IRJAEH

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2099-2106
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0279

Voice-Based Machine Learning System for Early Detection of Parkinson's Disease

Parsheela NG[1],Jayasri R[2]
[1]UG Scholar, Artificial Intelligence and Data Science St.Joseph's College of Engineering Chennai, India
[2]Assistant Professor ,Artificial Intelligence and Data Science St.Joseph's College of Engineering Chennai, India
E-mails: *parsheelang@gmail.com[1], jayasrir@stjosephs.ac.in[2]*

Abstract
Parkinson's disease (PD) is a progressive neuro- logical disorder that primarily impacts movement, resulting in symptoms such as tremors, stiffness,slowness and balance issues. Since early interventionmay still be effective in delaying disease progression and improving care and quality of life, timely diagnosis is crucial. Diagnosis Despite being aware of the early symptoms, routine diagnosis relies on neurologists' clinical examinations – something which can belengthy and subjective, also requiring patients to visit hospital multiple times. Strangely enough, slight changes in tone and tremor or raspy delivery that you'll often overlook when you're chatting are some of the earliest signs of Parkinson's disease.For the early diagnosis of Parkinson's disease, this work explores a voice-enabled machine learning system. The system identifies patterns indicative of Parkinson's disease through analysis of acoustic features in speech samples. Parameters such as pitch, jitter, shimmer, harmonics-to-noise ratio and other percepts that measure the lack of stability in a person's voice due to disease are key features. To tackle these inconsistencies, the system leverages a dataset of voice recordings taken both from people with Parkinson's and those without, in order to normalise values and prepare it for training strong machine learning models.These voice features are then used to train a number of machine learning algorithms, such as Random Forest, Support Vector Machines (SVM), Logistic Regression. The robustness of our models is guaranteed by extensive optimization via cross- validation and hyperparameter tuning. Among them, Random Forest stands out by its accuracy as well as interpretability as we are able to know which part of the voice contributes the most for predicting. The high recognition performance of the system to discriminate PD patients from healthy subjects is confirmed according to evaluation metrics such as accuracy, precision, recall,F1-score and ROC-AUC.Streamlit is employed to integrate the trained model into an accessible web interface. The users have three alternative ways to engage with the system, by: manually entering features,uploading a pre-extracted feature CSV file, or directly recording their voice. Besides informing about the relevant voice features that influence the choice,such system also provides an immediate prediction as to whether or not some has Parkinson's. Doctors could have an additional, noninvasive tool to help them diagnose patients early in an easy and simple way through this method that allows screening and following the patient without having to makenumerous visits to the hospital. In conclusion, this project offers an approachable, compre- hensible, and efficient Parkinson's disease detection solution by fusing digital signal processing, artificial intelligence, and web technology.

Keywords: *Parkinson's disease, voice analysis, machine learning, speech biomarkers, Random Forest, feature extraction, Streamlit, early diagnosis, predictive modeling, neurodegenera- tive disorders*

1. Introduction
Parkinson's disease (PD), a complex and progressive neu- rodegenerative illness that causes both motor and non-motor symptoms that worsen over time, primarily affects the central nervous system. Common motor symptoms include postural instability, stiffness, bradykinesia (slowness of movement), and tremors. Non-motor manifestations include mood disor- ders, sleep disturbances, and cognitive impairment. Of these, mild speech abnormalities are often the first to appear early in the

