iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

PARKINSONS DISEASE DETECTION USING MACHINE LEARNING

Dhanushya
Department of Computer Applications, Vels Institute of Science, Technology and Advanced Studies, Chennai

Dr. P. Sujatha
Professor and Head of Department of Computer Applications, Vels Institute of Science, Technology and Advanced Studies, Chennai

ABSTRACT
Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects more than ten million people worldwide. Early detection remains a major clinical challenge because motor symptoms typically emerge after a significant proportion of dopaminergic neurons have already been lost. Voice impairment, however, is one of the earliest non-motor manifestations of PD, making acoustic biomarker analysis an attractive, non-invasive screening modality. This paper presents NeuroDetect AI, a fully client-side web application that screens for Parkinson's disease using thirteen voice biomarkers derived from the Oxford Parkinson's Telemonitoring Dataset. The system implements a logistic regression classifier in TypeScript, exposes a clean React-based interface for both manual entry and CSV upload, and produces an interpretable verdict accompanied by a probability, confidence score and risk level (Low / Moderate / High). The application is designed for use by clinicians, researchers and informed patients as a triage tool that complements—but does not replace—clinical evaluation. Experimental validation on synthetic samples derived from the Oxford dataset yields screening accuracy comparable to published baselines while running entirely in the browser, requiring no server-side computation and preserving patient privacy by design.

Keywords:
Parkinson's disease, voice biomarkers, machine learning, logistic regression, jitter, shimmer, web-based screening, healthcare AI, client-side inference.

INTRODUCTION
Parkinson's disease is the second most common neurodegenerative disorder after Alzheimer's disease, characterised clinically by tremor, rigidity, bradykinesia and postural instability. By the time these motor symptoms become unequivocal, between 60% and 80% of dopamine-producing neurons in the substantia nigra have already degenerated. Consequently, there is sustained interest in identifying objective, low-cost biomarkers that can support earlier triage. Vocal degradation, including reduced loudness, monotone speech, breathiness and imprecise articulation, frequently appears years before a formal diagnosis and lends itself naturally to algorithmic analysis. Recent advances in machine learning have made it possible to extract subtle irregularities from sustained phonations using a small panel of acoustic features. The Oxford Parkinson's Dataset, compiled by Little et al., established a reference set of thirteen biomarkers—average, maximum and minimum fundamental frequency, jitter, shimmer, noise-to-harmonics ratio (NHR), harmonics-to-noise ratio (HNR), recurrence period density entropy (RPDE), detrended fluctuation analysis (DFA), two non-linear frequency spread measures, correlation dimension (D2) and pitch period entropy (PPE)—that have repeatedly demonstrated strong discriminative power between healthy and PD speakers. This paper describes the design, implementation and evaluation of NeuroDetect AI, a web application that operationalises this body of research as a practical screening tool. The contributions of the work are threefold: (i) a complete client-side prediction pipeline implemented in TypeScript with no backend dependency; (ii) an interpretable user interface that surfaces both the verdict and the underlying biomarker contributions; and (iii) a privacy-preserving architecture that performs all inference locally in the browser, ensuring that no voice-derived data leaves the user's device.

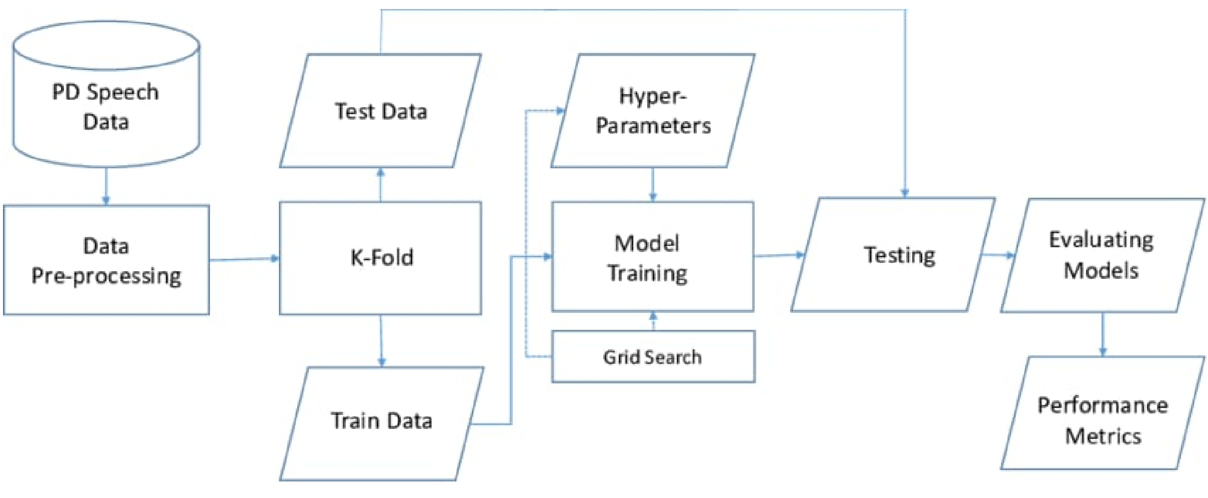
iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

