Franklin Open 17 (2026) 100759

Contents lists available at ScienceDirect

Franklin Open

journal homepage: www.elsevier.com/locate/fraope

Complexity-driven feature selection for enhancing tuberculosis detection

Sana Ben Mahjouba [a,b], Johannes C. Ayena [a,b,*], Youssef Ouakrim [a,b], Karim Loulou [a,b], Simon Grandjean Lapierre [c,d], Mihaja Raberahona [e,f], Neila Mezghani [a,b]

[a] Applied Artificial Intelligence Institute (I2A), TELUQ University, Montréal, Québec, Canada
[b] Open Innovation Laboratory in Health Technologies, Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM), Montréal, Québec, Canada
[c] Immunopathology Axis, CRCHUM, Montréal, Québec, Canada
[d] Department of Microbiology, Infectious Diseases and Immunology, Université de Montréal, Montréal, Québec, Canada
[e] CHU Joseph Rasera Befelatanana, Antananarivo, 101, Analamanga, Madagascar
[f] Équipe de Recherche Clinique en Maladies Infectieuses (ERCMI), Madagascar

ARTICLE INFO

Keywords:
Tuberculosis screening
Cough acoustics
Feature complexity
Computational efficiency
Interpretable AI
Digital health

ABSTRACT

Most existing machine learning approaches for tuberculosis (TB) screening typically utilize large, high-dimensional acoustic feature sets without examining their intrinsic discriminative power. To address this gap, we introduce a complexity-based feature selection approach that evaluates temporal and spectral descriptors using Fisher score (F1), class-distribution overlap (F2), and Shannon entropy (F4). Applied to the CODA-TB dataset (9772 audio recordings from 1105 participants), the proposed method identified 7 highly informative features from the original 26 features, primarily consisting of mel-frequency cepstral coefficients (MFCC) derivatives and spectral-shape measures. The resulting model achieved performance comparable to full-feature baselines while reducing feature dimensionality by 73% and computational cost by up to 14×. Comparative evaluation against four established feature selection techniques, supported by ablation and statistical analyses, confirmed the efficiency and robustness of the complexity-driven strategy, with no statistically significant loss in performance. These findings highlight the potential of lightweight, interpretable, and computationally efficient models for TB cough-based screening in resource-constrained environments.

1. Introduction

Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, remains one of the world's leading infectious killers despite decades of research and diagnostic progress [1,2]. While TB can manifest in extrapulmonary sites including kidneys, spine, brain, and lymph nodes, this study focuses exclusively on pulmonary TB (PTB). PTB accounts for approximately 83% of all TB cases and is the most contagious form of the disease [3]. Early and accurate detection of active PTB is critical for interrupting transmission chains and improving patient outcomes. Despite its importance, significant diagnostic challenges persist in resource-limited settings [4,5].

Conventional diagnostic modalities include sputum smear microscopy, chest radiography, and nucleic acid amplification tests such as Xpert MTB/RIF. However, these methods face substantial barriers related to cost, invasiveness, technical requirements, and accessibility in primary care settings [6,7]. Such limitations are particularly acute in high-burden, low-resource regions where TB prevalence is highest, underscoring the urgent need for affordable, rapid, and non-invasive alternatives for TB screening and triage [8].

Cough sound analysis has emerged as a promising digital biomarker for PTB [9–14], leveraging the fact that persistent cough is both a cardinal symptom and a primary transmission mechanism. These previous studies have demonstrated the potential of acoustic features such as mel-frequency cepstral coefficients (MFCCs), spectral centroid, zero-crossing rate (ZCR), etc., in machine learning models for TB detection. More recently, deep learning approaches based on convolutional and recurrent neural networks have further improved TB cough classification by automatically learning discriminative representations directly from cough signals [15–19]. For instance, Rajasekar et al. [15] reported precision values up to 97% in classifying TB coughs using bidirectional long short-term memory (BiLSTM) models. Despite these advances, deep learning models generally require substantial computational resources and provide limited interpretability compared with feature-based approaches. Relying on either empirically chosen feature sets or automatically learned deep representations, existing TB cough

* Corresponding author at: Applied Artificial Intelligence Institute (I2A), TELUQ University, Montréal, Québec, Canada.
E-mail address: johannes.ayena@teluq.ca (J.C. Ayena).

https://doi.org/10.1016/j.fraope.2026.100759
Received 22 December 2025; Received in revised form 10 August 2026; Accepted 26 August 2026
Available online 2 September 2026
2773-1863/© 2026 The Authors. Published by Elsevier Inc. on behalf of The Franklin Institute. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

detection approaches have not explicitly investigated the intrinsic discriminative complexity of individual features or their robustness to inter-patient variability. Indeed, a key limitation is that many methods still depend on basic acoustic descriptors that may be insufficiently sensitive to the subtle, non-linear patterns characteristic of TB-associated coughs.

To overcome these limitations, our study introduces a complexity-driven strategy that comprehensively evaluates both temporal and spectral features using rigorous mathematical metrics such as Fisher score, Shannon entropy, and class-overlap volume. Beyond standard features, we incorporate higher-order characteristics including MFCC temporal derivatives, spectral skewness, and kurtosis, which may more effectively capture the pathological acoustic signatures of PTB.

The primary contributions of this work are threefold: (1) introduce a systematic complexity analysis framework for cough acoustic features in TB detection; (2) validate the discriminative power of complexity-ranked features against traditional selection methods through comprehensive comparative evaluation; and (3) demonstrate the clinical utility of the proposed approach for building interpretable, robust, and scalable screening tools deployable in low-resource environments.

Our work builds upon information-theoretic foundations in feature selection but extends them to the specific challenges of respiratory acoustics. While some studies [20,21] have applied individual complexity metrics in isolation, our combined approach addresses the multi-faceted nature of cough discrimination: Fisher score allows class separability, overlap minimization reduces ambiguity in borderline cases, and entropy maximization captures the rich acoustic diversity of pathological coughs. This tripartite framework is particularly suited for TB detection, where cough characteristics vary substantially across disease stages, comorbidities, and individual patient factors.

The remainder of this paper is organized as follows: In Section 2, we describe the proposed methodology, including the dataset, preprocessing pipeline, feature extraction, complexity-based metrics, comparative feature selection strategies, and validation procedure. In Section 3, we present the experimental results, covering feature complexity analyses, ranked subset evaluation, classifier performance, comparative benchmarking, and computational efficiency, along with a discussion of key findings, limitations, and implications for real-world deployment. Finally, in Section 4, we conclude the paper and outline directions for future work.

2. Methods

The evaluation of cough sound features for TB detection is performed through a multi-step methodology. We begin this section by describing the dataset and the extraction of temporal and spectral features from segmented audio recordings. Next, we report the calculation of feature complexity metrics, including Fisher score, class distribution overlap, and Shannon entropy, which form the basis of our complexity-driven feature selection. Participant-level data partitioning allows realistic training and testing splits, while comparative feature selection methods and ranked feature subsets are employed to identify the most discriminative features. Finally, six classifiers are applied to assess the predictive performance of selected features, and a performance validation framework compares the effectiveness of different selection strategies in clinically relevant scenarios. All main components of this methodology are summarized in Fig. 1, and the following subsections describe these components in detail.

2.1. Study dataset

We used the COugh Diagnostic Algorithm for Tuberculosis (CODA-TB) dataset, an openly available under data use agreement collection of cough recordings [22]. It comprises 9772 samples from seven countries (1105 participants aged 18 years and older, both TB-positive and TB-negative cases confirmed through clinical testing). Of the 9772 recordings, 6842 (70.0%) were TB-negative and 2930 (30.0%) were TB-positive, reflecting a moderate class imbalance. This dataset provides a representative sample base in order to evaluate the discriminative potential of acoustic features in TB detection. Our analysis focused on acoustic properties extracted from $0.5\,\mathrm{s}$ cough segments. This duration was selected based on the work of Rajasekar et al. [15], who demonstrated that the explosive phase of the cough contains the most discriminative acoustic information for respiratory disease analysis. Focusing on this phase enables consistent extraction of the most informative acoustic characteristics while avoiding less informative portions of the cough signal.

