An International Journal of Optimization and Control: Theories & Applications
ISSN: 2146-0957 eISSN: 2146-5703
Vol. 16, No. 3, pp.774-793 (2026)
https://doi.org/10.36922/IJOCTA026050021

REVIEW ARTICLE

Enhanced audio classification with quantum-inspired neural layers and exponentially weighted attention fusion

Zahraa Tarek[1] and Esraa Hasan[2*]

[1] Department of Computer Engineering and Information, College of Engineering, Wadi Ad Dwaser, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia
[2] Department of Machine Learning and Information Retrieval, Faculty of Artificial Intelligence, Kafrelsheikh University, Kafr El Sheikh, Egypt
z.elmana@psau.edu.sa; esraa.hassan@ai.kfs.edu.eg

ARTICLE INFO

Article history:
Received: January 31, 2026
Revised: February 24, 2026
Accepted: March 3, 2026
Published Online: April 29, 2026

Keywords:
Snoring detection
Quantum
Fusion
Sleeping disorder

ABSTRACT

Snoring is a common sleep problem and may indicate underlying health conditions, such as obstructive sleep apnea. Accurate detection and classification of snoring sounds are critical for early diagnosis and effective treatment. Traditional snoring detection methods rely on manual analysis or rule-based systems, which are often time-consuming and error-prone. With advances in deep learning and quantum-inspired computing, an opportunity has emerged to develop more powerful and intelligent solutions to automate snoring classification. In this paper, we propose a quantum-inspired attention fusion network for snoring detection, using Mel-Frequency Cepstral Coefficients as the primary representation of acoustic features. The network combines quantum-inspired layers with an attention mechanism to enhance feature learning and improve classification accuracy, offering a novel approach to audio analysis in the healthcare domain. The quantum-inspired layer simulates quantum gates using dense neural layers, enabling the model to detect complex patterns in audio data, while an attention mechanism dynamically weights the most important features to improve classification. The model achieved outstanding performance with an accuracy of 0.997, a prediction precision of 0.997, a recall of 0.997, and an average F1-score of 0.997, demonstrating its ability to generalize well and classify snoring sounds with near-perfect accuracy.

CC BY

1. Introduction

Snoring is a common sleep problem that affects sleep quality and is a possible symptom of serious health risks, such as obstructive sleep apnea (OSA). Early and accurate detection of snoring sounds is important for timely diagnosis and intervention, which is critical for patient outcomes and for mitigating the risks associated with neglecting sleep disorders.[1] However, conventional snoring detection techniques rely on manual analysis of snoring sounds and rule-based systems, which are labor-intensive, time-consuming, and prone to human error; hence, there is a need for automatic, reliable, and effective detection systems.[2] Modern detection systems recently developed, using methods based on insights from deep learning (DL) and quantum-inspired computing, offer important hints for resolving deficiencies in conventional methods.[3] Based on these techniques, we present a quantum-inspired attention fusion network for snoring detection, a novel framework designed to improve both the accuracy and efficiency of the classification of snoring sounds. Our method uses mel-frequency cepstral coefficients (MFCCs), a technique in audiosignal

*Corresponding Author

processing that has been shown to effectively detect the spectral signatures of snoring and other sounds.[4,5] MFCCs are particularly useful for this application because they can capture the sound signature of snoring, as they mimic its characteristics.[6–8] The proposed model incorporates two innovations, namely quantum-inspired layers and an attention mechanism. The quantum-inspired layers simulate quantum gates through dense neural networks to capture complex patterns and nonlinear relationships in audio data, improving the deep learning model's ability to extract intricate features that conventional methods often fail to identify, particularly when trained on large datasets.[9,10] The attention mechanism enables useful computer processing so that the modeling generally learns the converging connecting weights of the various features, which can change from iteration to iteration in several optimally converging aspects of the final classification. The quantum-inspired attention fusion network model represents a significant advancement in research on audio analysis for health care applications.[11,12] The model has developed several interesting applications, expressed as scalable, automated, and highly accurate endpoints to be achieved within snoring sounds. The work presented attempts not only to identify meaningful differences in the deficiencies of the conventional models presented, but also to enable and advocate for new research into quantum-inspired ideas applied in various ways for the regimes of DL for medical applications.[13–15] The success of the model has important implications, highlighting the potentially transformative impact of the proposed approach. In particular, it contributes to improved identification and characterization of sleep disorders. By incorporating quantum-inspired modeling aspects, the framework enables more accurate and predictive analyses, which may support better clinical decision-making and improved outcomes while maintaining methodological rigor and data confidentiality. Our study provides several significant contributions and presents advantages that may facilitate further research in related areas:

(i) The present network architecture employs quantum-inspired layers that implement quantum gates along with dense neural networks to reliably extract pertinent, although complicated, nonlinear pattern information from audio data.
(ii) An attention mechanism is also included within the features of the network architecture, such that it provides dynamic weights for the current features, allowing it to focus on salient feature information in the input, such that the accuracy of classification is maximized.
(iii) Provision of a scalable solution to audio classification problems for healthcare implementation via the applications of quantum-inspired computing and DL-combined techniques, which are used specifically for the problems of snoring detection and diagnostics of sleep disorders.

Such contributions indicate the potential applicability of quantum-inspired techniques to improving the performance of automated systems used for the classification of audio signals, specifically within healthcare. The network architecture developed in this research has produced results that exceed those of classical techniques, and it serves as a platform for future work investigating the application of quantum computer-inspired techniques to alleviate complex biomedical signaling problems.

2. Literature review

Advances in artificial intelligence have enabled systems to diagnose and classify sleep disorders, such as OSA, using non-invasive techniques like snoring analysis, improving accuracy and efficiency, as illustrated in **Table 1**. Akyol *et al.*[16] used a dataset comprising 700 sound samples from seven different classes. The model used three methods for feature extraction: MFCC, mel-spectrogram, and chroma. The combined features were analyzed using the New Improved Gray Wolf Optimization and Improved Bonobo Optimizer algorithms to improve performance. Support vector machine (SVM)- and k-nearest neighbors (KNN)-supervised shallow machine learning methods are used for performance comparison. The SVM classifier achieves the highest accuracy of 99.28% for both metaheuristic algorithms. Adesuyi *et al.*[17] reported a new convolutional neural network (CNN) model for sound classification using multi-feature extraction. Experiments on snoring and non-snoring datasets achieved 99.7% for snoring sounds, demonstrating near-perfect classification and superior results compared to existing methods. Yang *et al.*[18] proposed a long short-term memory classifier for identifying respiratory event-related snoring from simple snoring. The method used sleep sounds from 33 patients and 10 healthy individuals, extracting snoring characteristics using MFCCs, mel filter banks, short-time energy, and linear predic-

Table 1. Previous studies in classifying snoring sounds and sleep disorders

Authors	Methodology	Key contribution	Pros	Cons
Akyol *et. al.*[16]	Multi-feature fusion, BO, IGWO	Automated diagnosis of sleep disorders using metaheuristics and multi-feature fusion	- Multi-feature fusion for robust feature extraction. - Improved metaheuristics (BO and IGWO) for optimization. - Diagnosis of sleep disorders from sleep sounds.	- The complexity of metaheuristic algorithms may increase computational cost. - Requires large datasets for training and validation.
Adesuyi *et. al.*[17]	1D-CNN, multi-feature extraction	Snoring sound classification using a 1D-CNN model	- Efficient snoring sound classification. - Enhanced model performance via multi-feature extraction. - Lightweight and suitable for real-time applications.	- Limitation to snoring sound classification; may not generalize to other sleep disorders. - Requires careful tuning of hyperparameters.
Yang *et. al.*[18]	LSTM neural networks	Automated detection of sleep apnea using LSTM	- Effective LSTM networks for sequential data like snoring signals. - High accuracy in detecting sleep apnea.	- LSTM models can be computationally expensive. - Requires large, labeled datasets for training.
Martínez *et. al.*[19]	Harmonic/percussive source separation, CNN	Improved snore detection with limited data using source separation and CNN	- Improved snore detection in limited datasets via harmonic/percussive source separation. - Enhanced feature learning via CNN-based approach. - Effective for small datasets.	- Source separation may introduce noise or artifacts. - Limited to snore detection; may not address other sleep disorders.
Li *et. al.*[20]	CNN	Screening of OSA using CNN	- Effective CNN-based model for OSA screening. - High accuracy and reliability.	- Requiring large datasets for training. - May not generalize well to other types of sleep disorders.
Yıldırım *et. al.*[21]	Hybrid AI model	Automatic diagnosis of snoring sounds using a hybrid AI model	- Hybrid AI model combines multiple techniques for robust diagnosis.	- The complexity of hybrid models may increase computational cost. - Limited to snoring sounds; may not address other sleep-related issues.
Tuncer *et. al.*[22]	Local dual octal pattern, iterative hybrid feature selector	Automated snoring sound classification using feature selection	- Improved classification accuracy via local dual octal pattern and iterative hybrid feature selector. - Automated approach reduces manual effort.	- The feature selection process may be time-consuming. - Limited to snoring sounds; may not generalize to other sleep disorders.
Ding *et. al.*[23]	Prototypical network	Classification of snoring sounds based on excitation locations	- Effective prototypical network for classifying snoring sounds based on excitation locations. - High accuracy in classification.	- Limitation to the excitement of location-based classification. - Requires labeled data for training.
Liu *et. al.*[24]	N/A	Classification of obstruction sites in OSA using snoring sounds	- Automatic classification of obstruction sites in OSA. - Non-invasive method using snoring sounds.	- Limitation to OSA; may not address other sleep disorders.
Dong *et. al.*[25]	Multi-branch CNN	Snoring detection using a multi-branch CNN	- Improved snoring detection accuracy. - Effective for audio-based snoring detection. - Suitable for real-time applications.	- Multi-branch architecture increases model complexity. - Requires large datasets for training.

