

Explainable AI-Based Audio Tampering Detection and AI Voice Identification

Mrs. Marry Leo Sherlin, Atharsha RJ, Kamala M, Jashika J, Abinaya M

¹ Assistant professor, ^{2,3,4,5} students Department of AI&DS, Loyola Institute of Technology and Science, Thovalai, Kanyakumari Dist, Tamil Nadu.

Abstract—Modern forensic audio analysis must distinguish authentic recordings from AI-synthesized and human-mimicked speech, yet most voice-spoofing detectors are binary, treat the trained model as a black box, and give an analyst no evidence for why a recording was flagged. This paper presents RFMD-XAI, an explainable forensic audio framework that keeps the Real-Fake-Mimic Detection (RFMD) classifier intact—a lightweight one-dimensional convolutional neural network (1D CNN) with a three-unit SoftMax output—and appends a post-hoc explainability stage. SHAP (SHapley Additive exPlanations) is the primary attribution method and estimates which temporal regions of the input push the prediction toward or away from each class, while LIME (Local Interpretable Model-agnostic Explanations) provides a perturbation-based local cross-check. The system is organized into four cooperating modules—Audio Preprocessing, Classification, Explainable AI, and Forensic Reporting—coordinated through a central inference engine. We describe the architecture, the algorithmic foundation of each module, the end-to-end workflow, a sample explanation output, a comparative analysis against typical spoofing detectors across fourteen parameters, and a reproducible protocol for evaluating explanation faithfulness, stability, sparsity, class contrast, and human interpretability. Section X reports the baseline classifier results of the reference study and defines XAI acceptance criteria; no new XAI performance values are claimed pending full-scale evaluation.

Keywords—Forensic Audio Analysis, Voice Spoofing Detection, Deepfake Speech, Human Voice Mimicry, One-Dimensional CNN, Explainable AI, SHAP, LIME, Audio Forensics.

I. INTRODUCTION

Speech carries linguistic content as well as speaker-specific acoustic characteristics such as pitch, rhythm, and intonation. Modern speech synthesis, voice conversion, and skilled human imitation can alter these characteristics enough to make forensic verification difficult. The reference RFMD study addresses this by classifying a recording into three categories—Real, Fake (AI-synthesized), and Mimic (human-mimicked)—using a lightweight 1D CNN, and reports 93% overall test accuracy with class-wise F1-scores of 0.88, 0.98, and 0.92 [1].

A high-performing classifier can still be hard to inspect. It outputs a class and a probability vector without showing which portions of the signal were influential, which matters in forensic workflows where a decision may be reviewed, challenged, or compared with other evidence. Recent work on explainable deepfake detection emphasizes that interpretability helps connect model decisions to observable evidences such as temporal regions and acoustic patterns [22], [24], [25].

This paper proposes RFMD-XAI, a forensic audio framework that integrates the original three-class detector with post-hoc SHAP and LIME explanations, a temporal attribution display, and an analyst-oriented evidence report within a single modular architecture. Rather than treating explainability as a cosmetic add-on, RFMD-XAI is designed as a research contribution centered on architectural preservation, algorithmic justification, and measurable evaluation of the explanations themselves.

The remainder of this paper is organized as follows. Section II reviews the limitations of existing systems. Section III identifies the research gap. Section IV presents the proposed system and its modules. Section V describes the algorithmic

foundation. Section VI details the system workflow and Section VII the architecture. Section VIII explains the novel contribution. Section IX outlines the experimental setup and evaluation metrics. Section X presents baseline performance and XAI evaluation criteria. Section XI presents a sample output. Section XII gives a comparative analysis, Section XIII discusses advantages and limitations, Section XIV future scope, and Section XV concludes the paper.

