Snoring Analysis System for Automatic Diagnosis of Obstructive Sleep Apnea Syndrome during VR Sleep Therapy

Dragoș Datcu
Independent Research
The Netherlands
email@dragosdatcu.eu

Florin Stoicescu
Independent Research
Romania
fstoicescu@gmail.com

Cătălin Ciobanu
Transilvania University of Brașov
Romania
catalin.ciobanu@unitbv.ro

Dorin Mircea Popovici
Ovidius University of Constanta
Romania
dmpopovici@univ-ovidius.ro

ABSTRACT
While Virtual Reality (VR) sleep applications are not directly related to sleep apnea detection, they demonstrate the potential for technology to influence sleep behaviors and environments. As research in this field progresses, it may be valuable to consider how immersive technologies could be integrated with sleep monitoring and diagnostic tools, potentially enhancing patient comfort and compliance in sleep studies. Obstructive Sleep Apnea (OSA) is a prevalent and serious sleep disorder affecting an estimated 936 million adults worldwide. By leveraging the high prevalence of snoring in OSA patients, this approach has the potential to significantly improve early diagnosis rates and, consequently, patient outcomes. This study proposes an innovative method for OSA detection through automatic snoring analysis by fine-tuning Wav2vec 2.0 speech model to support VR sleeping therapy, aiming to provide a more accessible and cost-effective alternative to traditional polysomnography. Additionally, three non-deep learning techniques are presented together with an ESP32S-based edge system prototype as support for VR sleeping therapy.

Author Keywords
VR, Sleep Apnea; Audio processing.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., USI): Miscellaneous.

General Terms
Human Factors; Design; Measurement.

DOI: 10.37789/ icusi.2024.30

INTRODUCTION
Recent trends in virtual reality (VR) technology have sparked interest in its potential applications for sleep improvement. While not directly related to sleep apnea detection, these developments highlight the growing intersection of technology and sleep science.

The emergence of virtual communities centered around sleep-related activities in VR environments has garnered significant interest [21]. Two notable examples include a group with over 15.000 members focused on virtual co-sleeping and social bonding, and another community of more than 4.000 participants dedicated to exploring sleep experiences in virtual reality[1]. These developments highlight the potential for technology to address social and psychological aspects of sleep, which may be relevant to future sleep research and interventions. Individuals are exploring the use of VR headsets (as illustrated in Figure 1) as an alternative to traditional sleep aids.

Figure 1. VR sleeper with VR HMD on[2].

Users report entering calming digital environments designed for relaxation and sleep, complete with ambient

[1] VRCHAT, https://vrchat.com/
[2] Image generated with OpenArt.AI (https://openart.ai/)

sounds and visually soothing elements. This approach bears similarities to established sleep aids like white noise machines, but with an added immersive visual component.

Evidence suggests that VR sleep environments may benefit some individuals with sleep difficulties, including insomnia [5]. Users report that the change in perceived environment helps alleviate stress associated with bedtime.

It is important to emphasize that while VR applications show promise, they should not be considered a replacement for proper medical diagnosis and treatment of sleep disorders such as sleep apnea. Further research is needed to understand the potential benefits and risks of using VR technology in sleep contexts, particularly for individuals with diagnosed sleep disorders.

Obstructive Sleep Apnea (OSA) is a sleep disorder characterized by repeated episodes of complete or partial upper airway obstruction during sleep. The global prevalence of OSA is alarmingly high, with an estimated 936 million adults worldwide affected by mild to severe forms of the condition. In Europe alone, approximately 175 million people (44.0% of the population) are believed to have OSA, with about 90 million suffering from moderate to severe cases, experiencing at least 15 breathing events per hour during sleep.

The health implications of untreated OSA are significant and wide-ranging. Patients with unmanaged OSA are at increased risk for various complications, including 1) Cardiovascular issues: Hypertension, heart disease, and stroke, 2) Renal dysfunction: Chronic kidney disease and related complications, 3) Metabolic disorders: Type 2 diabetes and metabolic syndrome. Conversely, proper treatment of OSA has been shown to improve patients' quality of life and help regulate blood pressure, underscoring the importance of early and accurate diagnosis.

Currently, the gold standard for OSA diagnosis is polysomnography [23], a comprehensive sleep study that monitors various physiological parameters overnight. While effective, this method is labor-intensive, expensive, and often inaccessible to many patients due to limited availability of sleep laboratories and specialized personnel.

Snoring, a common symptom present in up to 94% of OSA patients, has emerged as a promising diagnostic indicator. Previous research has demonstrated the effectiveness of snoring analysis in OSA diagnosis, suggesting its potential to replace traditional polysomnography. This approach leverages a highly prevalent symptom to create a more accessible and cost-effective diagnostic tool.

