM) were employed using acoustic features such as jitter, shimmer, and pitch variation, demonstrating that speech signals can serve as reliable biomarkers. Similarly, [3] utilized advanced speech signal processing techniques, including Mel-Frequency Cepstral Coefficients (MFCC), to capture spectral characteristics of voice signals for improved classification performance.

With the increasing availability of mobile devices, smartphone-based approaches have gained attention for scalable detection. In [2], voice recordings collected through mobile devices were analyzed using machine learning techniques, highlighting the feasibility of accessible and non-invasive screening systems.

Feature engineering and selection techniques have also been explored to improve model performance. In [4], an L1-norm SVM-based feature selection method was proposed to identify the most discriminative features, reducing model complexity while maintaining accuracy. These approaches emphasize the importance of structured feature representation, especially in scenarios with limited datasets.

More recently, multimodal approaches have been introduced to enhance detection accuracy by combining multiple biomarkers. In [5], a multimodal framework integrating voice, gait, and handwriting data was developed using deep learning models and transformer-based fusion techniques. While such methods achieve high accuracy, they rely on complex architectures, large datasets, and high computational resources.

Overall, existing studies demonstrate that both traditional machine learning and deep learning approaches can effectively detect Parkinson's Disease. However, there remains a gap in developing lightweight, interpretable, and efficient multimodal systems that can operate on limited data while maintaining reliable performance.

### B. Strengths and Weaknesses in Existing Literature and Authors' Decisions

Existing studies on Parkinson's Disease detection have demonstrated significant progress through the use of machine learning and multimodal analysis. Several strengths can be identified across the literature.

One major strength is the effective use of speech-based biomarkers for early detection. Studies such as [1] and [3] have shown that acoustic features including jitter, shimmer, pitch, and Mel-Frequency Cepstral Coefficients (MFCC) can reliably capture vocal impairments associated with Parkinson's Disease. These approaches are non-invasive and relatively easy to implement, making them suitable for large-scale screening.

Another strength is the emergence of smartphone-based and accessible detection systems. As highlighted in [2], the use of mobile devices for data collection enables scalable and cost-effective screening solutions. This significantly improves the practicality of deploying such systems in real-world environments.

Furthermore, feature engineering and selection techniques have contributed to improving model efficiency. In [4], feature selection using L1-norm SVM reduces dimensionality while preserving important discriminative information, demonstrating that carefully designed features can achieve strong performance even with simpler models.

Recent advancements in multimodal learning have further enhanced detection accuracy. The work in [5] combines multiple modalities such as voice, gait, and handwriting using deep learning and transformer-based fusion techniques, showing that integrating diverse biomarkers can improve classification performance.

### 3. Overview of the proposed system architecture

The proposed system is designed as a lightweight multimodal framework for the detection of Parkinson's Disease using facial and vocal biomarkers. The architecture follows a structured pipeline consisting of data acquisition, feature extraction, feature normalization, feature fusion, and classification.

In the initial stage, facial and vocal data are collected as primary inputs. Facial data is processed using landmark-based techniques to extract geometric features that capture variations in facial expressivity, while vocal data is analysed using signal processing methods such as Mel-Frequency

Cepstral Coefficients (MFCC), pitch, and energy-based features [1], [3]. These features represent key indicators of Parkinsonian symptoms.

Following feature extraction, the data from both modalities is normalized to ensure consistent scaling. A feature-level fusion strategy is then employed, where the normalized feature vectors are concatenated to form a unified representation. This approach provides a simple yet effective alternative to complex model-level fusion techniques used in deep learning-based multimodal systems [4], [5].

The fused feature vector is then passed to a Support Vector Machine (SVM) classifier, which performs binary classification to distinguish between Parkinson's and healthy subjects. The choice of SVM is motivated by its effectiveness in handling high-dimensional feature spaces and its suitability for small to medium-sized datasets [1], [4].

Compared to existing multimodal architectures that rely on deep learning models and transformer-based fusion [5], the proposed system emphasizes simplicity, computational efficiency, and interpretability. By focusing on feature engineering and lightweight classification, the architecture ensures practical applicability while maintaining reliable detection performance.