II. EXISTING SYSTEM AND LIMITATIONS

Existing voice-spoofing and audio-deepfake detectors, including academic prototypes built on CNNs, MFCC features with SVMs, VGG-style networks, self-supervised learning, and deep neural networks [3], [9], [10], [12], [13], exhibit the following recurring limitations:

- 1) **Binary Decisions:** Many systems are designed as real-versus-fake classifiers and do not treat human-mimicked speech as a separate class [1].
- 2) **Black-Box Predictions:** The model returns a label and score without indicating which parts of the signal drove the decision.
- 3) **No Temporal Localization:** Analysts cannot see which time regions of a recording contained the evidence.
- 4) **Heavy Input Representations:** Spectrogram-based 2D CNN pipelines add computational cost and can be sensitive to noise [1].
- 5) **Unvalidated Explanations:** Where saliency or attribution maps are shown, their faithfulness to the model is often not tested; a visually attractive heatmap is not proof of causal influence [24].
- 6) **Single-Method Explanations:** Different explanation methods can highlight different regions [24], yet a single method is usually reported without cross-checking.

7) **NoForensicReporting:** Outputsarerarelypackagedasa reviewable report with the probability vector, evidence regions, and stated limitations.

8) **Limited Analyst Support:** Disagreement or low-confidence cases are not flagged for human inspection.

9) **RetrainingCoupling:** Someinterpretabledesignsmodify the classifier or its loss, which changes the validated detection performance.

10) **UnclearRobustnessReporting:**Noiseroobustness,cross-datasetbehavior,andlatencyarenotalwaysreportedtogether with interpretability.

III. RESEARCHGAP

The literature on audio deepfake detection has largely progressed along two tracks: (i) detection-accuracy-focused work on binary spoof detection using ever stronger backbones, and (ii) a smaller body of explainability work applying SHAP, LIME, Grad-CAM, or time-domain attribution to those detectors [21], [24], [25], [26]. Comparativelylittleworkcombinesathree-classformulation that explicitly includes human mimicry, a lightweight detector left unmodified, two complementary post-hoc explainers, and a quantitative evaluation protocol within a singlecohesiveframework.Furthermore,explanationquality is often judged visually rather than through faithfulness, stability, and sparsity tests. This fragmentation leaves a gap for a lightweight, three-class, explainable, and auditable forensic audio framework—the gap this paper addresses through RFMD-XAI.

IV. PROPOSEDSYSTEM

RFMD-XAI is organized into four cooperating modules, coordinated by a central inference engine and a shared audio/model store, as illustrated in Fig. 1. The engine mediates all cross-module communication: it passes the preprocessed waveform to the classifier, the trained model and the same waveform to the explainers, and the probabilities, attributions, and agreement status to the reporting module, fusing them into a single explained response.

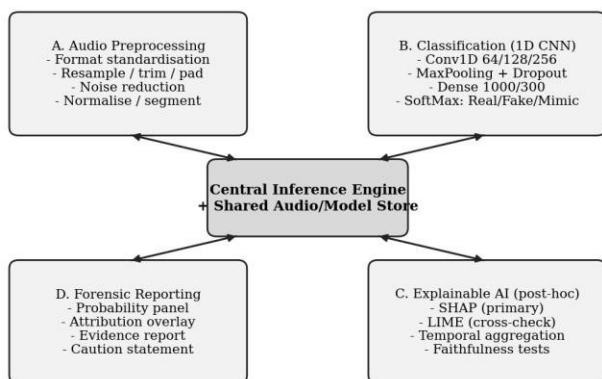


Fig. 1. Framework diagram of the proposed RFMD-XAI system: fourcooperatingmodulescoordinatedbyacentralinferenceengineandsharedstore.

A. AudioPreprocessingModule

This module prepares raw recordings exactly as in the referenceRFMDpipeline[1];nostepisremovedorreplaced. Itstandardizesfileextensions,resamplesaudiotoacommon rate(Fourier-basedresamplingto60,000Hzinthereference study), trims or pads recordings to a consistent length, reduces noise through energy-based filtering with spectral gating,normalizesamplitude,andsegmentslongrecordings. Forreadabilityoftheexplanation,theresultingwaveformcan additionally be partitioned into fixed-duration temporal regions (forexample,100–250 ms windows). This grouping is only an explanation interface and does not change the classifierinputortocreateanewclassifierfeature.Interaction: themodulesuppliesthesamewaveformtoboththeclassifier and the XAI module.

