A Deep Learning Model for Snoring Detection In Indoor Noise Using Spectrogram and 2D CNN Model

1st Sasidhar Reddy Navuluri
122101025
Electrical Engineering
IIT Palakkad

2nd Manish Reddy Nakkala
122101023
Electrical Engineering
IIT Palakkad

***Abstract*—Snoring is a common sleep issue that, if left untreated, can have severe negative effects on one's cardiovascular health. With the use of spectrogram analysis and two-dimensional convolutional neural networks (CNN), we provide a unique method for the automatic recognition of snoring in environments with a lot of background noise called indoor noise environments. The transformation of audio impulses into visual representations, which capture both temporal and frequency information, is accomplished through the utilization of spectrograms. After that, a two-dimensional convolutional neural network (CNN) architecture is utilized in order to acquire discriminative characteristics straight from the spectrogram images. A huge dataset consisting of labeled snoring and non-snoring audio recordings gathered from a variety of indoor environments is used to train the model through the process of training. The results of the experiments reveal that the suggested method is effective, as it achieves a high degree of accuracy in detecting snoring across a wide range of noise levels and environmental conditions. The approach that has been developed shows potential for real-world applications in healthcare diagnostics and sleep monitoring systems. It will allow for early diagnosis and intervention for individuals who are at risk of developing sleep-related problems.**

Keywords-Spectrogram, 2D CNN

I. Introduction

Sleep disorders, which include snoring, present substantial difficulties to the health of the general population all over the world. Snoring is a problem that affects a significant section of the population and can be a signal of underlying conditions such as obstructive sleep apnea (OSA). Snoring is characterized by the vibrating of respiratory structures when the individual or individuals are sleeping. Snoring and obstructive sleep apnea (OSA) can result in a wide range of health issues if they are not treated. These complications include cardiovascular diseases, cognitive impairment, and a decline in quality of life.

Although these diseases are common and can have serious consequences, their diagnosis and monitoring typically depend on expensive and time-consuming techniques, such as polysomnography performed in sleep laboratories. Furthermore, the subjective character of symptoms reported by individuals themselves and the inconvenience of traditional methods of monitoring create obstacles to prompt intervention and treatment.

To address these difficulties, there is an increasing demand for automated and non-intrusive techniques to detect and monitor sleep disorders. The progress made in machine learning, namely in the area of deep learning, presents great opportunities for the creation of such solutions. By utilizing methods such as spectrogram analysis and convolutional neural networks (CNNs), it is feasible to extract significant characteristics from audio recordings and detect patterns that suggest snoring in noisy indoor settings.

A. Significance

Our project aims to meet this requirement by suggesting an innovative method for identifying snoring in indoor noise through the use of spectrogram analysis and 2D convolutional neural networks (CNNs). Our objective is to improve the availability of sleep disorder diagnosis by automating the snoring identification process. This will allow for earlier intervention and treatment. The project's significance resides in its capacity to:

1) Enhance the effectiveness and precision of identifying sleep disorders, resulting in improved health outcomes for individuals who are susceptible.
2) Simplify the workload on healthcare systems by providing scalable and economical monitoring solutions.
3) Increase awareness of sleep health and encourage proactive management of sleep disorders.

B. Applications

The proposed method has a wide range of applications across several areas, such as:

1) Healthcare: Incorporation of sleep monitoring devices into clinical settings to promptly identify and treat sleep disorders.
2) Consumer electronics are being integrated into wearable devices and smartphone applications to provide personal sleep tracking and health monitoring.

3) Research: Application in extensive epidemiological research to examine the frequency and consequences of sleep problems on public health.

C. Existing Methods and Technologies

Polysomnography (PSG) is still considered the most reliable method for identifying sleep problems, such as snoring and obstructive sleep apnea (OSA). Polysomnography (PSG) entails thorough monitoring of many physiological indicators while an individual is asleep, including brain waves, eye movements, muscle activity, and breathing patterns. Nevertheless, PSG is expensive, requires a significant amount of time, and is usually performed in specialized sleep laboratories, which restricts its availability and ability to be scaled up.

Snore detection has utilized feature extraction techniques, such as Mel-Frequency Cepstral Coefficients (MFCCs) and Linear Predictive Coding (LPC). Although feature-based techniques are successful in capturing essential attributes of snoring sounds, they may encounter difficulties in loud settings and necessitate meticulous parameter selection and adjustment.

Machine learning and deep learning approaches, such as Support Vector Machines (SVMs), Random Forests, and Convolutional Neural Networks (CNNs), have shown potential for automated snoring identification. CNNs, in particular, excel at extracting hierarchical characteristics from raw data, making them ideal for applications such as spectrogram analysis in snoring detection.

Acoustic analysis methods have traditionally used manually or semi-automatically processing audio recordings to detect snoring events based on distinct sound patterns. These solutions frequently rely on heuristic rules or threshold-based algorithms, which may be insufficiently robust and generalizable across varied noise environments.

