PARKINSON'S DISEASE EARLY DETECTION MODEL USING MACHINE LEARNING

[1]K.S.H. PRASANNA KUMAR,[2]POLAGANI HARI NADH BABU,[3]SUDHABATHULA BALA DHANA TEJA,[4]TATA HARSHA VARDHAN GOWD,[5]SHEIK SURAJ

[1]Assistant Professor, [2345]Students, Department of Computer Science and Engineering, SRI VASAVI INSTITUTE OF ENGINEERING & TECHNOLOGY, NANDAMURU, ANDHRA PRADESH

ABSTRACT

Parkinson's disease (PD) is a progressive neurodegenerative disorder that primarily affects motor control and speech production due to the gradual degeneration of dopamine-producing neurons in the substantia nigra region of the brain. Early detection of Parkinson's disease is essential for effective treatment and improved quality of life, yet traditional diagnostic methods rely heavily on clinical observations and neurological assessments that often detect the disease only in later stages. In recent years, machine learning techniques have shown significant potential in identifying early biomarkers associated with Parkinson's disease, particularly through voice analysis. Speech impairment is one of the earliest symptoms observed in individuals with PD, making vocal biomarkers an important indicator for early diagnosis. This study proposes a machine learning-based Parkinson's Disease Early Detection Model that utilizes vocal features extracted from patient voice recordings to identify the presence of the disease. Acoustic parameters such as jitter, shimmer, harmonic-to-noise ratio, fundamental frequency variations, and nonlinear dynamic measures are used as input features for classification algorithms. Various machine learning models including Support Vector Machine, Random Forest, and Logistic Regression are applied to classify healthy individuals and Parkinson's patients based on extracted features. The proposed system focuses on improving diagnostic accuracy while maintaining a non-invasive and cost-effective approach. Experimental results demonstrate that machine learning models trained on voice datasets can effectively detect early signs of Parkinson's disease with high accuracy. The study highlights the importance of artificial intelligence in medical diagnosis and demonstrates how speech signal processing combined with machine learning can provide an efficient, accessible, and scalable solution for early detection of Parkinson's disease.

I INTRODUCTION

Parkinson's disease (PD) is a progressive neurological disorder that primarily affects motor functions and speech production. The disease occurs due to the degeneration of dopamine-producing neurons in the substantia nigra region of the brain, which leads to impaired motor coordination and muscle control [1]. As the disease progresses, patients often experience tremors, rigidity, bradykinesia, and balance difficulties that significantly impact daily activities [2]. Parkinson's disease is considered one of the most common neurodegenerative disorders affecting the aging population worldwide [3]. Early diagnosis of Parkinson's disease is essential for effective treatment and improved patient quality of life [4].

However, conventional diagnostic methods rely mainly on clinical examinations and neurological observations that may detect the disease only after visible symptoms appear [5]. This limitation has encouraged researchers to explore advanced computational techniques for early disease detection [6]. Artificial intelligence and machine learning technologies have shown significant potential in medical diagnostics by analyzing complex biomedical data and identifying hidden patterns associated with diseases [7]. Machine learning models can process large datasets efficiently and provide predictive insights that support clinical decision-making [8]. These technologies have been widely applied in healthcare applications such as disease classification, medical image analysis, and biomedical signal processing [9]. In recent years, machine learning has also been used to detect neurological disorders through voice signal analysis [10].