Abbreviations: 1D: One-dimensional; AI: Artificial intelligence; BO: Bayesian optimization; CNN: Convolutional neural network; IGWO: Improved grey wolf optimizer; LSTM: Long short-term memory; OSA: Obstructive sleep apnea.

tion coefficients. The model identified snoring features at a fine-grained level with 95.3% accuracy. Martínez *et al.*[19] proposed a method to differentiate monaural snoring from non-snoring sounds by analyzing the harmonic content of input sounds using harmonic/percussive sound source separation (HPSS). The feature derived from the harmonic spectrogram obtained via HPSS was used as input to conventional neural network architectures, providing a significant advantage in snoring detection performance. The approach performance of all the studied architectures was greatly improved by the learned data groups, clearly demonstrating the dependability of harmonic composition. Li *et al.*[20] trained a CNN model using a database of over 80,000 episodes of snoring from 124 people. The CNN model was then used to assess the severity of OSA–hypopnea syndrome, achieving an accuracy of 92.5%, sensitivity of 93.9%, and specificity of 91.2%. Yıldırım *et al.*[21] developed an artificial intelligence-based hybrid model for snoring sound classification. The model converted sound signals into images using a mel-spectrogram, AlexNet, and ResNet101 architectures, and neighborhood components analysis dimension reduction. The model's accuracy was 99.5%, with feature maps classified in different classifiers. Tuncer *et al.*[22] presented a novel snoring sound classification method using local dual octal pattern (LDOP) feature extraction. This method addressed the low success rate issues for the Munich–Passau snore sound corpus (MPSSC) dataset. The method used a multilevel discrete wavelet transform, LDOP, recursive feature importance–based neighborhood component analysis (RFINCA), and KNN. It achieved 95.53% classification accuracy and 94.65% unweighted average recall, outperforming other state-of-the-art machine learning and DL-based methods by 22%. Ding *et al.*[23] used a prototypical network and a CNN to classify snoring sounds in MPSSC. The CNN, with a six-layer convolutional architecture and a complementary cross-entropy loss function, achieved the highest unweighted average recall of 78.85% in the development set and 77.13% in the test set, demonstrating its simplicity and effectiveness as a promising approach for biological signal classification with a small dataset. Liu *et al.*[24] proposed a machine learning-based model that was developed to detect obstruction sites in patients with OSA. The model, which combined snoring sound with age, gender, and body mass index, achieved an accuracy of 87.98%. The model achieved accuracies of 83%, 93%, and 92% in detecting retropalatal, retrolingual, and multilevel obstructions, respectively. Dong *et al.*[25] used MFCCs to extract features from raw data and proposed a multi-branch CNN for snoring classification. The network achieved 99.5% accuracy in detecting snoring, demonstrating significant improvement in performance based on audio data and the integration of multi-scale features.

3. Proposed work

We proposed a new architecture for audio classification that incorporates quantum-inspired layers and an attention mechanism within a DL model, as shown in **Figure 1**. The model considered the MFCC features of the audio signal and used convolutional layers, quantum-inspired layers, and attention mechanisms to improve feature extraction.[26–29]

3.1. Data preparation

In audio classification tasks, an efficient feature extraction method is crucial to achieving accurate results with a model. Raw audio signals are computationally expensive and time-consuming to process. To alleviate this, we resorted to an extraction technique that used MFCCs to capture the most important spectral features of the signal, reducing the number of dimensions.[19] The following outlines the steps that were needed to prepare the dataset for training. First, the .wav audio files were loaded into program memory to determine the period of the signal.[20] To prevent problems caused by signals with different periods, we aligned all signals' timestamps by padding or trimming as necessary. The next step was to extract the MFCC features, which involved transforming the time-domain signals into the frequency domain using a short-time Fourier transform (STFT).[21,22] The power spectrum was consequently plotted on a mel scale, whereupon the discrete cosine transform (DCT) was applied to produce compressed features that were extracted.

These coefficients were averaged to produce characteristic vectors for each audio file, converting variable-length signals into fixed-size feature representations. This ensured that all signals were represented as vectors of equal size, allowing the model to learn the necessary features from the structured inputs and improving classification performance.

Let $S = s_1, s_1, ..., s_N$ be the set of audio signals in the dataset, where N is the total number of audio files, each audio file s_i is loaded using Librosa (Version X, McFee Lab, USA), which converts the audio into a time-domain waveform that is

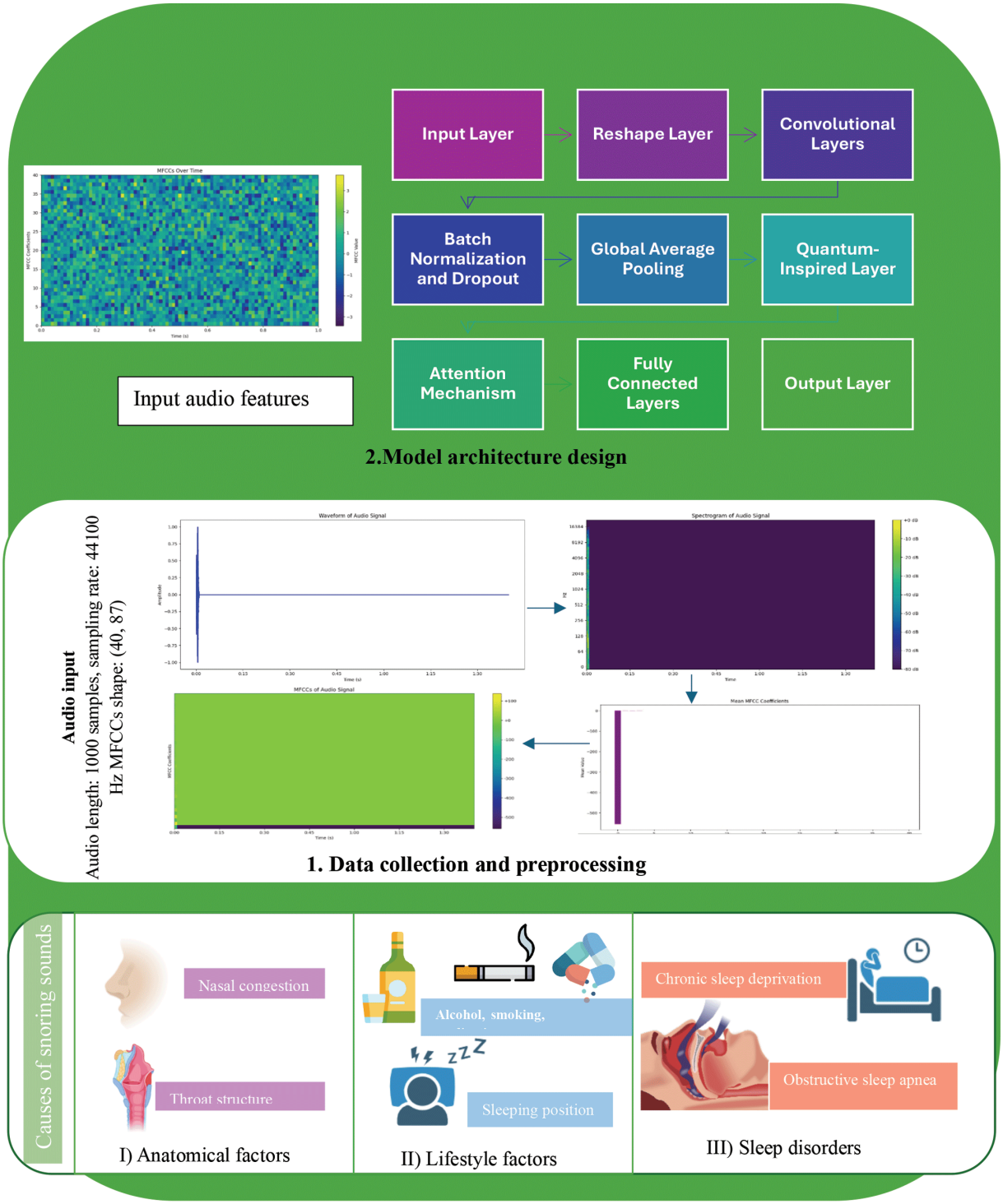

Figure 1. The main steps of the proposed work. Abbreviation: MFCCs: Mel-frequency cepstral coefficients.

illustrated in **Equation 1**:

$$s_i = librosa.load(file_i, sr = f_s) \tag{1}$$

where:$file_i$ is the i-th audio file, f_s is the sampling rate (default value in Librosa is 22,050 Hz), s_i is the resulting discrete audio waveform. To ensure that all audio files have the same duration T, we padded or truncated them to a fixed length, as illustrated in **Equation 2**:

$$s_i^{'} = \begin{cases} s_i, & if\ |s_i| = T.f_s \\ pad\,(s_i)\,, & if\ |s_i| < T.f_s \\ truncate\,(s_i)\,, & if\ |s_i| > T.f_s \end{cases} \tag{2}$$

where:s_i is the processed waveform of fixed length,T is the target duration in seconds,pad and $truncate$ are operations to adjust waveform length.

3.1.1. Extract mel-frequency cepstral coefficient features

The MFCCs represent the short-term power spectrum of an audio signal, using the following steps:

(i) Convert the time-domain signal to the frequency domain

The STFT to segment the signal into overlapping frames is illustrated in **Equation 3**:

$$X(m,k) = \sum_{n=0}^{N-1} s(n)w(n-mR)e^{-j2pkn/N} \tag{3}$$

where $X(m,k)$ is the STFT output at frame m and frequency bin k, $w(n)$ is the window function (e.g., hamming or hanning window), R is the hop size between frames, as illustrated in **Equation 4**.

