Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.12 n.57, p. 798-806. January/February, 2026.

DOI: https://doi.org/10.5935/jetia. v12i57.3075

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

VOICE DISORDER DETECTION: HYBRID GTCC-MFCC FUSION WITH SOURCE-BASED FEATURES OPTIMIZED FOR THE MALE SUB-COHORT

Aboubakr Missaoui[*1], Fatima Chouireb[2], Boubakeur Latreche[3], Abdelkerim Souahlia[4], Messaoud Linani[5] and Abdelaziz Rabhi[6]

[1, 2]Telecommunications, Signals & Systems Laboratory, University of Laghouat, Laghouat 03000, Algeria.
[3, 5]Laboratory of Computer Science and Applied Artificial Intelligence, University of Djelfa PO Box 3117, Djelfa 17000, Algeria.
[4, 6]Laboratory of Telecommunications and Smart Systems, Faculty of Sciences and Technology, University of Djelfa, Djelfa 17000, Algeria.

[1]https://orcid.org/0009-0008-9478-0015, [2]https://orcid.org/0000-0002-6049-6218, [3]https://orcid.org/0009-0007-9367-3368
[4]https://orcid.org/0000-0002-3393-1608, [5]https://orcid.org/0009-0001-6022-6648, [6]https://orcid.org/0000-0001-8684-4754

Email: *missboub@gmail.com, f.chouireb@lagh-univ.dz, b.latreche@univ-djelfa.dz, messaoud.linani@univ-djelfa.dz
abdelkerim.souahlia@univ-djelfa.dz, a.rabehi@univ-djelfa.dz

ARTICLE INFO

Article History
Received: December 7, 2025
Revised: January 10, 2026
Accepted: January 15, 2026
Published: February 28, 2026

Keywords:
Voice disorder diagnosis,
Gammatone coefficients,
Cepstral coefficients,
Gender-specific modeling,
Feature engineering,
Recursive Feature Elimination,
Clinical stability.

ABSTRACT

Voice disorders present a significant clinical challenge, adversely impacting communication and quality of life, thereby necessitating the development of reliable, non-invasive diagnostic systems. This research proposes an advanced diagnostic framework designed to overcome the limitations of traditional methodologies that rely exclusively on single-source spectral information. To achieve this, a systematic optimization methodology was applied to extract acoustic features from sustained vowel /a/ signals obtained from the male sub-cohort of the Saarbrücken Voice Database (SVD). The feature engineering phase integrated a comprehensive set of acoustic descriptors, combining advanced spectral coefficients (GTCC and MFCC) with traditional source-based features. To identify the most potent and non-redundant feature subset, a Recursive Feature Elimination (RFE) algorithm was rigorously employed across 100 iterative experiments, guaranteeing high statistical stability. This work substantiates two critical findings: First, that a hybrid strategy which intelligently combines auditory-inspired spectral features with traditional source-based biomarkers is necessary for maximizing diagnostic stability. Second, the RFE process validated the indispensability of key source-based metrics (CPP and GNR), which achieved high ranks in the final feature vector. The proposed framework achieved a peak Accuracy of 84.99%±4.50% and demonstrated good clinical stability in the early detection of voice disorders, confirming the necessity of integrating source-based biomarkers into advanced spectral analysis frameworks.

I. INTRODUCTION

The human voice serves as a complex bio-mechanical signal. Impairments in the vocal mechanism manifest as voice disorders, necessitating the development of reliable, non-invasive diagnostic systems. Current clinical protocols rely heavily on subjective auditory-perceptual assessments and invasive endoscopic examinations[1], [2]. However, these traditional methods suffer from inter-rater variability and a lack of standardization [3], [4]. Consequently, there is an urgent demand for automated decision support systems capable of providing objective, quantitative, and repeatable assessments. The field of automated voice pathology detection has recently shifted towards advanced spectral analysis (e.g., MFCCs) [5], [6]. While accurate, these methods face critical limitations that compromise clinical stability:

- Loss of Source Information: Spectral transformations focus primarily on the vocal tract and may inadvertently smooth out fine-grained micro-perturbations in the vocal folds (the source), which are critical markers for pathologies like dysphonia[7], [8].

- Black-Box Instability: Complex deep models are prone to overfitting when applied to small medical datasets (SVD), resulting in high performance variability [9], [10].

This research addresses these gaps by proposing a Hybrid Diagnostic Framework optimized specifically for the male sub-cohort of the SVD. We hypothesize that optimal diagnostic stability requires re-integrating "Source-Based" biomarkers (like CPP and GNR) with advanced "Auditory-Based" spectral features (GTCC and MFCC). The main contributions of this paper are:

1. Hybrid Feature Fusion: Systematic integration of GTCC/MFCC with traditional source-perturbation features to fully capture glottal instability.
2. Focused Gender Optimization: Development of a highly optimized diagnostic model for the male vocal mechanism to establish a robust performance benchmark.
3. Rigorous Feature Selection: Application of Recursive Feature Elimination (RFE) [11] to identify a compact, clinically stable feature subset.

Validation of Source Features: Statistical evidence proving that traditional features (CPP and GNR) are indispensable for high-specificity diagnosis.

II. RELATED WORK

The landscape of pathological voice assessment has undergone a significant transformation, evolving from the analysis of isolated acoustic markers to complex Deep Learning frameworks driven by spectral imagery. The existing literature can be delineated into three primary research streams. Early research centered on quantifying the biological dysfunction of the vocal folds (the Source). These foundational studies prioritized **perturbation metrics** such as Jitter (frequency instability) and Shimmer (amplitude instability) alongside Fundamental Frequency (F0) and noise indices like Harmonics-to-Noise Ratio (HNR) [4], [7], [8]. These descriptors were pivotal because they directly measured the physical irregularity of vocal vibration. Concurrently, cepstral analysis gained prominence, with Mel-Frequency Cepstral Coefficients (MFCCs) [5], [6] and Linear Prediction Cepstral Coefficients (LPCCs) [12], [13] serving as the standard inputs for classical algorithms. Among these, Support Vector Machines (SVMs)))[14], [15] and Random Forests (RFs) [16]) were widely adopted for their ability to handle tabular acoustic data.