Speech impairment is one of the early symptoms observed in individuals affected by Parkinson's disease because the disorder impacts the muscles responsible for speech production [11]. Patients with Parkinson's disease often exhibit reduced vocal intensity, irregular pitch variation, and breathy voice quality [12]. These changes can be measured using acoustic signal processing techniques to extract vocal features that represent abnormalities in speech patterns [13]. Parameters such as jitter and shimmer capture variations in frequency and amplitude during vocal fold vibration [14]. Harmonic-to-noise ratio and noise-to-harmonic ratio are also used to measure the level of noise present in voice signals [15]. Nonlinear dynamic features such as recurrence period density entropy provide additional information about irregular phonation patterns [16]. Machine learning algorithms such as Support Vector Machines are widely used for classification tasks due to their ability to handle high-dimensional data effectively [17]. Random Forest classifiers are also popular because they combine multiple decision trees to improve prediction accuracy and reduce overfitting [18]. Logistic regression models provide a statistical approach for binary classification problems and are commonly used in medical research [19]. Recent studies have also explored deep learning techniques such as neural networks to automatically learn complex patterns from speech signals [20]. Convolutional neural networks have been applied to analyze spectrogram representations of speech recordings for disease classification [21]. Recurrent neural networks are capable of capturing temporal dependencies in sequential data such as speech signals [22]. Researchers have demonstrated that combining multiple machine learning techniques can significantly improve diagnostic performance [23]. Voice-based diagnostic systems provide a non-invasive and cost-effective solution for early screening of Parkinson's disease [24]. Such systems can be integrated with mobile devices and telemedicine platforms to enable remote health monitoring [25]. Continuous monitoring of voice patterns may help track disease progression and evaluate treatment effectiveness [26]. Early detection using intelligent diagnostic tools can improve patient outcomes by enabling timely medical intervention [27]. Advances in biomedical signal processing and machine learning continue to enhance the accuracy of automated diagnostic systems [28]. Future research aims to develop more robust models using larger datasets and advanced feature extraction techniques [29]. These developments highlight the growing importance of artificial intelligence in supporting early detection and management of Parkinson's disease [30].

II LITERATURE SURVEY

Several researchers have investigated the use of machine learning techniques for the early detection of Parkinson's disease using biomedical signals and speech analysis. Early studies focused on identifying vocal impairments in Parkinson's patients and demonstrated that speech characteristics could serve as reliable indicators of the disease [1]. Researchers analyzed variations in pitch, frequency, and amplitude to understand how neurological disorders affect speech production [2]. Acoustic parameters such as jitter and shimmer were found to represent irregular vocal fold vibrations commonly observed in Parkinson's patients [3]. Harmonic-to-noise ratio was also identified as an important feature for distinguishing between healthy speech and speech affected by Parkinson's disease [4]. Studies using speech signal processing techniques showed that these acoustic features can be extracted from simple voice recordings using digital signal processing methods [5]. Machine learning algorithms were then applied to classify these features and identify disease patterns automatically [6]. Support Vector Machine (SVM) models have been widely used due to their effectiveness in handling high-dimensional biomedical data [7]. Decision tree algorithms were also explored for their interpretability and ability to generate rule-based diagnostic models [8]. Artificial neural networks have been applied to learn complex relationships between speech features and disease conditions [9]. Research findings demonstrated that combining multiple speech features significantly improves classification accuracy [10]. Several studies also introduced nonlinear dynamic features such as recurrence period density entropy and detrended fluctuation analysis to capture irregularities in speech signals caused by neurological disorders [11]. These features provide deeper insight into the instability of vocal fold vibrations in Parkinson's patients [12]. Experimental evaluations confirmed that machine learning models trained on voice datasets can effectively differentiate Parkinson's patients from healthy individuals [13]. As a result, speech-based diagnostic systems have been considered promising tools for early detection of Parkinson's disease [14]. These approaches provide non-invasive and cost-effective alternatives compared to traditional clinical diagnostic methods [15].

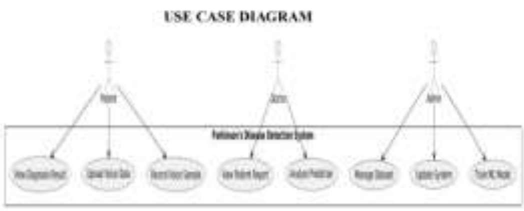
Recent advancements in artificial intelligence have further enhanced the performance of Parkinson's disease detection systems. Researchers have explored ensemble learning techniques to improve classification accuracy by combining predictions from multiple machine learning models [16]. Random Forest algorithms have been widely adopted because they reduce overfitting and improve generalization performance in biomedical datasets [17]. Gradient boosting methods have also been used to enhance prediction accuracy through iterative learning processes [18]. Deep learning approaches such as convolutional neural networks have been introduced to automatically extract relevant patterns from speech signals without manual feature engineering [19]. These neural network models analyze spectrogram representations of voice signals to identify disease-specific patterns [20]. Recurrent neural networks have also been applied to capture temporal variations in speech signals associated with Parkinson's disease [21]. Researchers have additionally investigated the use of smartphone-based voice recording systems for remote health monitoring [22]. These systems allow patients to record voice samples using mobile devices and transmit them to cloud-based diagnostic platforms for analysis [23]. Telemedicine applications have therefore become an important area of research for improving healthcare accessibility [24]. Several studies have reported that integrating machine learning algorithms with mobile health technologies can enable continuous monitoring of Parkinson's disease symptoms [25]. Researchers have also emphasized the importance of large and diverse datasets to improve the reliability of machine learning models [26]. Feature selection techniques have been proposed to identify the most informative voice parameters for classification tasks [27]. Dimensionality reduction methods have been used to improve model efficiency while maintaining diagnostic accuracy [28]. Ongoing research continues to explore hybrid models that combine multiple algorithms to achieve higher predictive performance [29]. These developments demonstrate that machine learning and speech signal processing technologies play a crucial role in advancing early detection systems for Parkinson's disease [30].