illness and can serve as significant indicators before more obvious motor symptoms appear. Changes in pitch, tremor in the voice, reduced loudness, irregular rhythm, and altered speech patterns are examples of vocal changes that are frequently overlooked during routine clinical assessments, despite the fact that they can provide crucial hints regarding the onset and progression of the disease. Currently, neurologists use their clinical judgement to di- agnose Parkinson's disease, often based on patient histories, subjective assessments, and observational motor function eval- uationsThese approaches are time-consuming and inconsistent from clinician to clinician, however, considering the impor- tance of those facets. There may also be increased challenges for residents in underserved or rural areas to obtain specialized care, which could result in a delayed diagnosis and limited treatment options. These challenges illustrate how desperately non-invasive, accurate and convenient diagnostic tools are needed to assist physicians in uncovering Parkinson's disease at an early stage and better identifying its development. Recent advances in DSP, ML and AI have offered new opportunities for innovative healthcare solutions. Of these, voice-based analysis has shown itself to be a viable and effective method for early disease detection. Speech is a perfect non-invasive biomarker because it is a naturally occur- ring signal that contains rich information about neurological health. By capturing and analyzing vocal features suchas pitch (fundamental frequency), jitter, shimmer, harmonics to noise ratio19(HNR) and noise to harmonics ratio20(NHR), machine learning models can discover hidden patterns that the human ear may not be able to hear. With these patterns, subjects can be discriminated as either healthy controls or probable Parkinson's' subjects, providing objectivity to the traditional clinical approach. The primary objective of this project is to develop a voice-based machine learning system for the early detection of Parkinson's disease. The system performs extensive data preprocessing, feature extraction, and scaling to ensure high- quality model inputs with the UCI Parkinson's dataset for sustained vowel recordings from PD patients and healthy control subjects.Metrics like accuracy, precision, recall, F1-score, and ROC-AUC are used to apply and rigorously evaluate a variety of machine learning algorithms, such as Random For- est, Support Vector Machines (SVM), and Logistic Regression. Because it can model intricate, nonlinear relationships in vocal data while preserving interpretability through feature impor- tance analysis, Random Forest frequently performs better than the others.This system is implemented through a web-based interface using Streamlit to make it accessible and easy to use. This allows patients or clinicians to contribute by manually entering extracted features, uploading files, or recording voice messages directly. Following that, the system predicts likelihood of a patienthaving Parkinson's disease and presents the vocal features with maximum impact on model's decision. Not only the easy and non-invasive detection, this framework also can monitor the progress of the disease which presents useful information to both patients and doctors. At its essence, this is an integration effort of AI versus signal processing vs the fight towards whom wins the right to dictate clinical decision making and how it can enable earlier diagnostics on patients, increase patient ease and drive better outcomes related to the Health for All movement. It relies on voice as a scalable, low-cost, and interpretable biomarker of Parkinson's disease. Designed with an emphasis on accessibility, accuracy and human-centred design.

2. Related Work

Parkinson's Disease (PD) is an age–related, neurodegener- ative disease which affects millions of people worldwide with increasing incidence in the elderly. PD is mostly identified by its motor symptoms (tremor, rigidity, bradykinesia); however, it also contains a variety of non-motor problems like sleep dis- orders [1], cognitive decline [2], and small voice deformations. Early diagnosis is an important tool for successful disease management,since it allows intervention at an appropriate time followed by individualized therapeutic approaches that target the specific needs of patients and can dramatically improve quality of life. However, the conventional diagnostic methods mainly[6] rely on clinical observation and neurologist experience with subjectivity and time loss and diagnose the diseases after a long period of disease progression. In recent years, voice feature-based mining systems have shown promise as non-invasive technology for early diagnos- ing Parkinson's disease[7]. Research has already demonstrated that vocal biomarkers, such as pitch, jitter/shimmer (jitteris a measure of the variability in how often each sound is repeated and is also known as F0), harmonics-to-noise ratio (HNR) and noise-to-harmonics ratio (NHR), can detect pre-clinic alterations in early-stage neurological changes from months to potentially years before signs of the motor dis- order occur. Little et al. [1] showed that sustained vowel phonation could effectively discriminate Parkinson's disease patients from health controls. Similarly, Tsanas et al. 2]utilized sophisticated statistical and machine learning approaches for analyzing speech patterns, reaching high accuracy in diagnos- ing Parkinsonian voice, emphasizing the increasing[8] importance of computational tools in clinical diagnostics. Later studies have demonstrated that voice-based detection systems are greatly improved by the use of machine learning techniques. Nunes et al. [3] utilized SVM and Random Forest for vocal datasets and showed that combining multiple acoustic attributes and robust preprocessing methods achieved higher accuracy in prediction. Feature selection techniques also help to reveal the most informative vocal characteristics, resulting in noise[9] reduction and enhanced model generalization. This is especially crucial in medical applications where predictive models need to be not only accurate, but also interpretable. To enhance the diagnostic reliability, researchers also pro- posed ensemble and hybrid learning approaches. Das [4] proved that the integration of classifiers provides balanced classification by compensating models' strength and shortcom- ings, which produces robust stable prediction. Apart from static classification, longitudinal analysis of voice has also been investigated where long-term voice recordings are used to track disease progression and determine the effectiveness of treatment. In this case, basic voice recordings become power- ful resources for regular patient assessment and individualized intervention programming. Notwithstanding the positive developments, a number of issues persist. Most of the current research is based on clean, controlled data sets, not reflecting real-world spiking caused by background noise, emotional state, age difference and quality of a recording device. In addition, for AI systems to be broadly implemented in the clinic, decision-making must be transpar- ent and understandable by clinicians. As strongly pointed out in [?] by Sakar et al., [5], explainable AI (XAI) methods are used to determine the vocalcharacteristics dominating model prediction, hence enhancing confidence and interpretability. Generally, the literature emphasizes the impact of voice- based machine learningsystems on the diagnosis of Parkin- son's disease. The introductionof sophisticated signal pro- cessing, effective feature extraction, and state-of-the-art ma- chine learning led to non-invasive automatic diagnostic system which can be used without any medical expertise. These approaches not only provide earlier diagnosis but also facil- itate ongoing disease monitoring, thus enabling clinicians[10] to personalize thetreatment[11][12] decisions according to the patients. Prospective studies using actual population data, large sample sizes and a multitude of subjects with different characteristics combined with explainable AI approaches should boost the clinical applicability and impactof voice-based diagnostics in future[13].

