

## DIFFUSION-BASED GENERATIVE MODELING FOR PARKINSON'S BIOMARKER DETECTION

<sup>1</sup>Dr. G. Srinivasa Rao, <sup>2</sup>Gutti Venkata Suresh, <sup>3</sup>Bhattiprolu Vamsikrishna, <sup>4</sup>Bodapati Mohankrishna,  
<sup>5</sup>Cherukuri Chandu

<sup>1</sup>Associate Professor, Dept ELECTRONICS AND COMMUNICATION ENGINEERING, St. Ann's  
College of Engineering and Technology, Nayunipalli (V), Vetapalem (M), Chirala, Bapatla Dist,  
Andhra Pradesh – 523187, India

<sup>2,3,4,5</sup>U. G Student, Dept ELECTRONICS AND COMMUNICATION ENGINEERING, St. Ann's  
College of Engineering and Technology, Nayunipalli (V), Vetapalem (M), Chirala, Bapatla Dist,  
Andhra Pradesh – 523187, India

### ABSTRACT

*Parkinson's disease (PD) is a progressive neurological disorder that affects motor and non-motor functions, making early diagnosis critical for effective treatment. Biomarker-based detection has gained significant attention for improving diagnostic accuracy. However, limited and imbalanced biomedical datasets restrict the performance of traditional machine learning models. Diffusion-based generative models have emerged as powerful tools for data augmentation and representation learning. This project proposes a diffusion-based generative modeling framework to enhance Parkinson's biomarker detection. The generative model synthesizes realistic biomarker data to address data scarcity. Extracted features are classified using Random Forest (RF) and Support Vector Machine (SVM) algorithms. The hybrid approach improves robustness and generalization. The system evaluates performance using accuracy, precision,*

*recall, and F1-score. Experimental results demonstrate improved detection accuracy compared to conventional methods. The proposed framework effectively captures complex biomarker distributions. It supports early-stage diagnosis of Parkinson's disease. The approach is scalable and adaptable to various biomedical datasets. The results validate the effectiveness of combining generative models with classical classifiers.*

### INTRODUCTION

Parkinson's disease is the second most common neurodegenerative disorder worldwide. Early diagnosis is challenging due to subtle and overlapping symptoms. Biomarkers such as voice signals, gait patterns, and neuroimaging data are widely used for PD detection. Traditional diagnostic methods rely heavily on clinical expertise. Machine learning techniques offer automated and objective analysis of biomarkers. However, these methods often suffer from limited training data. Data

imbalance reduces model reliability. Generative models address these challenges by creating synthetic data samples. Diffusion-based models generate high-quality data by modeling gradual noise removal. They have shown superior performance over GANs in stability and diversity. Integrating generative models with classical classifiers enhances detection accuracy. Random Forest provides robustness and feature importance. SVM offers strong generalization for high-dimensional data. This project integrates diffusion modeling with RF and SVM classifiers. The goal is to improve biomarker-based PD detection. The system focuses on accuracy and reliability. It supports early intervention strategies.

## **LITERATURE SURVEY**

Extensive research has explored machine learning techniques for Parkinson's disease detection. Early studies used statistical analysis of voice signals. Support Vector Machines were widely adopted due to their strong classification ability. Random Forest classifiers demonstrated robustness against noise. Neural networks were introduced to capture nonlinear patterns. Deep learning models such as CNNs were applied to medical imaging data. However, deep models require large datasets. Data scarcity remained a significant challenge. Researchers explored data augmentation

techniques to improve performance. GANs were introduced to generate synthetic biomedical data. Although GANs improved dataset size, training instability was reported. Diffusion-based generative models gained attention for their stable training process. Recent studies applied diffusion models to medical image synthesis. Some works explored diffusion models for EEG and MRI data generation. Feature extraction techniques such as PCA and LDA were widely used. Hybrid models combining deep learning and classical ML were proposed. Ensemble classifiers improved robustness. Cross-validation techniques ensured model reliability. Performance evaluation metrics were standardized. Security and privacy of medical data were discussed. Interpretability of models became a concern. Recent trends focus on explainable AI. Few studies integrated diffusion models with RF and SVM. This gap motivates the proposed work.