Wearable devices that have accelerometers or microphones provide a practical and unobtrusive method for monitoring sleep patterns, including instances of snoring. These devices commonly use uncomplicated threshold-based algorithms or rudimentary machine learning models to identify instances of snoring, although they may not have the same level of precision and complexity as laboratory-grade systems.

D. Motivation

The reason for tackling the difficulties of snoring detection is rooted in the significant influence that sleep disturbances can exert on individuals' health and overall state of being. Snoring, frequently considered bothersome, might indicate significant underlying disorders including obstructive sleep apnea (OSA), which, if not addressed, can result in cardiovascular diseases, cognitive decline, and diminished quality of life.

Our objective is to enhance the early identification and management of sleep disorders by creating precise and easily accessible techniques for detecting snoring. This will ultimately lead to improved health outcomes and quality of life for individuals. Timely management can halt the advancement of sleep-related disorders and lessen the strain on healthcare systems.

Moreover, the progress in technology, namely in machine learning and signal processing, has promising prospects to transform the domain of sleep medicine. By utilizing these technologies, we may create automated systems that can accurately and reliably detect instances of snoring. This will enable the development of scalable and cost-efficient solutions for diagnosing and monitoring sleep disorders.

Our motivation is ultimately driven by a dedication to enhancing public health through the utilization of cutting-edge technologies to tackle urgent healthcare issues. Our primary objective is to have a significant impact on the lives of those suffering from sleep disorders by concentrating on the identification of snoring. We wish to empower them to promptly seek treatment and live healthier and more satisfying lives.

E. Major Objectives

The main objective of this research is to create an automated snoring detection system that can precisely recognize instances of snoring in indoor noise situations. By utilizing sophisticated signal processing techniques, such as spectrogram analysis, and employing machine learning algorithms like Convolutional Neural Networks (CNNs), the system will accurately extract and categorize snoring sounds from audio recordings.

Another primary goal is to improve the system's resilience to interference and fluctuations in the environment that are frequently experienced in indoor environments. Our objective is to enhance the system's capability to differentiate snoring sounds from background noise and other non-snoring occurrences by improving signal processing techniques and optimizing model topologies. This will ensure consistent and dependable performance in various settings.

Validation experiments will be carried out to evaluate the performance and precision of the developed snoring detection system. By using a variety of datasets that include recordings from various noisy environments and demographic groupings, these experiments will confirm the system's capacity to work well in varied situations and with different people. This will give us crucial information to make the system even better.

The study will begin by doing a comprehensive literature analysis and preliminary research phase to create a solid grasp of current snoring detection methods and associated technologies. Subsequently, the phase of system development and implementation will encompass the creation and execution of the snoring detection system, which will entail tasks such as designing and implementing data preprocessing, feature extraction, and machine learning model construction.

The system's accuracy and reliability will be assessed through validation and performance evaluation activities, which will be done in diverse settings. The results of these activities will feed iterative modifications aimed at enhancing the system's performance. Ultimately, the optimization and deployment endeavors will prioritize improving the computing efficiency of the system and readying it for integration into healthcare systems and consumer devices.

II. Materials and Methods

A. System Architecture

In this paper, a deep learning classifier to detect snoring sound is modeled first, They are briefly described below.

1) Data extraction: It is important to note that the selection of WAV audio files as the input data format is critical because of the fact that these files are uncompressed, which allows the audio signal to maintain its fidelity. Every audio file contains a one-second clip, which is an adequate amount of time for collecting snoring incidents while maintaining an acceptable level of computing expenditures. For the purposes of training and evaluation, the entire size of the dataset, which consists of one thousand audio samples, guarantees a wide variety of snoring examples.

2) Spectrogram Analysis: Spectrogram analysis is a strong approach that is applied in the field of signal processing, particularly in the field of audio analysis. In its most fundamental form, spectrogram analysis is characterized by the transformation of temporal audio data into a representation in the frequency domain. This transformation is accomplished by the use of a mathematical procedure known as the Fourier transform, which breaks down the audio signal into the frequencies that make up its constituent parts.

After the audio signal has been converted into the frequency domain, spectrograms provide a visual depiction of the way in which the amplitudes of various frequencies fluctuate over the course of time. The intensity or amplitude of each frequency component is represented by spectrograms through the use of color or grayscale shading. Spectrograms are typically shown in the form of a two-dimensional graphic, with time representing the horizontal axis and frequency representing the vertical axis.

When it comes to recognizing the specific patterns that are related to snoring, spectrograms are very useful. There are various acoustic signatures that are produced by snoring, which is characterized by the recurrent vibrations of soft tissue in the throat that occur during sleeping. Specific frequency bands or temporal fluctuations in the spectrogram are the manifestations of these signatures.