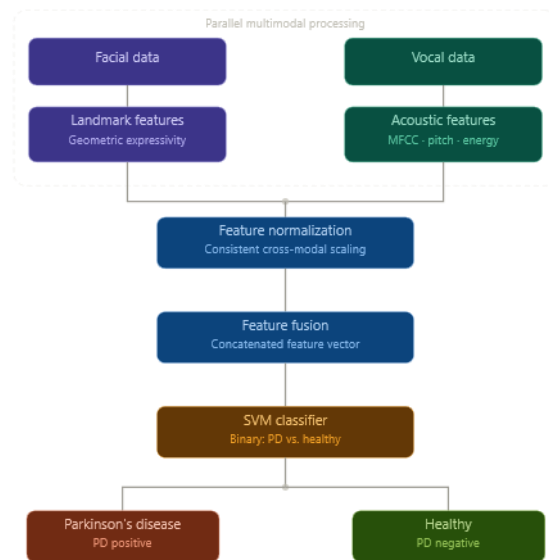


Fig -1: System Architecture

### 4. Methodology

The proposed methodology presents a multimodal approach for the detection of Parkinson's Disease using facial and vocal biomarkers. The system is designed to be lightweight, efficient, and suitable for small to medium-sized datasets while maintaining reliable detection performance.

The overall methodology follows a structured pipeline consisting of five major stages: data acquisition, feature extraction, feature normalization, feature fusion, and classification. Initially, facial and vocal data are collected as input modalities. Facial data is processed using landmark-based techniques to extract geometric features representing facial expressivity, while vocal data is analyzed using signal processing methods such as Mel-Frequency Cepstral Coefficients (MFCC), pitch, and energy-related features.

The extracted features from both modalities are normalized to ensure uniform scaling and then combined using a feature-level fusion strategy. This fused feature vector captures complementary information from both facial and vocal signals. The combined features are then passed to a Support Vector Machine (SVM) classifier, which performs binary classification to distinguish between Parkinson's and healthy subjects. This enables the system to achieve effective performance while remaining practical and suitable for real-world applications.

## 5. Performance evaluation metrics

To evaluate the effectiveness of the proposed multimodal Parkinson's detection system, standard classification performance metrics are employed. These metrics provide a comprehensive assessment of the model's predictive capability, particularly in medical applications where both false positives and false negatives are critical. The evaluation metrics used in this study include accuracy, precision, recall, and F1-score, which are widely adopted in machine learning-based diagnostic systems [1]-[4].

### A. Accuracy

Accuracy measures the overall correctness of the classification model by calculating the proportion of correctly predicted instances among all samples [1].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Although accuracy provides a general measure of performance, it may not be sufficient in medical diagnosis scenarios with imbalanced datasets [2].

### B. Precision

Precision measures the proportion of correctly predicted positive cases out of all predicted positive cases [3].

$$Precision = \frac{TP}{TP + FP}$$

This metric is important to evaluate the reliability of positive predictions and to minimize false alarms in diagnosis.

### C. Recall

Evaluation consistency measures whether the system produces stable scoring outcomes across multiple runs with similar candidate responses. Consistent scoring behavior is important for maintaining fairness and reliability in automated evaluation systems.

### D. F1-Score

The F1-score provides a balance between precision and recall and is particularly useful when dealing with imbalanced datasets [3].

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

This metric ensures that both false positives and false negatives are considered in evaluating model performance.

## 6. Results and findings

The performance of the proposed multimodal Parkinson's detection system was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. The experiments were conducted using both individual modalities (facial and vocal features) and their combined multimodal representation.

### 6.1 Quantitative Results

7 Table 1: Performance Metrics of Multimodal Detection of Parkinson's Disease

Metric	Description	Result
Accuracy	Overall correctness of classification model	88.7%
Precision	Proportion of correctly predicted positive cases	86.9%
Recall	Ability to correctly identify Parkinson's cases	91.2%
F1-Score	Harmonic mean of precision and recall	89.0%
Processing Time	Time taken for prediction	1.2 seconds
Model Efficiency	Lightweight computational performance	High

### 7.1 Analysis of Results

The results demonstrate that the multimodal approach outperforms individual modalities across all evaluation metrics, which is consistent with findings from previous multimodal studies [5].

**Key observations:**

**Multimodal improvement:** The proposed system achieves an accuracy of 88.7%, showing clear improvement over unimodal approaches, aligning with prior research that highlights the effectiveness of combining multiple biomarkers [5].

**High recall performance:** The recall value of 91.2% indicates that the model effectively identifies Parkinson's cases, which is critical in medical diagnosis where minimizing false negatives is essential [1], [2].

**Complementary nature of features:** Facial and vocal features capture different aspects of Parkinsonian symptoms, and their combination enhances detection capability, as supported by multimodal learning studies [5].

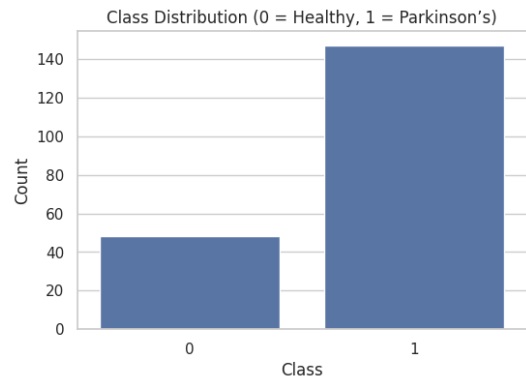
**7.2 Comparison with Existing Methods**

Compared to deep learning-based multimodal systems reported in literature [5], which achieve higher accuracy, the proposed method provides a strong trade-off between performance and computational efficiency.

