FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 16, June, 2026, pp 589 – 593
DOI : https://doi.org/10.33003/fjs-2026-1016-5865

Analysis of Prosodic Feature Vectors from Speeches of Severity-based Major Depression Disorder Patients in Nigeria

[1]Stephen Gbenga Akinbode, [2]Iyinoluwa Tioluwani Idowu, [1]Olayinka Olufunmilayo Olusanya, [3]Temilade Aderounmu and *[4]Peter Adebayo Idowu

[1]Department of Computer Science, Tai Solarin Federal University of Education, Ijagun, Ogun State, Nigeria.
[2]Department of Nursing Science Elizade University, Ilara Mokin, Ondo State, Nigeria.
[3]Department of Pediatrics and Child Health Care, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria.
[4]Faculty of Computing Science and Engineering, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria.

*Corresponding authors' email: paidowu@oauife.edu.ng

ABSTRACT

Diabetes mellitus remains a major metabolic disorder requiring safer and affordable therapeutic alternatives. This study evaluated the safety profile, in vitro antidiabetic activity, and phytochemical composition of a formulated polyherbal syrup (PHS) prepared from *Vernonia amygdalina*, *Moringa oleifera*, *Zingiber officinale*, *Allium cepa*, and *Allium sativum*. Aqueous extracts of onion, garlic, and ginger were obtained by decoction of 20 g powdered samples in 100 mL distilled water at 40 °C for 3 h, while fresh leaves of *Moringa oleifera* and *Vernonia amygdalina* were extracted by manual juice expression and filtration. The extracts were combined with honey to produce the polyherbal syrup. Acute toxicity evaluation in albino rats using oral doses of 1000–16000 mg/kg body weight revealed no mortality or visible signs of toxicity, indicating an LD_{50} greater than 16000 mg/kg. Sub-acute toxicity studies conducted at doses of 25–400 mg/kg also showed no adverse behavioral or physiological changes. In vitro antidiabetic assays revealed that PHS is concentration-dependent. The PHS exhibited IC_{50} values (µg/mL) of 8.88 for glucose uptake, 2.55 for α-amylase inhibition, and 3.50 for α-glucosidase inhibition, compared with the standard drug acarbose, which showed IC_{50} values of 3.55, 1.89, and 3.39 µg/mL, respectively. Although the standard drug showed slightly higher potency, the PHS demonstrated comparable α-glucosidase inhibitory activity. GC–MS profiling revealed numerous bioactive phytochemicals including 2,3-dihydro-3,5-dihydroxy-6-methyl-4H-pyran-4-one,
5-hydroxymethylfurfural, thymine, n-hexadecenoic acid, and octadec-9-enoic acid. The combined presence of these compounds suggests possible synergistic pharmacological interactions responsible for the observed antidiabetic effects. These findings indicate that the formulated polyherbal syrup is safe and possesses promising antidiabetic potentialThis exploratory data analysis establishes a foundational benchmark for culturally tailored computational psychiatry by investigating acoustic and prosodic feature vectors within an underrepresented African cohort. Utilizing speech samples from 53 SCID-confirmed Nigerian participants across five PHQ-9-aligned severity levels, high-level descriptors encompassing rhythm, intonation, stress, and pause architectures were extracted to assess their viability as non-invasive biomarkers.

Statistical distributions reveal clinically plausible yet highly volatile acoustic signatures across depression strata. Temporal metrics demonstrate profound overlap and intra-class variance; notably, speech rate fluctuates drastically across all cohorts, with upper-whisker limits spanning approximately 0.09 to 0.14. Similarly, pause architectures exhibit extreme deviations, highlighted by prolonged silent outliers in the mild (~0.74 s) and none/minimal (~0.71 s) cohorts.

Conversely, frequency-domain and qualitative metrics manifest distinct ordinal stratifications. The severe category is characterized by a significantly elevated median fundamental frequency centre (f0_mean ≈ 243 Hz) coupled with restricted pitch modulation (f0_std median ≈ 85 Hz), indicating a highly compressed, monotone vocal delivery. In contrast, the moderately severe cohort represents a volatile linguistic transition zone, exhibiting maximum pitch variability (f0_std bounding box: ~88 Hz to ~151 Hz) alongside degraded voice quality, marked by the lowest median harmonics-to-noise ratio (~10.3).