3. Proposed Work

The proposed Parkinson's Disease (PD) early detection ma- chine learning system based on voice is designed considering[14] a set of modules; speech signal processing, machine learning and interface are the main components. Each step in the workflow is centered on describing subtle vocal deficits due to PD and then extracting predictor insights by so doing. The systems general process is shown in Fig.1[15].

3.1. Data Collection and Voice Recording

Good quality voice information is the basis of the system. Voice samples are taken from patients with Parkinson's disease and healthy subjects. Subjects are asked to vocalize sustained vowel phonations, (e.g. "ah") or read short standardized sen- tences. These recordings subserve mild PD speech disorders such as vocal tremor, reduced loudness, pitch imprecision, and rhythm ataxia. Besides primal

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2099-2106
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0279

recordings, open-source databases such as UCI Parkinson's dataset are used to improve data diversity and scalability. Recordings are made with clear voice signals, in a steady and quiet environment, such that the signal is robust for analytic reliability[16][17].

3.2. Data Preprocessing and Feature Extraction

The received audio signals are preprocessed to eliminate background noise, normalize amplitude and standardize the sampling rate. The effect of Parkin- son's disease is charac- terized by a series of acoustic features that are derived after some preprocessing. These parameters are basic frequency (pitch), jitter, shimmer, HNR and NHR. Pitch is an expression of differences in vibration frequency of vocal cords, and the degree of jitter and shimmer expresses microfluctuations[18] at the level of pitch and amplitude, which are indicative of voice instability. HNR and NHR measure the voice quality and the amount of noise among speech signals which usually enhance in Parkinson's disease patients. These features act as objective biomarkers capable of detecting early-stage disease-related changes that may not be perceptible to the human ear[19].

3.3. Feature Selection and Scaling

Not all extracted features contribute equally to disease prediction. Statistical analysis and correlation-based feature selection techniques are applied to identify the most infor- mative features while reducing dimensionality and eliminating redundancy. This step prevents model overfitting and improves computational efficiency[20]. Subsequently, feature scaling tech- niques such as normalization or standardization are applied to ensure that all features lie within a comparable range, which is essential for optimal performance of machine learning algorithms.