2.2. Feature extraction

Based on previous studies [23–25], the cough recordings we extracted were resampled to 16 kHz prior to feature extraction (a common choice for cough/speech acoustic analysis that reduces computational cost while preserving the frequency range relevant to cough acoustics) [15,16]. Feature extraction was performed using a $25\,\mathrm{ms}$ Hamming window with 50% overlap, resulting in a consistent segmentation procedure across all recordings. No additional explicit noise-filtering step was applied beyond this standard preprocessing, and the recordings were analyzed as provided by the CODA-TB dataset. This configuration provided an optimal balance between temporal resolution and spectral stability for cough analysis. Afterwards, a comprehensive set of 26 acoustic features was computed (Table 1), capturing both temporal and spectral characteristics of cough signals. Temporal features characterize the cough waveform in the time domain, providing insights into amplitude variations and signal dynamics, whereas spectral features characterize the cough signal in the frequency domain, revealing resonant properties and spectral energy distribution [26].

2.3. Complexity-driven feature analysis

This subsection presents the complexity metrics and how the features were ranked according to established criteria.

2.3.1. Complexity metrics

Complexity analysis was performed on both temporal and frequency-domain features (Table 1) to assess their discriminative power for differentiating TB-positive (Class 1) from TB-negative (Class 2). We employed the following metrics to quantify class separability:

- The Fisher Score (F1) compares inter-class variance to intra-class variance [27]. It reflects how well a feature separates TB-positive from TB-negative samples, with higher values indicating better class separability.

$$F1 = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \tag{1}$$

  where μ_1, μ_2 are respectively the feature means for Class 1 and Class 2; σ_1^2, σ_2^2 are the corresponding variances.
- The Overlap Volume (F2) measures the overlap between class distributions [28]. It reflects the degree of overlap between the TB-positive and TB-negative feature distributions, with lower values indicating better class separability.

$$F2 = \int \min\left(p_1(f), p_2(f)\right)\, df \tag{2}$$

  where $p_1(f)$ and $p_2(f)$ are respectively the probability density functions of feature f for Class 1 and Class 2.
- The Feature Entropy (F4) calculates Shannon entropy of the feature distribution [29]. It reflects the amount of information or variability contained in a feature across the dataset, with higher values indicating richer discriminatory information.

$$F4 = -\sum_{i=1}^{M} p(f_i) \log_2 p(f_i) \tag{3}$$

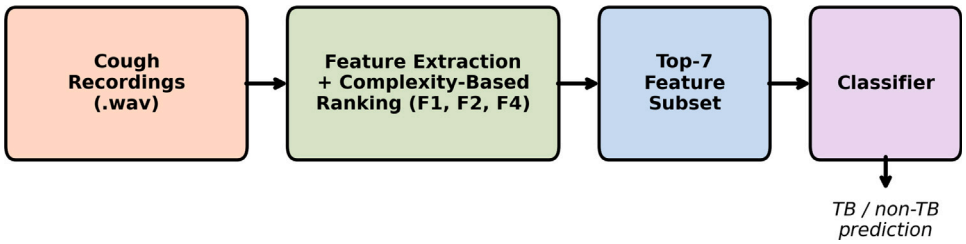

Fig. 1. Signal processing pipeline for cough-based TB detection.

Table 1
Temporal and spectral features extracted from cough signals.

Temporal Features		Spectral Features	
Name (label)	Description	Name (label)	Description
RMS (f1)	Signal energy, representing cough intensity and amplitude variations.	Spectral Centroid (f7)	Frequency indicating signal brightness or dominant frequency region.
ZCR (f2)	Rate of sign changes in the waveform, indicating periodicity.	Spectral Bandwidth (f8)	Spread of the spectrum around the centroid.
Duration (f3)	Total length of the cough event, indicating severity and effort.	Spectral Rolloff (f9)	Frequency below which 85% of spectral energy is contained, representing bandwidth and high-frequency content.
Skewness (f4)	Measure of asymmetry in the waveform, reflecting temporal distribution.	Spectral Flatness (f10)	Quantifies noisiness vs tonality, distinguishing harmonic/noisy components.
Kurtosis (f5)	Peakiness of the amplitude distribution, indicating burst-like characteristics.	MFCCs (f11-f22)	Twelve features (including delta/delta-delta for dynamic variations), each containing 13 coefficients representing the spectral envelope. f11 to f22 correspond to Mean, SD, Min, Max of MFCC, MFCC Delta, and MFCC Delta2.
Variance (f6)	Signal variability, capturing dynamic range and stability.	Minimum Spectrogram Values (f23-f26)	Captures baseline energy levels across frequency bands, reflecting consistency in spectral energy.

Note. RMS: Root Mean Square, ZCR: Zero Crossing Rate, MFCCs: Mel-Frequency Cepstral Coefficients, SD: Standard Deviation, Min: Minimum, Max: Maximum.

where M is the number of discrete feature states, and $p(f_i)$ is the probability of the feature state f_i.

2.3.2. Feature selection criterion

Each feature (i) is assigned a normalized combined complexity score (S_i) defined as:

$$S_i = \frac{F1_{i,\text{norm}} + (1 - F2_{i,\text{norm}}) + F4_{i,\text{norm}}}{3}, \tag{4}$$

where the subscript *norm* denotes min–max normalization across all 26 features.

Rather than using a learned or tuned weighting scheme, the three complexity metrics are assigned equal weights (1/3 each). This design choice maintains methodological simplicity, avoids introducing an additional hyperparameter requiring optimization on a limited dataset, and reduces the risk of overfitting the feature selection criterion [30, 31]. To evaluate the robustness of this design choice, we conducted a supplementary sensitivity analysis by increasing, in turn, the relative contribution of each metric, while keeping the remaining metrics equally weighted. Across all tested weighting schemes, the selected feature subsets and the resulting classification performance remained qualitatively stable around the equal-weight configuration, indicating that the proposed criterion is not overly sensitive to moderate changes in the weighting strategy.

A feature i is retained if S_i exceeds a data-driven threshold β, defined as the 70th percentile of the training distribution. In our experiments, this corresponds to $\beta = 0.61$. To further support this threshold, we conducted a sensitivity analysis by varying the number of retained features from k=4 to k=26 (i.e., the complete feature set) using the Random Forest classifier. Classification accuracy remained within a narrow range (0.670–0.679), while the highest F1-scores were consistently obtained when retaining between k=4–7 features (F1 = 0.229–0.293). Beyond this range, performance gradually declined as additional, less informative or partially redundant features were included. These findings indicate that the selected threshold lies within a stable near-optimal operating region rather than representing an arbitrary empirical choice. Furthermore, the selected threshold consistently identified the same seven highly discriminative features across multiple training splits. The predictive effectiveness of this selected feature subset is further illustrated in Results section. The corresponding ROC curves show consistent performance across six evaluated classifiers, supporting the robustness of the proposed feature selection strategy.

2.4. Comparative feature selection methods

We compared our proposed method, complexity-driven feature selection, with four existing feature selection approaches: (a) SelectKBest with F-test ranking [32], (b) sequential forward selection [33], (c) sequential backward elimination [34], and (d) Random Forest importance-based ranking [35]. These four benchmarks were chosen because they are widely adopted, computationally tractable at this dataset scale, and jointly span the three main families of feature selection strategies: filter methods (SelectKBest), wrapper methods (forward/backward sequential selection), and embedded methods (Random Forest importance), providing a representative and diverse basis for comparison [36–38]. The next lines distinguish between two selection strategies. The first selection strategy evaluates classification

performance of each method using a predefined subset of seven top-ranked features selected by the proposed complexity-based criterion. Conversely, the second strategy selection allows algorithms to automatically determine its own optimal subset size based on their internal criteria. This dual perspective enables a balanced comparison between fixed-size feature selection and adaptive feature selection.

2.4.1. Fixed feature selection

This approach utilized fixed feature order derived from Fisher score, class overlap, and entropy analyses. A separate validation framework compared four distinct approaches: (i) baseline using all 26 features with Random Forest classifier, (ii) proposed method using the top 7 complexity-ranked features (MFCC derivatives, spectral rolloff, skewness, kurtosis), (iii) SelectKBest with top 7 F-test ranked features, and (iv) Random Forest importance with top 7 ranked features. All the four distinct approaches were evaluated under identical conditions to ensure fair comparison. The feature selection algorithms were trained exclusively on training set participants to prevent data leakage, with feature importance scores computed only from training data.