Ultimately, while extreme severity boundaries exhibit distinct prosodic markers, the dense distributional overlaps observed among intermediate classes reflect the complex, non-linear progression of depression. These findings underscore the inherent subtleties of automated diagnostic screening and highlight the necessity of deploying multi-feature acoustic clusters over isolated indices to enhance model generalizability within African clinical contexts.

Received: 15 Aug. 2026
Accepted: 31 Aug. 2026
Published: 08 Sep. 2026

Keywords: Major depressive disorder, speech prosody, exploratory data analysis, acoustic features, feature fusion

INTRODUCTION

Major Depressive Disorder (MDD) is a debilitating condition characterized by persistent depressed mood, anhedonia, cognitive impairment, and vegetative symptoms [1]. In Nigeria, lifetime prevalence is estimated at 3.1%, affecting approximately 2-2.5 million people annually, yet 85-90% of individuals with mental health conditions lack access to proper care due to stigma, high costs, and severe specialist shortages [2][3]. Traditional diagnosis relies on time-intensive clinical interviews and self-report tools such as the Patient Health Questionnaire-9 (PHQ-9), which are subjective, resource-heavy, and difficult to sustain in low-resource settings. To address these diagnostic limitations, automated speech analysis has emerged as a promising, objective alternative. Pathological vocal changes in depression are heavily tied to psychomotor retardation, which disrupts the fine motor coordination of the articulatory musculature. Across acoustic literature, specific frequency-domain and temporal high-level descriptors (HLDs) consistently emerge as robust biomarkers of depressive severity. Specifically, pitch variability (fundamental frequency standard deviation, or f_0_std) and vocal intonation ranges serve as primary indicators of flattened affect. Longitudinal and ambulatory monitoring demonstrates that acute depressive episodes strongly correlate with compressed pitch dynamics, whereas clinical recovery or therapeutic intervention response directly mirrors an expansion in pitch range and variation [9][10].

Concurrently, the temporal architecture of speech undergoes distinct modifications during depressive states. Cognitive and physical slowing typically manifest as a significant deceleration in overall speech rate alongside prolonged and more frequent silent pause durations [10]. While these prosodic markers show stable diagnostic utility, acoustic feature behaviour can fluctuate significantly across specialized or atypical clinical sub-populations; for example, spectral metrics like Cepstral Peak Prominence demonstrate superior class separability when isolating depressive speech within populations with co-occurring neurogenic communication disorders like aphasia [11]. Despite the established clinical plausibility of these handcrafted features, existing computational models suffer from severe geographic and linguistic sampling biases. Current benchmarks are trained almost exclusively on Western or Chinese corpora (e.g., DAIC-WOZ) using localized scoring metrics, completely excluding African phonetic, dialectal, and cultural variations. This raises critical generalizability concerns when deploying automated tools in non-Western environments, as natural variations in accents, language pacing, and the integration of localized speech forms such as Nigerian Pidgin English can easily be misclassified as pathological prosodic indicators. This paper directly addresses this gap by conducting an exploratory data analysis (EDA) of a newly compiled, clinically validated Nigerian speech corpus. Rather than training a discrete classifier, this study systematically examines the distribution, internal variability, and boundary separability of acoustic and prosodic feature vectors extracted from Structured Clinical Interview for DSM (SCID)-confirmed voice samples across five PHQ-9-aligned severity levels. Specifically, we investigate the following core research questions: (1) Do prosodic features extracted from Nigerian English and Pidgin English speech exhibit systematic, interpretable gradients that correlate with MDD severity? (2) Which specific features provide the highest discriminative power between extreme clinical classes (none-minimal vs. severe)? (3) To what extent do intermediate severity classes (mild, moderate, moderately severe) display overlapping acoustic signatures? By evaluating these feature distributions, this exploratory analysis establishes a foundational benchmark for culturally tailored feature engineering and selection within African computational psychiatry