B. ClassificationModule

The original 1D CNN is kept unchanged. It contains three convolutionalblocksusingConv1D(64,128,and256filters, kernel size 5, ReLU), MaxPooling1D (pool size 2), and dropout (rate 0.5); the feature maps are flattened and passed throughdenselayersof1000and300neurons,followedbya three-unitSoftMaxoutputforReal,Fake,andMimic[1].The loss remains categorical cross-entropy and the optimizer remains Adam; no XAI term is added to the training loss. Interaction: the module returns the probability vector to the engineandexposesthetrainednetworktotheXAImoduleas a fixed prediction function.

C. ExplainableAIModule

This moduleis composed ofthree components. (1) A SHAP explainer computes additive attributions for the predicted class (and optionally for all three classes) using gradient- or deep-learning-based explainers with a representative background set [22]. (2) A LIME explainer masks or attenuates selected temporal segments, observes the change in class probability, and fits a sparse local surrogate model [23]. (3) A Temporal Aggregation component groups sample-level attributions into readable temporal regions and checks SHAP–LIME agreement. Interaction: the module receives the trained model and preprocessed waveform and returns ranked temporal evidence and an agreement flag to the reporting module.

D. ForensicReportingModule

This module assembles a reviewable explanation report for each analyzed recording: (1) the Real/Fake/Mimic probability vector; (2) the predicted class; (3) the top contributing temporal segments; (4) a waveform plot with attributionoverlays;(5)theLIMElocalcontributionlist;and (6)acautiostatementthatmodelattributionisnot,byitself, proof of how an audio file was generated. Recordings with lowconfidenceorSHAP–LIMedisagreementareflaggedfor analyst inspection. Reports are stored for later forensic review.

E. Inter-ModuleInteraction

The four modules are not independent pipelines but a coordinated system in which no module calls another

directly. Fig. 2 summarizes the resulting data flow: waveform and sampling rate flow from preprocessing to the classifier; the model and input flow to the XAI module; attributions return through the engine to the reporting module; and the explained result is returned to the user.

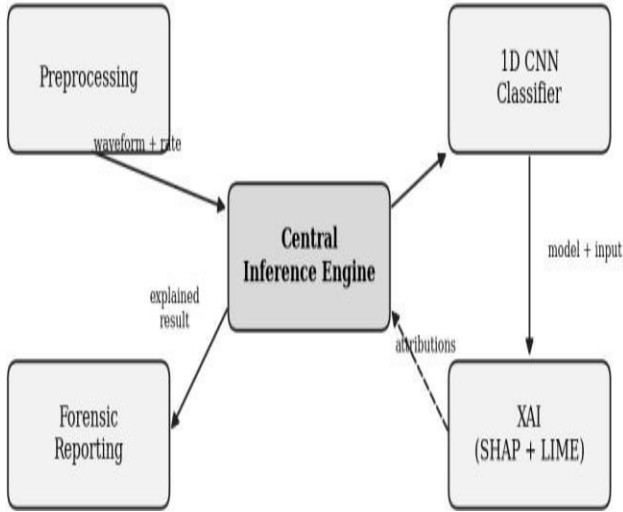


Fig. 2. Module interaction diagram: primary data flows (solid) and secondary flows (dashed) coordinated through the central inference engine.

V. ALGORITHMS

A. One-Dimensional CNN Classification

Purpose: To classify a preprocessed waveform into Real, Fake, or Mimic.

Working Principle: Convolutional filters slide along the time axis and learn local temporal and acoustic patterns; pooling reduces resolution; dense layers combine the features; SoftMax converts the three outputs into class probabilities,

$$p_c = \exp(z_c) / \sum_k \exp(z_k), c \in \{Real, Fake, Mimic\} \quad (1)$$

trained with categorical cross-entropy $L = -\sum_c y_c \log p_c(2)$.