In this paper, we propose a novel method for OSA detection through automatic snoring detection. Next to training three non deep learning models for classification, we propose fine-tuning the Wav2vec 2.0 speech model as a classifier to detect sleep snoring patterns. Also, we present the design and implementation of an affordable hardware platform to sense, process and classify the audio signal into snoring vs. non snoring patterns.

By utilizing advanced signal processing and machine learning techniques, our approach aims to:

1. Identify specific acoustic features of snoring that correlate with OSA presence and severity,
2. Develop an algorithm for automatic classification of snoring patterns,
3. Validate the accuracy and reliability of this method against polysomnography results.

The rest of the paper is structured as follows: the next section highlights relevant related research on apnea detection. The following section describes the architecture of our apnea prototype system, including the hardware components, the data preparation and the modeling of three non deep-learning techniques as well as the more modern fine-tuning of the Wav2vec 2.0 speech model. In the next section we present and discuss the results of our research. The final section of the paper presents the conclusions and the future work for our research.

RELATED WORK

Due to the relatively high prevalence and to the even higher societal awareness lately, studying apnea has gotten considerable attention from the research community. The following are some relevant scientific publications highlighting notable research milestones on the detection and classification of apnea by using at least acoustic-related analysis.

Ben-Israel el al. [2] implemented a Bayes classifier running on acoustic features correlated with the severity of the syndrome such as inter event silence, mel cepstability, energy running variance, Apneic phase ratio and pitch density, to achieve 80% correct classification for 5-fold cross validation.

The work of Kang et al. [12] used linear predictive coding (LPC) and Mel-Frequency Cepstral Coefficients (MFCC) features to classify different events such as snoring, apnea and silence from the sleep sound recordings. Their technique achieved an accuracy of 90.65% for detecting snoring events, 90.99% for Apnea, and 90.30% for silence.

Hayashi et al. [6] detected the severity of OSA by snoring sound and cluster analysis for which they used the MFCC, formant frequencies, and volume information.

Tuncer et al. [19] proposed a feature extractor named Local Dual Octal Pattern (LDOP) to solve the low success rate problems for the Munich-Passau Snore Sound Corpus (MPSSC) dataset. The authors emphasized that multilevel discrete wavelet transform (DWT) decomposition and the LDOP based feature generation, informative features selection with ReliefF and iterative neighborhood component analysis (RFINCA) and classification using k nearest neighbors (kNN) are fundamental phases of the proposed SSC method. In their study, they applied a seven-leveled DWT transform and LDOP together to generate low, medium, and high levels 4096 features out of which they selected 95 the most discriminative and informative ones to get 95.53% classification accuracy.

Luo et al. [15] presented five machine learning models and two OSA diagnostic schemes are used to classify night audio as non-snoring, snoring, or OSA-related snoring. Their systems achieved a diagnosis rate for OSA of about 97%.

Cheng et al. [3] proposed a classifier based on Long Short-Term Memory (LSTM) to identify the respiratory event-related snoring from simple snoring. The classification model runs on features related to Mel-frequency cepstrum coefficients (MFCC), Mel Filter Banks (Fbanks), Short-time Energy and Linear Prediction Coefficient(LPC), representing the different characteristics of snoring.

Huang et al. [8] proposed a segmentation model based on Transformer and multi-scale feature fusion, to effectively combine global information and multi-scale features to achieve localization of event start and end times for detecting apnea and hypopnea events using only audio signals.

Fang et a. [4] advanced a novel Snore Detection Cepstral Coefficient (SDCC) is proposed, based on Mel Frequency Cepstral Coefficients (MFCCs) and snore detection frequency division. Relief-F feature screening is then applied to SDCC and MFCC. The authors applied Canonical Correlation Analysis (CCA) on the fusion features obtained as a result and got an accuracy of 97.8% with Subspace KNN to effectively recognize and assess OSAHS as well as the severity of disease.

Li et al. [13] proposed a hybrid convolutional neural network (CNN) model for the automatic snore detection. The model had a one-dimensional (1D) CNN processing the original signal and a two-dimensional (2D) CNN representing images mapped by the visibility graph method. The algorithm achieved an average classification accuracy of 89.3% for the proposed snoring detection algorithm.