The last decade introduced a methodological pivot towards treating voice signals as visual time-frequency maps. Convolutional Neural Networks (CNNs) [9], [17], [18] and Recurrent Neural Networks (RNNs) [19], [20] became the dominant architectures, processing standard Spectrograms [21] as well as biologically motivated representations like Cochleagrams . [9] and Gammatone Spectral Lines (GTSL) [22]. However, while Deep Learning models have demonstrated superior accuracy in specific experimental setups, they face substantial hurdles in clinical applicability. Primarily, these models exhibit a propensity for overfitting when trained on limited, single-source medical datasets . [9], [10]. This limitation is particularly acute in multi-class classification tasks [23] , where the lack of generalization hinders their use as reliable diagnostic tools [23] .To mitigate data scarcity, recent studies have explored advanced techniques such as Semi-Supervised Learning (SSL) [14], [24] and Transfer Learning [21], [25]. Others have proposed multimodal fusion strategies[26] and dimensionality reduction methods like PCA and LDA [15], [27] to condense feature spaces.

Despite these innovations, a critical gap persists in validation standards. Methodologies fluctuate between static data splitting. [28], [29] and Cross-Validation [30],[31]. More critically, the widespread failure to report performance variability specifically the omission of Standard Deviation (σ) obscures the true stability of these systems [30], [25] . Furthermore, modest Balanced Accuracy scores suggest that differentiating between specific pathologies remains a resolved challenge[24]. While spectral representations (e.g., GTCC) simulate human auditory perception, a significant limitation in modern research is the exclusive reliance on these spectral envelopes at the expense of traditional source features. It is crucial to recognize that spectral transformations (like MFCC) were originally engineered for speech recognition; their inherent filtering processes tend to smooth out the temporal micro-perturbations and non-linearities that are diagnostic hallmarks of pathology [4], [7]. Relying solely on the spectral shape (the Filter) risks discarding vital information regarding vocal fold stability (the Source). Consequently, this study argues that optimal diagnosis requires an Integrative Hybrid Methodology that bridges these two domains. We validate this approach through:

1. **Comparative Analysis:** Systematically isolating and evaluating the diagnostic power of Traditional, GTCC, and MFCC features.
2. **Hybrid Fusion:** Merging macroscopic Spectral Shape (GTCC/MFCC) with microscopic Source Stability metrics (Jitter, CPP, GNR, Entropy).
3. **Statistical Optimization:** Utilizing Recursive Feature Elimination (RFE) to identify a mathematically stable feature subset.

Gender-Specific Focus: Developing a specialized.

III. PROPOSED METHODOLOGY

This study implements an integrated workflow for voice disorder classification, progressing from raw signal processing to optimized machine learning predictions.

III.1 SYSTEM INPUT AND PRE-PROCESSING

The system processes raw audio signals (sustained vowel /a/) in .wav format. To ensure high data quality prior to feature extraction, the signals undergo two critical pre-processing steps:

1. Silence Removal: An automated algorithm identifies and excises silent segments at the start and end of recordings to ensure analysis is restricted to active phonation.

2. Signal Normalization: Peak Normalization is applied to standardize the dynamic range, scaling maximum amplitude to a unified reference level (0 dB) to mitigate variations in recording gain.

III.2 FEATURE EXTRACTION AND ENGINEERING

A hybrid extraction strategy is adopted, integrating temporal, spectral, and cepstral features. These are divided into three distinct analytical pathways to capture different dimensions of the vocal signal.

III.2.1 Global Signal Analysis (Source-Based Features)

The signal is analyzed as a single unit to derive statistical indicators of overall vocal stability and quality. These metrics capture fundamental characteristics related to the vocal source mechanism (see Table 1).

Table 1: Source features and statistical parameters extracted from the overall signal.

Feature Category	Description & index	Diagnostic Relevance	References
Temporal Stability	Shimmer (1): Cycle-to-cycle amplitude variation.	Indicates irregular vocal fold closure and intensity variation; critical for detecting amplitude-related disorders.	[7], [8], [13]
Frequency Dynamics	Instantaneous Frequency: (2) SD of IF, (3) SD of its derivative, (4) mean zero-crossing intervals, and (5) SD of peak intervals.	Captures global instability, abrupt frequency shifts, and aperiodicity in vocal fold vibration.	[32], [33]
Voice Quality	(6) Glottal-to-Noise Ratio (GNR) & (7) Cepstral Peak Prominence (CPP)	Low GNR indicates breathiness (air leakage), while reduced CPP strongly correlates with hoarseness severity.	[13], [34]
Spectral Statistics	Spectral Measures: (8) Spectral Entropy, (9) SD of Spectral Energies, and (10) Bandwidth variability.	Quantifies randomness in spectral energy distribution, indicating roughness or turbulent phonation.	[35], [36], [37]

Source: Authors, (2026).

III.2.2. Multi-Window Analysis (Dynamic Features)

This approach monitors subtle dynamic changes over time, often missed by global static measurements. The signal is segmented into 200-sample windows with 50% overlap (see Table 2).

Table 2: Statistics of features across time windows.