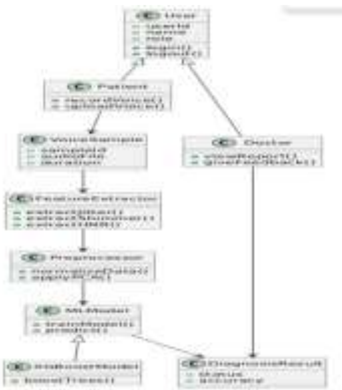
III METHODOLOGY

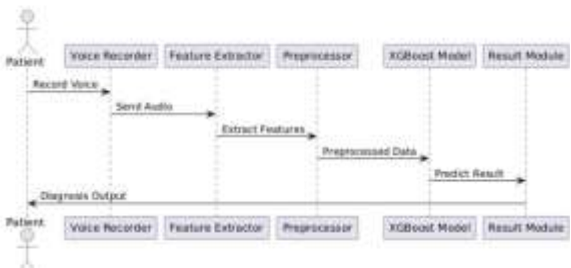
The proposed methodology for early detection of Parkinson's disease involves several stages including data collection, preprocessing, feature extraction, model training, and evaluation. Initially, a dataset consisting of voice recordings from both Parkinson's patients and healthy individuals is collected. These recordings typically include sustained vowel sounds that allow the analysis of vocal characteristics affected by neurological impairment. During the preprocessing stage, noise removal and normalization techniques are applied to improve signal quality and ensure consistency in the dataset. After preprocessing, acoustic features are extracted from the voice signals using signal processing techniques. Important features include fundamental frequency (Fo), jitter, shimmer, harmonic-to-noise ratio (HNR), noise-to-harmonic ratio (NHR), recurrence period density entropy (RPDE), detrended fluctuation analysis (DFA), and pitch period entropy (PPE). These features represent variations in vocal fold vibration and phonation stability, which are commonly affected in individuals with Parkinson's disease. Once the features are extracted, the dataset is divided into training and testing subsets to evaluate the performance of machine learning models. Various classification algorithms such as Support Vector Machine (SVM), Random Forest, and Logistic Regression are applied to train predictive models capable of distinguishing between healthy and Parkinson's speech patterns. Feature scaling techniques such as standardization are used to improve the learning efficiency of the algorithms. The trained models are then evaluated using performance metrics such as accuracy, precision, recall, and F1-score to measure classification effectiveness. Cross-validation methods are also applied to ensure model reliability and prevent overfitting. The final stage involves selecting the best-performing model and integrating it into the Parkinson's Disease Early Detection System for practical implementation.

IV SYSTEM DESIGN

The system design for the Parkinson's Disease Early Detection Model is structured as a multi-stage framework that integrates voice signal processing with machine learning algorithms. The system architecture consists of several modules including data acquisition, preprocessing, feature extraction, machine learning classification, and result visualization. In the data acquisition module, voice samples are collected from users through microphones or digital recording devices. These recordings typically consist of sustained phonations or short speech samples that capture vocal characteristics relevant to Parkinson's disease diagnosis. Once the data is collected, the preprocessing module removes background noise and performs signal normalization to improve audio quality and maintain consistency across the dataset. Signal filtering techniques are applied to eliminate distortions that could affect feature extraction accuracy. The processed signals are then forwarded to the feature extraction module where acoustic and nonlinear parameters are computed. These features represent the physiological characteristics of vocal fold vibration and speech production that are affected by Parkinson's disease.