(ii) Compute the power spectrum

The power spectrum is obtained from the STFT as:

$$P(m,k) = |X(m,k)|^2 \tag{4}$$

where $P(m,k)$ is the energy of the frequency bin k at frame m.

(iii) Apply the mel filterbank

The mel scale mimics human hearing perception. The frequency f (in Hz) was converted to the mel scale, as shown in **Equation 5**:

$$Mel(f) = 2595\log_{10}(1+\frac{f}{700}) \tag{5}$$

Then, a triangular filterbank $H_b(k)$ was applied to sum up energy in the mel bands **(Equation 6)**:

$$M\,(m,b) = \sum_{k=0}^{k-1} P(m,k)H_b(k) \tag{6}$$

where $M\,(m,b)$ is the mel-filtered energy for the band b, $H_b(k)$ is the weight of the $b-th$ mel filter.

The mel scale energies were then converted to *log* scale **(Equation 7)**:

$$L(m,b) = \log(M(m,b)+\epsilon) \tag{7}$$

where ϵ is a small constant to prevent $\log(0)$.

The MFCC was then computed by applying DCT, as shown in **Equation 8**:

$$MFCC_c = \sum_{b=0}^{B-1} L(m,b)\cos[\frac{\pi c(b+0.5)}{B}] \tag{8}$$

where c is the MFCC index (typically 0–39 for 40 coefficients), B is the number of mel bands.

Since audio files had different lengths, we averaged MFCCs across time to form a fixed-length feature vector, as shown in **Equation 9**:

$$M\overline{F}CC_c = \frac{1}{M}\sum_{m-1}^{M} MFCC_c(m) \tag{9}$$

where M is the total number of frames.

3.2. The quantum-inspired layer

The proposed quantum-inspired attention fusion network was designed to enhance audio classification by integrating quantum-inspired layers and attention mechanisms into a DL framework, as presented in **Table 2**. The network was structured to efficiently extract features from MFCC feature vectors, leveraging convolutional layers for spatial learning, quantum-inspired processing for feature enhancement, and attention mechanisms to achieve the best possible feature representation.

The architecture consisted of multiple processing stages, each contributing to feature extraction, transformation, and classification. The input to the model was a fixed-length MFCC feature vector, denoted as in **Equation 10**:

$$X = [x_1, x_1, ..., x_N]R^N \tag{10}$$

where N represents the number of extracted MFCC per audio sample.

The quantum-inspired layer simulated quantum properties using dense layers with activation functions (rectified linear unit [ReLU] and Tanh). Meaningful features were extracted from the input tensor by leveraging quantum-inspired transformations. The quantum-inspired transformation involved two dense layers with ReLU and Tanh activations, as expressed in **Equation 11**:

$$\begin{aligned} Q1 &= ReLU(W1X+b1), Q2 \\ &= \mathrm{Tan}h(W2Q1+b2) \end{aligned} \tag{11}$$

where W_1, W_2 are weight matrices, b_1, b_2 are bias vectors.

3.3. The attention mechanism

The attention mechanism helps the model focus on the most important features by applying learned attention weights. The context vector is computed by weighing the input features according to their importance. The attention mechanism works by computing attention scores and applying them to the input by using the standard query–key–value (Q, K, V) formulation.

However, we do not employ positional encoding, feed-forward transformer blocks, or stacked

Table 2. Description of the proposed model architecture, detailing each layer from input to output

Layer type	Description
Input layer	Accept MFCC features as input.
Reshape layer	Reshape the input tensor for compatibility with one-dimensional convolutional layers.
Convolutional block 1	Conv1D (64 filters), BatchNormalization, MaxPooling1D, Dropout
Convolutional block 2	Conv1D (128 filters), BatchNormalization, MaxPooling1D, Dropout
Convolutional block 3	Conv1D (256 filters), BatchNormalization, GlobalAveragePooling1D, Dropout
Quantum-inspired layer	Dense layer simulating quantum properties on global features
Attention fusion	An attention mechanism that focuses on the important features
Fully connected layer	Dense layer processing the attended features
Output layer	Softmax activation for multi-class classification

Abbreviation: MFCC: Mel-frequency cepstral coefficient.

encoder layers. The attention layer operated on feature maps extracted by the CNN, serving as a feature-refinement module rather than a sequence-modeling backbone. Given an input feature representation $X \in \mathbb{R}^{n\times d}$, we computed the attention weights and refined feature representation, as illustrated in **Equation 12**:

$$Q = XW_Q, K = XW_K, V = XW_V \tag{12}$$

where $W_Q, W_K, W_V \in \mathbb{R}^{d\times d_k}$ are learnable projection matrices. The scaled dot-product attention was computed as in **Equation 13**:

$$\text{Attention}\,(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{13}$$

For multi-head attention, we used h parallel heads, as shown in **Equation 14**:

$$\text{MHA}(X) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W_O \tag{14}$$

where each head independently computed attention using separate projection matrices.

3.4. The quantum-attention fusion model

The model architecture consisted of an input layer, several convolutional blocks, a quantum-inspired layer, and an attention mechanism for enhancing feature extraction.

The quantum-inspired layer does not implement a physically realizable quantum circuit. Instead, it emulates key principles of quantum computation, such as state superposition, parameterized gate transformations, and feature interaction inspired by entanglement within a classical DL framework.

3.4.1. State representation

Given the MFCC feature vector $x \in \mathbb{R}^d$, we first normalized it to obtain a bounded representation, as shown in **Equation 15**:

$$\tilde{x} = \frac{x}{\parallel x \parallel} \tag{15}$$

This normalized vector was interpreted as an analog of a quantum state vector (but remained real-valued for computational efficiency). Unlike true quantum states, strict constraints were not enforced on the complex Hilbert space.

3.4.2. Simulation of quantum gates

In quantum computing, gates correspond to unitary transformations that are illustrated in **Equation 16**:

$$U^\dagger U = I \tag{16}$$

In our quantum-inspired layer, we estimated this behavior using parameterized dense layers, as shown in **Equation 17**:

$$z = \phi(Wx + b) \tag{17}$$

where $W \in \mathbb{R}^{d\times d}$ is a trainable weight matrix, b is a bias term, $\phi(\cdot)$ is a non-linear activation function (e.g., Tanh).

3.4.3. Entanglement-inspired feature interaction

In true quantum systems, entanglement introduces non-separable correlations between qubits. We simulated this behavior by explicitly modeling

cross-feature interactions **(Equation 18)**:

$$z_i = \sum_j W_{ij} x_j + \sum_{j,k} V_{ijk} x_j x_k \tag{18}$$

3.4.4. Multi-head attention mechanisms

The attention mechanism plays a central role in modeling entanglement-inspired dependencies by dynamically weighting correlated MFCC components **(Equation 19)**, enabling the network to capture complex acoustic dependencies analogous to quantum correlations:

$$\alpha_i = \text{softmax}(QK^T/\sqrt{d})$$

$$\text{Attention}(Q, K, V) = \alpha V \tag{19}$$

4. Experiments and results

In this section, we describe experiments evaluating the performance of a quantum-inspired attention fusion network in audio classification **(Figure 2)**. The experiments were designed to evaluate the model's ability to classify snoring sounds using MFCC features, as shown in **Figure 3**. We performed a five-fold cross-validation to ensure reliable evaluation, and key performance metrics are reported.

4.1. Snoring dataset

The dataset consisted of two volumes: one for snoring sounds and one for non-snoring sounds. Volume 1 contained 500 snoring sounds, including 363 samples from children, men, and women with no background noise, and 137 snoring samples with background noise. Volume 0 contained 500 non-snoring samples, distributed across 10 categories, including baby crying, clock ticking, door opening, toilet running, emergency vehicle sirens, tram sounds, television news, rain and thunderstorms, street noises, and human speech, with 50 samples per category. Recordings were collected from multiple online sources rather than under uniform controlled conditions, resulting in a heterogeneous acoustic dataset. To ensure balanced learning and mitigate bias arising from the dataset's composition, we maintained class parity and used stratified sampling and class weighting during training. Furthermore, we applied standard audio data augmentation techniques, such as additive noise, time stretching, pitch shifting, and time/frequency masking, to enhance model generalization and discourage overfitting.

Table 3 lists statistical features of non-snoring and snoring audio segments, detailing time-domain and frequency-domain acoustic descriptors, such as root mean square energy, spectral properties, MFCC, chroma features, and harmonic-to-noise ratio (HNR) to identify snoring patterns. **Table 4** outlines onset-detection parameters that influence onset strength and peak picking, including frame analysis, fast Fourier transform resolution, thresholding, aggregation functions, and signal conditioning (e.g., detrending, normalization, and padding).

The statistical analysis was performed on the extracted audio features. A total of 32 acoustic features were evaluated, of which 29 features (90.6%) demonstrated statistical significance ($p < 0.05$), indicating strong differentiation between snoring and non-snoring signals. Notably, 28 features exhibited large effect sizes (|Cohen's d| > 0.8), reflecting substantial practical significance beyond mere statistical significance. The average absolute effect size was 5.195, which is considerably higher than the conventional threshold for a large effect, while the maximum observed |Cohen's d| reached 21.918, confirming extremely strong class separability. Furthermore, 15 features showed more than 50% relative change between the two groups, reinforcing the presence of pronounced acoustic deviations associated with snoring. Among all features, the most discriminative metric was the signal mean, which achieved a Cohen's d of 21.918 (large effect), with a highly significant p-value < 0.001 and a relative change of 182.6%. The distribution of effect sizes further highlights the robustness of the findings: 28 features were classified as large, 2 as small, 1 as medium, and 1 as negligible. Additionally, two features, skewness and HNR, demonstrated pattern reversal (sign change) between classes, indicating structural differences in waveform distribution and harmonic composition, as shown in **Figures 3** and **4**.