In a different approach, a sleep apnea classification model based on Bi-LSTM with attention mechanism [14] was built on Mel-Fbank features extracted from snore signals. The model focused on four types of snore signals namely hypopnea, normal condition, obstructive sleep apnea, and central sleep apnea and obtained about 62.31% subject-independent accuracy.
Singtothong and Siriborvornratanakul [18] introduced a multimodal deep learning-based sleep apnea detection model which uses sleep sounds, oxygen saturation (SpO2), and pulse rate. Their combined model achieved 96% accuracy in inferring apnea severity, outperforming individual models using SpO2 and pulse rate (79%) and sleep sound (83%).
T
he work [22] proposed a snoring sound detection algorithm using a multi-channel spectrogram and convolutional neural network (CNN). The authors derived four different feature maps including spectrogram, Mel-spectrogram, continuous wavelet transform (CWT), and multi-channel spectrogram composed of the three single-channel maps. The study explored the superior feature learning capability of the deep learning model, providing a more effective feature map for snoring detection.

Serrano et al. [17] proposed a stacked model which uses a combination of a pretrained VGG-like audio classification network and a bidirectional long short-term memory (bi-LSTM) network to facilitate OSAHS diagnoses by means of low-cost devices such as smartphones.

Xiu et al. [20] presented a wearable sleep monitoring system using eight frequency-domain features, with 59% average accuracy in identifying the severity of the four kinds of OSAS categories.

The more recent work of Hu et al. [7] made use of novel sound features to differentiate OSA and hypopnea from the normal snores. Such features account for percussive enhancing and positional encoding as the snores exhibit different percussive properties and temporal characteristics due to the disease generation mechanisms. The authors advance a multi-task learning framework to aid the main classification task by simultaneously learning two related simple tasks.

Figure 2. ESP32S NodeMCU Module Dev Kit C development board equipped with CH340 and features 2.4GHz dual-mode.

Another modern snoring detection approach is by Zhang et al. [23] which proposed a technique based on a long short-term memory based spiking neural network (LSTM-SNN) that is appropriate for large-scale home detection for snoring. The LSTM-SNN model classified automatically the non-snoring vs. snoring sounds by checking on the Mel frequency cepstral coefficients (MFCCs) extracted from sound signals and encoded into spike trains by a threshold encoding approach.

Jacob et al. [10] proposed an embedded system running on Arduino nano 33 BLE to capture the audio signal via a MP34DT05 sensor, to compute Mel-filter bank energy features, Mel Frequency Cepstral Coefficients and Spectrogram features and classify normal, snoring and OSA snoring.

Qiu et al. [16] presented a novel data-driven Audio-Semantic Multi-Modal model for OSAHS severity classification - ASMM-OSA based on patient snoring sound characteristics. The authors apply an augmentation of the audio features via PubMedBERT to enrich their diversity and detail. The classification of OSAHS by severity ie. as normal, mild, moderate, and severe, was realized by using XGBoost based on the number of sleep apnea events.

Even though older non-deep learning approaches generally scale well on edge computing for snore detection, they do not show robustness and good performance for generalization. Conversely, more modern approaches which make use of deep-learning modeling commonly have great generalization scores but lack on portability towards low-power devices.

ARCHITECTURE

Hardware platform

In our research, we have designed a prototype embedded system for the audio signal acquisition, storing, processing and identifying snore patterns, to be used in VR sleep scenarios.
The prototype hardware includes an ESP32S NodeMCU (Figure 2) together with additional sensors and modules such as microphone, storage module, clock module and an 1.3 inch OLED output screen.
Additionally, we have experimented with the substantially more powerful AMB82-MINI module (than ESP32S) which shows high potential for robust video processing in a multimodal data fusion setup.

ESP32S NodeMCU

The ESP32S NodeMCU Module Dev Kit C development board (Figure 2) is equipped with CH340 and features 2.4GHz dual-mode Wi-Fi and Bluetooth chips, as well as 40nm low-power technology.
The ESP32S NodeMCU Module Dev Kit C development board allows two cores for running code in parallel.
We have programmed this board to run two tasks, namely 1) the audio signal acquisition and storage on a microSD memory card, and 2) audio feature extraction (by running Fast Fourier Transform) and the classification of the snoring - no snoring audio patterns. We configured the first task to run on the first core of the ESP32S board, while the second task was set to run in parallel on the second core.
With optimization of the code, the two tasks running separately at the same time allow for continuous audio signal acquisition, processing and snoring pattern analysis ie. for any processing session, by the time the first task finishes acquisition and storing for the current data chunk running on the first core, the second core is finished with the classification of the previous data chunk. Then, the new session is ready to start again, beginning by copying the newly collected audio data chunk, passing it from the first core to the second core and running the snore pattern detection.
The hardware components of our prototype system were compatible with Arduino boards. We used EverywhereML[3] for porting the non deep learning snoring classification models to ESP32S.
The list of hardware components for the prototype is as follows:

- MAX9814 microphone AGC amplifier module,
- SPI reader micro memory SD TF module card memory shield,
- Real Time Clock RTC compatible with DS3231 I2C,
- 1.3 inch OLED SSH1106 (128x64 pixels) I2C display.