Feature Category	Description & index	Diagnostic Relevance	References
Formant Analysis	Vocal Tract Resonance: (11) (12) Mean and SD of F1, (13) (14) Mean and SD of F2(15) (16) Mean and SD of F3.	Mean reflects vocal tract shape configuration; SD reveals articulatory motor instability (e.g., Ataxia).	[38], [39]
F0 Variability	Fundamental Frequency Stats (2): (17) SD of F0 and (18) SD of its derivative across windows.	Precise measure of F0 dispersion; effective for detecting vocal tremor and laryngeal spasms.	[7], [8]
Waveform Regularity	Regularity Indices (4): (19) SD of Shimmer, (20) SD of peak/valley intervals, (21) SD of distances peaks and (22) SD of distances valleys	Reveals inconsistent amplitude control and higher-order irregularities in periodic patterns.	[40], [41]
Signal Complexity	Non-linear Dynamics (4) (23) SD of Skewness, (24) SD of Energy, (25) SD of Entropy, and (26) SD of Signal Quality within windows.	Quantifies local changes in waveform shape and randomness, reflecting disordered vocal dynamics.	[42], [35], [43]
Noise Balance	HNR Variability (3): (27-29) Sum, Mean, and SD of HNR across windows.	Provides insight into the dynamic balance between harmonics and noise, highlighting intermittent breathy segments.	[34], [44]
Micro-Tremor	MT-Jitter Variability: (30-35) SD of MT-Jitter values calculated at 6 different thresholds (0%, ±25%, \dots).	Elevated SD values suggest instability in frequency production, enhancing detection of neuromuscular dysfunction.	[45], [46]

Source: Authors, (2026).

III.2.3Advanced Cepstral Analysis (MFCC & GTCC)

This pathway extracts perceptual features that simulate the human auditory mechanism.

- MFCC (Mel-Frequency Cepstral Coefficients): Processes the signal on the non-linear Mel frequency scale.
- GTCC (Gammatone Cepstral Coefficients): Utilizes Gammatone filters to provide a more precise simulation of human cochlear filters, enhancing sensitivity to pathology-related characteristics.
- Dimensionality Reduction: Instead of utilizing full high-dimensional matrices, the Standard Deviation (SD) of each coefficient is calculated over time. This yields reduced vectors representing "Spectral Variability," a robust indicator of voice instability.

Outcome: All features (Global, Multi-Window, and Reduced Cepstral Vectors) are combined to form the complete **Initial Feature Vector (91 features)**.

III.3 FEATURE SELECTION AND OPTIMIZATION

To mitigate the "Curse of Dimensionality" and enhance classifier generalization:

- Recursive Feature Elimination (RFE): RFE was executed over 100 iterations. This rigorous repetition ensures Statistical Robustness, ranking features by their consistent discriminative power rather than chance.
- Normalization: Min-Max Normalization was applied to scale feature values to the [0, 1] range.

III.4 MODELING AND EVALUATION

Model Configuration and Optimization: The classification core relies on a Cost-Sensitive Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel.To explicitly address dataset imbalance, a class-weighting scheme was implemented where penalty weights are assigned inversely proportional to class frequencies, ensuring the minority class is prioritized during training. Feature vectors were scaled using Min-Max Normalization to the [0, 1] range prior to input. Hyperparameters (specifically *Box Constraint* and *Kernel Scale*) were fine-tuned using Bayesian Optimization utilizing the Expected Improvement Plus acquisition function over 20 iterations. The system's robustness was validated using a Stratified 5-Fold Cross-Validation protocol, ensuring consistent class distribution across all folds.

III.5 VALIDATION FOR SYSTEM ROBUSTNESS

The study relies on rigorous statistical validation to ensure the model's reliability and clinical generalizability, utilizing four key metrics reported with both the Mean (μ) and Standard Deviation(±σ).

- **Overall Accuracy:** Measures the percentage of correctly classified samples across all classes.
- **Specificity:** Quantifies the model's ability to exclude healthy cases (minimizing false positives), which is crucial for clinical confidence.
- **Sensitivity:** Measures the model's ability to correctly detect pathological cases (minimizing false negatives).
- **F1-Score:** Provides a balanced harmonic mean of precision and sensitivity, offering a single measure of performance especially under data imbalance.

A low Standard Deviation (±σ) accompanying each metric indicates that the system exhibits high stability and good generalization capability when applied to unseen data.

III.6 DATABASE

This study utilizes the Saarbrücken Voice Database (SVD), a primary reference in voice pathology research known for its statistical diversity spanning both genders and various age groups. To focus the analysis on voice disorders with complex or non-specific etiologies (e.g., tumor diseases, inflammations, and neurological disorders), a strict filtering methodology was applied. This process excluded cases attributable to clear, direct structural trauma, such as injuries resulting from surgical interventions (e.g., intubation damage). Crucially, the analysis was restricted to male subjects only to develop a gender-optimized diagnostic model. The filtering process resulted in a final male-specific dataset totaling 526 voice samples, organized into a binary diagnostic task:

- 218 Healthy Samples
- 308 Pathological Samples (comprising 15 distinct pathological subcategories, confirming the comprehensive coverage of the complex disorders analyzed in this research).

IV. RESULTS AND DISCUSSIONS

This section is dedicated to presenting the comprehensive findings of the study, structured to rigorously validate the necessity of the proposed hybrid methodology. The analysis begins by evaluating the effectiveness of the Recursive Feature Elimination (RFE) algorithm in identifying a unique, physiologically relevant feature subset that validates the core hypothesis of integrating **Source** and **Filter** features. Following the internal validation of the feature optimization path, the discussion then proceeds to contextualize the final model's performance within the broader research landscape by comparing our results against key state-of-the-art studies that utilized the Saarbrücken Voice Database (SVD). This process affirms the system's stability and competitive edge based on rigorous statistical protocols.

IV.1 RFE RANKING VECTOR

The feature selection process, executed via RFE over 100 iterations, yielded a robust ranking vector that determines the optimal feature subset. This vector, comprising all 91 features (Traditional, Multi-Window Aggregated, and Reduced Cepstral), is presented below, ordered from most significant (Rank 1) to least significant (Rank 91). Features highlighted in red represent the Source-Based Features base signal features.