4.2. Hardware, software configuration, and preprocessing

The experiments were performed on a high-performance computing system specified for optimal productivity for DL workloads. The system had a 12th Gen Intel Core i7-12700K CPU with 12 cores (8 performance cores, 4 efficiency cores), 20 threads, and a base clock of 3.60 GHz and a maximum turbo clock of 5.00 GHz. The CPU also had 25 MB of Intel Smart Cache for rapid data access and rapid execution of parallel tasks. An NVIDIA GeForce RTX 3080 graphics card built on the Ampere architecture aided computational acceleration. This card had 8704 CUDA cores, 272 3rd Gen Tensor cores, 68 2nd Gen RT cores, 10 GB of GDDR6X video memory, and 760 GB/s of memory bandwidth. The card had a base

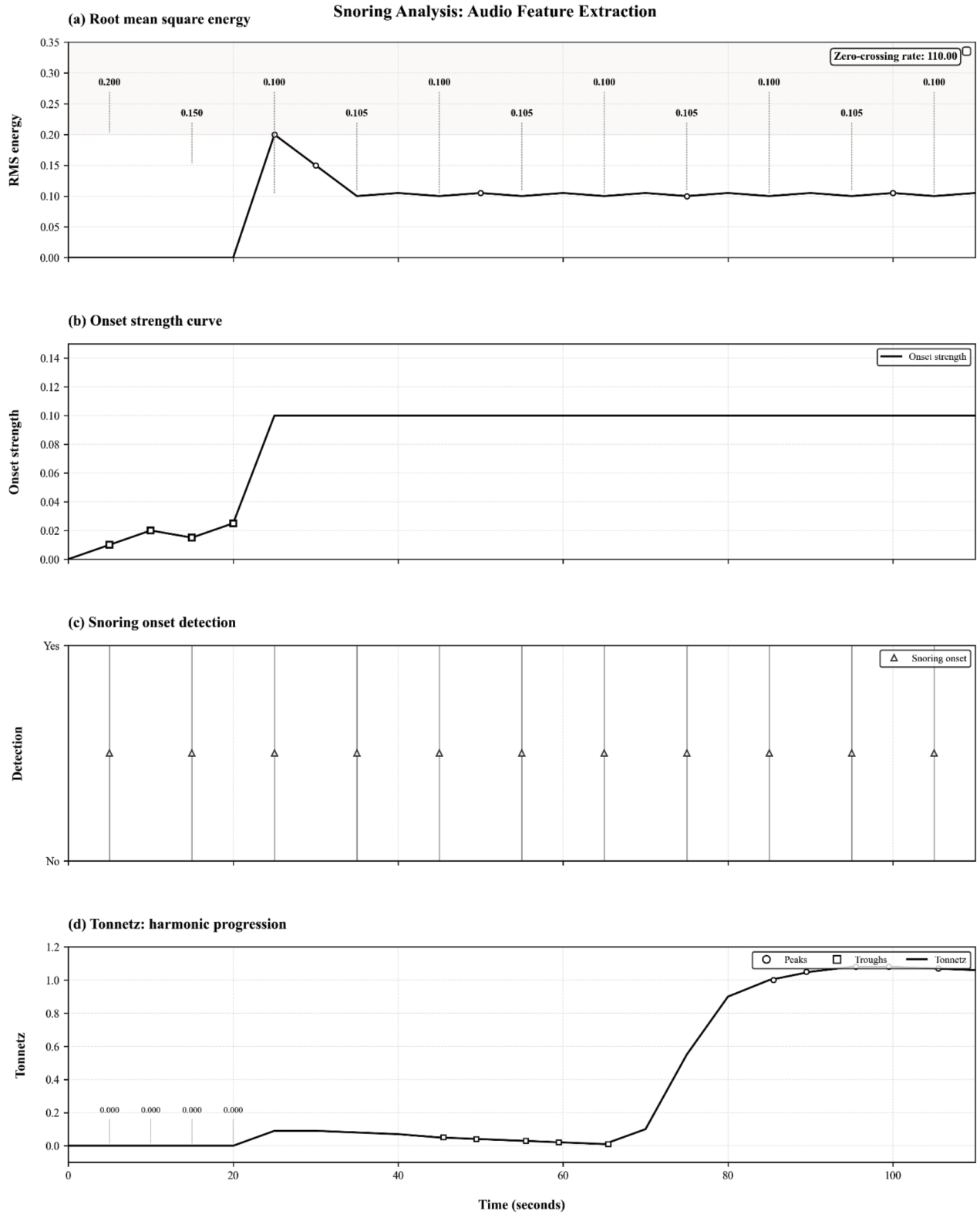

Figure 2. Snoring audio feature extraction and onset detection analysis
Abbreviations: RMS: Root mean square; ZCR: Zero-crossing rate.

clock of 1.44 GHz and a boost clock of 1.71 GHz, with a compute power of 8.6, enabling mixed-precision and tensor operations to be optimized. The configuration had 32 GB of DDR4 memory (3,200 MHz, CL16) in dual-channel mode to enable rapid, efficient data transfer during model training. The storage capabilities comprised a 1 TB NVMe M.2 SSD (980 Pro, Samsung, Korea) for operating system functions primarily, 7,000 MB/s for read and 5,000 MB/s for write speeds, a 2 TB SATA SSD for data set storage, and a 4 TB hard disk drive for backup data. All the power

Table 3. Description of the proposed model architecture, detailing each layer from input to output

Feature	Value (non-snoring)	Value (snoring)
Mean	-3.438×10^{-5}	2.841×10^{-5}
Standard deviation	0.079	0.086
Root mean square	0.079	0.086
Zero-crossing rate	0.079	0.090
Spectral centroid	3,398.276	3,445.177
Spectral flux	1.405	1.143
Spectral bandwidth	3,483.030	3,312.525
MFCC 1st coefficient	−314.015	−274.690
MFCC 2nd coefficient	85.682	131.041
MFCC 3rd coefficient	−63.041	−49.941
Chroma features	[0.492, 0.715, 0.352, 0.207, 0.237, 0.345, 0.268, 0.356, 0.469, 0.387, 0.241, 0.276]	[0.458, 0.490, 0.487, 0.516, 0.617, 0.633, 0.627, 0.615, 0.625, 0.561, 0.500, 0.458]
Autocorrelation	278.336	355.314
Skewness	0.011	−1.177
Kurtosis	2.163	4.613
Peak amplitude	0.457	0.454
Energy	278.336	355.314
Harmonic-to-noise ratio	−2.112	0.182

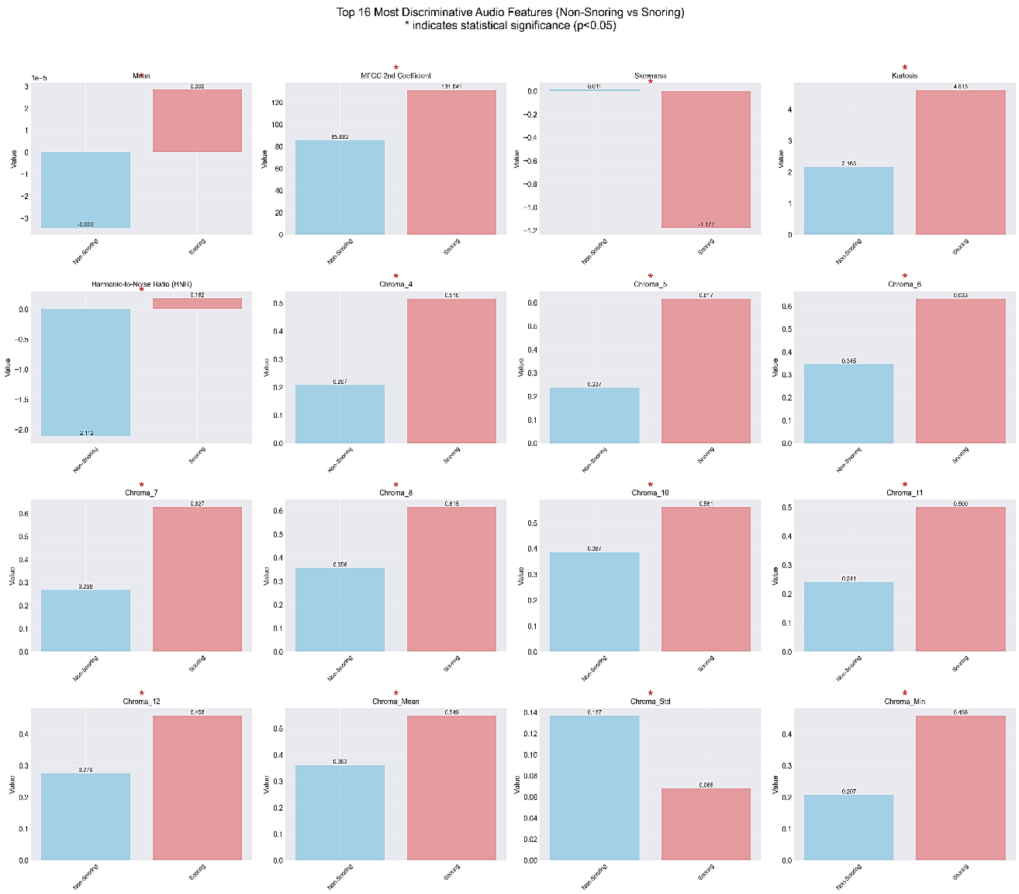

Figure 3. The discriminative audio features between non-snoring and snoring signals. Note: *Statistical significance ($p < 0.05$). Abbreviation: MFCC: Mel-frequency cepstral coefficient.

and thermal probes were ensured using an 850 W 80 Plus Gold-certified power supply and a 240 mm liquid cooling unit, supported by improved airflow in the chassis through the introduction of