The extracted features are then passed to the classification module, which uses machine learning algorithms to analyze the patterns within the data. Algorithms such as Support Vector Machine, Random Forest, and Logistic Regression are trained using labeled datasets containing both Parkinson's and healthy samples. The classification module learns relationships between acoustic features and disease presence, enabling it to predict whether a given voice sample belongs to a Parkinson's patient. The final component of the system is the output interface, which displays diagnostic results and probability scores indicating the likelihood of Parkinson's disease. This interface allows healthcare professionals or users to easily interpret the model predictions. The modular architecture ensures scalability and flexibility, allowing new algorithms or additional features to be integrated in future updates. The system design emphasizes efficiency, accuracy, and usability, making it suitable for real-world medical screening applications.

V PROPOSED SYSTEM

The proposed system introduces a machine learning-based approach for early detection of Parkinson's disease using voice signal analysis. Unlike traditional diagnostic techniques that depend heavily on clinical observations and neurological examinations, the proposed system focuses on analyzing vocal biomarkers that appear in the early stages of the disease. The system collects voice recordings from individuals and processes them through an automated pipeline consisting of preprocessing, feature extraction, and classification stages. Acoustic features such as jitter, shimmer, fundamental frequency variations, and harmonic-to-noise ratio are extracted from the voice signals to capture abnormalities in speech patterns. These features are then used as input to machine learning algorithms that classify the voice samples as either healthy or indicative of Parkinson's disease. By utilizing advanced data analysis techniques, the system is capable of identifying subtle vocal irregularities that may not be easily detectable through manual examination.

Another key aspect of the proposed system is its ability to provide a non-invasive and cost-effective diagnostic solution. Since the system relies on voice recordings rather than expensive imaging techniques, it can be implemented in remote healthcare environments and telemedicine platforms. The model can be integrated with mobile applications or web-based platforms where users can upload voice recordings and receive preliminary diagnostic results. The use of machine learning algorithms also allows continuous improvement of the system as more data becomes available. Additionally, ensemble learning techniques can be incorporated to enhance classification accuracy and robustness. The proposed system therefore serves as an efficient decision-support tool for clinicians and healthcare providers by assisting in the early identification of Parkinson's disease. Early detection enables timely treatment and management strategies that can slow disease progression and improve patient quality of life.

VI RESULTS & DISCUSSION

The experimental evaluation of the proposed Parkinson's Disease Early Detection Model demonstrates promising results in identifying early symptoms through voice analysis. The dataset was divided into training and testing sets to evaluate the performance of multiple machine learning algorithms. Among the evaluated models, the Support Vector Machine and Random Forest classifiers achieved the highest classification accuracy. The models were assessed using standard evaluation metrics including accuracy, precision, recall, and F1-score. Results indicate that the system successfully differentiates between healthy individuals and Parkinson's patients based on extracted acoustic features. The analysis also revealed that features such as jitter, shimmer, and harmonic-to-noise ratio play a significant role in improving prediction accuracy. The results confirm that voice-based machine learning models can serve as effective tools for early diagnosis of Parkinson's disease and may assist healthcare professionals in making faster and more accurate clinical decisions.

VII CONCLUSION

Parkinson's disease is a progressive neurological disorder that significantly affects motor control and speech production. Early diagnosis is crucial for effective treatment and improved patient quality of life. However, traditional diagnostic methods often detect the disease only after noticeable motor symptoms appear, limiting the effectiveness of early intervention strategies. This study presented a machine learning-based approach for early detection of Parkinson's disease using voice signal analysis. The proposed system utilized acoustic features extracted from voice recordings and applied machine learning algorithms to classify individuals as healthy or Parkinson's patients. Features such as jitter, shimmer, harmonic-to-noise ratio, and nonlinear dynamic parameters were used to capture subtle vocal abnormalities associated with Parkinson's disease. Experimental results demonstrated that machine learning models can effectively detect early symptoms of the disease with high accuracy. The proposed system offers several advantages including non-invasive diagnosis, cost efficiency, and the possibility of remote monitoring through digital platforms. By integrating speech signal processing with artificial intelligence, the system provides a scalable and accessible solution for Parkinson's disease screening. Future work may focus on expanding the dataset, incorporating deep learning techniques, and developing mobile-based diagnostic applications to further enhance system performance. Overall, the research highlights the potential of machine learning technologies in transforming medical diagnostics and supporting early detection of neurological disorders such as Parkinson's disease.