3) 2D Convolutional Neural Network (CNN): Convolutional Neural Networks (CNNs) have revolutionized image identification tasks by autonomously acquiring hierarchical representations of visual data. Within the field of snoring detection systems, the use of a 2D CNN is a deliberate implementation of CNNs' capacity to efficiently analyze spectrogram data.

Regarding the identification of snoring, the spectrogram is considered a 2D image that visually represents the frequency-domain information obtained from the audio stream. This method leverages the intrinsic arrangement of spectrograms, where time is represented on one axis and frequency on the other. CNNs can effectively capture spatial and temporal patterns in the spectrogram by utilizing its 2D structure.

Utilizing a 2D CNN enables the examination of complex connections within the spectrogram data. As the CNN progresses through each layer, it acquires the ability to identify and analyze features at various levels of complexity. It begins with basic elements such as edges and corners and gradually moves towards more intricate structures that are indicative of snoring.

4) Pooling Layers: Pooling layers are essential in Convolutional Neural Networks (CNNs) as they decrease the spatial dimensions of feature maps produced by convolutional layers. The downsampling procedure has multiple crucial functions inside the network, aiding in fast computing and successful feature extraction.

Pooling layers assist in managing the computational complexity of the network by decreasing the spatial dimensions of feature maps. When Convolutional Neural Networks (CNNs) handle larger input pictures or spectrograms, the amount of parameters and calculations needed can rapidly become overwhelming. Pooling layers address this problem by reducing the spatial resolution of feature maps, hence reducing the number of parameters and computational processes required in the following layers.

Furthermore, pooling layers contribute to the preservation of the most distinguishing traits found in the input data. Pooling layers preserve the most important features while eliminating unnecessary or less interesting details by summing the information from the local neighborhoods of the feature maps. This technique guarantees that the network concentrates on capturing high-level abstract representations of the input data, which are crucial for producing precise predictions.

5) Fully Connected Layers: The fully connected layers located at the conclusion of a Convolutional Neural Network (CNN) function as the last step in combining the extracted high-level characteristics from the spectrogram data for the purpose of classification. The CNN analyzes the spectrogram using convolutional and pooling layers to gradually transform the input data into more intricate representations. This allows it to identify significant patterns that indicate instances of snoring.

After the feature extraction procedure is finished, the fully connected layers are responsible for performing a nonlinear mapping from the acquired feature space to the output classes. In the case of snoring detection, these classes usually indicate whether snoring episodes are present or absent. These layers utilize activation functions like sigmoid or softmax to generate probability scores for each class, indicating the probability of the input spectrogram belonging to a specific category.

Using the information that has been compiled across all of the spatial locations in the feature maps that were generated by earlier layers, the fully linked layers give the CNN the ability to make accurate predictions regarding the presence or absence of snoring events in the audio that is being input. In order to ensure that the network takes into account all of the pertinent features that were retrieved from the spectrogram when it comes to reaching its ultimate conclusion, this comprehensive integration of spatial information is utilized.