MATERIALS AND METHODS

Participants and Data Collection

Fifty-three participants (37 male, 16 female) were recruited from Yaba Neuro-Psychiatric Hospital, Lagos State, comprising both in-patients (n=5) and out-patients (n=48). All participants met inclusion criteria based on clinical diagnosis using the Structured Clinical Interview for DSM (SCID) within a two-week period. The study was approved by the Ethics Review Board of each participating institution. A psychiatrist conducted one-on-one, semi-structured interviews based on the PHQ-9, allowing spontaneous follow-up conversation. The interview was conducted in a low-noise environment with no special acoustic control facility to ensure that the model would be robust to acoustic variations encountered in deployment settings. Participants were given the PHQ-9 questionnaire and asked to read each question aloud before answering; the interviewer then engaged in natural dialogue. In cases where participants could not read legibly, the interviewer read the questions aloud. Interviewer speech was later edited out of the recordings. All participants spoke English or Nigerian Pidgin English to ensure linguistic consistency and avoid overbearing confounding language prosody with depression prosody. Average recording duration per participant ranged from 10 to 16 minutes. Severity levels were assigned based on PHQ-9 scores as follows (from the thesis operational definitions, Section 1.6): none-minimal (0-4), mild (5-9), moderate (10-14), moderately-severe (15-19), and severe (20-27). The distribution is shown in Table 1. Speech samples were recorded using a Galaxy A05 smartphone in M4A format, transcoded to WAV (16-bit, 44.1 kHz)

Table 1: Population and Gender Data Distribution

Severity Level	Male	Female	Total
None-minimal	6	1	7
Mild	17	4	21
Moderate	10	6	16
Moderately-Severe	4	0	4
Severe	0	5	5
Total	37	16	53

Source: Researcher’s Field Work (2025)

Data Preprocessing

Each raw audio file was loaded using librosa.load(), followed by spectral-subtraction-based noise reduction. Amplitude normalization was applied to the range $[-1, 1]$. To confirm that no crucial information was lost, features were extracted before and after preprocessing, distributions were compared using statistical measures (mean, variance, correlation coefficients). Pre-processing did not distort acoustic characteristics. Figure 1 shows the audio waveform and spectrogram of speech sample of patient 1 before and after audio cleaning.

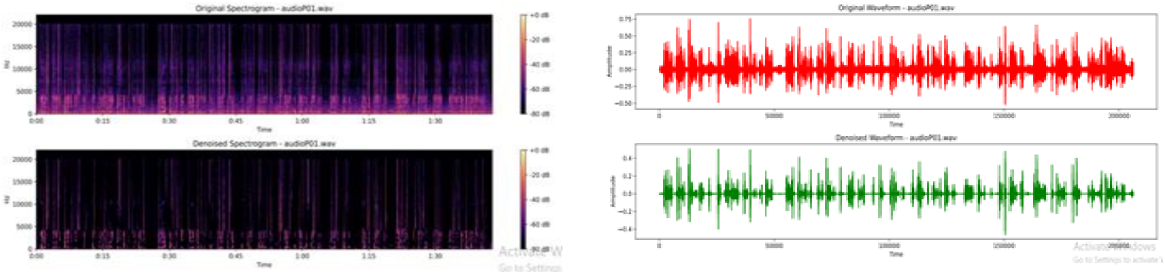

Figure 1: Waveform and Spectrogram Characteristics of Participant 1, Pre and Post Processing.
Source: Researcher's Work (2025)

Feature Extraction (Handcrafted)

All prosodic feature extraction was conducted using Librosa and Praat via Parselmouth. Each audio file of 44.1kHz was normalized, and trimmed to remove non-speech regions using librosa.effects.split(). Speech segments were divided into 25 ms frames with 50% overlap, and each frame was multiplied by a Hamming window to minimize spectral leakage.