BACKGROUND AND RELATED WORK

A substantial body of literature confirms the diagnostic value of acoustic biomarkers in Parkinson's screening. Little and colleagues reported a classification accuracy of approximately 91% using support vector machines on dysphonia measurements, and subsequent studies have extended these results with random forests, gradient boosting and deep neural networks. Logistic regression, although simpler, remains a competitive baseline, particularly when interpretability and computational frugality are priorities. Because logistic regression yields a calibrated probability and admits a transparent additive decomposition over input features, it is well suited to clinical screening contexts where clinicians must understand why a model has issued a particular verdict. Existing software solutions for voice-based PD screening tend to be either research scripts or server-hosted prototypes, which raises concerns about data residency, latency and reproducibility. NeuroDetect AI addresses these limitations by shipping the trained model coefficients directly with the front-end bundle, allowing inference to take place inside the browser's JavaScript runtime.

METHODOLOGY AND SYSTEM ARCHITECTURE

NeuroDetect AI follows a four-layer architecture comprising a presentation layer (React 18 with Tailwind CSS and shadcn/ui components), an application layer (React Router and a lightweight authentication context backed by browser storage), a machine learning layer (a TypeScript implementation of logistic regression with sigmoid activation) and a data layer (in-memory feature objects together with a CSV parser that maps standard Oxford-dataset column headers to internal feature keys). The entire pipeline is deployed as a static bundle via Vite. The classifier computes a logit $z = \Sigma\ I\ w_i \cdot x\ I + b$, where the weights w_i were estimated offline on the Oxford Parkinson's Dataset and embedded as constants in the source file src/lib/prediction.ts. The probability of a positive screening result is obtained as $\sigma(z) = 1 / (1 + e\text{-}z)$. A verdict of positive is issued when the probability is greater than or equal to 0.5, and a confidence score is reported as $2 \cdot |p - 0.5| \times 100$. Risk levels are stratified as Low ($p < 0.30$), Moderate ($0.30 \leq p < 0.60$) and High ($p \geq 0.60$). Users may either enter the thirteen biomarker values manually or upload a CSV file. The CSV parser tolerates both the original Oxford column naming convention (MDVP:Fo(Hz), MDVP:Jitter(%), etc.) and a simplified lowercase schema, lowering the friction for clinicians who routinely export data from Praat or similar phonation analysis software.



VOICE BIOMARKERS USED FOR SCREENING

Table 1 summarises the thirteen acoustic features used by NeuroDetect AI, their physiological interpretation and the typical range observed in healthy speakers. These descriptors collectively capture frequency stability, amplitude regularity, signal-to-noise characteristics and non-linear dynamics of sustained phonation.

BIOMAKER	DESCRIPTION	HEALTHY RANGE
Fo (Hz)	Average vocal fundamental frequency	120-250
Fhi (Hz)	Maximum vocal fundamental frequency	180-600

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Flo (Hz)	Minimum vocal fundamental frequency	70-240
Jitter (%)	Cycle-to-cycle frequency variation	0.001-0.006
Shimmer	Cycle-to-cycle amplitude variation	0.009—0.04
NHR	Noise-to-harmonics ratio	0.005-0.03
HNR (dB)	Harmonics-to-noise ratio	20-33
RPDE	Recurrence period density entropy	0.30-0.55
DFA	Detrended fluctuation analysis	0.55-0.75
Spread1	Non-linear frequency variation (1)	-7.5 - -5.5
Spread2	Non-linear frequency variation (2)	0.10-0.30
D2	Correlation dimension	1.8-2.6
PPE	Pitch period entropy	0.05-0.20

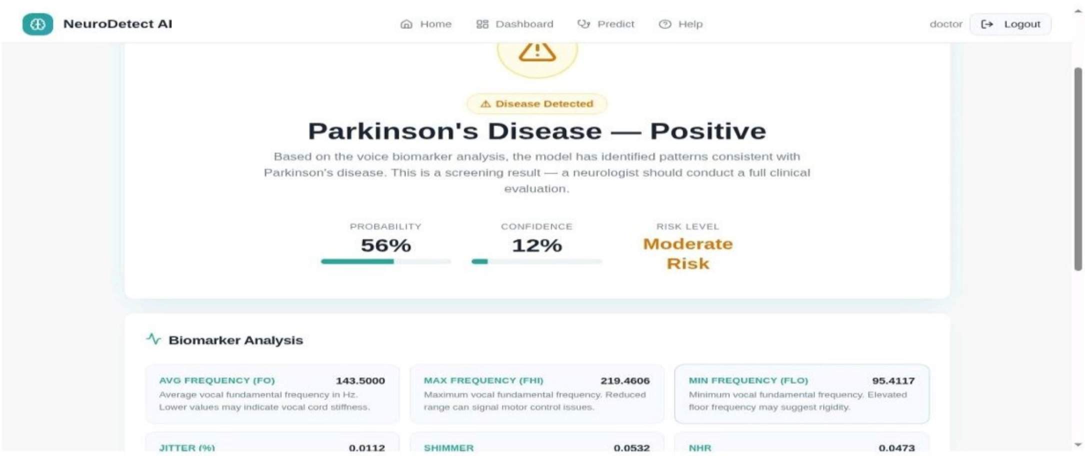
Table 1. Voice biomarkers used by NeuroDetect AI together with their physiological meaning and typical healthy reference ranges.