RFE Ranking Vector (Rank 1 to 91): vector = [84, 85, 52, 7, 62, 83, 80, 82, 38, 6, 54, 51, 37, 41, 8, 35, 87, 90, 91, 11, 53, 70, 79, 5, 14, 66, 23, 63, 65, 12, 50, 55, 71, 40, 78, 76, 13, 17, 30, 88, 60, 42, 81, 67, 56, 58, 25, 31, 61, 39, 34, 20, 33, 28, 10, 86, 57, 89, 32, 16, 29, 45, 77, 27, 22, 75, 26, 4, 24, 59, 48, 44, 15, 24, 47, 18, 74, 9, 46, 73, 49, 68, 2, 64, 36, 69, 3, 72, 43, 19, 1]

The RFE results provide crucial evidence supporting the core hypothesis: optimal classification relies on the intelligent hybrid integration of advanced spectral features and specialized source perturbation biomarkers. This validation is demonstrated by the high ranking of key traditional features.

IV.1.1 Prominence of Source Quality Measures (CPP and GNR).

The intrinsic diagnostic power of source-based quality metrics was not nullified by the dominance of the GTCC/MFCC coefficients, but rather confirmed:

1. Feature 7 (Cepstral Peak Prominence - CPP): This feature was ranked remarkably high at Position 4. Had the spectral transforms been capable of fully encoding the information related to roughness and vocal quality, CPP would have been expected to drop to significantly lower ranks. Its high placement unequivocally proves that CPP represents a unique traditional measure of Source Quality that the Vocal Tract-focused spectral transforms cannot substitute.
2. Feature 6 (Glottal-to-Noise Ratio - GNR): This feature was also secured within the top tier at Position 10. As GNR (the ratio of the periodic component to noise) is essential for evaluating breathiness, its inclusion among the top ten features demonstrates that these carefully selected traditional features directly reflect fundamental physical aspects of voice production, essential for clinical differentiation.

IV.1.2 Validation of the Core Claim

The rigorous RFE ranking validates the core claim of the study, providing statistical backing for the necessity of the hybrid framework:

1. Selected Traditional Features are Not Inferior: Traditional features chosen based on their intrinsic diagnostic importance (CPP, GNR) demonstrate an added value superior to spectral transforms in capturing specific, micro-level aspects of Source Quality.
2. Optimal Stability Requires Hybridization: While GTCC/MFCC coefficients contribute significantly due to their comprehensive representation of the Vocal Tract (Filter), the system achieves maximal classification stability and specificity through the mandatory inclusion of these specialized source perturbation biomarkers.

This robust evidence supports the final claim: optimal diagnostic stability and accuracy are achieved exclusively through the intelligent hybrid integration of both advanced spectral shape features and specialized source perturbation biomarkers.

IV.2 VALIDATION OF RFE OPTIMIZATION PATH

Table 3: Performance Trajectory of Spectral and Hybrid Models by Feature Count (N).

N	Composition	Acc (μ±σ) %	Spec (μ±σ) %	Sens (μ±σ) %	F1-Score (μ±σ) %
10	1	83.09±6.54	85.85±11.80	81.19±7.48	84.87±5.93
	2	83.66±4.17	85.33±9.16	82.46±7.72	85.43±4.01
20	1	82.52±4.03	84.87±7.27	80.86±6.35	84.35±3.84
	2	84.42±2.91	88.10±6.47	81.83±5.37	85.97±2.89
	1	82.71±5.02	86.71±9.54	79.89±7.97	84.30±4.83
	2	70.92±3.78	96.81±5.92	52.62±8.33	67.62±6.11
25	1	83.65±3.52	86.72±8.06	81.50±3.51	85.39±2.92
	2	84.99±4.50	89.95±7.62	81.49±5.91	86.36±4.23
30	1	82.51±4.37	86.71±12.27	79.56±7.07	84.17±3.83
	2	84.04±2.00	85.32±8.62	83.13±6.22	85.86±2.06
35	1	81.95±5.08	86.71±10.09	78.59±7.82	83.51±4.90
	2	84.04±4.93	87.20±9.54	81.81±6.43	85.68±4.43
	1	82.51±3.58	86.28±10.24	79.87±6.55	84.20±3.36
	2	83.27±4.87	85.81±11.49	81.50±5.30	85.12±3.99

Source: Authors, (2026).

The primary goal of the feature selection phase was to rigorously validate the necessity of the hybrid approach by comparing its performance trajectory against the purely spectral baseline model. Table 3 explicitly illustrates the dynamic changes in classification metrics for both the Spectral Baseline 1(GTCC+ MFCC) and the Hybrid Winner 2 (FTRS+ GTCC+ MFCC) across critical number feature(N). This provides crucial evidence for the effectiveness of the RFE process and the added value of the Filtered Traditional Features (FTRS). The detailed trajectory presented in Table 3 offers several critical insights into the feature space structure and the role of the hybrid strategy.

IV.2.1 Superiority of the Hybrid Trajectory

The performance trajectory, as illustrated in Table 3, demonstrates the **consistent superiority** of the Hybrid Winner (FTRS + GTCC + MFCC) over the Spectral Baseline (GTCC + MFCC) throughout the entire Recursive Feature Elimination (RFE) path (from N=10 up to N=50) in terms of both Accuracy and Specificity. This superiority culminated in the **Final Peak Validation** at N=30, where the Hybrid model achieved its absolute maximum accuracy of **84.99% ± 4.50%**.