2.4.2. Adaptive feature selection

This approach utilized all the features highlighted in Table 1. It allows algorithms to automatically determine the optimal subset size based on their internal criteria.

The feature methods selection employed in this study were evaluated under identical initial conditions, meaning that each method operated on the same training/testing partitions, the same normalized feature space, and the same incremental subset evaluation protocol.

2.5. Model training and validation

To achieve clinically realistic evaluation and prevent data leakage, we implemented participant-level data partitioning using *GroupShuffleSplit*. Since each participant contributes multiple cough recordings, traditional random splitting could place recordings from the same individual in both training and test sets [39]. This would artificially inflate performance metrics, as the model might learn participant-specific acoustic patterns rather than genuine TB-discriminative features.

2.5.1. Group-wise data partitioning

Our rigorous approach guarantees that all cough recordings from a single participant are exclusively assigned to either training (80% of participants) or test sets (20% of participants). The final dataset partitioning resulted in 7825 recordings for training and 1947 recordings for testing, allowing complete participant-level separation between sets. This partitioning strategy mimics real-world deployment scenarios where the system encounters completely new patients, providing a reliable estimate of generalization performance. The GroupShuffleSplit strategy was employed with participant IDs as grouping labels, guaranteeing that models are evaluated on their ability to generalize to unseen individuals rather than recognizing patterns from partially seen patients.

2.5.2. Classifier setup and evaluation

Six machine learning classifiers were employed for comprehensive evaluation: (i) Decision Tree with balanced class weights, (ii) Random Forest, (iii) k-Nearest Neighbors (KNN), (iv) Gaussian Naïve Bayes, (v) Linear Support Vector Machine (SVM), and (vi) non-linear SVM with RBF kernel. These six classifiers were selected to span complementary modeling paradigms: a single interpretable decision boundary (Decision Tree), an ensemble of trees (Random Forest), a non-parametric instance-based method (KNN), a probabilistic generative model (Naive Bayes), and both linear and non-linear margin-based discriminative models (Linear SVM and RBF-SVM). This diversity allows that the observed robustness of the selected features is not tied to a single class of learning algorithm. To facilitate a fair comparison across all feature selection strategies, the classifiers were evaluated using standard, widely adopted hyperparameter settings rather than performing an exhaustive hyperparameter optimization for each feature selection method. Random Forest was configured with 200 estimators and KNN with k=5, while the remaining classifiers used standard configurations unless otherwise specified. This design isolates the effect of the proposed feature selection strategy from classifier-specific hyperparameter tuning and provides a consistent evaluation framework across all methods. Extensive hyperparameter optimization could further improve the absolute performance of individual classifiers and is therefore left for future work. Before classifier training, all input features were standardized to zero mean and unit variance using a StandardScaler fitted exclusively on the training partition and subsequently applied to the corresponding test partition. This procedure was repeated independently for each group-wise split to prevent data leakage while ensuring consistent feature scaling across classifiers.

The evaluation followed the group-wise cross-validation scheme. Each model was trained with incrementally increasing feature subsets (k=1 to 26 features) to assess cumulative performance progression. Performance was quantified using accuracy, ROC-AUC (Receiver Operating Characteristic, Area Under the Curve), precision, recall, and F1-score across all folds.

Statistical significance between methods was assessed through paired t-tests at each feature count level using repeated group-wise validation scores. Given the relatively small number of repeated group-wise splits used for these comparisons, formal assessment of the normality assumption has limited statistical power [40,41]. Therefore, the paired t-test results are interpreted as indicative rather than definitive and are used to complement the overall comparative analysis rather than serve as the sole basis for the conclusions. The magnitude and consistent direction of the observed differences across classifiers further support the conclusion that no practically meaningful performance degradation was introduced by the proposed feature selection strategy.

Computational efficiency was evaluated by measuring execution times across all method-classifier combinations, while Kendall's τ correlation quantified the alignment between complexity rankings and Random Forest importance orders.

2.5.3. Ablation study on individual complexity metrics

To assess whether the combined criterion S_i (Eq. (4)) provides an advantage over any single complexity metric, we repeated the full feature selection and evaluation pipeline using each metric individually to rank features (Fisher score alone, Overlap volume alone, Shannon entropy alone), selecting the respective top-7 features for each. Averaged across all six classifiers, the combined criterion achieved Accuracy (Acc) = 0.653, F1 = 0.375, AUC = 0.601; Fisher-only selection achieved Acc = 0.646, F1 = 0.366, AUC = 0.598; Overlap-only selection achieved Acc = 0.616, F1 = 0.352, AUC = 0.574; and Entropy-only selection achieved Acc = 0.653, F1 = 0.367, AUC = 0.598. The combined score matched or slightly outperformed every individual metric and notably outperformed Overlap-volume alone by a clear margin. These results demonstrate that the complementary information provided by Fisher score, overlap volume, and Shannon entropy yields a more robust feature ranking than any individual complexity metric alone. This ablation study therefore provides direct empirical support for the proposed combined complexity score and further distinguishes the proposed approach from feature selection methods based on a single discriminative criterion.

3. Results and discussion

Fig. 2 shows the evaluation of feature complexity using discriminative metrics (F1, F2, F4). The features f20 and f9 achieved high F1 scores (471.7 and 175.5, respectively), indicating strong TB/non-TB separation (Fig. 2(a)). Low F2 values with f3 (F2 = 0.13) suggest