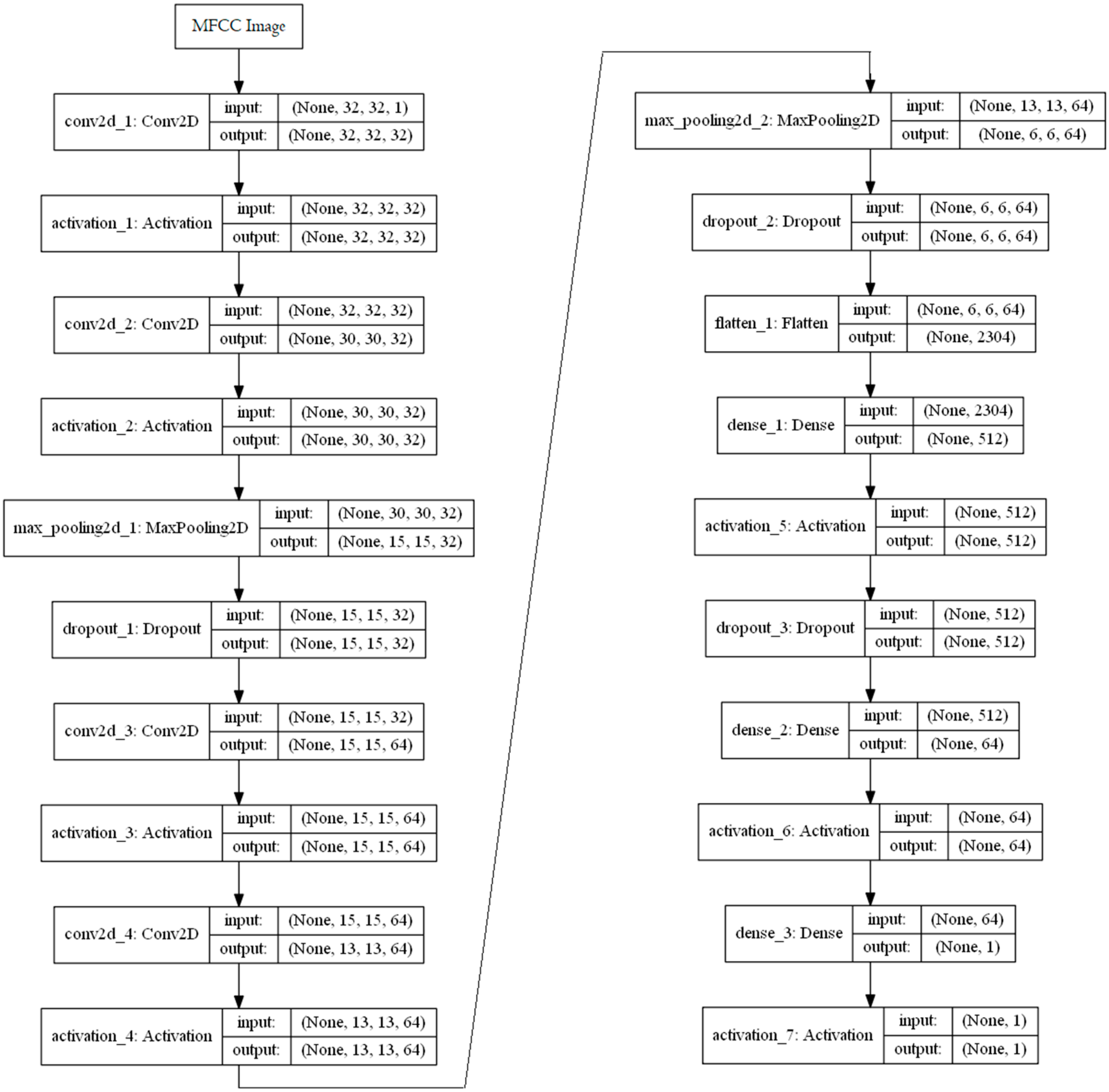

Fig. 1. The architecture of the convolutional neural network

B. Sensors or Other Modules Related to the Project

- Microphones
  - These sensors are responsible for recording auditory impulses, such as the sounds of snoring when sleeping.
  - Specifications such as sensitivity and noise levels are crucial for achieving high-quality recording with accurate reproduction.
- Data Acquisition Systems
  - An apparatus used to gather and convert audio signals obtained from microphones into digital format.
  - The quality of data is influenced by specifications such as sampling rate and resolution.
- Signal Processing Modules
  - Preprocess raw audio data before inputting it into the CNN model.
  - The tasks involve reducing noise and segmenting the raw data to improve it for analysis.

C. Different Modules of the Proposed Methods

- Feature Extraction Modules-These modules pull essential information from raw audio data. Alternative methods can extract pitch, formants, and cepstral coefficients from audio signals via signal processing methods, unlike CNN-based methods, which directly process spectrogram pictures.
- Classification Algorithms-Instead of CNNs, alternative classification methods may use SVMs, Random Forests, k-NN, or Gaussian Mixture Models. These algorithms use extracted information to characterize audio segments as snoring or not.
- Deep Learning Architectures Other Than Besides CNNs for image-based tasks like spectrogram analysis, RNNs, LSTM networks, and Transformer-based models like BERT can be employed for sequential or temporal data processing in snoring detection.

D. Mathematical Expressions

1) Fourier Transform (FT): The Fourier Transform converts a time-domain signal x(t) into its frequency-domain representation X(f), expressed mathematically as

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\, dt$$

Where X(f) represents the frequency-domain representation of the signal,x(t) is the time-domain signal, f is the frequency, and j is the imaginary unit.

2) Spectrogram Calculation: The Short-Time Fourier Transform (STFT) is used to calculate the spectrogram of a signal x(t) at successive overlapping windows. The spectrogram S(f,t) is defined as:

$$S(f,t) = \left|\left|\int_{-\infty}^{\infty} x(\tau)w(\tau - t)e^{-j2\pi f(\tau - t)}d\tau\right|\right|^2$$

S(f,t) represents the spectrogram, w(t) is the window function centered at time t, and is the time variable for integration.

3) Convolution Operation: In a CNN, the convolution operation between the input image I and a filter K is expressed as:

$$(I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n) \quad (1)$$

Where (IK)(i,j) represents the output of the convolution operation at position (i,j), I(m,n) is the pixel intensity of the input image at position (m,n), and K(im,jn) is the corresponding filter coefficient.