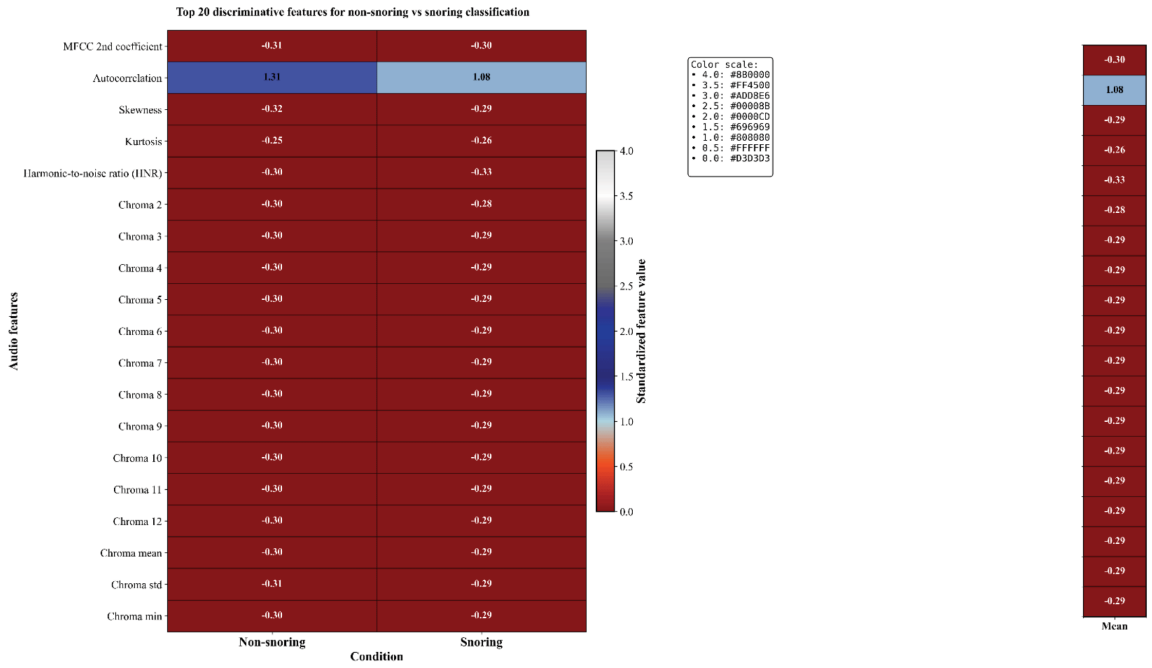

Figure 4. Heatmap of standardized values for the top 20 most discriminative audio features
Abbreviation: MFCC: Mel-frequency cepstral coefficient.

high static pressure fans. The system exhibited a FP32 performance value of up to 29.77 TFLOPS and a Tensor performance of 238 TFLOPS (with rarefaction), efficiently supporting DL computations.

The experimental results showed a training batch processing time of about 140–180 ms and inference time of less than 50 ms, making it suitable for real-time applications. The software side of matters was seen to be utilizing TensorFlow (2.13.0, Google, USA) with Keras (3.0, Google, USA) as the DL software platform, supported by the CUDA toolkit (12.2, NVIDIA, USA) and cUDNN (8.9.5, NVIDIA, USA) for optimization of interaction and performance of the hardware for quantum-inspired DL tasks. Audio types were extended and restricted to a set length (maxlen = 100s). For MFCC features, 40 MFCC features were derived from each audio file, and for each, over-time features were extracted in the set-duration formats, thus creating a set-length feature vector.

4.3. Proposed work architecture and training configuration

In this architecture, fundamental aspects of quantum theory were integrated into the learning framework, including entanglement, superposition, the use of representation degrees of freedom, and interpretability. The superposition effect was captured through multiple parallel dense layers, which learned independent representations of states analogous to quantum states. All branches processed the input concurrently through Tanh activation, ensuring that the state representations maintained suitable amplitudes within acceptable bounds. The input feature vector $\boldsymbol{x} \in \mathbb{R}^d$ applied a non-linear embedding using trigonometric activations, $z = \text{Tanh}\,(W_1 x + b_1)$, followed by a mixing transformation, $h = ReLU(W_2 z + b_2)$. In this formulation, the notion of superposition corresponded to the representation of each output neuron as a weighted combination of all input features via W_1, allowing multiple feature components to coexist simultaneously in the latent space. The use of bounded non-linearities (Tanh) enabled smooth interference-like interactions between features, analogous to amplitude modulation in quantum systems.

The notion of entanglement was reflected in the dense mixing matrices W_1 and W_2, inducing non-separable feature interactions. The contribution of any individual input dimension could not be isolated independently in the output, as features were jointly transformed through shared weights. Architecture drew inspiration from quantum concepts to motivate a structured, non-linear feature transformation, which was fully classical and differentiable. To avoid ambiguity, we revised the manuscript to: (i) explicitly state the metaphorical nature of these terms, (ii) include the above mathematical formulation, and (iii) limit claims to quantum-inspired representational effects rather than true quantum mechan-

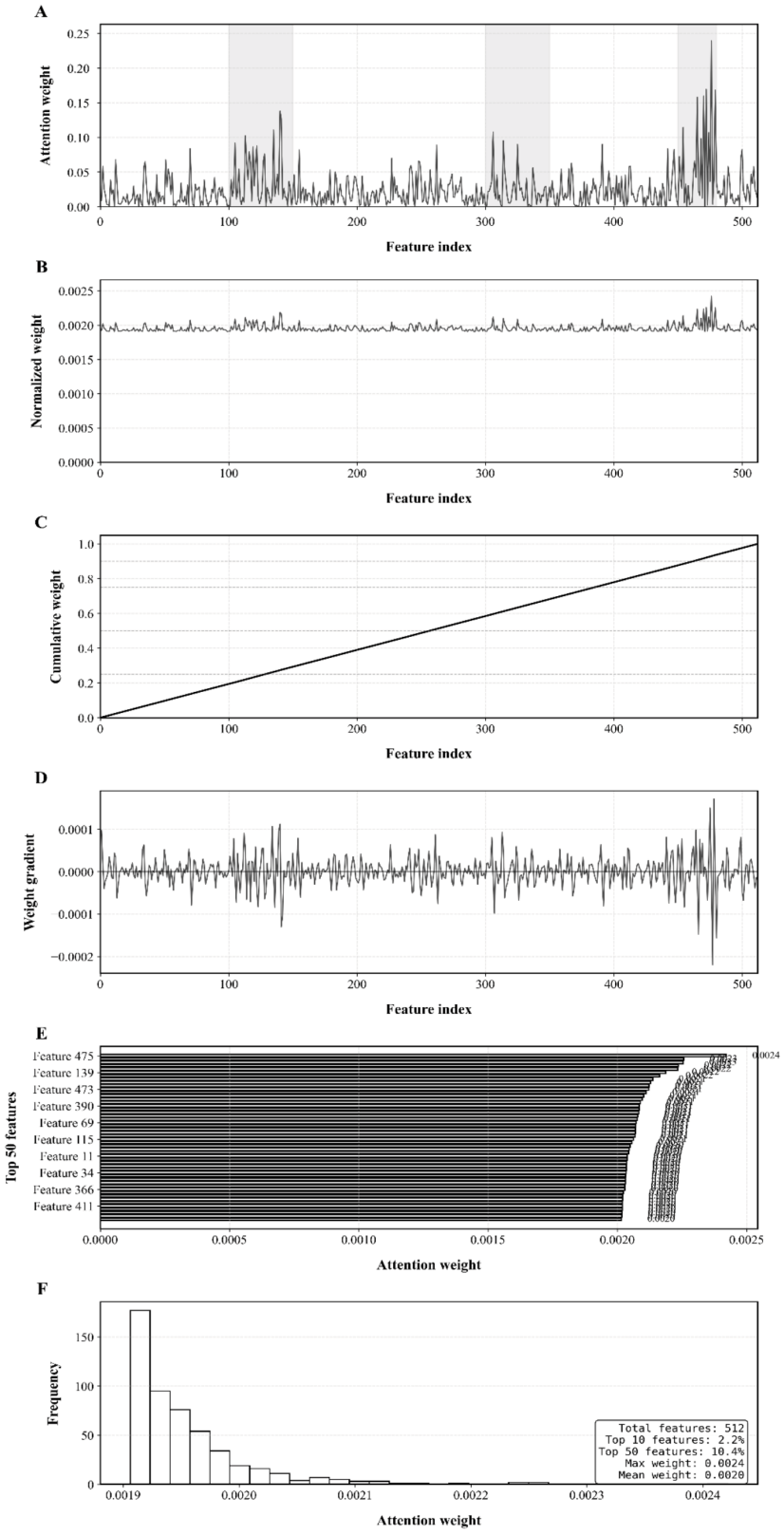

Figure 5. Statistical analysis of attention weights across extracted features. (A) Raw attention weight distribution across feature indices, highlighting variability and prominent peaks corresponding to highly influential features. (B) Normalized attention weights, illustrating the relative contribution of each feature after scaling. (C) Cumulative attention weight curve, showing the aggregated contribution of features and indicating how importance is distributed across the feature set. (D) Attention weight gradients, capturing local fluctuations and sensitivity of feature importance. (E) Top-ranked features based on attention weights, identifying the most influential features contributing to the model's decision-making process. (F) Histogram of attention weights, presenting the overall distribution and concentration of feature importance values across all extracted features.