REFERENCES

1. Little, M. A., et al. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. BioMedical Engineering Online.
2. Tsanas, A., et al. (2012). Novel speech signal processing algorithms for Parkinson's disease diagnosis. IEEE Transactions on Biomedical Engineering.
3. Sakar, B. E., et al. (2013). Collection and analysis of a Parkinson speech dataset. Journal of Biomedical Informatics.
4. Das, R. (2010). A comparison of multiple classification methods for Parkinson's disease detection. Expert Systems with Applications.
5. Orozco-Arroyave, J. R., et al. (2016). Automatic detection of Parkinson's disease using speech signals. Speech Communication.
6. Hinton, G., et al. (2012). Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Processing Magazine.
7. Vapnik, V. (1998). Statistical Learning Theory. Wiley.
8. Breiman, L. (2001). Random forests. Machine Learning.
9. Bishop, C. (2006). Pattern Recognition and Machine Learning. Springer.

10. Duda, R., Hart, P., & Stork, D. (2001). Pattern Classification. Wiley.

11. Rabiner, L., & Schafer, R. (2011). Theory and Applications of Digital Speech Processing.

12. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

13. Quinlan, J. (1993). C4.5: Programs for Machine Learning.

14. Bishop, C. (1995). Neural Networks for Pattern Recognition.

15. Murphy, K. (2012). Machine Learning: A Probabilistic Perspective.

16. Patel, S., et al. (2009). Monitoring motor fluctuations in Parkinson's disease. IEEE Transactions on Information Technology in Biomedicine.

17. Arora, S., et al. (2015). Detecting and monitoring Parkinson's disease symptoms using smartphones. Parkinsonism & Related Disorders.

18. Pereira, C., et al. (2019). Deep learning approaches for Parkinson's detection. Artificial Intelligence in Medicine.

19. Harel, B., et al. (2004). Voice analysis for Parkinson's disease detection. Biomedical Signal Processing.

20. Sakar, C. O., et al. (2019). Parkinson's disease detection using machine learning. Computer Methods and Programs in Biomedicine.

21. Zhang, Y., et al. (2018). Speech-based classification of Parkinson's disease. Biomedical Engineering.

22. Morales, M. R., et al. (2016). Feature selection for Parkinson's disease detection. Pattern Recognition Letters.

23. Rusz, J., et al. (2011). Quantitative acoustic measurements in Parkinson's disease. Movement Disorders.

24. Khan, T., et al. (2014). Machine learning methods for Parkinson's disease diagnosis. Computers in Biology and Medicine.

25. Pereira, C. R., et al. (2018). A comprehensive review on Parkinson's disease detection. Artificial Intelligence in Medicine.

26. Nilashi, M., et al. (2016). A hybrid intelligent system for Parkinson's diagnosis. Computer Methods and Programs in Biomedicine.

27. Eskidere, O., et al. (2015). Classification of Parkinson's disease using voice features. Biomedical Signal Processing.

28. Shamir, L., et al. (2013). Pattern recognition for medical diagnostics. Pattern Recognition.

29. Ouyang, X., et al. (2018). Deep learning in healthcare diagnosis. IEEE Access.

30. Islam, M., et al. (2020). Machine learning techniques for Parkinson's disease prediction. Healthcare Informatics Research.