4) Pooling Operation: The max-pooling operation in a CNN reduces the spatial dimensions of the feature maps by selecting the maximum value within each pooling window. Mathematically, it is expressed as:
argmax
MaxPooling(x, y) = $^{k}_{i=1}{}^{k}_{j=1}$Input$(x+i, y+j)$

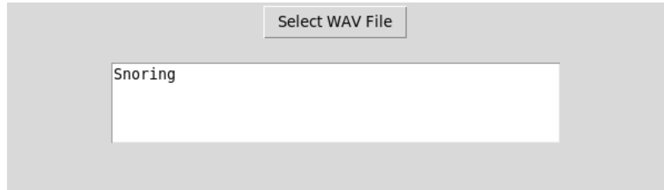

Fig. 2. A basic GUI

E. Performance Metrics

Several performance metrics are very important for correctly measuring how well a system works when judging how good a method is for finding snoring. One important metric is accuracy, which shows what percentage of instances were properly classified out of all the instances. It gives a general idea of how well the predictor can tell the difference between snoring and non-snoring events. Precision, on the other hand, looks at how many actually positive cases there are out of all the positive cases. It shows how well the system can reduce false positives, which is very important for making sure that snoring events are correctly identified.
Another important metric is recall, which is also called sensitivity. It shows what percentage of true positives were properly classified out of all actual positives. It shows that the system can successfully find snoring events while minimizing false negatives. The F1 score, which is the harmonic mean of accuracy and recall, gives a fair picture of how well a classifier is doing because it takes into account both false positives and false negatives. It gives a full assessment of how well the classifier works at snoring detection jobs.

F. User Interface

The user interface has a file upload area that allows users to effortlessly upload audio samples for analysis. After being uploaded, users can utilize an audio player interface to listen to the audio recordings. In addition, a visual representation is presented to show the presence of snoring in the uploaded file, offering distinct outcomes through the use of text or color-coded markers. Users have the option to access supplementary information such as the length of snoring occurrences or the level of certainty indicated by confidence scores.

III. Results

A. Experimental Setup And Database Collection

1) Dataset Selection and Preparation: A dataset including 1000 sound samples has been created as part of this study. The dataset has two distinct categories: snoring sounds and non-snoring sounds. Every class contains a total of 500 examples. The snoring sounds were gathered from several online sources. The non-snoring noises were also gathered from comparable online sources. Next, the periods of quiet were removed from the sound files, and the files were divided into smaller files of identical duration, each lasting one second. This was done using WavePad Sound Editor. Therefore, each sample has a period of one second.

Out of the 500 snoring samples, 363 samples contain snoring noises produced by children, adult males, and adult women, with no accompanying background noise. There are 137 samples left, which contain snoring sounds accompanied by non-snoring sounds in the background. Ambient noises were blended with the snoring sound. The 500 non-snoring samples comprise ambient noises that are likely to be present in the vicinity of the snorer. A total of ten distinct categories of non-snoring noises are gathered, with each category consisting of precisely 50 individual samples. The 10 categories include the sounds of a baby wailing, the ticking of a clock, the opening and closing of a door, complete silence with only the faint sound of a gadget's vibration motor, the flushing of a toilet, the siren of an emergency vehicle, the sounds of rain and thunderstorm, streetcar noises, people talking, and the background noise of a television news broadcast.

2) Data Splitting: Divide the dataset into three subsets: training, validation, and testing sets. Ensure that each subset contains a representative distribution of snoring and non-snoring instances to prevent class imbalance. Use a stratified sampling strategy to maintain class proportions across subsets.

3) Feature Extraction: Utilize the Short-Time Fourier Transform (STFT) to derive spectrogram representations from the audio recordings. Optimize the quality and resolution of the spectrogram by configuring settings such as window size, overlap, and frequency resolution. Transform spectrograms into appropriate input forms for the 2D CNN model, such as picture arrays or tensors.

4) Model Architecture Design: Create the architectural blueprint for the 2D Convolutional Neural Network (CNN) model used for snoring detection. Indicate the number of convolutional layers, dimensions of the filters, number of pooling layers, and quantity of fully linked layers. Conduct experiments using various architectures, including varied levels of depth, width, and complexity, in order to determine the most effective configurations.

5) Training Procedure: Initialize the 2D Convolutional Neural Network (CNN) model using either randomly generated weights or pre-trained weights obtained from similar tasks, often known as transfer learning. Optimize the model by training it with iterative methods like stochastic gradient descent (SGD) or Adam, using the training set. Employ strategies like as early stopping, learning rate scheduling, and batch normalization to improve the convergence of the model and mitigate overfitting.

6) Hyperparameter Tuning: Perform hyperparameter adjustment by utilizing the validation set to enhance the performance of the model. Investigate various hyperparameter configurations, encompassing learning rate, batch size, dropout rate, and regularization strength. Utilize grid search or random search methodologies to systematically investigate the hyperparameter space and determine the most optimal configurations.

7) Baseline Comparison: Examine how well the suggested strategy (spectrogram and 2D CNN) performs in comparison to baseline or current snoring detection techniques. To determine whether the suggested approach is better or more effective than the baseline methods, compare it using the same dataset and evaluation metrics.