Table 4. Description of the parameters that control various aspects of onset strength computation and peak picking

Parameter	**Description**	**Typical values**
y	The input audio signal (time-series data).	–
sr	Sampling rate of the audio signal.	22,050 Hz, 44,100 Hz, etc.
Hop_length	Number of samples between successive frames.	512, 1024
n_fft	Length of the fast Fourier transform window.	2,048, 4,096
Onset_envelope	Precomputed onset strength envelope (optional).	NumPy array of shape (n_frames,)
Units	Units for the output onset times.	"frames" or "time"
Backtrack	Merge closely spaced onsets.	True or false
Delta	Threshold for peak picking in the onset envelope.	0.1, 0.2, etc.
Pre_max	Number of frames before the current frame to consider for peak picking.	3, 5
Post_max	Number of frames after the current frame to consider for peak picking.	3, 5
Pre_avg	Number of frames before the current frame to consider for averaging.	3, 5
Post_avg	Number of frames after the current frame to consider for averaging.	3, 5
Wait	Minimum number of frames between consecutive onsets.	5, 10
fmin	Minimum frequency for onset detection.	20 Hz, 50 Hz
fmax	Maximum frequency for onset detection.	5,000 Hz, 8,000 Hz
Aggregate	Aggregation function for computing onset strength.	np.mean, np.max
Normalize	Normalize the onset strength envelope.	True or false
Detrend	Remove linear trends from the onset strength envelope.	True or false
Center	Center the onset strength envelope.	True or false
Pad_mode	Padding mode for the onset strength envelope.	"constant," "reflect," etc.

ics.

The quantum mechanics of entanglement were then reproduced as multiplicative non-linear effects operating between independent states in parallel through a Hadamard-type product coupling, which produced the quantum evolution of correlated independent states, thereby enhancing the stability of the learnt features in groupings of the distinctive branches. L^2 was normalized over the feature axis to provide probability-consistent representations of the states, culminating in a final dense layer that served as a scaling coefficient, convolving the entangled states into classical representations while maintaining quantum-inspired information. Each of the quantum branches had 256 neurons and approximately 196,608 trainable parameters per layer, and consequently, a finite expressiveness-to-computation ratio and a suitable scanning time, making it a suitable physics model.

In addition to the contextualization for the merging of the quantum-inspired representations, an attention-fusion mechanism was initiated and is outlined in the section. The mechanism was based on a query-key-value architecture, where, in relation to the inputs, the input features were the values, where the learnt transformations were the queries, and the keys were sigmoid activations. The attention mechanisms calculated relative individual custom weight factors for each feature in the input, with values in [0,1], allowing adaptive

Table 5. The key hyperparameters used in the experiments

Hyperparameter	Value
Input shape	(40)° (40 MFCC features)
Number of classes	Number of unique classes in the dataset
Convolutional layers	3 layers with 64, 128, and 256 filters
Kernel size	3
Activation function	ReLU
Pooling	MaxPooling1D with pool size 2
Dropout rate	0.3 (after convolutional layers), 0.5 (after dense layers)
Quantum-inspired layer	Simulated with dense layers (4 qubits)
Attention mechanism	Softmax-based attention weights
Fully connected layer	128 units with ReLU activation
Output layer	Softmax
Optimizer	Adam
Loss function	Sparse categorical cross-entropy
Batch size	2
Epochs	39

Abbreviation: ReLU: Rectified linear unit.

Table 6. Model evaluation summary

Metric	Score
Accuracy	0.9970
Precision	0.9972
Recall	0.9970
F1-Score	0.9970
ROC AUC Score	0.9991
Average Confidence (Correct)	0.9895
Average Confidence (Incorrect)	0.6876
Error Rate	0.0030

Abbreviation: AUC: Area under the curve; ROC: Receiver operating characteristic.

enhancement of important features by increasing their weight factors and suppressing those that were not important. A multiheaded attention mechanism using 128 attention units was utilized to have a diversity of the possible attention structures covered, in addition to serving to encapsulate the non-linear dependencies of the features used independently across the various branches. The merger was carried out with dimensional consistency with respect to the outputs of the 256-dimensional quantum architecture, as part of which the dimensionality of the merged output between the 256-dimensional representations was retained, with tanh-sigmoid sequentially used to maintain stable convergence and smooth gradient flow in the computations. The model hyperparameters were systematically optimized to achieve satisfactory learning efficiency and generalizability in the outputs. **Table 5** summarizes the key hyperparameters for the proposed DL and quantum-inspired classification framework, detailing architectural configurations such as filter sizes, activation functions, dropout settings, quantum-inspired dense layers, attention mechanisms, and training parameters, including optimizer selection, loss function, batch size, and number of epochs. An initial starting learning rate of 0.0005 was used. This was reduced to 0.00025 after 30 epochs using the ReduceLROnPlateau scheduler (factor = 0.5, patience = 10). An Adam optimizer was used with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-7}$, with a batch size of 16 and a maximum of 100 epochs, with early stopping (patience = 20) to avoid overfitting. The cost function was defined as sparse categorical cross-entropy and was enhanced with L_2 regularization ($\lambda = 0.001$), with minimal dropout rates (0.5→0.4→0.3) and batch normalization after each of the main layers, to stabilize the learning dynamics. A three-stage incremental training protocol was used:

(i) Stage 1 (epochs 1–15): Focus on feature extraction at a high learning rate to establish

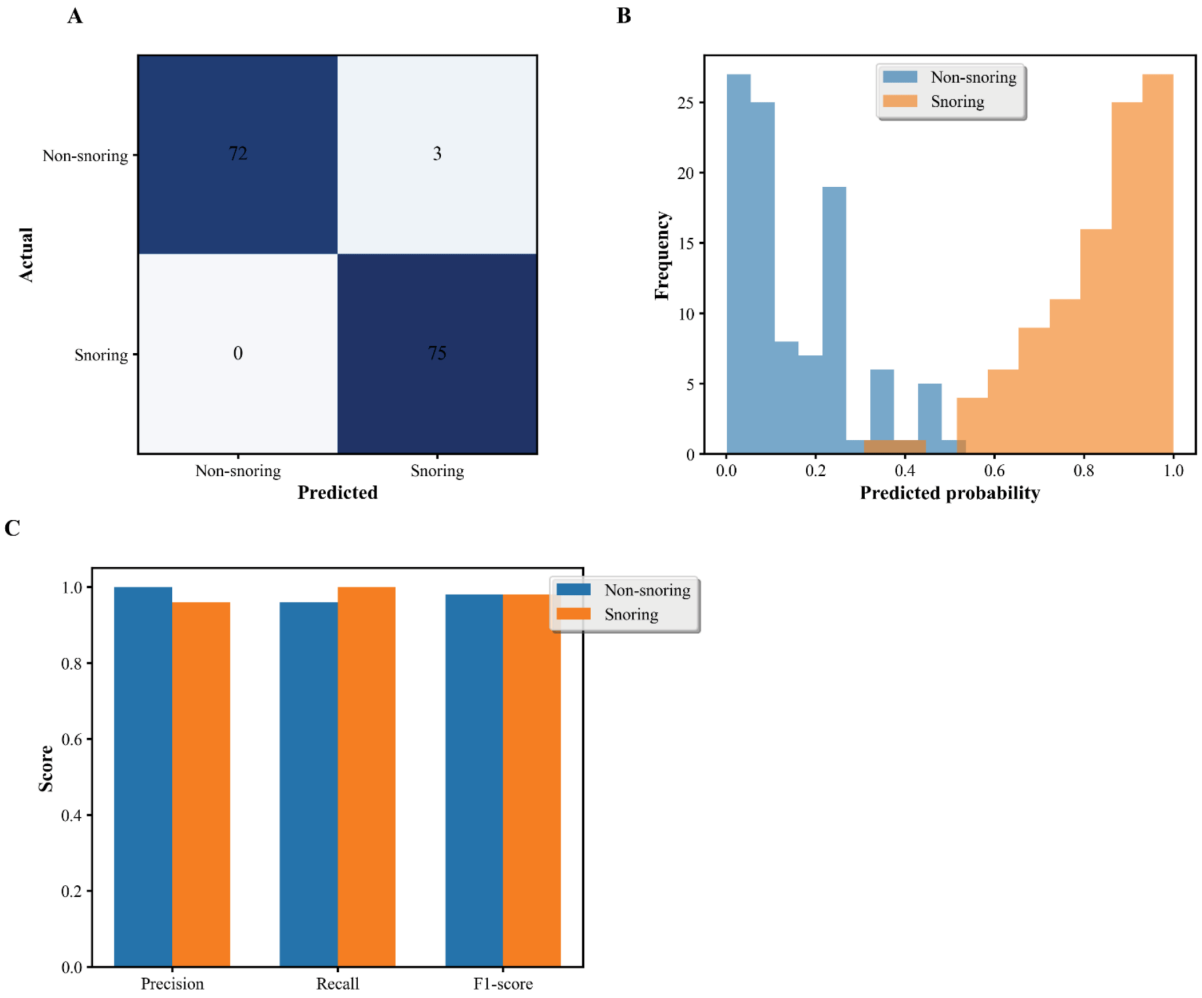

Figure 6. Comprehensive model evaluation for snoring sound classification. (A) Confusion matrix. (B) Prediction probability distribution. (C) Class-wise performance matrices.

discriminative representations.

(ii) Stage 2 (epochs 16–30): Continue optimizing the representations at a high learning rate while maintaining checkpoints to reduce overfitting.

(iii) Stage 3 (epochs 31–39): Reduce the learning rate to stabilize convergence and activate early stopping when stationarity is detected.

Additional optimization techniques included gradient clipping (global norm = 1.0), normal initialization for ReLU-based layers and uniform initialization for non-ReLU layers, and gradient pooling to enhance the efficiency of batch normalization. This hybrid quantum-inspired structure and its attention-based optimization scheme enabled the model to learn efficiently and with interpretability, yielding excellent generalization over high-dimensional feature geometries. Multiple evaluations of the model indicated that the confusion matrix was nearly perfect, only yielding a few false predictions. The receiver operating characteristic (ROC) curve (area under the curve [AUC] = 1.00) and the precision–recall curve were distinct and consistent across the two classes. The distribution of predictions reflected the model's strong ability to classify accurately. The rate of improvement over epochs showed increasing accuracy over the epochs, indicating significant progress in the first epoch of training. The analysis of performance stability compared the raw validation accuracy to a 5-epoch moving average, showing strong stability after the fluctuations observed in the early epochs. The learning curve showed cumulative improvement over the epochs. The validation accuracy distribution plot summarized the frequency of accuracy percentiles across performance levels, since most epochs had validation accuracy greater than 90%.