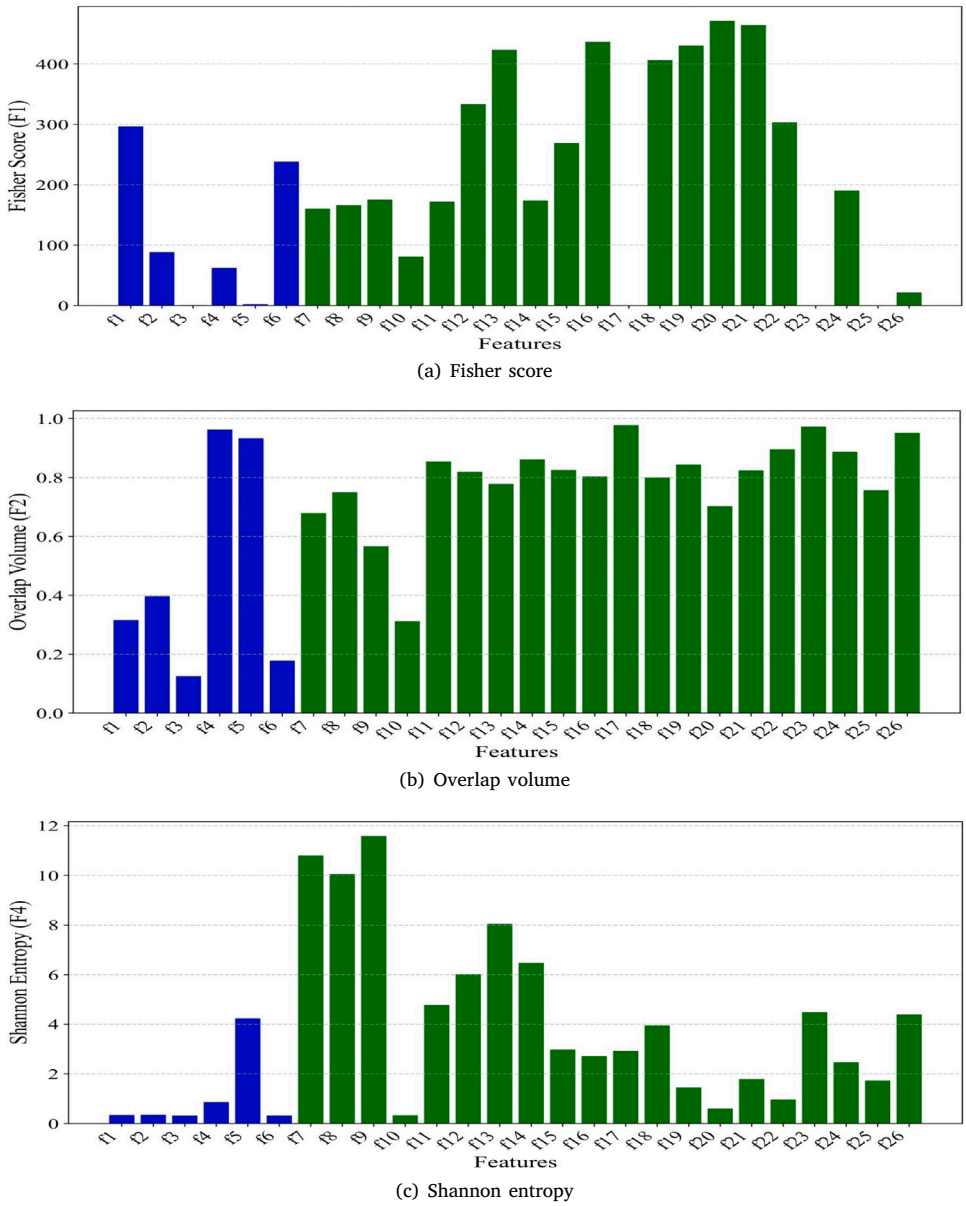

Fig. 2. Comparison of three complexity metrics for temporal (blue bars) and spectral (green bars) features. Feature labels *f1* to *f26* and their corresponding meanings are shown in Table 1.

minimal overlap (Fig. 2(b)), while spectral features (f7–f26) had high entropy (F4 > 4.5) (Fig. 2(c)). Combining high F1, low F2 and high F4, features like f20 (MFCC delta2 std) and f9 (spectral rolloff) could enhance TB screening robustness. Overall, Fig. 2 demonstrates that the most discriminative features consistently combine high class separability, low class overlap, and high information content, thereby motivating their selection for the subsequent classification experiments. Our findings show that complexity-informed features can improve TB cough detection by capturing nonlinear acoustic cues linked to airway pathology. MFCC derivatives (delta and delta-delta) and spectral rolloff showed strong separability (F1 > 100) and moderate overlap (F2 ∼ 0.6–0.8) outperforming traditional features like RMS and ZCR.

Unlike traditional black-box models, our approach focuses on highlighting the individual contribution of each acoustic feature. This significantly enhances interpretability and supports more transparent diagnostic decision-making. This level of explainability is especially valuable in medical contexts, where trust and clarity are essential. The CODA-TB dataset's diversity supports generalization to real-world settings, aligned with WHO's goals for AI in global health [42]. Nonetheless, variable recording conditions may affect feature consistency.

3.1. Feature ranking based on complexity metrics

Table 2 presents the ranking of all 26 acoustic features based on the combined complexity score S_i, following the selection procedure described in Section 2.3. The top 7 features (highlighted in bold) were selected using the 70th percentile threshold ($\beta = 0.61$) and represent the most discriminative characteristics for TB cough classification.

3.2. Feature set analysis

The analysis of feature selection during the two strategies reveals that complexity-based selection achieves peak performance with fewer features than sequential methods, demonstrating superior feature efficiency.

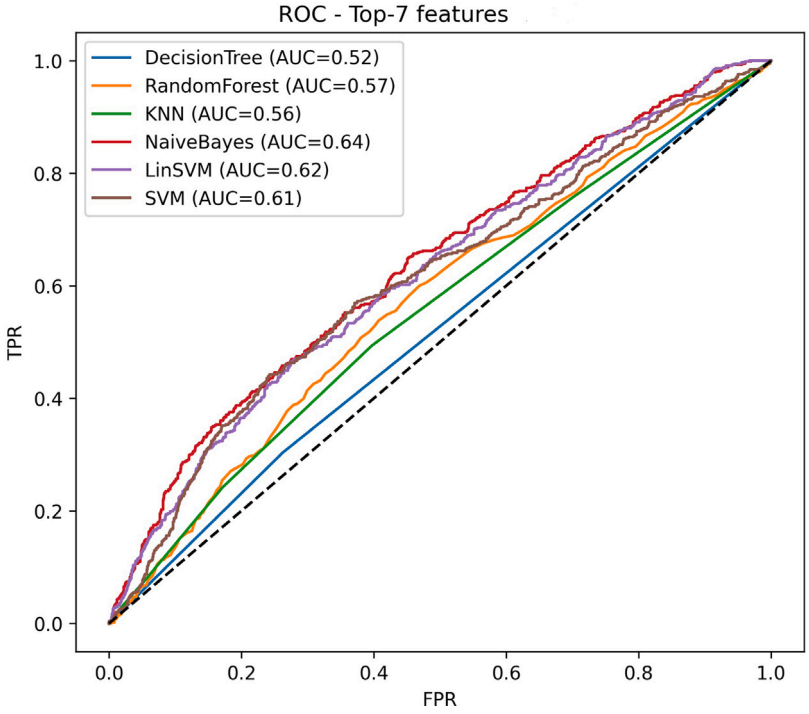

Fig. 3. ROC curves for the six classifiers using the top-7 complexity-selected features.

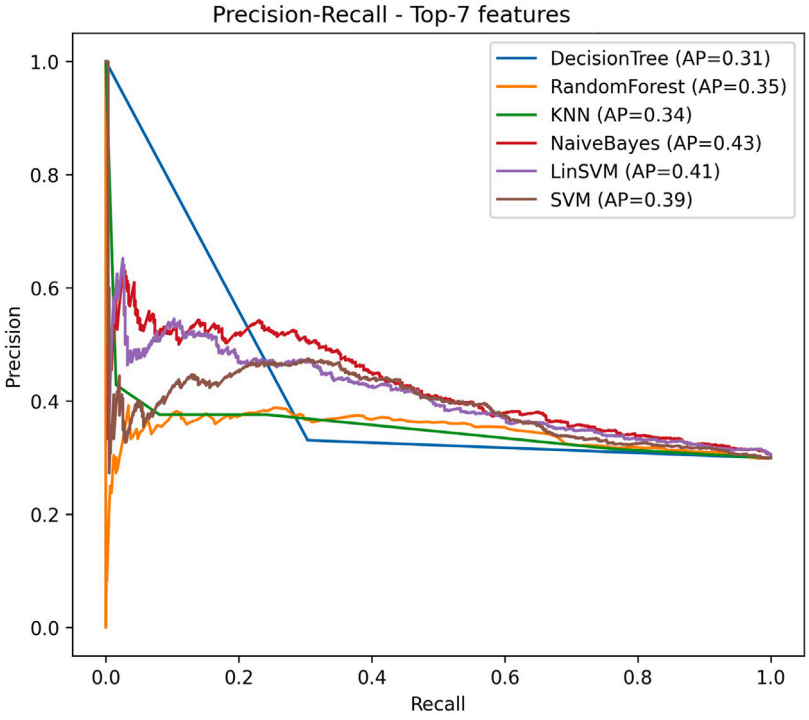

Fig. 4. Precision–Recall curves for the six classifiers using the top-7 complexity-selected features (AP: Average Precision).

Table 2
Complete feature ranking with combined complexity score.

Rank	Feature	Fisher Score	Overlap	Entropy
1	**f20: Std of MFCC Delta-Delta**	471.71	0.70	0.61
2	**f9: Spectral Rolloff**	175.52	0.57	11.59
3	**f19: Mean of MFCC Delta-Delta**	430.46	0.84	1.46
4	**f16: Std of MFCC Delta**	436.72	0.80	2.72
5	**f21: Min of MFCC Delta-Delta**	464.59	0.82	1.80
6	**f5: Kurtosis**	2.26	0.93	4.24
7	**f18: Max of MFCC Delta**	406.47	0.80	3.95
8	f15: Mean of MFCC Delta	268.99	0.83	2.98
9	f13: Min of MFCC	423.48	0.78	8.04
10	f12: Std of MFCC	333.64	0.82	6.02
11	f22: Max of MFCC Delta-Delta	303.44	0.90	0.97
12	f6: Variance	238.39	0.18	0.33
13	f8: Spectral Bandwidth	166.23	0.75	10.05
14	f7: Spectral Centroid	160.43	0.68	10.80
15	f1: RMS	296.52	0.32	0.34
16	f14: Max of MFCC	174.18	0.86	6.48
17	f11: Mean of MFCC	172.01	0.85	4.78
18	f10: Spectral Flatness	81.00	0.31	0.34
19	f2: ZCR	88.69	0.40	0.35
20	f25: Spectrogram Min	0.10	0.76	1.73
21	f24: Spectrogram Std	190.37	0.89	2.47
22	f26: Spectrogram Max	21.90	0.95	4.39
23	f23: Spectrogram Mean	0.13	0.97	4.49
24	f4: Skewness	62.29	0.96	0.86
25	f3: Duration	0.81	0.13	0.33
26	f17: Min of MFCC Delta	0.03	0.98	2.93

Note. Features in bold represent the top 7 selected features based on the 70th percentile threshold ($S_i \geq 0.61$). The combined complexity score S_i is computed as presented in Eq. (4).