This peak contrasts sharply with the best Spectral Baseline performance of **83.65% ± 3.52%** (also at N=30), demonstrating a **+1.34% absolute gain in accuracy** attributed directly to the inclusion of the Filtered Traditional Features (FTRS). Furthermore, the **Role of Hybridization** is clearly validated by the most significant gain observed in **Specificity** (↑ 3.23% at the optimum). This confirms the hypothesis that traditional, source-based biomarkers (like CPP and GNR) offer a unique layer of information necessary to distinguish between subtle, non-pathological acoustic variations and genuine, clinically relevant vocal damage.

IV.2.2 Feature Space Stability and Critical Instability

Analysis of the performance trajectory provides a clear view of the feature subset's structural stability as dimensions are removed. Both models exhibited strong performance and low Standard Deviation (σ) within the range N in [10, 20], indicating that the initial features selected are highly robust and non-redundant. However, the data reveals a **Critical Instability Point** for the Hybrid Winner at N=25. Here, the Sensitivity plummeted dramatically (**52.62%**) while Specificity surged (**96.81%**). This signifies that the RFE process, upon removing a single critical feature between N=30 and N=25, caused the model to become overly conservative, resulting in a high rate of false negatives. This critical point underscores the importance of precisely identifying the stable optimum (N=30) and avoiding feature subsets that compromise diagnostic reliability.

IV.2.3 Optimizing for Computational Efficiency.

The performance peak at N=30 for both models, followed by a general decay beyond N=40, confirms the efficiency of the selection process. The research successfully identified a minimal, high-yield feature subset that maximizes diagnostic power while minimizing computational overhead, which is essential for clinical deployment.

IV.3 CONTEXTUALIZATION AND COMPARATIVE ANALYSIS

The reliability of our diagnostic model is primarily rooted in its rigorous internal validation protocol. Utilizing the Stratified 5-Fold Cross-Validation approach and consistently reporting the Mean and Standard Deviation for all metrics ensures a stable, statistically defensible assessment of the system's quality. This internal consistency is paramount for clinical adoption. The purpose of the subsequent comparison (Table 4) is to contextualize this internally validated performance within the broader research landscape. We compare our results specifically against studies using the SVD (Saarbrücken Voice Database), a fundamental benchmark in the field. We must acknowledge the inherent limitations of cross-study comparisons; the absence of a unified protocol results in significant heterogeneity across dataset sizes, pathology types, and evaluation metrics.

Therefore, our aim is not to claim absolute superiority, but rather to demonstrate that our feature processing and classification methodology is highly competitive and stands alongside the best reported state-of-the-art performances using the same reference database. The comparative analysis demonstrates that our Hybrid Model not only achieves high accuracy but does so under rigorous statistical validation protocols. Outperformance over Fixed Split: Our Hybrid Model achieved an accuracy of 84.99% ± 4.50%, surpassing the highest reported results using the less robust Fixed Split protocol (e.g., [25] at 81.6% and [21]at 82%). Fixed Split results are often optimistically biased and fail to reflect the model's stability when exposed to unseen data.

Best-in-Class Cross-Validation (CV) Performance: When compared against studies employing the more reliable Cross-Validation (CV) protocol, our model's superiority is evident. The highest reported CV accuracy was 82.58% ± 3.02% by Yagnavajjula et al. [6]. Our model achieves an absolute gain of +2.41% over this benchmark, which validates the efficacy of our Hybrid Fusion strategy and the RFE feature selection process. Several recent studies (e.g., [25] and [21]) employ Deep Learning (CNN) models on spectral representations (Spectrograms), yet their results do not exceed 82%.

- Validation of Study Hypothesis: Our performance proves that a methodology based on selective Feature Engineering and Hybrid Fusion can outperform deep learning approaches in the context of a small, specialized medical dataset like the SVD, where DL is susceptible to overfitting and instability.
- Power of Hybrid Features: The significant gain in accuracy (84.99%) is directly attributable to the intelligent integration of Source features (CPP, GNR) which cannot be effectively captured by purely spectral features (Spectrograms MFCCs). One of the most critical aspects of this comparison is the explicit reporting of the Standard Deviation. The majority of studies relying on Fixed Split omit \sigma, diminishing confidence in the clinical repeatability of their results.

Confirmation of Stability: Our use of the 5-Fold Cross-Validation protocol and reporting of ± 4.50% ensures an honest assessment of the system's stability. While our standard deviation is slightly higher than that reported by Yagnavajjula et al. (± 3.02%), the absolute gain in accuracy (+2.41%) justifies this trade-off, especially since we successfully identified the Optimal Stability Point (N=30) which maximizes diagnostic reliability.

Table 4: Comparison of State-of-the-Art Voice Pathology Classification Studies on the SVD.

Study	# Of Healthy voices	# Of Pathologies voices and types	Vocal Tasks	Feature Domains	Classifier	Statistical Validation	Acc%
[25]	869Augmented by using MUSAN and MIT IR Survey	520 (105 Laryngitis, 41 Leukoplakia, 63 Edema, 205 Paralysis, 62 SD, 44 Polyp) (Augmented using MUSAN and MIT IR Survey)	/a/	Mel Spectrogram	CNN Few-shot Transfer Learning (Pretrained ResNet-18)	Few-Shot Meta-Testing	73.7
[28]	687	1354 (71 Pathologies)	/a/, /i/, /u/	MFCCs	OSELM	Fixed Split	81.48%
[17], [47]	482	482 (140 Laryngitis, 41 Leukoplakia, 68 Reinke's Edema, 213 RLNP, 22 Carcinoma, 45 Polyps)	/a/	spectrograms	CNN **CNN-CDBN**	Fixed Split	77 71
[23]	595	1090	/a/	MFSC	CNN (DCA ResNet)	Fixed Split	81.6
[21]	506	506 Organic Dysphonia (Laryngitis, Leukoplakia, Reinke's Edema, Carcinoma, VFP)	/a/, /i/, /u/	spectrograms	CNN VGG16	Fixed Split	82
[31]	60	60 SD, 60 RLNP	/a/, /i/, /u/, Sentence	(WST)-based Features	feed-forward NN	10-Fold CV	82.58 ± 3.02 78.64 ± 3.55
[14]	587	231 (146 Hyperfunctional Dysphonia, 85 VFP)	/a/	wav2vec 2.0	SVM	5-Fold CV	M. 75.65 F. 74.50
[6]	357	357	/a/	Mel Spectrogram	2D CNN + specAugmenter	4-Fold CV	73,4
Ours	218	308	/a/	MFCC+GTCC+ based Features	SVM	5-Fold CV	**84.99±4.50**

Source: Authors, (2026).