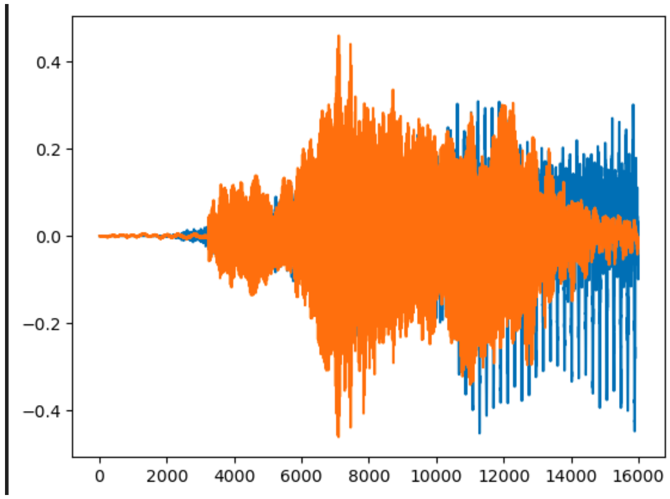

Fig. 3. waveforms of snoring and non snoring(orange-snoring,blue-non snoring)

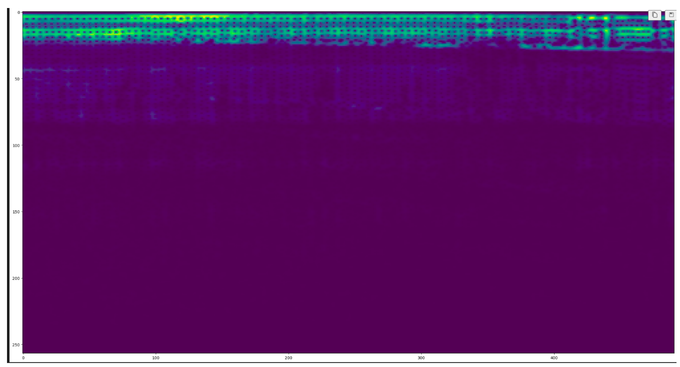

Fig. 4. spectrogram of the audio wave

Layer (type)	Output Shape	Param #
Conv2D	(None, 489, 255, 16)	160
Conv2D	(None, 487, 253, 16)	2320
Flatten	(None, 1971376)	0
Dense	(None, 128)	252336256
Dense	(None, 1)	129

TABLE I
MODEL ARCHITECTURE

The detection of snoring with a spectrogram analysis accuracy of 91.86% using a 2D CNN is a significant improvement in the monitoring of sleep disorders. This high degree of accuracy shows that the dataset has a strong ability to distinguish between snoring and non-snoring events. This kind of finding highlights how well the model captures the unique patterns typical of snoring events in the spectrogram data.

Dataset	Training	Validation
Sensitivity (Recall)	0.9738	0.9947
Accuracy	0.9186	0.9639