Table 3
Performance with top-7 selected features (group-wise validation).

Method	Acc.	F1	AUC	CV F1
All 26 features	0.67	0.31	0.61	0.31 ± 0.07
Complexity-based	0.66	**0.34**	0.59	**0.34 ± 0.06**
SelectKBest	0.67	0.34	0.60	0.34 ± 0.03
RF Importance	0.66	0.34	0.60	0.34 ± 0.07

Note. Top 7 features in Table 2. Acc.: Accuracy, AUC: Area Under Curve, CV: Cross-validation.

3.2.1. Fixed feature selection

Figure 3 illustrates the discriminative performance of the proposed top-7 feature subset across the six evaluated classifiers through the corresponding ROC curves. Complexity-based selection demonstrated competitive performance while reducing the original feature set from 26 to 7 features (73% reduction), making it particularly suitable for practical deployment scenarios (Table 3). Although the performance improvement over the baseline methods is modest, the proposed approach achieves comparable predictive performance with a substantially smaller and more interpretable feature subset, reducing feature extraction complexity and facilitating deployment in resource-constrained TB screening settings. Notably, complexity-based selection also achieved a high cross-validation F1-score (0.34 ± 0.06), indicating better balance between precision and recall (Fig. 4), a crucial consideration for medical screening applications where both false positives and false negatives carry significant consequences. Considering this F1-score in a clinical-deployment context, we position the proposed model as a lightweight pre-screening/triage aid intended to prioritize patients for confirmatory testing (e.g., Xpert MTB/RIF) rather than as a standalone diagnostic tool, consistent with WHO guidance on target product profiles for community-based TB triage tests.

3.2.2. Adaptive feature selection

In this strategy, complexity-based method achieved optimal performance with only 9 features for RandomForest (accuracy: 0.71) and 10 features for SVM (accuracy: 0.67), representing 65% and 62% feature reduction respectively while maintaining competitive performance (Table 4). It consistently achieves peak or near-peak performance with markedly fewer features than competing methods. Fig. 5 presents the Random Forest sensitivity analysis as the number of selected features increases. The results show that near-optimal performance is achieved with only a small subset of features, supporting the robustness of the proposed feature selection criterion.

The advantage was particularly evident for the SVM classifier, as illustrated in Fig. 6. Fig. 6 shows how SVM accuracy evolves as the number of selected features increases for each feature selection method. The proposed complexity-based method reached its highest accuracy (0.67) with only 10 selected features, whereas the Forward and Backward sequential methods remained below 0.7. Although RF Importance achieved comparable performance for larger feature subsets, the proposed method reached its optimum with fewer features, demonstrating more efficient identification of informative acoustic descriptors (Table 4). These findings confirm that the complexity criterion provides both an efficient and robust selection strategy, allowing high predictive performance with minimal feature subsets.

3.3. Overall performance comparison

The superior performance of complexity-based feature selection stems from its foundation in information-theoretic principles. By prioritizing features with: (1) high Fisher scores indicating strong class separability, (2) low overlap volume ensuring minimal distribution confusion, and (3) high feature entropy capturing rich information content, the method identifies characteristics that are inherently discriminative and robust to inter-patient variability. This theoretical foundation explains why complexity-based selection maintains competitive performance while offering substantial computational advantages.

Paired t-tests across all folds and classifiers (Table 5) confirm no statistically significant accuracy degradation against any of the four benchmarks. To complement Table 5, we additionally report per-classifier paired t-test results (on F1-scores) comparing the proposed top-7 feature set against freshly recomputed top-7 subsets for all four benchmark methods (SelectKBest, Random Forest importance, sequential forward selection, sequential backward elimination). Out of 24 classifier/method comparisons, 22 were non-significant at α=0.05; two were statistically significant (KNN vs. Forward, p = 0.014; Linear SVM vs. Backward, p = 0.047). This 2/24 (approximately 8%) rate is consistent with the aggregated per-method proportions reported in Table 5 (0.23 for Forward and Backward, 0.08 for KBest, 0.12 for RF Importance) and does not alter the overall conclusion of no broad, systematic performance degradation. This makes complexity-driven selection the method of choice for deployment on resource-constrained devices. Fig. 7 presents the confusion matrices for the six evaluated classifiers using the proposed top-7 feature subset. The matrices show a balanced classification of TB-positive and TB-negative recordings, consistent with the quantitative performance metrics reported above.

The ability to maintain competitive performance with only 7 features vs. 26 (Table 3) reduces system complexity, memory footprint, and inference latency while enhancing interpretability for clinical stakeholders. Beyond computational efficiency, this compact feature set facilitates deployment on low-cost edge devices and point-of-care systems in resource-constrained settings, where memory and processing capabilities are limited. Furthermore, focusing on a small set of physiologically meaningful acoustic descriptors provides greater clinical interpretability, enabling a clearer understanding of the acoustic signatures associated with PTB and supporting more transparent decision-making. Kurtosis (f5) captures the leptokurtic (peaked) amplitude distribution characteristic of the explosive phase in TB coughs, potentially reflecting rapid airway opening against increased mucus viscosity. Spectral rolloff (f9) shifts may indicate altered harmonic structure due to pulmonary consolidation and cavitation. MFCC delta features (f16, f18–f21) likely capture the irregular timing and spectral evolution of cough sequences, which clinical studies associate with

Table 4
Optimal feature counts and corresponding maximum accuracy.

Method	Classifier					
	DT	RF	KNN	NB	LinSVM	SVM
Complexity	19 (0.63)	**9 (0.71)**	**9 (0.68)**	**1 (0.72)**	13 (0.64)	**10 (0.67)**
KBest	7 (0.62)	13 (0.70)	13 (0.67)	**1 (0.72)**	18 (0.64)	7 (0.65)
Forward	24 (0.62)	26 (0.70)	11 (0.67)	1 (0.71)	8 (0.65)	7 (0.60)
Backward	24 (0.62)	26 (0.70)	11 (0.67)	1 (0.71)	8 (0.65)	7 (0.60)
RF Importance	17 (0.63)	**8 (0.71)**	11 (0.67)	4 (0.72)	**7 (0.66)**	11 (0.65)

Note. DT: Decision Tree, RF: Random Forest, NB: Naive Bayes, LinSVM: Linear SVM. Values show optimal feature count (accuracy). Bold indicates methods achieving highest accuracy with fewest features.

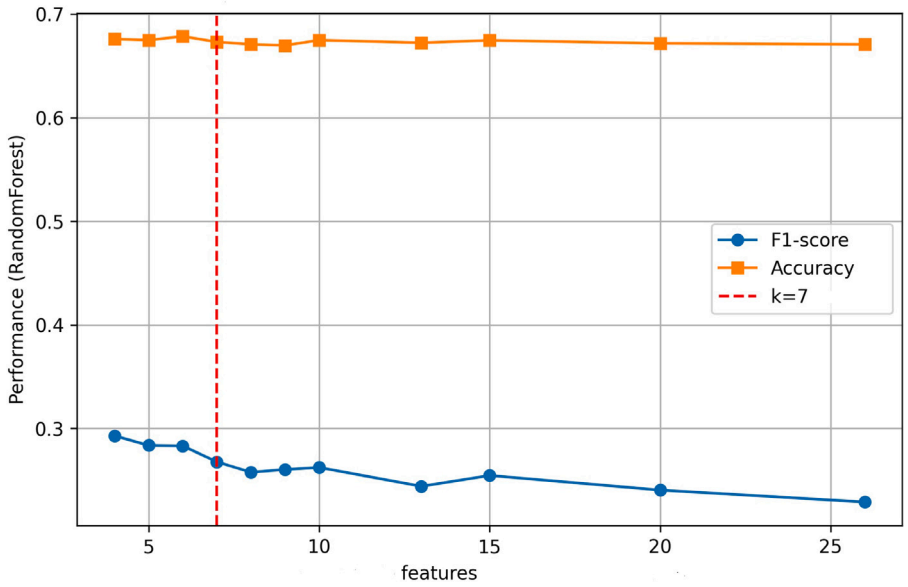

Fig. 5. Sensitivity of Random Forest performance to the number of retained features (published ranking order), supporting the choice of the 70th-percentile threshold (k=7).