The final design of the prototype system is depicted in Figure 3.
We have designed and built the hardware prototype of the snoring detection device with the purpose to support VR sleep therapy. However, the device can be used in different scenarios as well ie. as a standalone sleep analytics device without the VR component.

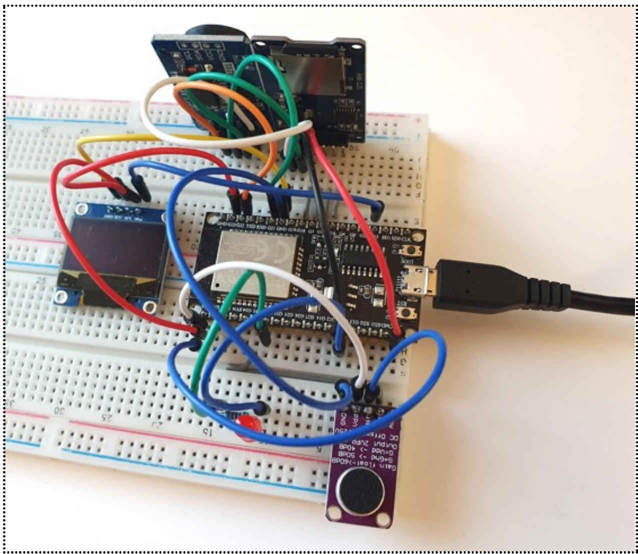

Figure 3. The snoring detection system prototype running on ESP32S and MAX9814 microfoon.

AMB82-MINI

The AMB82-MINI board (Figure 4) can make use of its Realtek RTL8735BDM internal NN engine to deploy edge AI devices, interesting intelligent equipment, object detection, audio recognition, facial recognition, AI models by Yolov4-Tiny, Tensorflow-Lite.

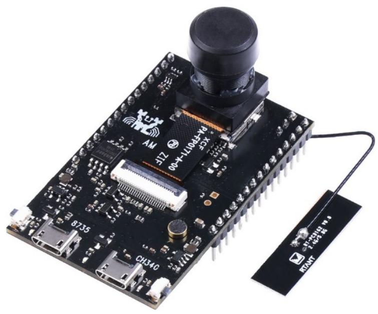

Figure 4. Realtek AMB82-Mini IoT AI Camera Arduino Dev. board[4].

[3] https://github.com/eloquentarduino/everywhereml

[4] AMB82-MINI development board https://www.amebaiot.com/en/amebapro2-amb82-mini-arduino-getting-started/

The board is equipped with an ARMv8M (up to 500MHz) and an Intelligent Engine NPU (0.4 TOPS). The board has 128MB internal DDR2 (on SoC) and 16MB external SPI. Data communication with the board is facilitated via its built-in support for Wi-Fi and Bluetooth.

Snore Classification Models

Dataset preparation

For the research, we have used the Khan dataset [12]. The dataset is organized into two primary folders, each containing audio samples of snoring and non-snoring sounds as follows:

Snoring Sounds Folder (folder 1):

- This folder comprises a total of 500 audio samples, each with a duration of 1 second.
- Out of these 500 samples, 363 are pure snoring sounds, which include recordings from children, adult men, and adult women, captured without any background noise.
- The remaining 137 samples feature snoring sounds with various background noises.

Non-Snoring Sounds Folder (folder 0):

- This folder also contains 500 audio samples, each 1 second in length.
- The non-snoring samples represent various background sounds that might be present near a snorer.
- These samples are categorized into ten distinct types of non-snoring sounds, with each category containing 50 samples. The categories are:
  - Baby crying,
  - Clock ticking,
  - Door opening and closing,
  - Silence with minor gadget vibration noise,
  - Toilet flushing,
  - Emergency vehicle siren,
  - Rain and thunderstorm,
  - Streetcar sounds,
  - People talking,
  - Background television news.

This structured dataset is designed to facilitate the analysis and classification of snoring versus non-snoring sounds in various acoustic environments.

Design of the classification models

Classification in machine learning is a supervised learning task that entails predicting a categorical label for a given input data point. The process involves training an algorithm on a labeled dataset, where the input features are utilized to learn the mapping between these features and their corresponding class labels. Once trained, the model can be applied to predict the class labels of new, unseen data points.