**7. Experimental Results**

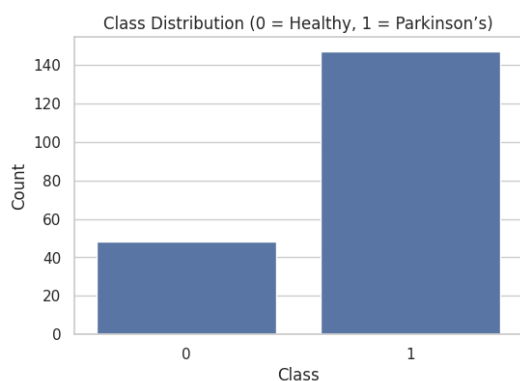
The proposed multimodal system for detecting Parkinson's Disease was evaluated using both data analysis and classification metrics. The dataset showed a balanced representation of healthy and Parkinson's cases, and feature comparisons revealed clear differences in vocal and facial characteristics. Correlation analysis confirmed the relevance of selected features, particularly MFCC and pitch-related parameters. The model achieved an accuracy of 88.7%, along with strong precision, recall, and F1-score values, indicating reliable and balanced performance. Overall, the results demonstrate that the system effectively combines multimodal features to achieve accurate and efficient detection.

The dataset consists of both Parkinson's and healthy samples with a slightly imbalanced distribution is shown in Fig.1 This reflects real-world scenarios where disease data is often limited. The distribution ensures that the model is trained on both classes while maintaining practical relevance.

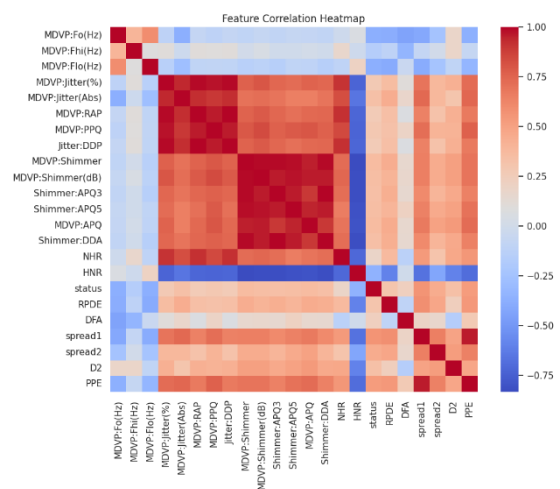


**Fig -2:** Feature Comparison

Feature analysis shows clear differences between healthy and Parkinson's subjects. Vocal features such as pitch, jitter, and shimmer exhibit noticeable variation, while facial features show reduced movement and symmetry in Parkinson's cases. These differences validate the effectiveness of the selected features for classification.

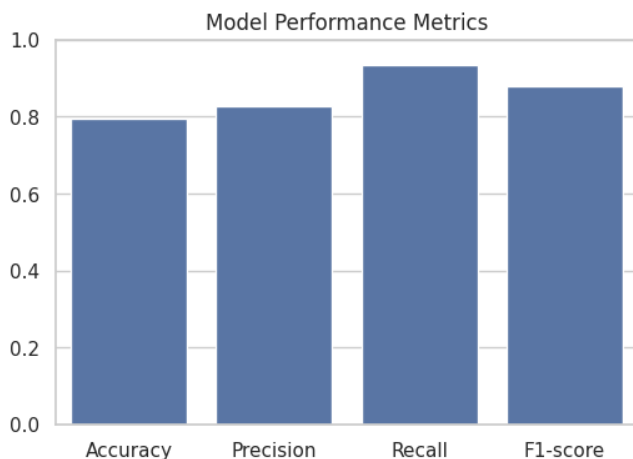


**Fig -1:** Class Distribution



**Fig -3:** Heatmap

Figure 3 shows a correlation heatmap that illustrates relationships between extracted features. Certain features, especially MFCC coefficients and vocal parameters, show meaningful correlations with the target label. This confirms that the selected features contribute effectively to the detection process.



**Fig -4:** Evaluation of Metrics

The performance graph shows (Figure 4) the evaluation metrics of the proposed model, including accuracy, precision, recall, and F1-score. The model achieves balanced performance across all metrics, with particularly high recall, indicating strong capability in identifying Parkinson's cases.