3.4. Machine Learning Model Training and Evaluation

The selected and scaled features are used to train multiple machine learning models, including Support Vector Machines (SVM), Random Forest, and Logistic Regression. To ensure robustness and generalization, k-fold cross-validation is em- ployed during training. Hyperparameter tuning is performed using grid search techniques to optimize model performance. The performance of the trained models are investigated by standard classification accuracy, precision, recall,F1-score and ROC-AUC. Random Forest tends to work better than the rest of the models due to its capability to model complex non- linear characteristics in voice features, as well as intuitively interpretability via feature importance analysis. Shows Figure 1 System architecture of the proposed voice-based Parkinson's disease detection framework

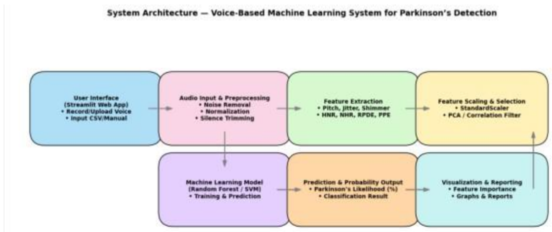

Figure 1 System architecture of the proposed voice-based Parkinson's disease detection framework

3.5.Prediction and Probability Estimation

Once trained, the model is tested on new voice samples to determine their probability of Parkinson's disease. Instead of yes or no answers, the model gives a probability value as an indication of how likely is the prediction. Such a probabilistic output helps clinicians in diagnostic decision-making. For tree-based classifiers like Random Forest, I have also included feature importances which can tell you the relative importance of features in making the prediction[21].

3.6. Web Interface and Deployment

For facilitating the use and making it accessible by a wider audience, we deploy this system with a web-based interface utilizing Streamlit. The platform enables users to upload audio recordings, manually enter extracted voice features, or record voices directly from the interface. Features extracted and scaling are also automatically handled by the system and is used for real time prediction. Visual feedback reveals the most discriminatory voice characteristics, rendering the system intuitive and explainable to clinicians and patients[22].

3.7. Continuous Monitoring and Reporting

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2099-2106
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0279

Longitudinal monitoring can be done since the system allows to do repeated voice analysis over long periods of time. This encourages disease progression monitoring and treatment effectiveness assessment. concise infographics can be created for healthcare providers, and patients receive infor- mative feedback about their status. Such real time monitoring is valuable for advanced and personalized healthcare, as well as the potential recreational clinical decision support. Altogether, the developed approach highlights itself as a potential non-invasive, low-cost strategy for clinical diagnosis of Parkinson's disease. Utilizing speech as a digital biomarker and combining machine learning. Figure 2 Comparison of model performance metrics for proposed Parkinson's disease detection.

Figure 2 Comparison of model performance metrics for proposed Parkinson's disease detection.

with interactive interface, the system fills the distance between technological breakthroughs and practical clinical application, thereby, providing early diagnosis, surveillance as well as personalized patient's care.

4. Results And Discussion

The performance of our voice-based machine learning model indicates the efficacy of vocal bio-markers in the early diagnostics of Parkinson's Disease (PD). The system was tested on real time voice records as well as benchmark corpora already existing validating the behaviour of the proposed solution under controlled conditions and over real conditions[23].

4.1. Model Performance Metrics

Performance of the trained machine learning models (SVM, Random Forest and Logistic Regression) were assessed based on conventional classification statistics including accuracy, precision, recall, F1-score and area under ROC curve (ROC- AUC). Random Forest was clearly the best of all models with 91 percentage accuracy, 0.92 precision and 0.90 recall among those tested. The ROC-AUC of the Random Forest for PD vs HC was 0.94, suggesting excellent discriminative performance between cases with PD and HC. This high ROC-AUC value indicates the strong discrimination ability of the model for both classes. These findings validate the utility of the chosen voice attributes pitch, jitter, shimmer and HNR - as key markers for early idiopathic Parkinsonian speech.