Advantages: Lightweight, avoids spectrogram-heavy 2D processing, and reports about 0.12 s inference latency per sample in the reference study [1].

Why Selected: It is the validated baseline; keeping it unchanged preserves the reported performance.

B. SHAP Attribution (Primary)

Purpose: To estimate each temporal group's contribution to the prediction.

Working Principle: SHAP assigns each input feature (here, each temporal group) a Shapley-value-based importance ϕ_i such that the prediction is decomposed additively [22]:

$$f(x) = \phi_0 + \sum_i \phi_i(3)$$

Positive ϕ_i indicates contribution toward the selected class; negative ϕ_i indicates contribution away from it. For a

differentiable Keras network, gradient- or deep-learning explainers are used with a representative background set. Attributions are aggregated per temporal group.

Why Selected: It provides additive, per-class, signed attributions and fits differentiable networks.

C. LIME Local Cross-Check

Purpose: To provide an independent, perturbation-based explanation of the same prediction.

Working Principle: LIME perturbs the input by masking or attenuating temporal segments, records the classifier's probability, and fits a sparse interpretable model g in the neighborhood of x [23]:

$$\zeta(x) = \argmin_g L(f, g, \pi_x) + \Omega(g)(4)$$

where π_x weights samples by proximity to x and $\Omega(g)$ penalizes complexity. Agreement between SHAP and LIME is not treated as proof, but disagreement identifies cases that deserve analyst inspection.

D. Temporal Aggregation and Agreement

Purpose: To present explanations in human-readable time regions and quantify method agreement.

Working Principle: Attributions are summed within each window, ranked by magnitude, and the overlap of the SHAP and LIME top- k sets is computed:

$$Agreement@k = |Top-k(SHAP) \cap Top-k(LIME)| / k(5)$$

E. Faithfulness, Stability, and Sparsity Metrics

To support reproducible evaluation (Section IX), explanation quality is formalized as follows. Faithfulness measures the probability drop when the top- k attributed segments S_k are masked, compared with random masking of the same size:

$$\Delta F = P_c(x) - P_c(x \ominus S_k)(6)$$

Stability is the rank correlation between attribution vectors of an input and a slightly perturbed input (7). Sparsity is the smallest number of segments that account for a chosen share (for example 80%) of the total attribution magnitude (8).

F. Classification Metric Formulations

$$Precision = TP / (TP + FP)(9)$$

$$Recall = TP / (TP + FN)(10)$$

$$F1 = 2 \cdot (Precision \cdot Recall) / (Precision + Recall)(11)$$

VI. SYSTEM WORKFLOW

Fig. 3 traces the end-to-end workflow of RFMD-XAI from recording upload to stored forensic report. The user supplies a recording; the unchanged preprocessing pipeline prepares the waveform; the unchanged 1D CNN produces the Real/Fake/Mimic probability vector; SHAP computes temporal attribution for the selected class while LIME performs local perturbation tests; the system aggregates attributions into readable temporal regions, checks agreement, and displays the prediction with its explanation; the report is stored for later forensic review.