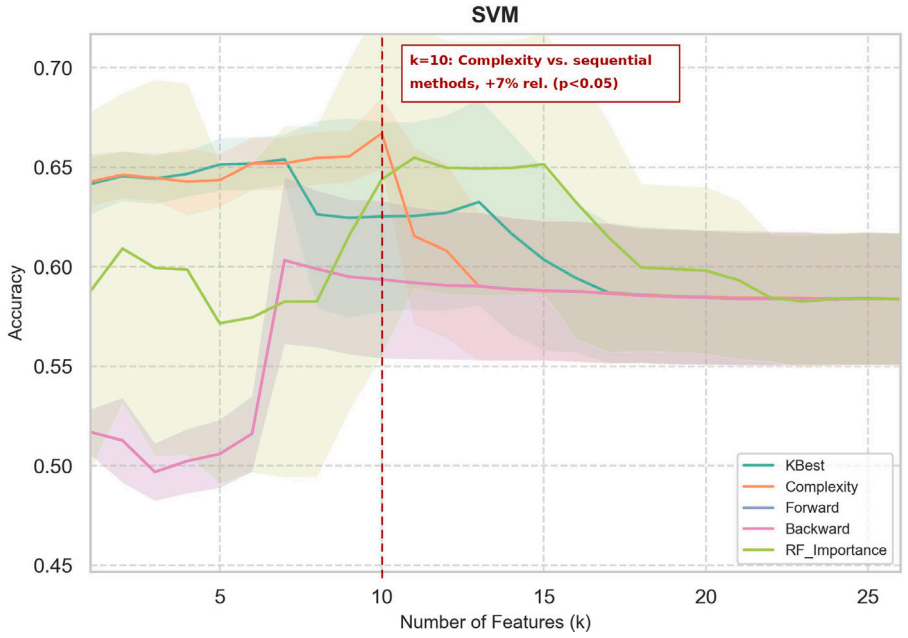

Fig. 6. Evolution of Support Vector Machine (SVM) accuracy according to the number of selected features for different feature selection methods. The complexity-based selection consistently outperforms sequential approaches across all feature counts, achieving the highest accuracy (0.67) with only 10 selected features, with the statistically significant advantage of the complexity-based method at k=10 highlighted.

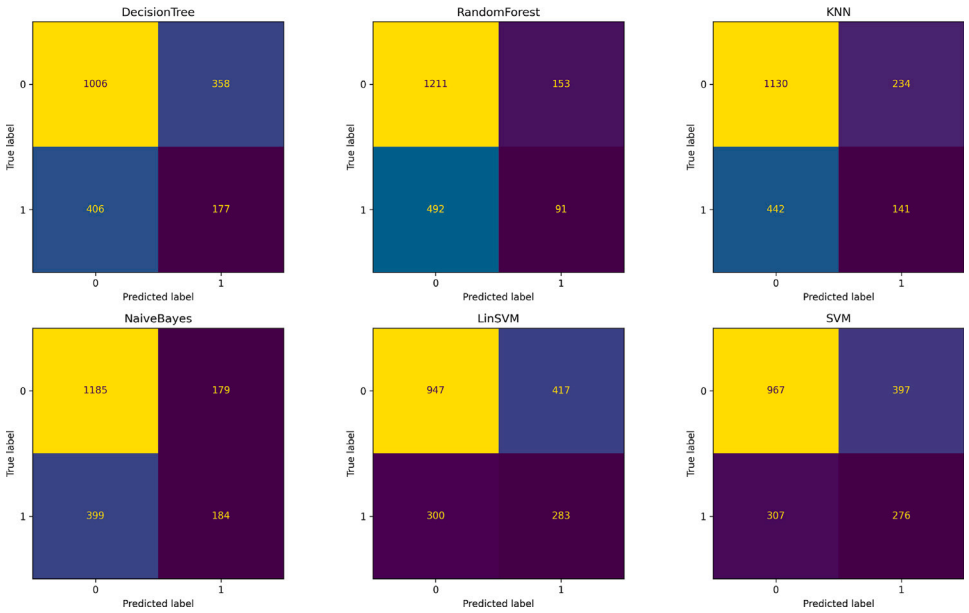

Fig. 7. Confusion matrices for the six classifiers using the top-7 complexity-selected features (participant-level held-out test set).

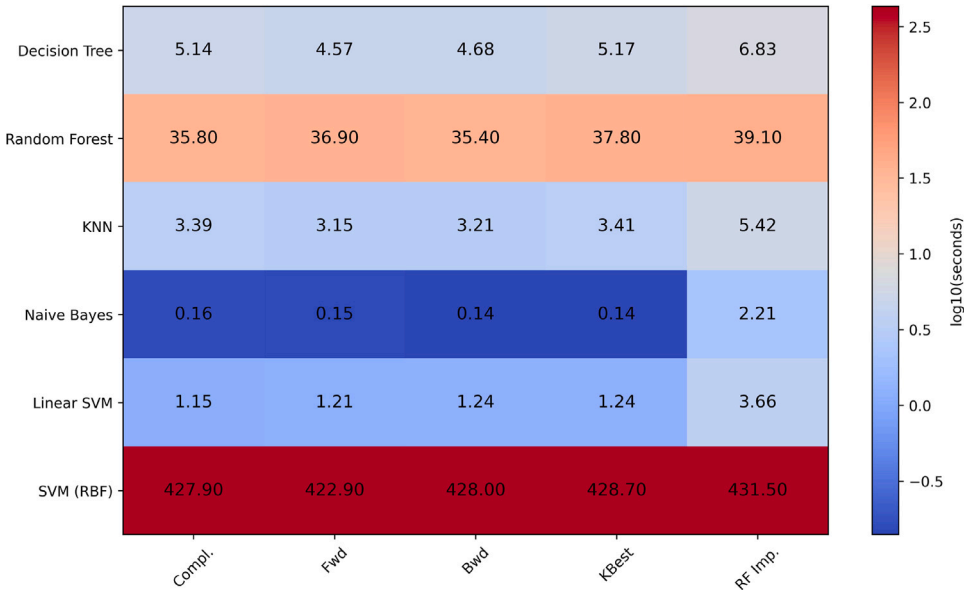

Fig. 8. Average computation time (selection + training) in seconds. Compl: Complexity-based method, Fwd: sequential forward selection, Bwd: sequential backward elimination, KBest: SelectKBest, RF Imp: Random Forest importance-based. The values displayed in each cell correspond to execution times in seconds, whereas the color scale represents $\log_{10}$ of the execution time to improve visualization across methods.

Table 5
Statistical significance ($p < 0.05$ proportion).

vs →	Fwd	Bwd	KBest	RF Imp.
6 classifiers	.23	.23	.08	.12

Note. Compl. Complexity-based method, Fwd: sequential forward selection, Bwd: sequential backward elimination, KBest: SelectKBest, RF Imp.: Random Forest importance-based.

bronchial wall thickening and parenchymal involvement in TB [43]. These acoustic correlates align with known TB pathophysiology, including caseous necrosis, granuloma formation, and bronchial distortion, suggesting that the complexity-driven selection identifies features with genuine physiological relevance rather than statistical artifacts.