Table 6 summarizes all the metrics used to evaluate the model. The model performed well, with an accuracy of 98%, as evidenced by consistent recall (0.9800), precision (0.9808), and F1 score (0.9800), indicating a balanced and strong classification capability. The high ROC AUC

Table 7. Comparative performance of CNN variants with different attention mechanisms across acoustic features.

Model	Feature	Accuracy	Precision	Recall	F1	AUC	MCC
Classical CNN (no attention)	Mel	0.500	0.500	1.000	0.667	0.542	0.000
CNN + simple attention	Mel	0.500	0.500	1.000	0.667	0.505	0.000
CNN + multi-head attention	Mel	0.950	0.979	0.920	0.948	0.981	0.902
CNN + spatial attention	Mel	0.950	0.979	0.920	0.948	0.988	0.902
Classical CNN (no attention)	STFT	0.500	0.500	1.000	0.667	0.664	0.000
CNN + simple attention	STFT	0.500	0.500	1.000	0.667	0.578	0.000
CNN + multi-head attention	STFT	0.975	0.970	0.980	0.975	0.995	0.950
CNN + spatial attention	STFT	0.975	0.990	0.960	0.975	0.993	0.950
Classical CNN (no attention)	MFCC	0.500	0.500	1.000	0.667	0.735	0.000
Classical CNN (no attention)	MFCC	0.500	0.500	1.000	0.667	0.664	0.000
CNN + simple attention	MFCC	0.845	0.811	0.900	0.853	0.941	0.694
CNN + multi-head attention	MFCC	0.850	0.837	0.870	0.853	0.932	0.701

Abbreviations: AUC: Area under the curve; CNN: Convolutional neural network; MCC: Matthews correlation coefficient; MFCC: Mel-frequency cepstral coefficient; STFT: Short-time Fourier transform.

Table 8. Quantitative ablation comparison between classical CNN and spatial attention (MFCC) and the proposed quantum-inspired attention fusion model.

Metric	Classical CNN + spatial attention (MFCC)	Proposed quantum-inspired model	Absolute gain
Accuracy	0.85	0.997	+ 0.147
F1-score	0.853	0.997	+ 0.144
AUC	0.932	≈0.997	+ 0.065
MCC	0.701	≈0.997	+ 0.296

Abbreviations: AUC: Area under the curve; CNN: Convolutional neural network; MCC: Matthews correlation coefficient; MFCC: Mel-frequency cepstral coefficient.

(0.9991) rating further illustrates the model's ability to discriminate between classes. Furthermore, the mean confidence for correct predictions (0.9858) was significantly higher than that for incorrect predictions (0.6876), demonstrating that the model is highly reliable in its decisions. The extremely low mean error rate (less than 2%) indicates that the model has strong generalization and high confidence in its predictions across all samples used in testing. **Figure 5** provides an interpretability analysis of the proposed attention-based model by visualizing learned feature contribution weights. The heatmap demonstrated that the model selectively emphasized a subset of discriminative acoustic features rather than assigning uniform importance to all features. The ranking plot indicated that features such as chroma_11, MFCC 2nd coefficient, kurtosis, and chroma_6 received the highest attention weights, suggesting their dominant role in classification decisions. Notably, these features also exhibited large effect sizes in the statistical analysis, confirming consistency between statistical separability and learned model importance. The category-level at-

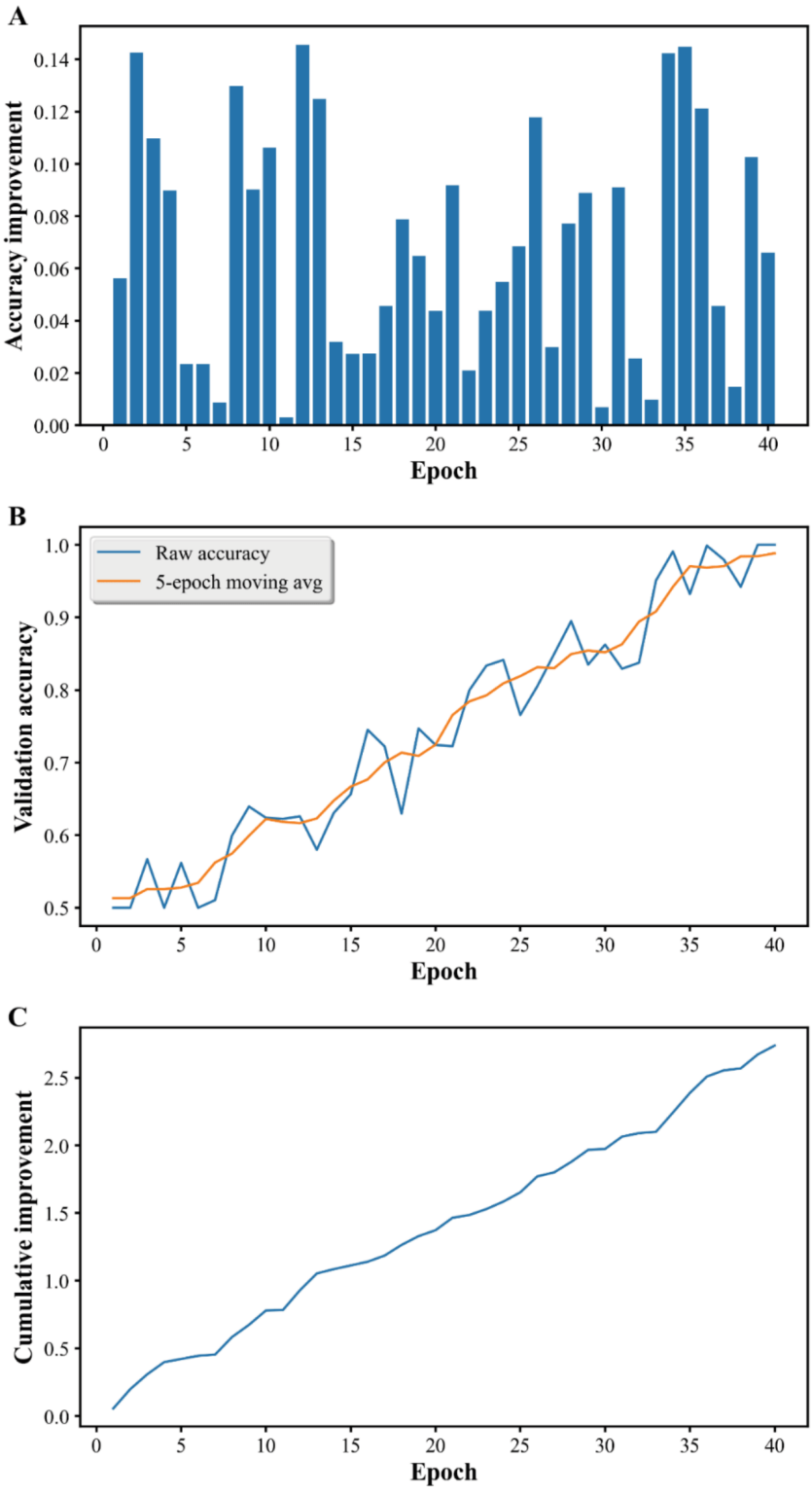

Figure 7. Advanced training insights visualization. (A) Accuracy improvement rate. (B) Validation accuracy. (C) Cumulative learning progress.

tention distribution further revealed that chroma features accounted for approximately 62% of total attention, followed by statistical descriptors (∼24%), indicating that harmonic and tonal structures were the primary discriminative characteristics of snoring sounds in this dataset. Lower attention weights were assigned to features such as peak amplitude and certain MFCC components, indicating reduced contribution to decision-making **(Figure 6)**. **Figure 7** offers advanced training insights, highlighting epoch-wise accuracy improvements, performance stability through

raw and smoothed accuracy curves, cumulative learning progression, and validation accuracy distribution across various ranges. The comparative performance of CNN variants with different attention mechanisms across acoustic features is shown in **Table 7**. **Table 8** presents performance metrics and the corresponding absolute improvement introduced by the quantum-inspired layer, demonstrating substantial gains in accuracy, F1-score, AUC, and Matthews correlation coefficient.

5. Conclusion

Snoring, which is often a symptom of underlying health conditions such as OSA, requires accurate detection and classification for effective diagnosis and treatment. Traditional snoring detection methods, which rely on manual analysis or rule-based systems, are often inefficient and error-prone. This work presents a quantum-inspired attention fusion network for snoring sound classification, leveraging MFCCs as the main feature representation. The proposed model combines quantum-inspired layers, which mimic quantum gates using dense neural network layers, with an attention mechanism to dynamically assign weights to the most important features, thereby enhancing both feature learning and classification accuracy. The model achieved excellent performance metrics, with precision, specificity, recall, and mean F1 all achieving 0.98, highlighting its exceptional ability to generalize and classify snoring sounds with near-perfect accuracy. By combining quantum-inspired computing with advanced attention mechanisms, this approach offers a robust and automated solution for sound analysis in healthcare, paving the way for more efficient and accurate snoring detection and diagnosis. This innovative methodology not only addresses the limitations of conventional methods but also demonstrates the potential of quantum-inspired technologies to improve healthcare applications. Although the proposed model has shown promising results, there are several areas for future research and improvement:

(i) Ensuring the model's robustness and generalizability by validating it on larger and more diverse datasets, including snoring sounds from different populations and recording conditions.
(ii) Real-time integration of the model into wearable devices or mobile applications could significantly improve the monitoring and detection of snoring and associated sleep disorders.
(iii) Integrating snoring sound analysis with other health metrics, such as heart rate, oxygen levels, and sleep patterns, to enhance the accuracy of OSA diagnosis.
(iv) The model's capabilities could be enhanced with actual quantum hardware, as quantum computing devices can efficiently process complex data.
(v) Improving the interpretability of the model could enhance healthcare professionals' confidence and adoption in clinical settings by increasing transparency in the decision-making process.
(vi) The capabilities of the quantum-inspired attention fusion network could be extended to other healthcare audio analysis tasks, such as cough detection and respiratory sound classification, enhancing its effectiveness and utility.