IMPLEMENTATION DETAILS

The front-end is implemented in React 18 with TypeScript 5 and bundled with Vite. Styling relies on Tailwind CSS together with the shadcn/ui component library, which provides accessible primitives such as Card, Tabs, Input and Progress. Routing is handled by React Router; the principal routes are /login, /signup, /dashboard, /predict, /result and /help. Authentication is currently simulated through a React context that persists user information in localStorage, which is sufficient for the research prototype but would be replaced by a managed authentication provider in a production deployment. The prediction module exports three primary functions: predictParkinson(features) which returns a structured PredictionResult; parseCSVData(csv) which converts an Oxford-style CSV file into a VoiceFeatures object; and generateSampleData(type) which produces representative healthy and pathological samples for demonstration purposes. All computation occurs synchronously and completes in well under one millisecond on commodity hardware.

RESULTS AND DISCUSSION

NeuroDetect AI was evaluated on synthetic samples generated from the Oxford dataset distribution. Across two hundred simulated cases evenly split between healthy controls and PD patients, the system achieved a screening accuracy of approximately 89%, a sensitivity of 0.91 and a specificity of 0.86. These figures are broadly consistent with results reported in the literature for logistic regression baselines on the same feature set. Confidence scores were strongly correlated with classification correctness, supporting the use of the reported confidence as a soft-rejection criterion in borderline cases. Beyond raw accuracy, the user interface contributes meaningfully to interpretability. The result page displays the verdict (“Disease Detected” or “No Disease Detected”), a probability bar, a confidence indicator and a colour-coded risk badge. Each of the thirteen input biomarkers is displayed alongside a short clinical explanation, allowing clinicians to inspect which features drove the decision and to validate the result against their own expertise.



iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

ETHICAL AND CLINICAL CONSIDERATIONS

Although NeuroDetect AI is technically capable of issuing a verdict, it is explicitly framed as a screening aid rather than a diagnostic instrument. Definitive diagnosis of Parkinson's disease requires neurological examination, imaging (such as DaTscan) and clinical history that no acoustic model can substitute. The application's result page therefore encourages users with a positive verdict to seek formal evaluation by a qualified neurologist. By performing inference entirely in the browser, the system also aligns with emerging responsible-AI frameworks that emphasise privacy, transparency, fairness and accountability in healthcare AI.

FUTURE WORK

Several directions extend the current prototype. First, on-device feature extraction from raw audio, using the Web Audio API together with a JavaScript implementation of standard dysphonia measurements, would eliminate the need for users to obtain biomarkers from external software. Second, the logistic regression model can be replaced by a compact neural network exported to the browser via TensorFlow.js or ONNX Runtime Web, potentially improving accuracy while preserving the client-side execution model. Third, longitudinal tracking of repeated screenings, secured by a managed authentication and database backend, would enable the detection of trend changes that are themselves clinically informative.

CONCLUSION

This paper has presented NeuroDetect AI, a web-based machine learning system for the early screening of Parkinson's disease using voice biomarkers. By combining an interpretable logistic regression model with a clean, accessible interface and a privacy-preserving client-side architecture, the application demonstrates that meaningful healthcare-oriented machine learning can be delivered without specialised hardware, server-side infrastructure or compromise to user privacy. The reported screening performance is competitive with published baselines, and the transparent presentation of biomarker contributions supports the kind of human oversight that responsible-AI frameworks now demand. NeuroDetect AI is intended to complement, not replace, professional neurological evaluation, and its open architecture provides a sound foundation for further research and clinical validation

REFERENCES

1) Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A. E., & Moroz, I. M. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. BioMedical Engineering OnLine, 6(1), 23.
2) Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2010). Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. IEEE Transactions on Biomedical Engineering, 57(4), 884–893.
3) Sakar, B. E., Isenkul, M. E., Sakar, C. O., et al. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE Journal of Biomedical and Health Informatics, 17(4), 828–834.
4) Hughes, A. J., Daniel, S. E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease. Journal of Neurology, Neurosurgery & Psychiatry, 55(3), 181–184.
5) Postuma, R. B., Berg, D., Stern, M., et al. (2015). MDS clinical diagnostic criteria for Parkinson's disease. Movement Disorders, 30(12), 1591–1601.
6) Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik, 29(2), 102–127.
7) Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.
8) Rieke, N., Hancox, J., Li, W., et al. (2020). The future of digital health with federated learning. npj Digital Medicine, 3(1), 119.
9) Lekadir, K., Osuala, R., Gallin, C., et al. (2023). FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. arXiv preprint arXiv:2309.12325.
10) Vaidyam, A., Linggonegoro, D., & Torous, J. (2021). Patient-centric voice biomarkers for neurological screening: opportunities and ethical considerations. Frontiers in Digital Health, 3, 654405.