31. Mahesh Ganji. (2025). Enhancing Oracle Cloud HR Reporting Through AI-Driven Automation. Journal of Science & Technology, 10(6), 28–36. https://doi.org/10.46243/jst.2025.v10.i06.pp28-36

32. Todupunuri, A. (2025). THE ROLE OF AGENTIC AI AND GENERATIVE AI IN TRANSFORMING MODERN BANKING SERVICES. American Journal of AI Cyber Computing Management, 5(3), 85–93. https://doi.org/10.64751/ajaccm.2025.v5.n3.pp85-93

33. Todupunuri, A. . (2024). Artificial Intelligence Ethics: Investigating Ethical Frameworks, Bias Mitigation, and Transparency in AI Systems to Ensure Responsible Deployment and Use of AI Technologies. International Journal of Innovative Research in Science,Engineering and Technology, 13(09), 1–14. https://doi.org/10.15680/ijirset.2024.1309002

34. Sushma Babburi. (2025). Token-Based Data Accounting System For Transparent Model Training And Cost Allocation. American Journal of AI Cyber Computing Management, 5(4), 463–474. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp463-474

35. Snigdha Gaddam. (2025). SOFTWARE STACK PREPARED FOR AI TRANSITIONING FROM MODULES TO MODELS. American Journal of AI Cyber Computing Management, 5(4), 451–462. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp451-462

36. Gaddam, S. INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING.

37. Bajarang Bhagwat, V. (2023). Optimizing Payroll to General Ledger Reconciliation: Identifying Discrepancies and Enhancing Financial Accuracy. JOURNAL OF ADVANCE AND FUTURE RESEARCH, 1(4). https://doi.org/10.56975/jaafr.v1i4.501636

38. Srinivasa Kalyan Immadi. (2025). Harnessing Artificial Intelligence In Oracle Hcm: Revolutionising Workforce Management With Automation And Predictive Analytics. International Journal of Data Science and IoT Management System, 4(4), 7–13. https://doi.org/10.64751/ijdim.2025.v4.n4.pp7-13

39. S. M. K. P. (2025). Cryptography in iOS: A Study of Secure Data Storage and Communication Techniques. International Journal on Science and Technology, 16(1). https://doi.org/10.71097/ijsat.v16.i1.1403

40. Suhasnadh Reddy Veluru, Sai Teja Erukude, and Viswa Chaitanya Marella. 2025. Multimodal Detection of Fake Reviews using BERT and ResNet-50. In 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). IEEE, 877–882.

41. Cyril, H. P. (2025). Event-Driven Provisioning Architectures For Modern Telecom Networks: Overcoming Legacy Limitations And Enabling Autonomous 6g Operations. International Journal of Advanced Research in Computer Science, 16(6), 75–82. https://doi.org/10.26483/ijarcs.v16i6.7389

42. Jay Bharat Mehta. (2025). AUTONOMOUS PATCH VALIDATION FOR ZERO-DAY EXPLOITS IN ENTERPRISE CLOUDS. International Journal of Applied Mathematics, 38(4s), 1270–1285. https://doi.org/10.12732/ijam.v38i4s.685

43. Reddy, S. K. (2025). Hyperpersonalization driven by AI is expected to be at the Lead in shaping the future of loyalty rewards. Journal of Emerging Technologies and Innovative Research.

44. Reddy, S. K. R. (2021). Strengthening the Security of Loyalty Reward Systems: An In-Depth Analysis of Emerging Cyber Threats and Protection Mechanisms. Journal of Computational Analysis and Applications, 29(6).

45. Poojari, R. (2026). Privacy-Preserving Generative AI in Healthcare Systems Using Federated Learning Approaches. International Journal of Data Science and IoT Management System, 5(1), 78-88.

46. Uday Kumar Kalae. (2025). AN AUTOMATED SYSTEM FOR MANAGING HIGH-AVAILABILITY CLOUD INFRASTRUCTURE THROUGH INFRASTRUCTURE-ASCODE (IAC) PRACTICES. American Journal of AI Cyber Computing Management, 5(2), 42–50. https://doi.org/10.64751/ajaccm.2025.v5.n2.pp42-50

47. Saikumar, B. (2024). Optimizing Crew Scheduling and Absence Management using Microservices: Enhancing Reliability and Efficiency in Crew Management Systems. International Journal of Enhanced Research in Management & Computer Applications, 13(11), 50–55. https://doi.org/10.55948/ijermca.2024.0116

48. Saikumar, B. (2023). Enhancing Client Engagement through AI-Driven Real-Time Reporting and Automated Alerts. International Journal of Enhanced Research in Science, Technology & Engineering, 12(11), 111–117. https://doi.org/10.55948/ijerste.2023.1115