TABLE II
PERFORMANCE METRICS

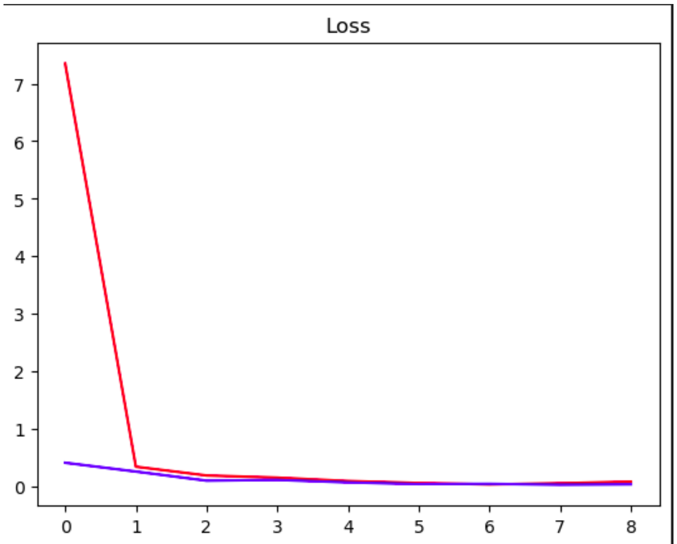

Fig. 5. loss of model

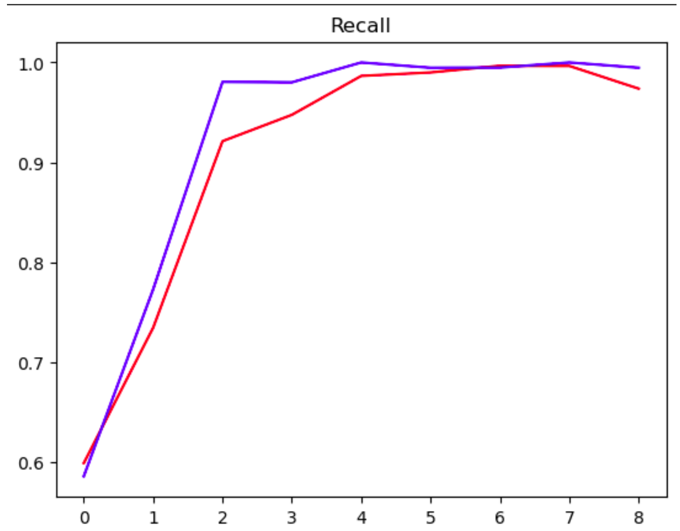

Fig. 7. recall of model

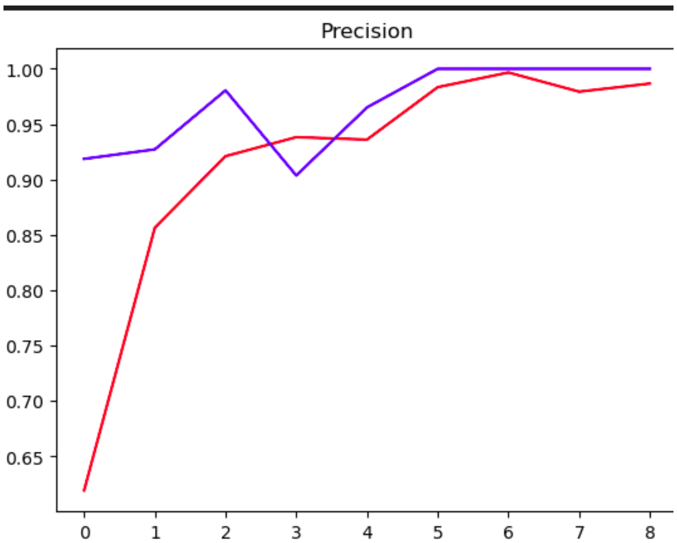

Fig. 6. precision of the model

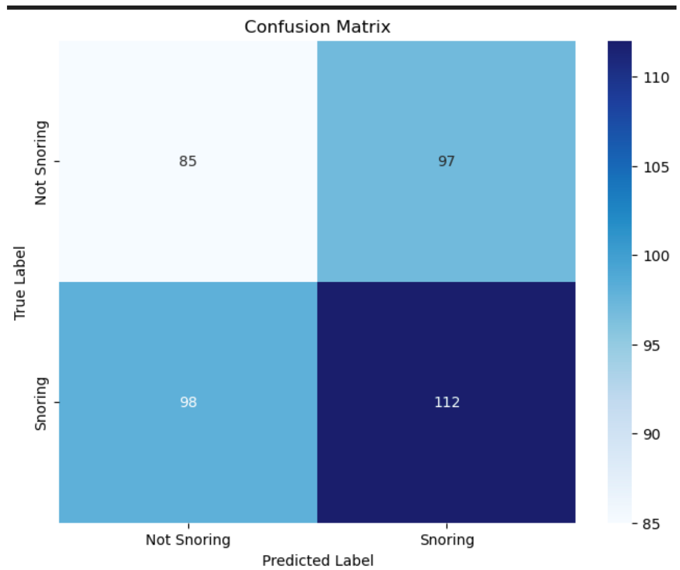

Fig. 8. confusion matrix

The high level of accuracy demonstrated by the model indicates its potential to be a reliable screening tool for detecting instances of snoring during sleep sessions. This model can be used by healthcare professionals and individuals to analyze audio recordings and quickly identify possible occurrences of snoring. This can help in detecting and intervening early, leading to better sleep quality and general health.

Furthermore, boasting a remarkable accuracy rate of 91.3%, the model has great potential for being used in clinical settings to diagnose and monitor sleep disorders. Healthcare providers can use it as a beneficial tool to identify symptoms associated with illnesses such as obstructive sleep apnea, where snoring is a major indicator. The model accurately identifies instances of snoring, offering physicians vital information about patients' sleep patterns. This information assists in creating customized treatment plans and treatments to effectively address sleep problems.

IV. Conclusion and Future works

A. Conclusion

This project studied the use of a 2D CNN model that uses spectrograms to detect snoring in indoor background noises and recognize snore sounds. The results show that the suggested method performs promisingly when it comes to separating snoring episodes from other types of background noise. The key findings are as follows;

- The distinctive time-frequency patterns of snoring noises are captured by the 2D CNN, which efficiently recovers spatial characteristics from spectrograms.

– The resilience of audio signals' temporal and frequency properties to changes in background noise is captured by employing 2D CNNs and spectrograms for snoring identification.
– The model shows potential for practical applications since it detects snores with good accuracy.