3.4. Computational efficiency

Fig. 8 reports the average computation time during the second strategy. It shows that complexity-based selection is systematically among the fastest methods and is 1.8–14× faster than RF Importance while being comparable to filter and wrapper methods. The empirical execution times reported in Fig. 8 are consistent with the theoretical computational complexity of the evaluated feature selection methods. The runtime complexity for the filter-based methods (complexity-based selection, SelectKBest) is O(n*d) in the number of samples n and features d, since each feature is scored independently. Random Forest importance additionally requires fitting an ensemble of trees (O(n*d*T*log n) for T trees), which explains its higher runtime, while sequential forward/backward selection is the most expensive at $O(n*d^2)$ in the worst case due to the repeated retraining over candidate feature subsets. The

computational efficiency analysis demonstrates substantial advantages for complexity-based selection. Complexity-based selection achieved competitive computation times comparable to sequential forward selection (0.99–1.07×) while maintaining equivalent or superior accuracy. This efficiency advantage is particularly valuable for real-time TB screening applications and resource-constrained environments [44,45].

3.5. Limitations of this study

While our complexity-driven feature selection approach showed promising efficiency and interpretability for TB cough detection, several limitations warrant consideration. First, the analysis relies on the CODA-TB dataset, which, despite its multinational diversity (seven countries, 1105 participants), primarily captures PTB cases confirmed via clinical testing. This may limit generalizability to extrapulmonary TB, pediatric populations, or regions with underrepresented demographics, such as high-altitude or extreme environmental conditions that could alter cough acoustics. Second, variable recording conditions such as ambient noise, microphone quality, etc., inherent to field-collected data may introduce inconsistencies in feature extraction, potentially affecting model robustness in uncontrolled real-world settings. Third, the dataset exhibits a moderate class imbalance (70.0% TB-negative vs. 30.0% TB-positive recordings), which we partially mitigated through class-balanced weighting in all classifiers, but which may still contribute to the relatively modest F1-scores observed and could bias the model toward the majority (TB-negative) class in deployment. Although group-wise partitioning mitigates data leakage, the achieved performance metrics (accuracy ∼0.67, F1-score ∼0.34) indicate room for improvement, possibly due to class imbalance or subtle acoustic overlaps between TB-positive and negative cases influenced by comorbidities. Finally, the present study relies exclusively on acoustic features extracted from cough recordings, without incorporating complementary multimodal information such as clinical metadata, symptom profiles, demographic characteristics, or medical imaging. Although this design enabled a focused evaluation of the proposed complexity-based feature selection strategy, integrating multimodal information could further improve model robustness and diagnostic performance. Future work will therefore investigate multimodal feature fusion and validate the proposed framework in prospective multi-center studies to enhance its generalizability and clinical applicability.

4. Conclusion and future work

This study demonstrates that complexity-driven feature selection achieves equivalent or superior classification performance using only 7–10 features instead of 26. It exhibits no statistically significant accuracy loss against four established methods and up to 14× lower computational cost than commonly used Random Forest Importance baselines. The significant efficiency gains combined with maintained performance position complexity-based selection as the optimal approach for scalable TB screening in resource-constrained environments. The consistency of the complexity-based rankings across multiple classifiers suggests that the identified feature subset captures fundamental acoustic properties of TB coughs rather than classifier-specific artifacts. This generalizability enhances the clinical utility of our approach, as the same feature set could be deployed with different modeling techniques depending on computational constraints or performance requirements in various healthcare settings. Furthermore, the predominance of temporal derivatives in our selected features indicates that dynamic cough characteristics may be more discriminative than static spectral properties for TB detection, an insight that could guide future feature engineering in respiratory sound analysis. Looking beyond this study, we envision the proposed lightweight feature set as a key component of a broader digital-health triage pipeline. Integrated into a smartphone-based cough-screening application, it could support community health workers in identifying and prioritizing individuals for confirmatory TB testing, particularly in settings with limited access to laboratory-based diagnostics. More broadly, such an approach could facilitate the integration of AI-assisted cough screening into primary healthcare systems, enabling earlier referral, more efficient resource allocation, and improved access to TB screening in underserved communities. Future work will focus on integrating clinical metadata with acoustic features and validating the approach in prospective multicenter studies to enhance generalizability across diverse populations and recording conditions.

CRediT authorship contribution statement

Sana Ben Mahjouba: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Johannes C. Ayena:** Writing – review & editing, Visualization, Validation, Project administration, Conceptualization. **Youssef Ouakrim:** Visualization, Validation, Software, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Karim Loulou:** Resources, Methodology, Formal analysis, Data curation. **Simon Grandjean Lapierre:** Supervision, Resources, Formal analysis, Data curation. **Mihaja Raberahona:** Supervision, Resources, Formal analysis, Data curation. **Neila Mezghani:** Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the Canada Research Chair on Biomedical Data Mining (950-231214). The datasets were provided by Drs. Cattamanchi and Grandjean Lapierre in collaboration with international research partners and funded by the U.S. NIH the Patrick J. McGovern Foundation and Global Health Labs. The authors declare no conflicts of interest. The datasets used for the analyses described were contributed by Dr. Adithya Cattamanchi at UCSF and Dr. Simon Grandjean Lapierre at University of Montreal and were generated in collaboration with researchers at Stellenbosch University (PI Grant Theron), Walimu (PIs William Worodria and Alfred Andama); De La Salle Medical and Health Sciences Institute (PI Charles Yu), Vietnam National Tuberculosis Program (PI Nguyen Viet Nhung), Christian Medical College (PI DJ Christopher), Centre Infectiologie Charles Merieux Madagascar (PIs Mihaja Raberahona& Rivonirina Rakotoarivelo), and Ifakara Health Institute (PIs Issa Lyimo & Omar Lweno) with funding from the U.S. National Institutes of Health(U01 AI152087), The Patrick J. McGovern Foundation and Global Health Labs.

Data availability

The CODA-TB dataset used in this study is available under the terms of its corresponding data use agreement. Due to these restrictions, the authors cannot redistribute the dataset. The code and configuration files required to reproduce the feature extraction, complexity analysis, and classification pipeline are available from the corresponding author upon request.

References

[1] World Health Organization, Global tuberculosis report 2023, 2023, https://www.who.int/teams/global-programme-on-tuberculosis-and-lung-health/tb-reports/global-tuberculosis-report-2023.

[2] H. Yang, X. Ruan, W. Li, J. Xiong, Y. Zheng, Global, regional, and national burden of tuberculosis and attributable risk factors for 204 countries and territories, 1990–2021: a systematic analysis for the global burden of diseases 2021 study, BMC Public Health 24 (1) (2024) 3111, http://dx.doi.org/10.1186/s12889-024-20664-w, Springer.

[3] World Health Organization, Global tuberculosis report 2022, 2022.

[4] C.M. Gill, L. Dolan, L.M. Piggott, A.M. McLaughlin, New developments in tuberculosis diagnosis and treatment, Breathe 18 (1) (2022).

[5] C.C. Madekwe, C.C. Madekwe, E.I. Obeagu, Inequality of monitoring in human immunodeficiency virus, tuberculosis and malaria: a review, Madonna Univ. J. Med. Health Sci. ISSN: 2814-3035 2 (3) (2022) 6–15.

[6] D. Capellán-Martín, J.J. Gómez-Valverde, D. Bermejo-Peláez, M.J. Ledesma-Carbayo, A lightweight, rapid and efficient deep convolutional network for chest X-Ray tuberculosis detection, 2023, http://dx.doi.org/10.48550/arXiv.2309.02140, ArXiv Preprint.

[7] M.A.-M. Hegazy, A.I. Abdelaziz, Revisiting tuberculosis diagnosis: the prospect of urine LAM in low-resource settings, Beni-Suef Univ. J. Basic Appl. Sci. 13 (1) (2024) 42, http://dx.doi.org/10.1186/s43088-024-00578-7, URL: https://bjbas.springeropen.com/articles/10.1186/s43088-024-00578-7.

[8] U. Ullah, S. Saleem, M. Farooq, B. Yameen, M.I. Cheema, Lipoarabinomannan-based tuberculosis diagnosis using a fiber cavity ring down biosensor, 2023, https://arxiv.org/abs/2401.00030.

[9] A.J. Zimmer, C. Ugarte-Gil, R. Pathri, P. Dewan, D. Jaganath, A. Cattamanchi, M. Pai, S. Grandjean Lapierre, Making cough count in tuberculosis care, Commun. Med. 2 (1) (2022) 83.