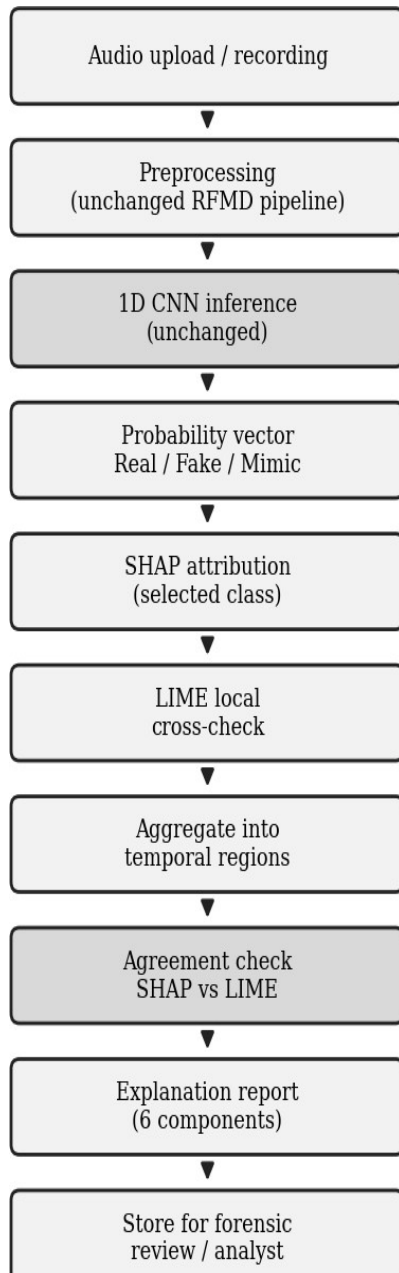


Fig. 3. End-to-end workflow of the RFMD-XAI system.

VII. SYSTEM ARCHITECTURE

Fig. 4 shows the layered architecture. The Frontend layer captures the uploaded or recorded audio and forwards it to the Orchestration layer, which routes requests to the Preprocessing, 1D CNN, XAI, and Report modules. All modules read from and write to a shared store holding audio, the trained model, attributions, and reports, and consolidated, explained results are rendered through the Output layer.

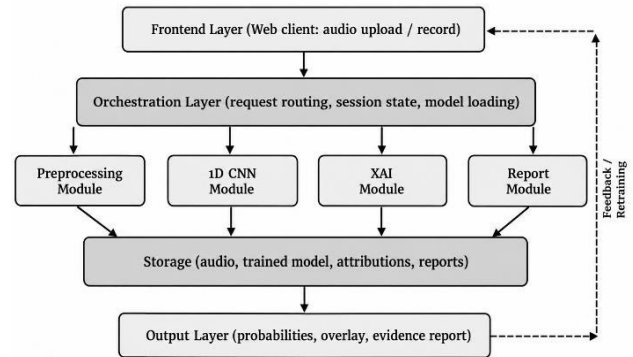


Fig. 4. Layered architecture of RFMD-XAI.

VIII. NOVEL CONTRIBUTION

RFMD-XAI's novelty lies not in any single algorithm but in their integration: (1) a post-hoc explanation stage that leaves the validated three-class RFMD classifier unchanged; (2) a three-class formulation, including human mimicry, that is explained rather than merely predicted; (3) SHAP as the primary attribution method with LIME as an independent cross-check, with disagreement used as an analyst-review signal; (4) time-localized evidence presented as waveform overlays over readable temporal regions; (5) an explicit, reproducible evaluation protocol for faithfulness, stability, sparsity, class contrast, and human interpretability; (6) a structured forensic explanation report with a built-in caution statement; and (7) a lightweight 1D waveform design that avoids added classifiers, attention modules, or generative components. Together, these distinguish RFMD-XAI from prior detection-only or single-explainer systems reviewed in Section II.

IX. EXPERIMENTAL SETUP

Dataset: As RFMD-XAI is at the design and implementation stage, this section specifies the retained dataset and a proposed explanation configuration rather than reporting new results. The reference implementation uses the DEEP-VOICE dataset enriched with additional human-mimic recordings to obtain the three target classes. The curated dataset contains 6,780 samples divided 70:15:15 into training, validation, and test partitions [1]; the reported test set contains 678 samples (221 Real, 221 Fake, 236 Mimic).

Training and Testing: The classifier follows the reference configuration: categorical cross-entropy, Adam optimizer, learning rate 0.001, batch size 32, up to 100 epochs, and early stopping on validation loss [1]. The explainers require no retraining. SHAP will be computed on a representative background subset drawn from the training partition; LIME will use temporal-segment masking over fixed windows (for example 200 ms). The number of background samples and LIME perturbations will be chosen during implementation and reported with the results.