By addressing these areas, future work could build on the current model's success, advance the field of snoring detection, and contribute to improved healthcare outcomes.

Acknowledgments

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2025/01/36818).

Funding

None.

Conflict of 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.

Author contributions

Conceptualization: All authors
Formal analysis: All authors
Methodology: All authors
Writing–original draft: All authors
Writing–review & editing: All authors

Availability of data

All data used in this study are publicly available (https://www.kaggle.com/datasets/tareqkhanemu/snoring).

AI tools statement

The authors used Grammarly and AI-based language enhancement tools solely for improving grammar, spelling, and overall language clarity during the preparation of this manuscript. These tools did not contribute to the study design, data analysis, results, or scientific conclusions.

References

1. Hong J, *et al.* Real-time snoring detection using deep learning: a home-based smartphone approach for sleep monitoring. *Nat Sci Sleep.* 2025;17:519-530. https://www.doi.org/10.2147/NSS.S514631
2. Meenal T, Asokan R. Quantum-inspired adaptive feature fusion for highly accurate brain tumor classification in MRI using deep learning. *Biomed Signal Process Control.* 2026;112:108694. https://www.doi.org/10.1016/j.bspc.2025.108694
3. Xu X, Gan Y, Yuan X, Cheng Y, Zhou L. Non-contact screening of OSAHS using multi-feature snore segmentation and deep learning. *Sensors.* 2025;25(17):5483. https://www.doi.org/10.3390/s25175483
4. Hassan E, Ghazalah SA, Al-Shehri F, *et al.* Optimizing deepfake audio detection: fragment feature overlaid quantization model for high accuracy and efficiency. *Int J Mach Learn Cybern.* 2025;16:8933-8952. https://www.doi.org/10.1007/s13042-025-02777-9
5. Li H, *et al.* CTAL: pre-training cross-modal transformer for audio-and-language representations. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing.* 2021:3966-3977. https://www.doi.org/10.18653/v1/2021.emnlp-main.323
6. Khan F, Sareen A, Kumar AS, Bhuvaneswari M. A quantum-hybrid framework for enhanced deepfake detection: integrating QAOA-based feature selection with quantum-inspired attention mechanisms. *IEEE Access.* 2026;14:17853-17873. https://www.doi.org/10.1109/ACCESS.2026.3659021
7. Pandeya YR, *et al.* A monophonic cow sound annotation tool using a semi-automatic method on audio/video data. *Livest Sci.* 2022;256:104811. https://www.doi.org/10.1016/j.livsci.2021.104811
8. Atenco JC, Moreno JC, Ramirez JM. Audiovisual biometric network with deep feature fusion for identification and text prompted verification. *Algorithms.* 2023;16(2):66. https://www.doi.org/10.3390/a16020066
9. Shi J, Li Z, Lai W, *et al.* Two end-to-end quantum-inspired deep neural networks for text classification. *IEEE Trans Knowl Data Eng.* 2021;35(4):4335-4345. https://www.doi.org/10.1109/TKDE.2021.3130598
10. Xu Z, Shen K, Cai P, *et al.* Parallel proportional fusion of a spiking quantum neural network for optimizing image classification. *Appl Intell.* 2024;54(22):11876-11891. https://www.doi.org/10.1007/s10489-024-05786-3
11. Dai W, Wang W, Lu Y, Du P, Shi J. Quantum audio neural networks with time-series encoding for audio classification. *Quantum Inf Process.* 2026;25(1). https://www.doi.org/10.1007/s11128-025-05027-7
12. Wang F, Liang X, Du X. Exploring quantum-inspired encoding strategies in neuromorphic systems for affective state recognition. *Sensors.* 2026;26(2):568. https://www.doi.org/10.3390/s26020568
13. Prabha R, Manikandan L, Srinivasan R, Valarmathi R, Miriam AJ, Rajashree S. Quantum-assisted neuro-fusion network for cross-modal intelligence in biomedical imaging. *Int J Adv Signal Image Sci.* 2026;12(1):274-285. https://www.doi.org/10.29284/10knqj25
14. Hassan E, Shams MY, Abd El-Hafeez T, *et al.* A novel model for expanding horizons in sign language recognition. *Sci Rep.* 2025;15(1):24358. https://www.doi.org/10.1038/s41598-025-09643-2
15. Saber A, Emara T, Elbedwehy S, *et al.* A novel approach for breast cancer detection using a Nesterov accelerated Adam optimizer with an attention mechanism. *Sci Rep.* 2025;15:27065. https://www.doi.org/10.1038/s41598-025-12070-y
16. Akyol S, Yildirim M, Alatas B. Multi-feature fusion and improved BO and IGWO metaheuristics-based models for automatically diagnosing sleep disorders from sleep sounds. *Comput Biol Med.* 2023;157:106768. https://www.doi.org/10.1016/j.compbiomed.2023.106768
17. Cheng S, Wang C, Yue K, Li R, Shen F, Shuai W, et al. Automated sleep apnea detection in snoring signals using long short-term memory neural networks. *Biomed Signal Process Control.* 2022;71:103238. https://www.doi.org/10.1016/j.bspc.2021.103238
18. Yang CH, Kuo YM, Chen IC, Lin FM, Chung PC. A machine learning-based detection method for snoring and coughing. *J Internet Technol.* 2022;23(6):1233-1244.
19. González-Martínez FD, Carabias-Orti JJ, Cañadas-Quesada FJ, Ruiz-Reyes N, Martínez-Muñoz D, García-Galán S. Improving snore detection under limited dataset through harmonic/percussive source separation and convolutional neural networks. *Appl Acoust.* 2024;216:109811.

https://www.doi.org/10.1016/j.apacoust.2023.109811
20. Li R, Li W, Yue K, Zhang R, Li Y. Automatic snoring detection using a hybrid 1D–2D convolutional neural network. *Sci Rep.* 2023;13(1):14009. https://www.doi.org/10.1038/s41598-023-41267-7
21. Yıldırım M. Automatic diagnosis of snoring sounds with the developed artificial intelligence-based hybrid model. *Turk J Sci Technol.* 2022;17(2):405-416.
22. Tuncer T, Akbal E, Dogan S. An automated snoring sound classification method based on local dual octal pattern and iterative hybrid feature selector. *Biomed Signal Process Control.* 2021;63:102173. https://www.doi.org/10.1016/j.bspc.2020.102173
23. Ding L, Peng J. Automatic classification of snoring sounds from excitation locations based on prototypical network. *Appl Acoust.* 2022;195:108799. https://www.doi.org/10.1016/j.apacoust.2022.108799
24. Liu Y, Feng Y, Li Y, Xu W, Wang X, Han D. Automatic classification of the obstruction site in obstructive sleep apnea based on snoring sounds. *Am J Otolaryngol.* 2022;43(6):103584. https://www.doi.org/10.1016/j.amjoto.2022.103584
25. Dong H, Wu H, Yang G, Zhang J, Wan K. A multi-branch convolutional neural network for snoring detection based on audio. *Comput Methods Biomech Biomed Engin.* 2025;28(8):1243-1254.
26. Hassan E, Hossain MS, Saber A, *et al.* A quantum convolutional network and ResNet(50)-based classification architecture for the MNIST medical dataset. *Biomed Signal Process Control.* 2024;87:105560. https://www.doi.org/10.1016/j.bspc.2023.105560
27. Lamichhane P, Rawat DB. Quantum machine learning: recent advances, challenges and perspectives. *IEEE Access.* 2025. https://www.doi.org/10.1109/ACCESS.2025.3573244
28. Hassan E, Talaat AS, Elsabagh MA. Intelligent text similarity assessment using RoBERTa with integrated chaotic perturbation optimization techniques. *J Big Data.* 2025;12:164. https://www.doi.org/10.1186/s40537-025-01233-3
29. Saber A, *et al.* Guardians of the voice: defending speech interfaces against deepfakes and adversarial attacks. In: *Advancements in Speech Processing for Human-Computer Interaction.* IGI Global Scientific Publishing; 2026:267-298. https://www.doi.org/10.4018/979-8-3373-3048-8.ch009

Zahraa Tarek *is a researcher in the field of Artificial Intelligence and Computer Vision. Her research interests include deep learning, medical image analysis, and intelligent systems for healthcare applications. She has worked on developing advanced machine learning models for classification and detection tasks, with a particular focus on improving performance and robustness in real-world datasets.*
https://orcid.org/0000-0001-9389-2850

Esraa Hassan *is a researcher in Artificial Intelligence and Computer Science, with a focus on deep learning and computer vision. Her work involves developing intelligent systems for data analysis and real-world applications, particularly in areas requiring robust and efficient machine learning solutions.*
https://orcid.org/0000-0001-8346-5724

This work is licensed under a Creative Commons Attribution 4.0 International License. The authors retain ownership of the copyright for their article, but they allow anyone to download, reuse, reprint, modify, distribute, and/or copy articles in IJOCTA, so long as the original authors and source are credited. To see the complete license contents, please visit http://creativecommons.org/licenses/by/4.0/.