[10] G. Botha, G. Theron, R. Warren, M. Klopper, K. Dheda, P. Van Helden, T. Niesler, Detection of tuberculosis by automatic cough sound analysis, Physiol. Meas. 39 (4) (2018) 045005.

[11] G.P. Kafentzis, S. Tetsing, J. Brew, L. Jover, M. Galvosas, C. Chaccour, P.M. Small, Predicting tuberculosis from real-world cough audio recordings and metadata, 2023, arXiv preprint arXiv:2307.04842.

[12] G. Rudraraju, S. Palreddy, B. Mamidgi, N.R. Sripada, Y.P. Sai, N.K. Vodnala, S.P. Haranath, Cough sound analysis and objective correlation with spirometry and clinical diagnosis, Informatics Med. Unlocked 19 (2020) 100319.

[13] T. Manyazewal, Y. Woldeamanuel, H.M. Blumberg, A. Fekadu, V.C. Marconi, The potential use of digital health technologies in the african context: a systematic review of evidence from ethiopia, NPJ Digit. Med. 4 (1) (2021) 125.

[14] G. Frost, G. Theron, T. Niesler, TB or not tb? Acoustic cough analysis for tuberculosis classification, 2022, ArXiv Preprint.

[15] S.J.S. Rajasekar, A.R. Balaraman, D.V. Balaraman, S.M. Ali, K. Narasimhan, N. Krishnasamy, V. Perumal, Detection of tuberculosis using cough audio analysis: a deep learning approach with capsule networks, Discov. Artif. Intell. 4 (1) (2024) 79, http://dx.doi.org/10.1007/s44163-024-00179-4, URL: https://link.springer.com/article/10.1007/s44163-024-00179-4.

[16] M. Sharma, V. Nduba, L.N. Njagi, W. Murithi, Z. Mwongera, T.R. Hawn, S.N. Patel, D.J. Horne, TBscreen: A passive cough classifier for tuberculosis screening with a controlled dataset, Sci. Adv. 10 (1) (2024) eadi0282.

[17] W. Xu, X. Bao, X. Lou, X. Liu, Y. Chen, X. Zhao, C. Zhang, C. Pan, W. Liu, F. Liu, Feature fusion method for pulmonary tuberculosis patient detection based on cough sound, Plos One 19 (5) (2024) e0302651.

[18] G.P. Kafentzis, E. Selisios, Tuberculosis screening from cough audio: Baseline models, clinical variables, and uncertainty quantification, Sensors 26 (4) (2026) 1223.

[19] A.J. Zimmer, P. Espinoza-Lopez, V. Ravi, S.K. Sieberts, S. Abbasgholizadeh Rahimi, M. Pai, C. Ugarte-Gil, S. Grandjean Lapierre, External validation of cough-based algorithms for pulmonary tuberculosis screening from the CODA TB DREAM challenge using cough data from peru, Sci. Rep. (2026).

[20] D.M. Chandrika, R. Gamage, D. Herath, I. Perera, A. Wickramasinghe, A. Niroshan, R. Munasinghe, C. Weerasingha, Automatic cough classification for tuberculosis screening in a real-world environment, IEEE J. Biomed. Health Informatics 26 (2) (2022) 693–703, http://dx.doi.org/10.1109/JBHI.2021.3138663, URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC8721487/.

[21] A. Sachowski, J. Jaworek-Korjakowska, R. Tadeusiewicz, On the effectiveness of signal decomposition, feature extraction and selection on lung sound classification, 2021, arXiv preprint arXiv:2012.11759.

[22] V. Rajasekar, Y. Zhang, et al., CODA-TB: A multi-country cough audio dataset for pulmonary tuberculosis screening, Sci. Data (2023) http://dx.doi.org/10.1038/s41597-023-02356-0, URL: https://www.nature.com/articles/s41597-023-02356-0.

[23] R.W. Hamming, Error detecting and error correcting codes, Bell Syst. Tech. J. 29 (2) (1950) 147–160, http://dx.doi.org/10.1002/j.1538-7305.1950.tb00463.x, URL: https://ieeexplore.ieee.org/document/6772729.

[24] L.R. Rabiner, R.W. Schafer, Digital Processing of Speech Signals, Prentice Hall, Englewood Cliffs, NJ, 1978.

[25] M. Pahar, A. Klopper, T.F. McCarthy, F.C. Bell, Mel frequency cepstral coefficients for the detection of COVID-19 from cough, breath and speech sounds, IEEE Access 9 (2021) 65227–65241, http://dx.doi.org/10.1109/ACCESS.2021.3074437, URL: https://ieeexplore.ieee.org/document/9471859.

[26] J. Zhao, Y.Y. Zhang, Z. Wang, H. Wang, W. Zhou, X. Zhang, A review of cough sound analysis for pulmonary disease diagnosis, IEEE Rev. Biomed. Eng. 15 (2022) 123–141, http://dx.doi.org/10.1109/RBME.2021.3122530, URL: https://ieeexplore.ieee.org/document/9429536.

[27] R.A. Fisher, The use of multiple measurements in taxonomic problems, Ann. Eugen. 7 (2) (1936) 179–188, http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x, URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1469-1809.1936.tb02137.x.

[28] A. Hinneburg, D.A. Keim, An efficient approach to clustering in large multimedia databases with noise, Proc. the Fourth Int. Conf. Knowl. Discov. Data Min. (KDD) (2000) 58–65, URL: https://dl.acm.org/doi/10.1145/347090.347107.

[29] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (3) (1948) 379–423, http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x, URL: https://ieeexplore.ieee.org/document/6773024.

[30] B. Bischl, M. Binder, M. Lang, T. Pielok, J. Richter, S. Coors, J. Thomas, T. Ullmann, M. Becker, A.-L. Boulesteix, et al., Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges, Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 13 (2) (2023) e1484.

[31] R. Wild, F. Wodaczek, V.D. Tatto, B. Cheng, A. Laio, Automatic feature selection and weighting in molecular systems using differentiable information imbalance, Nat. Commun. 16 (2025) 270, http://dx.doi.org/10.1038/s41467-024-55449-7.

[32] R.R. Hocking, A biometrics invited paper: The analysis and selection of variables in linear regression, Biometrics (1976).

[33] D.E. Whitney, A direct method of nonparametric measurement selection, IEEE Trans. Comput. (1971).

[34] T. Marill, D.M. Green, On the effectiveness of receptors in recognition systems, IEEE Trans. Inform. Theory (1963).

[35] L. Breiman, Random forests, Mach. Learn. (2001).

[36] M. Büyükkeçeci, M.C. Okur, A comprehensive review of feature selection and feature selection stability in machine learning, Gazi Univ. J. Sci. 36 (4) (2023) 1506–1520.

[37] Y. Akhiat, A. Zinedine, M. Chahhou, Using reinforcement learning to select an optimal feature set, J. Autom. Mob. Robot. Intell. Syst. 18 (1) (2024) 56–66.

[38] L. Zhang, et al., Serial filter–wrapper feature selection method with elite guided mutation strategy on cancer gene expression data, Artif. Intell. Rev. (2025) http://dx.doi.org/10.1007/s10462-024-11029-1.

[39] G. Varoquaux, et al., Assessing and tuning brain decoders: Cross-validation, caveats and guidelines, NeuroImage (2017).

[40] T.K. Kim, J.H. Park, More about the basic assumptions of t-test: normality and sample size, Korean J. Anesthesiol. 72 (4) (2019) 331–335.

[41] C. Avram, M. Măruşteri, Normality assessment, few paradigms and use cases, Rev. RomânĂ de MedicinĂ de Lab. 30 (3) (2022) 251–260, http://dx.doi.org/10.2478/rrlm-2022-0030.

[42] World Health Organization, Ethics and governance of artificial intelligence for health, 2021, (Accessed 13 June 2025), https://www.who.int/publications/i/item/9789240029200.

[43] W. Xu, et al., Feature fusion method for pulmonary tuberculosis patient classification using cough sounds, Sci. Rep. / Or J. List. PubMed Central (2024) URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11093322/, Open access (PubMed Central).

[44] World Health Organization, Global strategy on digital health 2020–2025, 2023, URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafac562051240.pdf.

[45] World Health Organization, Global tuberculosis report 2024, 2024, URL: https://www.who.int/teams/global-tuberculosis-programme/tb-reports.