Hardware and Software: Development is planned in Python with TensorFlow/Keras for the 1D CNN, the SHAP and LIME libraries for explanation, Librosa for audio processing, and a FastAPI backend with a web client for the interface, on

a GPU-enabled workstation or cloud instance. This is stated as the proposed setup pending full deployment.

Evaluation Metrics: Classification is evaluated with Accuracy, Precision, Recall, F1-Score, and inference latency. Explanations are evaluated with Faithfulness (ΔF against random masking), Stability (rank correlation under small perturbations), Sparsity, Class Contrast, SHAP-LIME Agreement@k, and a small structured Human Interpretability review.

X. BASELINE PERFORMANCE AND XAI EVALUATION CRITERIA

The classifier results below are retained from the reference study and describe the underlying RFMD model, not a newly trained XAI experiment [1]. No numerical SHAP/LIME score is claimed, because such values require running the actual trained model with the chosen explanation configuration; the criteria in Table 2 define how those values will be obtained and judged.

TABLE 1. REFERENCE RFMD CLASSIFICATION REPORT [1]

Class	Precision	Recall	F1	Support
Real	0.99	0.80	0.88	221
Fake	0.95	1.00	0.98	221
Mimic	0.87	0.99	0.92	236
Accuracy	—	—	0.93	678
MacroAvg.	0.94	0.93	0.93	678
WeightedAvg.	0.94	0.93	0.93	678

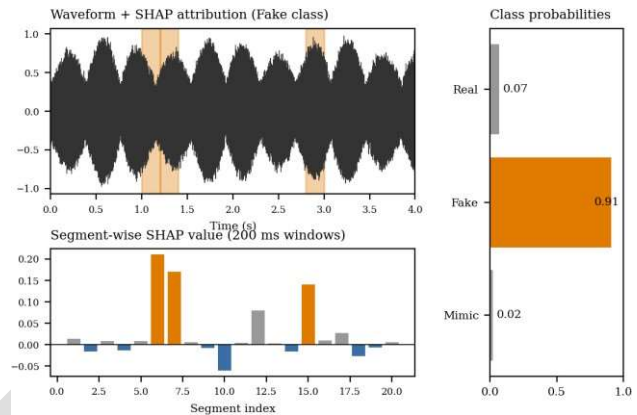
The reference study also reports above 89% accuracy under tested noise conditions of 5–20 dB SNR, 90.2% accuracy on a subset of ASVspoof 2019 for cross-dataset validation, and approximately 0.12 s inference latency per sample [1]. These values are retained only as baseline evidence for the detector.

TABLE 2. PROPOSED XAI EVALUATION CRITERIA

Metric	What tasks	Practical test/acceptance rule
Faithfulness	Do highlighted input truly affect the prediction?	Mask top-attributed segments; the target-class probability drop should exceed that of random masking.
Stability	Do similar inputs receive similar explanations?	Apply small permissible perturbations; compare attribution overlap or rank correlation.
Sparsity	Is the explanation focused?	Count segments covering a chosen share of attribution magnitude.
Class contrast	Do Real/Fake/Mimic explanations differ where expected?	Compare class-specific attributions for the same recording.
Method agreement	Do SHAP and LIME point to similar regions?	Compute Agreement@k; flag low agreement for review.
Human interpretability	Can an	Small structured review of clarity

XI. SAMPLE OUTPUT

Fig. 5 and Table 3 show the form of the explanation output produced for one recording. The values are illustrative and generated from synthetic data to demonstrate the interface and report structure; they are not measured outputs of the trained RFMD model. Here the model assigns the highest probability to Fake, the highlighted regions represent the strongest positive contributions for the Fake class, and the LIME check partly agrees with SHAP.



Illustrative example with synthetic data - not a measured output of the trained model

Fig. 5. Sample explanation interface (illustrative): waveform with SHAP overlay, segment-wise attribution, and class probabilities.