LLDs were extracted first while their HLDs were computed accordingly. Pitch (F0) Descriptors were extracted using Praat's autocorrelation-based pitch tracking (75-600 Hz). Computed metrics include mean, median, minimum, maximum, standard deviation, and pitch slope. Intensity Descriptors were quantified using Praat's to Intensity function; extracted mean, standard deviation, minimum, and maximum intensity. Voice Quality Descriptors such as Harmonics-to-Noise Ratio (HNR) was computed to quantify spectral purity. Finally, the rhythm and temporal features were derived using librosa.effects.split(): cumulative speech duration, cumulative pause duration, frequency of pauses, mean pause length, speech-to-pause ratio, and speech rate.

Following extraction, all feature vectors were compiled (see Figure 2).

Figure 2: Extracted Acoustic and Prosodic Feature Elements of Participants
Source: Researcher's Computation (2025)

Exploratory Data Analysis Procedure

To compare the distribution of prosodic features across depression severity levels, boxplots were generated for *all HLDs* grouped by depression category (mild, moderate, severe). Each patient contributed a single value per feature, and the data were arranged such that each row corresponded to a patient, each column to a feature, and the depression level was used as the grouping variable.

All numeric prosodic features were visualized using boxplots generated with Python libraries Pandas, Seaborn, and Matplotlib. The target variable (LABEL, representing depression severity) was converted to categorical. Only numeric feature columns were selected, excluding identifiers (PATIENT_ID, AUDIO_FILE) and constant columns. For each feature, a Seaborn boxplot was generated with depression severity as the categorical x-axis, color-coded by severity class.

The analysis examined:

i. Central tendency (median) across severity classes
ii. Interquartile range (spread)
iii. Presence of outliers

Overlap between adjacent severity classes

RESULTS AND DISCUSSION

The demographic distribution (37 male, 16 female) reflects practical challenges of gender balance in clinical data collection. The predominance of mild and moderate cases (35 of 53 participants) mirrors typical outpatient populations, considering the underrepresentation of moderately-severe (n=4) and severe (n=5) classes, these small sample sizes mean statistical inferences about these classes are tentative.

The boxplot analysis from Figure 3 to Figure 8, revealed the following systematic patterns. Pitch variability (f0_std) exhibits a non-monotonic, bell-shaped distribution across intermediate strata before collapsing at the extreme severity threshold. The none-minimal and severe groups showed minimal overlap. Mild and moderate groups exhibited considerable overlap, indicating that pitch variability alone cannot reliably separate these adjacent intermediate classes. The distribution of acoustic spectral purity demonstrates an irregular, non-monotonic profile across severity boundaries rather than a uniform decrease. Vocal clarity features a localized degradation within intermediate cohorts, dropping from a baseline median of ~11.2 (none-minimal) down to ~10.7 in mild and hitting a study-minimum median of ~10.3 within the moderately severe cohort, signalling a hoarser, more breathy speech structure during these transition phases. Conversely and unexpectedly, the severe cohort manifests the highest overall voice quality with a study-maximum median of ~11.5, showing that spectral clarity recovers sharply at the terminal severity limit. Because no numeric means or percentage significance testing was conducted, these cross-class variations are evaluated qualitatively through boxplot distribution characteristics. For Speech-to-Pause Ratio, Lower ratios (more pause time relative to speech) were associated with higher severity. The moderate and moderately-severe groups were nearly indistinguishable on this metric, suggesting that pause patterns change gradually rather than stepwise. Temporal architectures confirm that lower speech-to-pause ratios signifying a heavier accumulation of pause intervals relative to active articulation are associated with advanced clinical severity boundaries. Crucially, the moderate and moderately severe cohorts manifest near-complete distributional overlap across pause parameters (pause_duration and avg_pause), rendering them nearly indistinguishable on these metrics. This lack of distinct separation indicates that transitional vocal pausing features evolve continuously along a gradual gradient rather than progressing in discrete, stepwise intervals.