V. CONCLUSION

This study successfully developed and validated a Hybrid Diagnostic Framework optimized specifically for the male sub-cohort of the SVD to enhance the stability and accuracy of automated voice pathology detection. By addressing the critical limitation of spectral models namely, the loss of Source Information our work introduces a robust methodology capable of improving clinical diagnosis. The main contributions and findings are:

Validation of the Hybrid Hypothesis: The rigorous RFE process confirmed that the optimal feature subset requires the mandatory inclusion of carefully selected source-based biomarkers (CPP and GNR) alongside advanced spectral features (GTCC/MFCC). This proves that maximal specificity depends on integrating both the Source and Filter domains of vocal production.

Achieving Optimal Stability: The Hybrid Model successfully identified a minimal, high-yield feature subset at N=30, achieving a maximum accuracy of 84.99% ± 4.50%. This represents a statistically significant improvement over the purely spectral baseline, positioning our system competitively against the current state-of-the-art while ensuring low performance variability (\sigma).

Physiological Insight: The analysis of the Critical Instability Point at N=25 highlighted the unique diagnostic power of micro-perturbation features, underscoring the necessity of using the highly stable N=30 subset to avoid compromising diagnostic Sensitivity.

Future research will focus on two key areas:

1. Gender Comparative Analysis: Expanding the current model to include a fully optimized female sub-cohort framework to perform a comprehensive comparative study across genders, thereby fulfilling the initial goal of Gender-Specific Optimization.
2. Multi-Class Differentiation: Shifting the focus from binary classification (Pathological/Healthy) to a multi-class system capable of accurately differentiating between specific voice disorder types.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Aboubakr Missaoui.
Methodology: Aboubakr Missaoui, Fatima Chouireb.
Investigation: Aboubakr Missaoui, Boubakeur Latreche.
Discussion of results: Fatima Chouireb, Abdelkerim Souahlia, Boubakeur Latreche.
Writing – Original Draft: Aboubakr Missaoui.
Writing – Review and Editing: Boubakeur Latreche, Fatima Chouireb.
Ressources: Aboubakr Missaoui, Abdelaziz Rabhi, Messaoud Linani.
Supervision: Fatima Chouireb, Abdelkerim Souahlia.
Approval of the final text: Aboubakr Missaoui, Fatima Chouireb, Abdelkerim Souahlia, Boubakeur Latreche, Messaoud Linani. Abdelaziz Rabhi.

VII. REFERENCE

[1] A. am Zehnhoff-Dinnesen, K. Neumann, B. Wiskirska-Woźnica, et T. Nawka, Phoniatrics I. Springer, 2020. doi: 10.1007/978-3-662-46780-0.

[2] P. Barche, « Acoustic analysis of voice disorders from clinical perspective », PhD Thesis, International Institute of Information Technology Hyderabad, 2024.

[3] M. Ur Rehman, A. Shafique, Q.-U.-A. Azhar, S. S. Jamal, Y. Gheraibia, et A. B. Usman, « Voice disorder detection using machine learning algorithms: An application in speech and language pathology », Engineering Applications of Artificial Intelligence, vol. 133, p. 108047, juill. 2024, doi: 10.1016/j.engappai.2024.108047.

[4] H. Ankışhan et S. Ç. İnam, « Voice pathology detection by using the deep network architecture », Applied Soft Computing, vol. 106, p. 107310, juill. 2021, doi: 10.1016/j.asoc.2021.107310.

[5] N. Q. Abdulmajeed, B. Al-Khateeb, et M. A. Mohammed, « A review on voice pathology: Taxonomy, diagnosis, medical procedures and detection techniques, open challenges, limitations, and recommendations for future directions », Journal of Intelligent Systems, vol. 31, no 1, p. 855-875, juill. 2022, doi: 10.1515/jisys-2022-0058.

[6] F. Javanmardi, S. R. Kadiri, et P. Alku, « A comparison of data augmentation methods in voice pathology detection », Computer Speech & Language, vol. 83, p. 101552, janv. 2024, doi: 10.1016/j.csl.2023.101552.

[7] G. Li, Q. Hou, C. Zhang, Z. Jiang, et S. Gong, « Acoustic parameters for the evaluation of voice quality in patients with voice disorders », Ann Palliat Med, vol. 10, no 1, p. 130-136, janv. 2021, doi: 10.21037/apm-20-2102.

[8] L.-C. Keung, K. Richardson, D. Sharp Matheron, et V. Martel-Sauvageau, « A Comparison of Healthy and Disordered Voices Using Multi-Dimensional Voice Program, Praat, and TF32 », Journal of Voice, vol. 38, no 4, p. 963.e23-963.e38, juill. 2024, doi: 10.1016/j.jvoice.2022.01.010.

[9] R. Islam, E. Abdel-Raheem, et M. Tarique, « Cochleagram to Recognize Dysphonia: Auditory Perceptual Analysis for Health Informatics », IEEE Access, vol. 12, p. 59198-59210, 2024, doi: 10.1109/ACCESS.2024.3392808.