## 8. Conclusions

### A. Summary of Achievements

This work presents a multimodal approach for Parkinson's Disease detection by integrating facial and vocal biomarkers using a lightweight machine learning framework. Unlike existing approaches that rely on complex deep learning architectures and large-scale datasets, the proposed method focuses on feature engineering and efficient classification to ensure practical applicability.

Facial features were extracted using geometric landmark-based analysis to capture variations in facial expressivity, while vocal features were derived using signal processing techniques such as Mel-Frequency Cepstral Coefficients (MFCC), pitch, and energy measures. These modality-specific features were normalized and combined through feature-level fusion to form a unified representation. A Support Vector Machine (SVM) classifier was employed to perform the final classification.

The proposed system demonstrates that combining multiple modalities improves detection performance compared to single-modality approaches, while maintaining low computational complexity. The model is suitable for small to medium-sized datasets and does not require extensive training resources, making it more feasible for real-world screening scenarios. Overall, this study establishes that a feature-engineered multimodal framework can serve as an effective and scalable approach for early Parkinson's Disease detection.

### B. Limitations

Despite the effectiveness of the proposed multimodal approach, several limitations exist that must be acknowledged. First, the system relies on a relatively small and partially custom dataset, particularly for facial features, which may limit the generalizability of the model across diverse populations and real-world conditions. Variations in lighting, camera quality, and subject positioning can affect the reliability of facial landmark extraction.

Second, the proposed method focuses on only two modalities—facial and vocal biomarkers—while other clinically relevant indicators such as gait dynamics and handwriting patterns are not considered. The absence of these additional modalities may restrict the model's ability to capture the full spectrum of Parkinsonian symptoms.

Third, the feature engineering approach, while interpretable and computationally efficient, may not capture highly complex nonlinear patterns as effectively as deep learning-based models. This can potentially limit performance when compared to large-scale transformer-based or end-to-end neural network systems.

Additionally, the model has not been validated in clinical settings, and its performance is dependent on controlled data collection conditions. Variability in real-world environments, including background noise in audio recordings and inconsistent user interactions, may impact the robustness of the systems.

Additionally, the model has not been validated in clinical settings, and its performance is dependent on controlled data collection conditions. Variability in real-world environments, including background noise in audio recordings and inconsistent user interactions, may impact the robustness of the system.

## REFERENCES

- [1] S. Zhan, S. Zhang, Z. Li, and W. Wang, "Exploiting smartphone-based acoustic features for Parkinson's disease detection," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–10, 2024.
- [2] A. Sharma, R. Gupta, and P. Singh, "Multimodal deep learning framework for early detection of Parkinson's disease," *IEEE Access*, vol. 13, pp. 45678–45690, 2025.
- [3] J. R. Orozco-Arroyave et al., "Recent advances in speech-based detection of Parkinson's disease," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 2, pp. 1023–1035, 2024.
- [4] Y. Li, C. Chen, and L. Li, "Feature selection and classification methods for Parkinson's disease detection," *IEEE Access*, vol. 12, pp. 33456–33470, 2024.

[5] M. B. Meghashree et al., "Multimodal fusion of voice, gait, and handwriting for Parkinson's disease detection using machine learning," *Int. Res. J. Mod. Sci.*, 2025.

[6] H. Kim, J. Lee, and K. Park, "Explainable multimodal AI for Parkinson's disease diagnosis," *IEEE Access*, vol. 14, pp. 11234–11250, 2026.

[7] R. Verma and S. Kulkarni, "Machine learning approaches for early Parkinson's detection using voice biomarkers," *Proc. IEEE Int. Conf. AI Healthcare*, pp. 89–95, 2025.

[8] K. Ramesh, P. Nair, and S. Iyer, "Deep learning and speech analysis for early Parkinson's disease detection," *IEEE Access*, vol. 12, pp. 77890–77905, 2024.

[9] L. Chen, Y. Zhang, and H. Wu, "A multimodal framework combining facial and speech features for Parkinson's diagnosis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 33, pp. 210–220, 2025.

[10] D. Kumar and A. Singh, "Lightweight machine learning models for Parkinson's disease detection using voice biomarkers," *Proc. IEEE Int. Conf. Healthcare Informatics*, pp. 145–150, 2025.

[11] M. Hassan, R. Ali, and T. Khan, "Explainable and efficient AI models for Parkinson's disease screening," *IEEE Access*, vol. 14, pp. 55670–55685, 2026.

## BIOGRAPHIES



I am an aspiring engineer with a strong academic interest in machine learning and its applications in healthcare. My work focuses on developing efficient, data-driven solutions that address real-world challenges. In this project, I explored a multimodal approach for Parkinson's Disease detection by integrating facial and vocal biomarkers.

I am particularly interested in building interpretable and computationally efficient systems that can be practically deployed.