For the None-minimal vs Severe classes, there exist clear separation on multiple features (pitch frequency standard deviation, speech rate, Harmonics to Noise Ratio, average fundamental frequency). This finding is clinically important depicting ability to distinguish between healthy and severely depressed individuals. For the Mild vs. Moderate classes, there are significant overlap on all measured features. This mirrors clinical reality where clinicians also struggle to differentiate these levels without detailed history. For the Moderate vs. Moderately-severe, Overlaps are almost complete, likely exacerbated by the very small sample size for moderately-severe class (n=4).

The gradient effect (progressive change in pitch variability, HNR, and speech-to-pause ratio across severity levels) supports the validity of prosodic features as continuous biomarkers of depression. This aligns with Cummins *et al.* [6], who reported that depressed speech is characterized by reduced pitch variability and increased vocal instability. The overlapping distributions between adjacent intermediate classes are not a failure of the features but rather reflect the inherent subtlety of prosodic changes across mild-to-moderate depression. Clinicians using PHQ-9 also show highest misclassification between these levels.

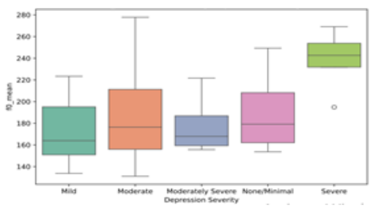

Figure 3: f0 mean distribution

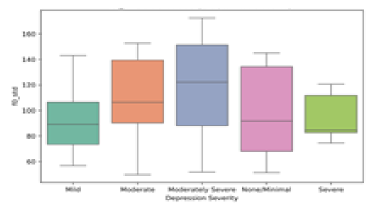

Figure 4: f0 std distribution

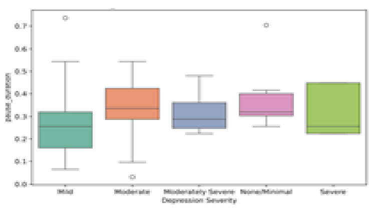

Figure 5: paused duration distribution

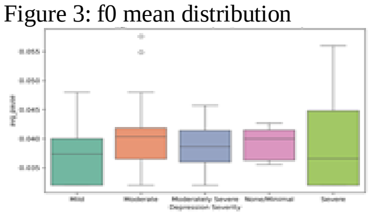

Figure 6: average pause distribution

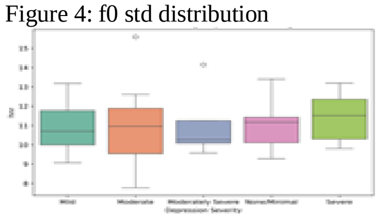

Figure 7: HNR distribution

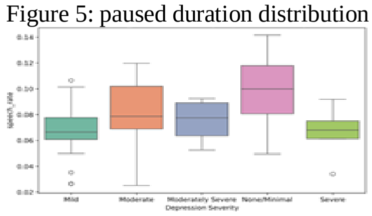

Figure 8:speech_rate distribution

This finding has important implications for future modelling. Ordinal regression loss functions may be more appropriate than standard categorical cross-entropy, as the latter treats each class as independent rather than ordered. The clear separability of extreme classes suggests that speech analysis may be most clinically useful for ruling out severe depression (high specificity) or flagging possible severe cases (high sensitivity) in primary care, rather than fine-grained discrimination among mild/moderate categories.

The gender imbalance (37 male, 16 female) limits generalizability, especially for female vocal characteristics (pitch norms differ by sex).

Finally, For the None-minimal vs Severe classes, clear vertical separation is maintained across frequency-domain metrics (f0_mean and f0_std) and spectral purity (HNR). However, temporal rate dimensions show a unidirectional boundary compression; while the None/Minimal cohort exhibits the highest overall vocal velocity ceiling (upper whisker ≈ 0.142), its lower whisker extends down to 0.050, demonstrating a low-end velocity overlap with the compressed box profile of the Severe cohort (~0.061 to ~0.075). This finding is clinically important, depicting that while extreme cohorts are highly discriminable via acoustic frequency-domain metrics, temporal speed tracking remains bounded by native baseline overlaps in localized dialects.