[10] H. Kim et al., « Convolutional Neural Network Classifies Pathological Voice Change in Laryngeal Cancer with High Accuracy », Journal of Clinical Medicine, vol. 9, no 11, p. 3415, oct. 2020, doi: 10.3390/jcm9113415.

[11] K. M. Alalayah, E. M. Senan, H. F. Atlam, I. A. Ahmed, et H. S. A. Shatnawi, « Automatic and Early Detection of Parkinson's Disease by Analyzing Acoustic Signals Using Classification Algorithms Based on Recursive Feature Elimination Method », Diagnostics, vol. 13, no 11, p. 1924, mai 2023, doi: 10.3390/diagnostics13111924.

[12] J. I. Godino-Llorente, S. Aguilera-Navarro, et P. Gómez-Vilda, « LPC, LPCC and MFCC parameterisation applied to the detection of voice impairments », in 6th International Conference on Spoken Language Processing (ICSLP 2000), ISCA, oct. 2000, vol. vol. 3, 965-968-0. doi: 10.21437/ICSLP.2000-695.

[13] R. Islam, M. Tarique, et E. Abdel-Raheem, « A Survey on Signal Processing Based Pathological Voice Detection Techniques », IEEE Access, vol. 8, p. 66749-66776, 2020, doi: 10.1109/ACCESS.2020.2985280.

[14] S. Tirronen, S. R. Kadiri, et P. Alku, « Hierarchical Multi-Class Classification of Voice Disorders Using Self-Supervised Models and Glottal Features », IEEE Open J. Signal Process., vol. 4, p. 80-88, 2023, doi: 10.1109/OJSP.2023.3242862.

[15] N. Souissi et A. Cherif, « Speech recognition system based on short-term cepstral parameters, feature reduction method and Artificial Neural Networks », in 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Monastir, Tunisia: IEEE, mars 2016, p. 667-671. doi: 10.1109/ATSIP.2016.7523163.

[16] C. Guo, F. Chen, Y. Chang, et J. Yan, « Applying Random Forest classification to diagnose autism using acoustical voice-quality parameters during lexical tone production », Biomedical Signal Processing and Control, vol. 77, p. 103811, août 2022, doi: 10.1016/j.bspc.2022.103811.

[17] H. Wu, J. Soraghan, A. Lowit, et G. Di Caterina, « Convolutional Neural Networks for Pathological Voice Detection », in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI: IEEE, juill. 2018, p. 1-4. doi: 10.1109/EMBC.2018.8513222.

[18] R. Islam, E. Abdel-Raheem, et M. Tarique, « Voice pathology detection using convolutional neural networks with electroglottographic (EGG) and speech signals », Computer Methods and Programs in Biomedicine Update, vol. 2, p. 100074, 2022, doi: 10.1016/j.cmpbup.2022.100074.

[19] H. A. A. E. Aal, S. A. Taie, et N. El-Bendary, « An optimized RNN-LSTM approach for parkinson's disease early detection using speech features », Bulletin EEI, vol. 10, no 5, p. 2503-2512, oct. 2021, doi: 10.11591/eei.v10i5.3128.

[20] S. Hidaka, Y. Lee, K. Wakamiya, T. Nakagawa, et T. Kaburagi, « Automatic Estimation of Pathological Voice Quality Based on Recurrent Neural Network Using Amplitude and Phase Spectrogram », in Interspeech 2020, ISCA: ISCA, oct. 2020, p. 3880-3884. doi: 10.21437/interspeech.2020-3228.

[21] L. Vavrek, M. Hires, D. Kumar, et P. Drotar, « Deep convolutional neural network for detection of pathological speech », in 2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI), Herl'any, Slovakia: IEEE, janv. 2021, p. 000245-000250. doi: 10.1109/sami50585.2021.9378656.

[22] C. Zhou, Y. Wu, Z. Fan, X. Zhang, D. Wu, et Z. Tao, « Gammatone spectral latitude features extraction for pathological voice detection and classification », Applied Acoustics, vol. 185, p. 108417, janv. 2022, doi: 10.1016/j.apacoust.2021.108417.

[23] H. Ding, Z. Gu, P. Dai, Z. Zhou, L. Wang, et X. Wu, « Deep connected attention (DCA) ResNet for robust voice pathology detection and classification », Biomedical Signal Processing and Control, vol. 70, p. 102973, sept. 2021, doi: 10.1016/j.bspc.2021.102973.

[24] D. Ribas, M. A. Pastor, A. Miguel, D. Martinez, A. Ortega, et E. Lleida, « Automatic Voice Disorder Detection Using Self-Supervised Representations », IEEE Access, vol. 11, p. 14915-14927, 2023, doi: 10.1109/access.2023.3243986.

[25] J.-H. Won et D.-H. Kim, « Metric-Based Few-Shot Transfer Learning Approach for Voice Pathology Detection », IEEE Access, vol. 12, p. 159226-159238, 2024, doi: 10.1109/access.2024.3480523.

[26] I. Kwon et al., « Diagnosis of Early Glottic Cancer Using Laryngeal Image and Voice Based on Ensemble Learning of Convolutional Neural Network Classifiers », Journal of Voice, vol. 39, no 1, p. 245-257, janv. 2025, doi: 10.1016/j.jvoice.2022.07.007.

[27] T. N. Sainath, A. Mohamed, B. Kingsbury, et B. Ramabhadran, « Deep convolutional neural networks for LVCSR », in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada: IEEE, mai 2013, p. 8614-8618. doi: 10.1109/ICASSP.2013.6639347.

[28] F. T. Al-Dhief, N. M. A. Latiff, N. N. N. Abd. Malik, M. M. Baki, N. Sabri, et M. A. A. Albadr, « Dysphonia Detection Based on Voice Signals Using Naive Bayes Classifier », in 2022 IEEE 6th International Symposium on Telecommunication Technologies (ISTT), Johor Bahru, Malaysia: IEEE, nov. 2022, p. 56-61. doi: 10.1109/istt56288.2022.9966535.