TABLE 3. SAMPLE EXPLANATION REPORT (ILLUSTRATIVE)

Field	Sample value
Input	sample_case_001.wav (4.0s, 200ms windows, 20 segments)
Probabilities	Real 0.07 Fake 0.91 Mimic 0.02
Predicted class	Fake
Top SHAP segments	S6 (1.00–1.20s): +0.21; S7 (1.20–1.40s): +0.17; S15 (2.80–3.00s): +0.14
Segment against class	S10 (1.80–2.00s): -0.06
LIME top-3	S6, S15, S12
SHAP-LIME agreement	Agreement@3 = 2/3 (moderate; review optional)
Caution	Attribution explains model behavior, not the physical origin of the audio; use as supporting evidence only.

A machine-readable version of the same report (sample_output.json) can be used by the web interface to render the probability panel, overlay, and evidence list.

XII. COMPARATIVE ANALYSIS

TABLE 4. TYPICAL SPOOFING DETECTORS VS. RFMD-XAI

Parameter	Typical detectors	RFMD-XAI
Output classes	Often binary (real/fake)	Three-class (Real/Fake/Mimic)
Human mimicry	Usually not a class	Explicit class
Input representation	Often spectrogram/features	Preprocessed 1D waveform

Parameter	Typical detectors	RFMD-XAI
Model weight	Often heavy backbones	Lightweight 1DCNN
Explainability	Often absent	Post-hoc SHAP+ LIME
Temporal localization	Rare	Segment-level evidence
Cross-method check	Rare	SHAP vs LIME agreement
Faithfulness testing	Rare	Defined protocol
Stability testing	Rare	Defined protocol
Forensic report	Not provided	Structured 6-part report
Analyst review flag	Not provided	Low-agreement/low-confidence flag
Classifier modification	May alter model or loss	None (post-hoc)
Noise/cross-dataset baseline	Variable	Retained from [1]
Trust/transparency	Low	Higher (evidence-backed)

XIII. ADVANTAGES AND LIMITATIONS

RFMD-XAI offers: (i) transparency without architectural disruption, since the classifier and its reported performance are untouched; (ii) time-localized evidence to support analyst review, debugging, and model auditing; (iii) a three-class formulation that covers human mimicry; (iv) two complementary explainers whose disagreement is itself informative; and (v) a defined evaluation protocol so explanation quality is measured rather than judged by appearance.

Limitations are also acknowledged. Attribution quality depends on the explainer, background data, segmentation strategy, and perturbation design. SHAP and LIME explain the behavior of the model, not the physical origin of the audio, and may highlight nuisance cues, recording artifacts, or dataset-specific shortcuts. XAI output should therefore be treated as supporting evidence for model inspection rather than an independent forensic conclusion [24], [27].

XIV. FUTURE SCOPE

Planned extensions include multilingual explanations, cross-dataset attribution stability studies, more rigorous faithfulness testing including partial-spoof evaluation [24], analyst-in-the-loop validation, a full web-based forensic dashboard, and evaluation on additional spoofing corpora, all without changing the Real/Fake/Mimic problem definition.

XV. CONCLUSION

This paper's principal contribution is architectural: RFMD-XAI extends a validated lightweight 1DCNN forensic audio detector with a post-hoc explainability stage while keeping the dataset organization, preprocessing pipeline, three-class output, and training objective intact. Concretely, the framework contributes (1) a modular design with SHAP as the primary attribution method and LIME as a local cross-check; (2) a structured explanation report and sample output

stability, sparsity, class contrast, and human interpretability. The reference classifier results remain those reported in [1]; no new XAI performance values are fabricated. The next implementation step is to train or load the RFMD model, connect the post-hoc explainers, validate the explanations with the proposed tests, and generate forensic-friendly reports.

format suited to forensic review; and (3) a reproducible protocol (Section IX, Table 2) for measuring faithfulness,