For further improvement of the system in the future, methods such as data augmentation, using deeper network architectures, and combining a 2D CNN model with RNNs can be implemented. The study concludes by demonstrating the viability of snoring detection in indoor conditions using a 2D CNN model with spectrograms. Future studies can improve the model's performance and open the door to useful applications in sleep monitoring and possible sleep disorder screening by implementing the recommended improvements.

B. Future works

Future work would entail improving the performance and accuracy of the entire system and being able to recognize live sounds as is. This can be done by improving upon the CNN model that is being used to detect the snores. It can be done in the following ways;

a) Augementing data: By including various forms of background noise (such as chatting, music, and traffic) in your training data, one can replicate real-world situations. This aids the model's ability to distinguish snoring from other types of interference. It can also be done by altering the spectrogram by either scaling or time-warping to boost data variability and enhance the model's capacity to generalize to unknown sounds.

b) Altering the model architecture: One way of doing this is by enhancing the 2D CNN with additional convolutional layers. As a result, the model can identify more intricate patterns in the spectrograms. Overfitting is one issue that may arise as a result of this. This can be tackled by using strategies such as dropout regularization. To investigate if altering the activation function in the convolutional layers helps the model acquire snoring characteristics, one can try experimenting with ReLU (Rectified Linear Unit) or Leaky ReLU.

c) Improvise on training techniques: Applying pre-trained models on analogous audio categorization tasks and utilizing the already existing snoring detection dataset, these models can be refined. By doing so, performance may be enhanced and the information gained from a larger audio feature area may be utilized.

d) Use advanced architectures: Merging a Recurrent Neural Network (RNN) to handle the sequential nature of audio data with a 2D CNN to capture spatial characteristics in spectrograms can prove beneficial as it can increase the precision of the model as a whole. Furthermore, if there are intricate temporal patterns in the background noise, this might be advantageous.

e) Improvising User Interface: In order to further improve the interface, several enhancements could be implemented. These include implementing a more elaborate feedback system to gather user input, refining error-handling procedures, incorporating interactive user guidance elements like tooltips and tutorials, providing comprehensive documentation, optimizing accessibility, ensuring cross-platform compatibility through responsive design, integrating with external services, and offering customization options for personalization. The purpose of these enhancements is to enhance the ease of use, offer improved assistance, and guarantee a smooth user experience on different devices and user preferences. Consistent feedback and recurrent testing will be essential in fine-tuning and progressing these improvements.

Dataset containing snoring samples and other indoor noises.

References

[1] Brian J. Taylor and Marjorie A. Darrah. Rule extraction as a formal method for the verification and validation of neural networks. In IEEE International Joint Conference on Neural Networks (IJCNN), volume 5, pages 2915–2920. IEEE, 2005.

[2] Tsui-Wei Weng, Huan Zhang, Pin-Yu Chen, Jinfeng Yi, Dong Su, Yupeng Gao, Cho-Jui Hsieh, and Luca Daniel. Evaluating the robustness of neural networks: An extreme value theory approach. 2018. arXiv:1801.10578.

[3] Seungil You, David Ding, Kevin Canini, Jan Pfeifer, and Maya Gupta. Deep lattice networks and partial monotonic functions. In Advances in Neural Information Processing Systems, 2017.

[4] Yuxuan Liu et al. Automatic snore sound detection using a deep neural network with spectrograms". (2019)

[5] Yuri Nishio et al. "Snore Detection with Deep Convolutional Neural Networks". (2018) IEEE Journal of Biomedical and Health Informatics

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] Xinjie Xu et al. "Deep Learning for Audio Event Detection: A Review" (2019).

[8] Abdoli, Sajjad Cardinal, Patrick Koerich, Alessandro. (2019). End-to-End Environmental Sound Classification using a 1D Convolutional Neural Network. Expert Systems with Applications. 136. 10.1016/j.eswa.2019.06.040. "End-to-End Environmental Sound Classification using a 1D Convolutional Neural Network"

[9] Lee, G.S.; Lee, L.A.; Wang, C.Y.; Chen, N.H.; Fang, T.J.; Huang, C.G.; Cheng, W.N.; Li, H.Y. The Frequency and Energy of Snoring Sounds Are Associated with Common Carotid Artery Intima-Media Thickness in Obstructive Sleep Apnea Patients. Sci. Rep. 2016, 6, 30559.

[10] Jiang, Yanmei Peng, Jianxin Zhang, Xiaowen. (2020). Automatic snoring sounds detection from sleep sounds based on deep learning. Physical and Engineering Sciences in Medicine. 43. 10.1007/s13246-020-00876-1.

[11] Sassi, Martina Pecchia, Leandro Haleem, Muhammad Salman. (2024). Spectrogram-Based Approach with Convolutional Neural Network for Human Activity Classification. 387-401. 10.1007/978-3-031-49068-2-40.

[12] T. H. Khan, "A deep learning model for snoring detection and vibration notification using a smart wearable gadget," Electronics, vol. 8, no. 9, article. 987, ISSN 2079-9292, 2019.