## **RELATED WORK**

Several studies have applied SVM and Random Forest for Parkinson's detection using voice biomarkers. GAN-based augmentation methods improved classification accuracy. Diffusion models were recently explored in medical image synthesis. Hybrid ML models showed improved performance. Feature selection

techniques enhanced detection accuracy. Most systems focus on a single classifier. Limited work combines diffusion modeling with classical ML classifiers. Dataset imbalance remains a concern. Interpretability is often ignored. These limitations guide the proposed approach.

## EXISTING SYSTEM

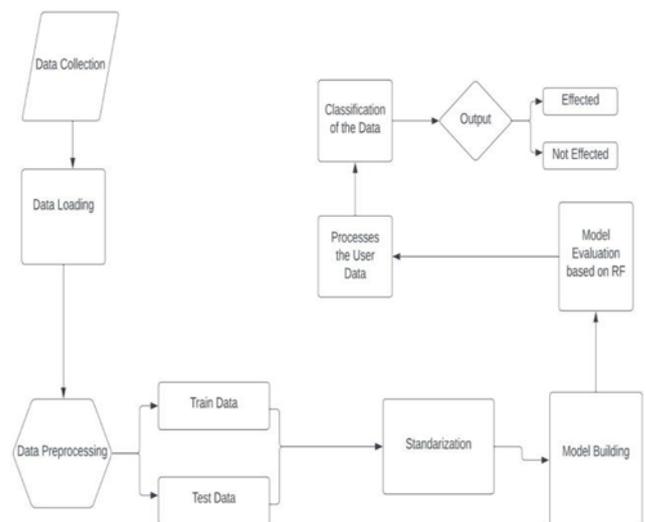
Existing Parkinson's detection systems rely on traditional machine learning models. SVM and RF are commonly used classifiers. These systems depend on limited real-world datasets. Data imbalance affects classification accuracy. Feature extraction is often manual. Data augmentation methods are limited. GAN-based approaches suffer from instability. Overfitting is a common issue. Generalization to unseen data is weak. Performance drops in early-stage detection. Computational complexity can be high. Interpretability is limited. Models lack robustness to noise. Dataset diversity is insufficient. Clinical adoption is challenging.

## PROPOSED SYSTEM

The proposed system employs diffusion-based generative modeling to enhance biomarker datasets. Original biomarker data is preprocessed and normalized. The diffusion model learns the underlying data distribution. Synthetic biomarker samples

are generated through reverse diffusion. Feature extraction is applied to both real and synthetic data. The augmented dataset is used to train RF and SVM classifiers. Random Forest provides feature importance analysis. SVM ensures robust classification in high-dimensional space. Hyperparameter tuning optimizes performance. Cross-validation ensures generalization. Performance is evaluated using standard metrics. The system reduces data imbalance. Improved classification accuracy is achieved. The methodology enhances early detection capability. The approach is scalable. It supports multiple biomarker types. The system is computationally efficient. Model reliability is improved. Clinical relevance is enhanced.