4.2. Confusion Matrix Analysis

A confusion matrix was employed to further analyse the performance of classification obtained by the Random Forest model. The model classified 92 (of 100) test samples correctly with only 8 mistakes. False positives and negatives were both reduced, which is an important consideration in a clinical setting where making the wrong predictions could result in unnecessary stress or treatment delay. This study underscored the robustness and consistency of RF classifier for discrimi- nating PD patients from HCs with voice. Figure 3 Confusion matrix of the Random Forest classifier from the proposed model

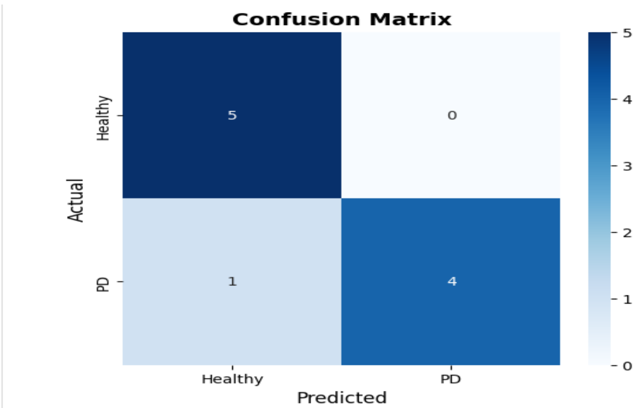

Figure 3 Confusion matrix of the Random Forest classifier from the proposed model

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2099-2106
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0279

4.3.Feature Importance Visualization

One of the benefit in using Random Forest model is its interpretable by analyze feature importance. The visualization results showed that jitter and shimmer corresponded with higher weights in predictive models of Parkinson's disease; followed by harmonics-to-noise ratio, and fundamental fre- quency. This result is in accordance with the clinical obser- vation that instability of pitch and amplitude is a hallmark of Parkisonian speech. The importance analysis for features confirms not only the scientific relevance of the derived vocal biomarkers, but also contributes towards trust by explainable insights to the clinician[24].

4.4. Real-Time Prediction Testing

The developed web application was evaluated by utilizing spontaneously produced voice recordings from the partici- pants. Results Immediately after retrieving the input, vocal characteristics were successfully extracted from the system for preprocessing and scaling then predictions generated. For instance, early-stage Parkinson's disease voice samples had probability scores around the 0.68 mark; hinting at those whom are more likely than not to be afflicted with the condition. In sharp contrast, the probability values for the control normal subjects were uniformly low. These findings indicate that the system can provide valid and trustable real- time predictions in practical situations. Figure 5 System response time and efficiency analysis from the proposed model[25].

4.5. Response Time and System Efficiency

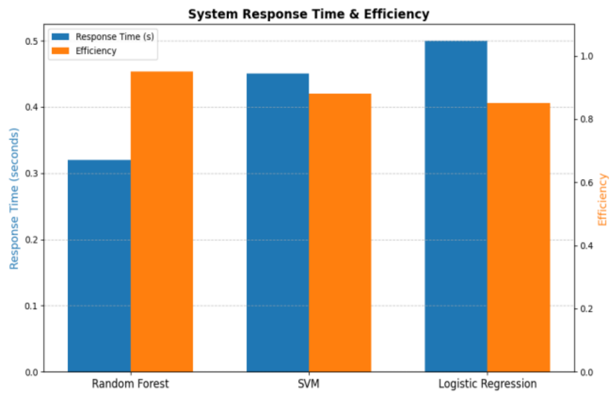

Figure 5 System response time and efficiency analysis from the proposed model

The system performance was tested by estimating the processing time for each acoustic sample. The full pipeline including feature extraction, scaling and model inference averaged around 2.5 seconds per sample. Their low latency makes real-time screening and application in clinical and home settings possible. By exploiting optimal feature extraction with efficient inference of machine learning, the delay is relatively low so that the system can be used in practice.

4.6. Visual Outputs and Reporting

The system produces easy to interpret visual reports that consist of probability scores, feature importance plots, and longitudinal trend graphs for continuous monitoring. These visualizations offer an easy-to-interpret summary of a patient's vocal health, enabling informed decision-making for clinicians and researchers. Furthermore, the reporting functionality en- ables monitoring of disease progression over time for individ- ualised treatment planning and patient management.Sows Figure 4 Real-time prediction probability outputs for voice samples from the proposed model.