[29] N. Mu'azzah Abdul Latiff et al., « Voice pathology detection using machine learning algorithms based on different voice databases », Results in Engineering, vol. 25, p. 103937, mars 2025, doi: 10.1016/j.rineng.2025.103937.

[30]L. Verde, G. De Pietro, et G. Sannino, « Voice Disorder Identification by Using Machine Learning Techniques », IEEE Access, vol. 6, p. 16246-16255, 2018, doi: 10.1109/ACCESS.2018.2816338.

[31] M. K. Yagnavajjula, K. R. Mittapalle, P. Alku, S. R. K., et P. Mitra, « Automatic classification of neurological voice disorders using wavelet scattering features », Speech Communication, vol. 157, p. 103040, févr. 2024, doi: 10.1016/j.specom.2024.103040.

[32] L. Cohen et C. Lee, « Standard deviation of instantaneous frequency », in International Conference on Acoustics, Speech, and Signal Processing, Glasgow, UK: IEEE, 1989, p. 2238-2241. doi: 10.1109/icassp.1989.266910.

[33] B. T. Nguyen, Y. Wakabayashi, K. Iwai, et T. Nishiura, « Analysis of derivative of instantaneous frequency and its application to voice activity detection », Applied Acoustics, vol. 181, p. 108116, oct. 2021, doi: 10.1016/j.apacoust.2021.108116.

[34] G. D. Krom, « A Cepstrum-Based Technique for Determining a Harmonics-to-Noise Ratio in Speech Signals », Journal of Speech, Language, and Hearing Research, vol. 36, no 2, p. 254-266, avr. 1993, doi: 10.1044/jshr.3602.254.

[35] A. Al-Nasheri et al., « Voice Pathology Detection and Classification Using Auto-Correlation and Entropy Features in Different Frequency Regions », IEEE Access, vol. 6, p. 6961-6974, 2018, doi: 10.1109/access.2017.2696056.

[36] M. Markaki et Y. Stylianou, « Voice Pathology Detection and Discrimination Based on Modulation Spectral Features », IEEE Trans. Audio Speech Lang. Process., vol. 19, no 7, p. 1938-1948, sept. 2011, doi: 10.1109/tasl.2010.2104141.

[37] H. Lathrop et Adams School of Dentistry, Department of Orthodontics, « Orthognathic Speech Pathology: Understanding How Class III Jaw Disharmonies Influence Speech Utilizing Spectral Moment Analysis. Masters Thesis », 2019, doi: 10.17615/WKF7-MV24.

[38] L. M. T. Jesus, J. Martinez, A. Hall, et A. Ferreira, « Acoustic Correlates of Compensatory Adjustments to the Glottic and Supraglottic Structures in Patients with Unilateral Vocal Fold Paralysis », BioMed Research International, vol. 2015, p. 1-9, 2015, doi: 10.1155/2015/704121.

[39] J.-W. Lee, H.-G. Kang, J.-Y. Choi, et Y.-I. Son, « An Investigation of Vocal Tract Characteristics for Acoustic Discrimination of Pathological Voices », BioMed Research International, vol. 2013, no 1, p. 758731, 2013, doi: 10.1155/2013/758731.

[40] V. Dellwo, A. Leemann, et M.-J. Kolly, « Speaker idiosyncratic rhythmic features in the speech signal », présenté à Interspeech 2012, Portland (OR), USA: Interspeech Conference Proceedings, sept. 2012, p. 1-4. doi: 10.5167/uzh-68554.

[41] K. E. Hajal, E. Hermann, S. Hovsepyan, et M. Magimai. -Doss, « Unsupervised Rhythm and Voice Conversion to Improve ASR on Dysarthric Speech », 2 juin 2025, arXiv: arXiv:2506.01618. doi: 10.48550/arXiv.2506.01618.

[42] A. Verikas, M. Bacauskiene, A. Gelzinis, E. Vaiciukynas, et V. Uloza, « Questionnaire- versus voice-based screening for laryngeal disorders », Expert Systems with Applications, vol. 39, no 6, p. 6254-6262, mai 2012, doi: 10.1016/j.eswa.2011.12.037.

[43] P. R. Scalassara, M. E. Dajer, C. D. Maciel, et J. C. Pereira, « VOICE SIGNALS CHARACTERIZATION THROUGH ENTROPY MEASURES », in Proceedings of the First International Conference on Bio-inspired Systems and Signal Processing, Funchal, Madeira, Portugal: SciTePress - Science and and Technology Publications, 2008, p. 163-170. doi: 10.5220/0001065401630170.

[44] J. F. T. Fernandes, D. Freitas, A. C. Junior, et J. P. Teixeira, « Determination of Harmonic Parameters in Pathological Voices—Efficient Algorithm », Applied Sciences, vol. 13, no 4, p. 2333, févr. 2023, doi: 10.3390/app13042333.

[45] N. Hanna, J. Smith, et J. Wolfe, « Frequencies, bandwidths and magnitudes of vocal tract and surrounding tissue resonances, measured through the lips during phonation », The Journal of the Acoustical Society of America, vol. 139, no 5, p. 2924-2936, mai 2016, doi: 10.1121/1.4948754.

[46] A. Geredakis, M. Karala, N. Ziavra, et E. Toki, « Preliminary measurements of voice parameters using Multi Dimensional Voice Program », World Journal of Research and Review, vol. 5, no 1, 2017.

[47] H. Wu, J. Soraghan, A. Lowit, et G. Di-Caterina, « A Deep Learning Method for Pathological Voice Detection Using Convolutional Deep Belief Networks », in Interspeech 2018, ISCA: ISCA, sept. 2018, p. 446-450. doi: 10.21437/interspeech.2018-1351.