REFERENCES

- [1] M. H. Abdulbas, R. K. Khalaf, A. T. M. Ali and N. D. Al-Shakarchy, "Detection of Real, AI-Synthesized, and Human-Mimicked Voices in Forensic Audio Using One-Dimensional Convolution Neural Networks," *Ingénierie des Systèmes d'Information*, vol. 31, no. 1, pp. 21–26, 2026, doi: 10.18280/isi.31.01.03.
- [2] Z. Khanjani, G. Watson and V. P. Janeja, "Audio deepfakes: A survey," *Frontiers in Big Data*, vol. 5, 1001063, 2023, doi: 10.3389/fdata.2022.1001063.
- [3] Z. M. Almutairi and H. Elgibreen, "Detecting fake audio of Arabic speakers using self-supervised deep learning," *IEEE Access*, vol. 11, pp. 72134–72147, 2023, doi: 10.1109/ACCESS.2023.3286864.
- [4] M. Todisco et al., "ASVspoof 2019: Future horizons in spoofed and fake audio detection," in *Proc. Interspeech*, 2019, pp. 1008–1012, doi: 10.21437/Interspeech.2019-2249.
- [5] X. Wang et al., "ASVspoof 2021: Spoofing countermeasures for automatic speaker verification," *arXiv:2109.00537*, doi: 10.48550/arXiv.2109.00537.
- [6] G. Lavrentyeva et al., "Audio replay attack detection with deep learning frameworks," in *Proc. Interspeech*, 2017, pp. 82–86, doi: 10.21437/Interspeech.2017-360.
- [7] F. Bencheikh, A. Cherif and H. Drias, "An intelligent information system for secure biometric authentication using deep learning," *Ingénierie des Systèmes d'Information*, vol. 28, no. 4, pp. 923–934, 2023, doi: 10.18280/isi.28.04.18.
- [8] D. E. Kouicem, A. Bouabdallah and H. Lakhlef, "Machine learning-based security solutions for intelligent information systems," *Ingénierie des Systèmes d'Information*, vol. 27, no. 6, pp. 1015–1026, 2022, doi: 10.18280/isi.27.06.15.
- [9] D. M. Ballesteros, Y. Rodriguez-Ortega, D. Renza and G. Arce, "Deep4SNet: Deep learning for fake speech classification," *Expert Systems with Applications*, vol. 184, 115465, 2021, doi: 10.1016/j.eswa.2021.115465.
- [10] S. A. AlAjmi, K. Hayat, A. M. AlObaidi et al., "Faked speech detection with zero prior knowledge," *Discover Applied Sciences*, vol. 6, no. 2, pp. 2086–2092, 2024, doi: 10.1007/s42452-024-05893-3.
- [11] M. H. Abdulbas and N. D. Al-Shakarchy, "Person identification based on facial biometrics in different lighting conditions," *IJECE*, vol. 13, no. 2, pp. 2086–2092, 2023, doi: 10.11591/ijece.v13i2.pp2086-2092.
- [12] T. Kanwal, R. Mahum, A. M. AlSalman et al.,

“Fake speech detection using VGGish with attention block,”
EURASIP J. Audio, Speech, Music Process., vol. 2024, no. 1, 35,
2024, doi: 10.1186/s13636-024-
00348-4.

- [13] A. Hamza, A. R. R. Javed, F. Iqbal et al., “Deepfake audio detection via MFCC features using machine learning,” IEEE Access, vol. 10, pp. 134018–134028, 2022, doi: 10.1109/ACCESS.2022.3231480.
- [14] B. Magnier, I. M. Ales and F. Ales, “Speech forensics: Blind voice mimicry detection,” arXiv:2209.12573, 2022.
- [15] R. K. Khalaf and N. D. Al-Shakarchy, “Verifying the facial kinship evidence to assist forensic investigation based on deep neural networks,” in Emerging Trends and Applications in Artificial Intelligence, vol. 960, pp. 493–504, 2024, doi: 10.1007/978-3-031-56728-5_41.
- [16] A. T. M. Ali, H. Hallawi and N. D. Al-Shakarchy, “Cyber-vandalism detection in Wikipedia using light architecture of 1D-CNN,” Karbala Int. J. Modern Science, vol. 9, no. 4, art. 3, 2023, doi: 10.33640/2405-609X.3321.

IJETS