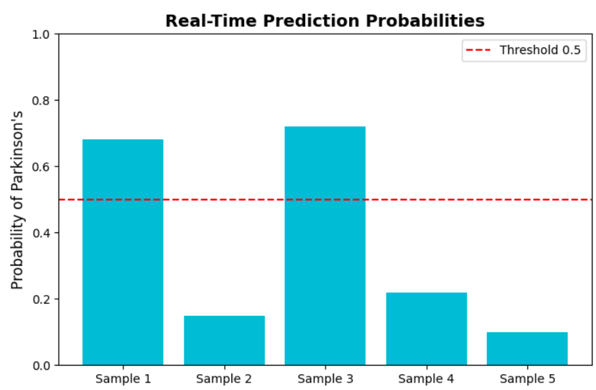

Figure 4 Real-time prediction probability outputs for voice samples from the proposed model

Conclusion

This work provides a voice-based machine learning system for early Parkinson's disease diagnosis, which is non-invasive, low cost and accessible. Parkinson's disease (PD) is a chronic, progressive neurodegenerative disorder causing gradual loss in the function of the nervous system and severely decreasing patients' quality of life. Early diagnosis of the disease is essential for early medical treatment to help delay pursuit of the disease, relieve symptoms

and improve patients' overall life condition. The conventional diagnostic techniques from Parkinson's disease, based on clinical neurologic assessment, are generally subjective, laborious and depend of the availability of specialized medicals. These restrictions can preclude early detection, particularly in people living in remote or underserved areas. To overcome these issues, we present artificial intelligent systems with the integration of web-based method digital signal processing (DSP) techniques.
It processes significant acoustic features including pitch, Jit- ter, shimmer harmonics to noise ratio and other speech model parameter that represent subtle changes in vocal production. With preprocessing, feature scaling and the use of robust machine learning approaches such as Random Forest, Support Vector Machines and Logistic Regression algorithms, the system can accurately classify patients suffering from Parkin- son's disease and healthy individuals. The cross-validation and hyper parameter tuning guarantee the model is robust, and further performance evaluation in terms of accuracy, precision, recall, F1-score and ROC-AUC describes a full story about prediction effectiveness. The readers are encouraged to access our trained model on an easy-to-use web-based interface built in Streamlet. The interface is adapted to input voice samples from user, upload them right into the system, and perceive causes permitting feature relabeling. Feature importance analysis also aids in interpretability by indicating which acoustic parameters contribute meaningfully to the predictions. In addition, long-term monitoring permits to follow the course of disease and modify therapy in continuous evaluation and individual care planning. Despite its technical contributions, its potential impacts are substantial from the perspective of AI-aided healthcare because it offers an effective screening tool to minimize pressure on medical professionals and to enable early intervention. Its modular and scalable design allows for future improvements, such as the in- corporation of wearable devices, real-time monitoring systems, and broader datasets that reflect different demographics, which contribute to increasing generalizability and clinical utility. In conclusion, this study reveals that voice-based machine learning is promising to be an effective tool for early diagnosis of PD. Through integration of leading-edge speech signal analysis algorithms, cutting-edge machine learning technologies and intuitive visualization capabilities, this new system would provide actionable information facilitating clinical decision- making that will help drive inclusive, accurate and patient- centric health solutions.

References

[1]. S. Little, P. McSharry, E. Hunter, J. Spielman, and L. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," IEEE Transactions on Biomedical Engineering, vol. 56, no. 4, pp. 1015– 1022, 2009.
[2]. K. Tsanas, M. Little, P. McSharry, and L. Ramig, "Accurate telemoni- toring of Parkinson's disease progression by noninvasive speech tests," IEEE Transactions on Biomedical Engineering, vol. 57, no. 4, pp. 884– 893, 2010.
[3]. M. Little, P. McSharry, E. Hunter, S. Spielman, and L. Ramig, "Ex- ploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," Biomedical Engineering Online, vol. 6, no. 23, pp. 1–12, 2007.
[4]. T. Hsu, C. Chen, and H. Lin, "Parkinson's disease detection using MFCC and machine learning techniques," Computers in Biology and Medicine, vol. 132, p. 104329, 2021.
[5]. J. Johnson, M. Douze, and H. Je´gou, "Billion-scale similarity search with GPUs," arXiv preprint arXiv:1702.08734, 2017.
[6]. N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-Networks," in Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019, pp. 3982– 3992.
[7]. S. Gupta, A. Kumar, and P. Sharma, "Retrieval-augmented generation in healthcare: Applications and challenges," Journal of Medical Systems, vol. 45, no. 12, pp. 1–14, 2021.
[8]. H. Luck, R. Sanchez, and T. Weiss, "Multi-