In our research we have created four models for snoring classification. The first three rely on more classical non-deep learning techniques, namely on KNeighbors classifier, on Random Forest classifier and on Logistic Regression classifier.

The fourth model relies on Wav2vec 2.0 [1], a deep-learning speech model stemming from a robust framework for self-supervised learning of speech representations.

The classification techniques we had employed in our study are as follows:

KNeighbors classifier

The k-nearest neighbors algorithm (k-NN) is a non-parametric supervised learning method utilized in both classification and regression tasks. k-NN operates by considering the k closest training examples in the dataset for a given input. For k-NN classification, the output is determined based on class membership. An object is classified by a plurality vote among its k nearest neighbors, with the object being assigned to the most common class among these neighbors. When k equals 1, the object is classified based on the single nearest neighbor's class.

Random Forest classifier

The Random Forest classifier is an ensemble learning method used for classification and regression tasks. It constructs multiple decision trees during training and outputs the mode of the classes (classification) or the mean prediction (regression) of the individual trees. This approach enhances predictive accuracy and controls overfitting by combining the results of numerous decision trees, each built on a random subset of the data and features.

Logistic Regression

Logistic Regression is a supervised learning algorithm commonly used for binary classification tasks. It models the probability of a binary outcome by fitting data to a logistic function (sigmoid curve). The algorithm estimates the parameters of the logistic function using maximum likelihood estimation, enabling the prediction of the probability that a given input belongs to a particular class. Logistic Regression is effective for problems where the relationship between the input features and the class probabilities can be linearly separated.

Wav2vec 2.0

Wav2vec 2.0 [1] is a model (by Facebook AI research labs) relying on learning powerful representations from speech audio alone followed by fine-tuning on transcribed speech.

The model consists of a multi-layer convolutional feature encoder which takes as input raw audio and outputs latent speech representations for specific number of time-steps. These are then fed to a Transformer to build representations capturing information from the entire sequence. The output of the feature encoder is discretized with a quantization module to represent the targets (Figure 5) in the self-supervised objective.

According to the benchmarks, Wav2vec 2.0 clearly outperforms the best alternative semi-supervised methods for speech oriented tasks (ie. speech recognition) while being conceptually simpler. The model masks the speech input in the latent space and solves a contrastive task defined over a quantization of the latent representations which are jointly learned.

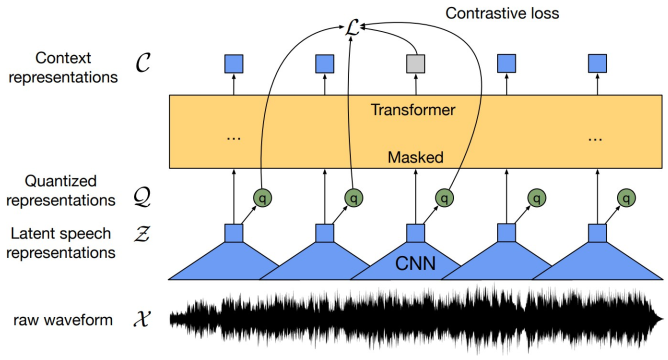

Figure 5. Wav2vec 2.0 framework [1] which jointly learns contextualized speech representations and an inventory of discretized speech units.

RESULTS AND DISCUSSION

First, we trained classifiers running on KNeighbors, random forest and logistic regression. For the first two classification models (KNeighbors and random forest classifiers) we had employed grid search to automatically determine the most optimal model parameters. Eventually, the three models were trained using features derived from the spectral components of the audio signal.

We run Fast Fourier Transform (FFT) on the audio signal and further applied data normalization. The 16kHz audio signal was applied to a shifting data processing window of 1024 samples in length, with 30% overlapping for consecutive windows. The upper cut-off frequency was 2kHz, leading to 131 features/window and 5240 features/sample. The normalization parameters were computed given the training dataset.

Next, we fine-tuned Wav2vec 2.0 on the Khan snore dataset and used the fine-tuned model for inference. The base model was pre-trained on 16kHz sampled speech audio and does not include a tokenizer, as it was trained exclusively on audio data. For utilizing this model in speech recognition tasks, we created a tokenizer and further fine-tuned the model using labeled text data.

The results for the three snoring classification are depicted in Figure 6-7. Figure 6 displays the ROC (Receiver Operating Characteristics) curve of the kNN classifier.

The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) measures the model's ability to distinguish between classes. A higher AUC indicates better performance in correctly classifying 0s as 0s and 1s as 1s. In a medical context, a higher AUC reflects the model's improved capability in differentiating between patients with and without the disease.