CONCLUSION

It must be acknowledged that the study possesses certain limitations i.e., small sample size, gender imbalance (37M/16F) and absence of feature distributions, gender-wise. Also, other limitations including the lack of statistical significance tests like p-values, effect sizes (Cohen's d), confidence intervals, ANOVA/Kruskal-Wallis results, lack of multivariate EDA, absent correlation matrix, lack of inter-rater reliability for feature extraction parameters and missing healthy control groups. However, This EDA of prosodic feature vectors from 53 Nigerian MDD patients demonstrates that handcrafted speech features exhibit systematic, clinically plausible gradients across PHQ-9 severity levels**.** Extreme classes (none-minimal vs. severe) are well-separated by pitch variability, HNR, and speech-to-pause ratio. Intermediate classes (mild vs. moderate, moderate vs. moderately-severe) show substantial overlap, reflecting the inherent continuity of depression symptomatology and the subtlety of prosodic changes in early-stage disease.

The gradient effect validates the use of prosodic features as continuous biomarkers for depression severity, supporting the ordinal nature of the PHQ-9 scale. The clear separability of severe depression suggests that speech analysis could be most clinically valuable as a screening support tool to flag high-risk individuals in primary care or community settings, where specialist access is limited. Future work should address limitations stated. This EDA establishes a foundation for future machine learning models by identifying which features carry the most discriminative signal and which severity pairs require more sophisticated modelling approaches (e.g., ordinal regression).

REFERENCES

Otte, C., Gold, S. M., Penninx, B. W., Pariante, C. M., Etkin, A., Fava, M., ... & Schatzberg, A. F. (2016). Major depressive disorder. *Nature reviews Disease primers*, *2*(1), 16065.

Patel, S. (2025, October 20). *How Many People Live with Depression?* Advance Study. Retrieved April 12, 2026, from https://advancestudy.org/how-many-people-live-with-depression/

Ugoeze, N. O. (2025). World Mental Health Day: 85% unreached as Nigeria buckles under massive care gap. *The Guardian*. https://guardian.ng/featured/world-mental-health-day-85-unreached-as-nigeria-buckles-under-massive-care-gap/#:~:text=Obindo%20informed%20that%20about%2085,treatment%20and%20inadequate%20medical%20facilities.

Lin, L., Chen, X., Shen, Y., & Zhang, L. (2020). Towards Automatic Depression Detection: A BiLSTM/1D CNN-Based Model. *Applied Science, 10*(23), 1-20

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech communication*, *71*, 10-49.

Seifpanahi, M. S., Ghaemi, T., Ghaleiha, A., Sobhani-Rad, D., & Zarabian, M. K. (2023). The association between depression severity, prosody, and voice acoustic features in women with depression. *The Scientific World Journal*, *2023*(1), 9928446.

Yang, Y., Fairbairn, C., & Cohn, J. F. (2012). Detecting depression severity from vocal prosody. *IEEE transactions on affective computing*, *4*(2), 142-150

Mengesha, Z., Heldreth, C., Lahav, M., Sublewski, Juliana., and Tuennerman, E. (2021). "I don't Think These Devices are Very Culturally Sensitive." Impact of Automated Speech Recognition Errors on African Americans Frontiers in Artificial Intelligence Volume 4 https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2021.725911/ https://doi.org/10.3389/frai.2021.725911 ISSN2624-8212

Mundt, J. C., Snyder, P. J., Cannizzaro, M. S., Chappie, K., & Geralts, D. S. (2007). Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of neurolinguistics*, *20*(1), 50-64.

Wadle, L. M., Ebner-Priemer, U. W., Foo, J. C., Yamamoto, Y., Streit, F., Witt, S. H., ... & Sirignano, L. (2024). Speech features as predictors of momentary depression severity in patients with depressive disorder undergoing sleep deprivation therapy: ambulatory assessment pilot study. *JMIR mental health*, *11*, e49222.

Gillespie, S., Moore, E., Laures-Gore, J., & Farina, M. (2016, March). Exploratory analysis of speech features related to depression in adults with Aphasia. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5815-5819). IEEE.