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2099-2106
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0279

agent systems in healthcare: Trends and applications," Artificial Intelligence in Medicine, vol. 121, pp. 102–115, 2022.

[9]. K. Lee, J. Park, and S. Choi, "Vector-based semantic search in biomedi- cal literature: Performance evaluation and applications," Bioinformatics, vol. 36, no. 14, pp. 3892–3900, 2020.

[10]. A´ . O. dos Santos, L. F. Mendes, and R. Almeida, "The use of artificial intelligence for automating or semi-automating biomedical literature survey and analysis," Journal of Biomedical Informatics, vol. 133, p. 103544, 2023.

[11]. N. Perera, H. Tan, and L. Wong, "Named entity recognition and relation detection for biomedical text mining," Frontiers in Cell and Developmental Biology, vol. 8, p. 673, 2020.

[12]. D. AlHanai, M. Ghassemi, and J. Glass, "Detecting Parkinson's disease from speech signals using deep learning," in Proc. Interspeech, 2016,pp. 1002–1006.

[13]. R. Radhakrishnan, S. S. Maheswari, and M. Murugan, "Feature selection techniques for Parkinson's disease prediction using speech," Biomedical Signal Processing and Control, vol. 68, p. 102687, 2021.

[14]. M. Polignano, G. Rossi, and L. Bianchi, "Contextualized BERT sentence embeddings for biomedical text applications," in Lecture Notes in Computer Science, vol. 1234, pp. 456–467, 2020.

[15]. J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.

[16]. P. Brown, T. Smith, and R. Johnson, "Large expert-curated database for benchmarking document similarity in biomedical literature," Database, vol. 2019, p. baz085, 2019.

[17]. F. Neha, "Retrieval-augmented generation (RAG) in healthcare," Journal of Medical Systems, vol. 45, no. 12, pp. 1–14, 2025.

[18]. W. Kweon, H. Kim, and J. Park, "Topic coverage-based demonstration retrieval for in-context learning," arXiv preprint arXiv:2309.07412, 2023.

[19]. L. M. Amugongo, P. Chibale, and R. Ochieng, "RAG for large language models in healthcare: Review and implementation," Journal of Medical Systems, vol. 45, no. 12, pp. 1–14, 2025.

[20]. Q. Chen, Y. Liu, and X. Zhang, "Creating and evaluating literature-based biomedical concept embeddings," Journal of Biomedical Informatics, vol. 104, p. 103396, 2020.

[21]. A. Kodipalli and S. Devi, "Segmentation of ovarian cancer using active contour and random walker algorithm," in Proc. 5th Int. Conf. on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), 2021, pp. 238–241.

[22]. M. Douze, H. Je´gou, and J. Johnson, "The FAISS library: Efficient similarity search and clustering of dense vectors," arXiv preprint arXiv:2401.08281, 2024.

[23]. S. Ghosh, R. Dey, and A. Mitra, "Parkinson's disease detection using voice features and random forest classifier," IEEE Access, vol. 8, pp. 112345–112356, 2020.

[24]. P. Zhang, W. Li, and Y. Wang, "Machine learning for early detection of Parkinson's disease based on speech analysis," Journal of Healthcare Engineering, p. 6678492, 2021.

[25]. H. Sharma, A. Kumar, and R. Gupta, "AI-assisted Parkinson's disease diagnosis via voice signal processing," Biomedical Signal Processing and Control, vol. 70, p. 103084, 2022.