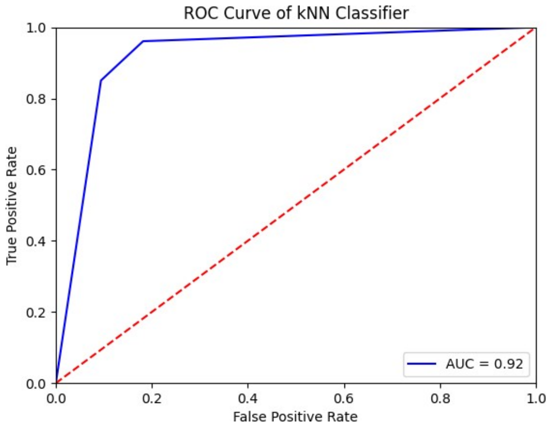

Figure 6. ROC Curve of kNN Classifier (AUC=0.92).

Figure 7 shows the ROC curve of the Random Forest classifier, which has AUC of 0.97. Similarly, Figure 8 highlights the ROC curve of the Logistic Regression Classifier (AUC 0.90).

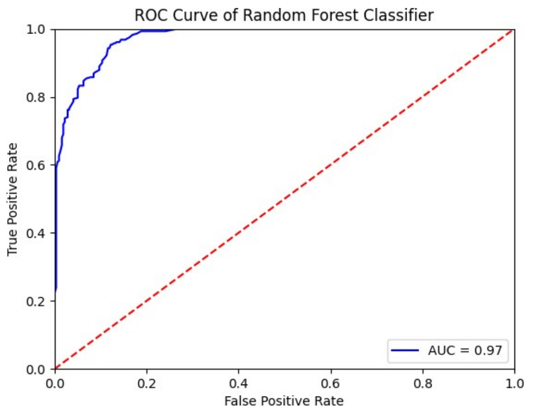

Figure 7. ROC Curve of Random Forest Classifier (AUC=0.97).

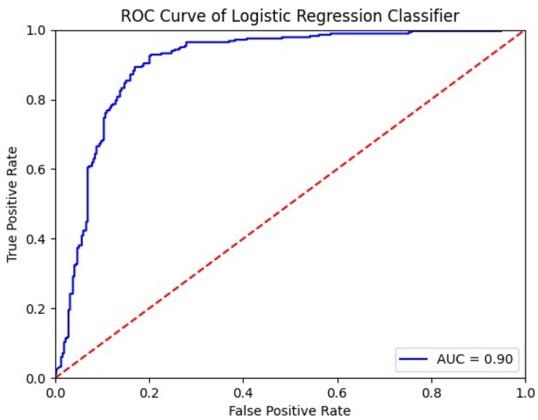

Figure 8. ROC Curve of Logistic Regression Classifier (AUC=0.90).

Out of the three classifier types, the results indicate the Random Forest Classifier has the best performance (AUC 0.97). The second best out of the three models is the kNN classifier, with 0.92 AUC.

Snoring Detection by Wav2vec 2.0

To fine-tune Wav2vec 2.0 for snoring detection, we used Google Colab[5] with the T4 GPU hardware accelerator. Important to notice that the fine-tuning can be done on a high-end computer (including laptop) as well, however the fine-tuning process takes a longer time.

For fine-tuning Wav2vec 2.0, we have used hyperparameters *warmup_ratio = 0.1*, *learning_rate = 3e-5*, and *accuracy* as metric to retain the best model (*metric_for_best_model*).

Figure 9 illustrates the training and validation loss graphs for fine-tuning Wav2vec 2.0. Figure 10 shows the validation accuracy for the fine-tuned models, given the training steps. The best Wav2vec 2.0 based fine-tuned model shows 99% evaluation accuracy (Figure 10).

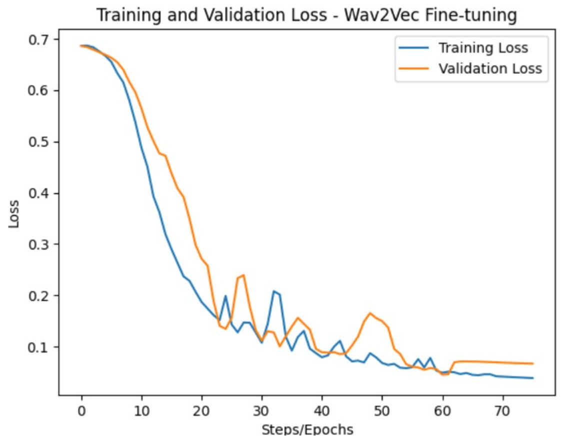

Figure 9. Wav2Vec 2.0 fine-tuning (training for 75 epochs), training and validation loss.