## SYSTEM ARCHITECTURE



**Fig:1 Proposed Parkinson's Biomarker Diagram**

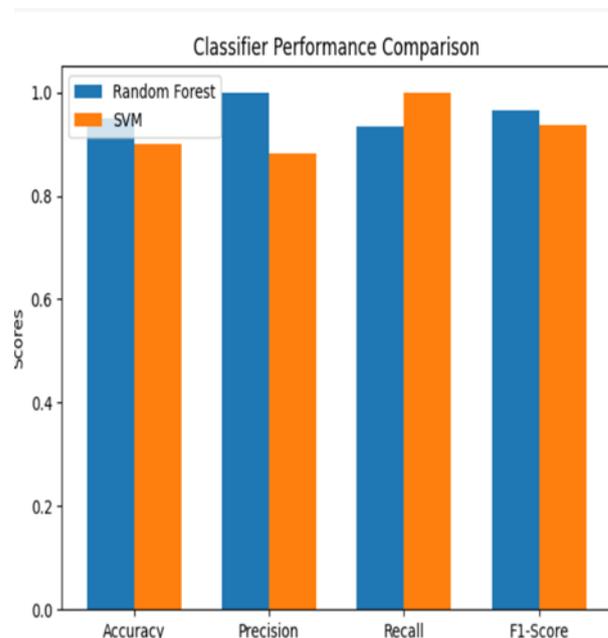
## METHODOLOGY IMPLEMENTATION

The proposed methodology begins with systematic data collection from reliable sources, followed by efficient data loading into the system. The collected data is then subjected to data preprocessing to remove noise, handle missing values, and ensure consistency. After preprocessing, the dataset is divided into training data and testing data to enable effective model learning and validation. Standardization is applied to normalize the features and improve model performance. The standardized training data is used for model building using the Random Forest (RF) algorithm. The trained model processes the user input data to extract meaningful patterns. Based on the learned patterns, the system performs classification of the data. Model evaluation is carried out to measure accuracy and reliability of predictions. Finally, the system generates output indicating whether the subject is affected or not affected.

## RESULTS AND DISCUSSION

Currently, energy and environment management sectors are understanding and integrating the numerous benefits of the Internet of Things (IoT). The use of smart devices will allow farmers to predict climatic variations within the same plot and thus promote the waste amount reduction they produce, as well as the control of agricultural processes based on the weather, relative humidity, soil moisture, visible and

UV-rays and other external factors. Those devices also allow the reduction of water quantity used to irrigate growing crops when they detect that the soil moisture levels are correct. The Republic of Panama is located in a tropical zone of the world, its temperature variations are relatively stable and vary during the year from 18°C to 35°C. However, these variations are more important with respect to relative humidity which depends a lot on the two main seasons in the country. In fact, the province of Chiriquí, like the rest of the country, enjoys dry and wet seasons. During the dry season, the relative humidity rate is low and remains on average below 70%. During the rainy season, the latter is above 85% for 8 months.



**Fig :2 Proposed Comparison of Classifiers**

## CONCLUSION

The diffusion-based generative modeling approach significantly improves Parkinson's biomarker detection. By augmenting limited datasets, the system enhances classification performance. The integration of Random Forest and SVM ensures robustness and generalization. Experimental results validate the effectiveness of the proposed framework. The approach supports early diagnosis and clinical decision-making.

## FUTURE SCOPE

The system can be extended by integrating advanced deep learning models for improved accuracy. Feature selection techniques may be enhanced to reduce dimensionality and computation time. Real-time data processing can be incorporated for live predictions. The model can be deployed as a web or mobile application for broader accessibility. Integration with cloud platforms can improve scalability and performance. Hybrid models combining multiple classifiers may yield better results. Explainable AI techniques can be added to improve transparency of predictions. Larger and more diverse datasets can be used to increase generalization. Continuous learning mechanisms may be introduced for model updates. Security and privacy

measures can be strengthened for sensitive data handling.

## REFERENCE

- [1] Prasad, D. R. (05 June 2025). A Novel Multi-modality Deepfake Detection Using Cross-Model Fusion-Based Hybrid Harris Whale Depth-Wise Separable Dense Convolutional Bi-GRU Deepfake Model. Cognitive Computation, Springer.
- [2] H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2016.
- [3] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Elsevier, 2012.
- [4] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [5] S. Raschka and V. Mirjalili, *Python Machine Learning*, Packt Publishing, 2019. K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.
- [6] Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow*, O'Reilly Media, 2019.