The advantage of the classifier models based on kNN, Random Forest and Logistic Regression is that they are lightweight and are consequently easily portable for edge computing setups. Development boards such as ESP32S or AMB82-Mini show full potential in supporting the computational load to run audio signal acquisition, data storing, feature extraction, and detection of snore audio patterns. Such edge computing systems turn out to be more accessible and cost-effective alternatives of the traditional polysomnography systems and can facilitate therapy in VR sleep scenarios.

On the other hand, the generalization capability of the models based on classical non deep-learning approaches is considerably reduced, when compared to the more complex snoring classification model based on Wav2vec 2.0. Conversely, at the moment the more robust Wav2vec model is not easily portable for edge computing.

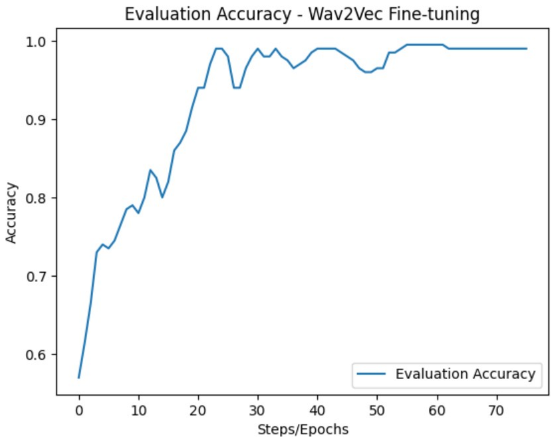

Figure 10. Wav2Vec 2.0 fine-tuning (training for 75 epochs), evaluation accuracy.

One considerable limitation of the snoring detection approaches presented in this paper relates to the lack of identifying OSA severities. The OSA severity estimation represents an extension of the current research and is the authors' next target.

CONCLUSIONS

The development of an accurate, accessible, and cost-effective method for OSA detection for VR sleepers through automatic snoring analysis has the potential to significantly improve diagnosis rates and, consequently, patient outcomes. By leveraging a common symptom of OSA, this approach could enable wider screening and earlier intervention, potentially reducing the global burden of OSA-related health complications.

Further research and clinical validation are necessary to establish this method as a reliable alternative or complement to traditional polysomnography in OSA diagnosis.

Our future research will include fine-tuning on much larger datasets, also the classification of non-OSA snoring patterns as well as patterns of different OSA severities. Additionally, we will focus more on optimization techniques such as ablation and quantization, to facilitate porting the deep-learning models like Wav2vec 2.0 fine-tuned models on more powerful boards such as AMB82-Mini. Moreover, we plan to explore multimodal approaches by adding video analysis for an enhanced OSA event detection.

REFERENCES

1. Baevski, A., Zhou, H., Mohamed, A., and Auli, M. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations, doi: 10.48550/arXiv.2006.11477, (2020).
2. Ben-Israel, N., Tarasiuk, A, and Zigel, Y. Nocturnal sound analysis for the diagnosis of obstructive sleep apnea, 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos

[5] https://colab.research.google.com/

Aires, Argentina, doi: 10.1109/IEMBS.2010.5627784, (2010), 6146-6149.

3. Cheng, S., Wang, C., Yue, K., Li, R., Shen, F., Shuai, W., Li, W., and Dai, L. Automated sleep apnea detection in snoring signal using long short-term memory neural networks, *Biomedical Signal Processing and Control* 71, Part B, 103238, ISSN 1746-8094, doi: 10.1016/j.bspc.2021.103238, (2022).

4. Fang, Y., Liu, D., Zhao, S., and Deng, D. Improving OSAHS Prevention Based on Multidimensional Feature Analysis of Snoring. *Electronics 2023, 12*, 4148, doi: 10.3390/electronics12194148, (2023).

5. Goldsworthy, A., Chawla, J., Birt, J., Baumann, O., and Gough, S. Use of extended reality in sleep health, medicine, and research: a scoping review, *Sleep, 46*, 11, November 2023, zsad201, doi: 10.1093/sleep/zsad201, (2023).

6. Hayashi, S., Tamaoka, M., Tateishi, T., Murota, Y, Handa, I., and Miyazaki, Y. A New Feature with the Potential to Detect the Severity of Obstructive Sleep Apnoea via Snoring Sound Analysis. *Int. J. Environ. Res. Public Health, 17*, 2951. doi: 10.3390/ijerph17082951, (2020).

7. Hu, A., Zhao, Q., Zhang, P., Zhang, X., Zhang, S., Lu, Y., Ye, P., and Yan, Y. Snore Sound Features Based on Percussive Enhancing and Positional Encoding Combined with Multi-Task Learning for Osahs Detection, ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea, Republic of, doi: 10.1109/ICASSP48485.2024, (2024), 901-905.

8. Huang, Y., Chen, L., and Huang, Q. (2023). Fine-Grained Detection of Apnea-Hypopnea Events Based on Transformer Network in Audio Recordings, pp. 580-585, doi: 10.1109/ICSP58490.2023.10248848.

9. Ibáñez, V., Silva, J., and Cauli, O. A survey on sleep assessment methods, PeerJ, doi: 10.7717/peerj.4849, (2018).

10. Jacob, D., Kokil, P., Subramanian, S., and Thiruvengadam, J. Decoding Sleep: Microphone-Based Snoring Analysis using Embedded Machine Learning for Obstructive Sleep Apnea Detection, 2024 Tenth International Conference on Bio Signals, Images, and Instrumentation (ICBSII), Chennai, India, doi: 10.1109/ICBSII61384. 2024.10564033, (2024), 1-6.

11. Kang, B., Dang, X., and Wei, R. Snoring and apnea detection based on hybrid neural networks, International Conference on Orange Technologies (ICOT), Singapore, doi: 10.1109/ICOT.2017.8336088, (2017), 57-60.

12. Khan, T. H. A deep learning model for snoring detection and vibration notification using a smart wearable gadget, *Electronics, 8*, 9, 987, ISSN 2079-9292, (2019).

13. Li, R., Li, W., Yue, K. et al. Automatic snoring detection using a hybrid 1D–2D convolutional neural network. Sci Rep 13, 14009, doi: 10.1038/s41598-023-41170-w, (2023).

14. Li, H., Lin, X., Lu, Y., Wang, M., and Cheng, H. Pilot study of contactless sleep apnea detection based on snore signals with hardware implementation. Institute of Physics and Engineering in Medicine, *Physiological Measurement 44*, 8, doi: 10.1088/1361-6579/acebb5, (2023).

15. Luo, H., Zhang, L., Zhou, L., Lin, X., Zhang, Z., and Wang, M. Design of Real-Time System Based on Machine Learning for Snoring and OSA Detection, ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, doi: 10.1109/ICASSP43922.2022.9747393, (2022), 1156-1160.

16. Qiu, X., Wang., C., Li, B, Tong, H., Tan, X., Yang. L., Tao, J., and Huang, J. An audio-semantic multimodal model for automatic obstructive sleep Apnea-Hypopnea Syndrome classification via multi-feature analysis of snoring sounds. Front Neurosci. 2024 May 10;18:1336307, doi: 10.3389/fnins.2024.1336307, (2024).

17. Serrano, S., Patanè, L., Serghini, O., and Scarpa, M. Detection and Classification of Obstructive Sleep Apnea Using Audio Spectrogram Analysis. *Electronics, 13*, 2567, doi: 10.3390/electronics13132567, (2024).

18. Singtothong, C., and Siriborvornratanakul, T. Deep-learning based sleep apnea detection using sleep sound, SpO2, and pulse rate. Int. J. Inf. Tecnol., doi: 10.1007/s41870-024-01906-x, (2024).

19. Tuncer, T, Akbal, E, and Dogan, S. An automated snoring sound classification method based on local dual octal pattern and iterative hybrid feature selector, *Biomedical Signal Processing and Control, 63*, 102173, ISSN 1746-8094, doi: 10.1016/j.bspc.2020.102173, (2021).

20. Xin, Y., Li, R., Song, X., Wang, Y., Zhang, H., and Chen, Z. Wearable Sleep Monitoring System Based on Machine Learning Using Snoring Sound Signal. ASME. *ASME J of Medical Diagnostics. 7(2)*: 021001, doi: 10.1115/1.4063395, (2024).

21. Yin, M., and Xiao, R. Drifting Off in Paradise: Why People Sleep in Virtual Reality. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 498, doi: 10.1145/3544548.3580947, (2023), 1–13.

22. Ye, Z, Jianxin, P., Zhang, X., and Song, L. Snoring Sound Recognition Using Multi-Channel Spectrograms, Archives of Acoustics, 49, 2, doi: 10.24425/aoa.2024.148775, (2024), 169–178.

23. Zhang, R., Li, R., Liang, J., Yue, K., Li, W., and Li, Y. Long Short-Term Memory Spiking Neural Networks for Classification of Snoring and Non-Snoring Sound Events, in *Chinese Journal of Electronics, 33*, 3, May 2024, doi: 10.23919/cje.2022.00.210, (2024), 793-802.