Yawning Detection using Earphone Inertial Measurement Units

Jacob Brown
University of Cambridge
United Kingdom

Yang Liu
University of Cambridge
United Kingdom

Cecilia Mascolo
University of Cambridge
United Kingdom

ABSTRACT

Yawning is a key and reliable indicator of fatigue, and detecting fatigue is vital in scenarios ranging from safety-critical situations for preventing performance impairment, to work environments for promoting timely breaks, leading to enhanced worker healthiness and productivity. State-of-the-art yawning detection studies face several limitations, such as privacy concerns, high costs, and lack of portability. In this paper, we conduct a feasibility study on enabling a privacy-resistant, low-cost, and portable solution to detect yawning by leveraging earphones equipped with inertial measurement units (IMUs), with the aim of benefiting future fatigue detection methods. We employ a range of preprocessing methods and develop 5 neural networks along with 3 classical machine learning (ML) approaches based on our initial research into the patterns within earphone IMU data from yawning and various activities. We collect data from 10 participants wearing a headphone with an IMU and evaluate the performance of our models on both the collected dataset and a public dataset. The results show F_1 scores of up to 0.90 on the collected dataset and 0.71 on the public dataset, which indicate the feasibility of yawning detection from earables.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing**.

KEYWORDS

Earables, IMU, Machine Learning, Yawning, Drowsiness

1 INTRODUCTION

Alertness is of paramount importance in safety-critical situations, such as long-haul driving, working with heavy machinery or in emergency response. Being tired can have a significant impact on a person's ability to perform these tasks safely and can lead to accidents; according to the National Highway Traffic Safety Administration (NHTSA), there were 91,000 crashes involving drowsy driving in 2017, leading to roughly 50,000 injuries and 800 deaths [23]. Moreover, excessive drowsiness among individuals could have detrimental effects on their work performance, resulting in decreased productivity, frequent lapses in work, and a negative impact on their overall mood and well-being [8, 26]. Therefore, the ability to detect when a person is becoming drowsy as to encourage taking breaks is of great importance.

Yawning is a reliable indicator of drowsiness since it often occurs when an individual is feeling tired or sleepy, serving as a visible sign of the state of alertness [10]. Further, previous studies show it is a commonly-used and successful method for inferring drowsiness [4, 15]. Existing studies have intensively investigated automatic and reliable yawning detection techniques using vision-based [4, 7, 21, 24, 28], wireless-based [27], and wearable-based [9] techniques. However, vision-based solutions could incur significant privacy issues and high system costs. Both vision-based and wireless-based approaches are not portable, being limited through a requirement for dedicated infrastructure (cameras and WiFi transmitters near the user, respectively), and they are also environment-dependent, being influenced by lighting conditions and multi-path reflections, respectively. The wearable-based solution is susceptible to various hand movements encountered in daily life since it uses wrist-worn photoplethysmography (PPG) and is designed specifically for driving scenarios with consistent hand movements.

To this end, this paper aims to explore a privacy-resistant, low-cost, portable solution of yawning detection in daily life. Particularly, we investigate this issue based on IMUs embedded into standard earphones, a promising candidate for this task due to the proximity of earphones to the jaw, thus being well-suited to capturing yawning-incurred movements [11] and being isolated from hand motions. IMUs additionally provide high privacy-resistance, are low-cost [20], are small and comfortable enabling portable, user-friendly, non-invasive long-term usage, and are increasingly widespread, with earphones such as the Apple Airpods [1], Google Pixel Buds [2], and Samsung Galaxy Buds [3] all shipping with IMUs.

SmartWear '23, October 6, 2023, Madrid, Spain

ACM ISBN 979-8-4007-0343-0/23/10.
https://doi.org/10.1145/3615592.3616854

We introduce a feasibility study for yawning detection by leveraging headphones equipped with 6-axis IMUs (*i.e.*, 3-axis accelerometer and 3-axis gyroscope). We employ a range of time- and frequency-domain preprocessing methods and develop 5 neural networks along with 3 classical ML approaches via our initial understanding of the patterns within earphone IMU data from yawning and different activities. IMU data is collected from 10 participants wearing the eSense headphones [22] while performing various activities including yawning, resting, walking, talking, making head movements, eating and making various facial expressions, aiming to distinguish yawning from other daily activities. In addition, we also leverage a public dataset collected in [12], a larger 21-user study with data obtained from similar activities to validate the feasibility of our study.

We show the LSTM-based model using the preprocessed raw data to be the most effective model with an average F_1 score of 0.90 on the collected dataset, and 0.71 on the public dataset. The CNN-based model using spectrograms of the preprocessed raw data is also proved effective, with an average F_1 score of 0.89 on the collected dataset, though limitations of the lengths of activities obtainable from the public dataset hindered the model on this dataset.

We summarize the contributions of this work as follows: **1).** We propose the utilization of earphones equipped with IMUs to enable a privacy-resistant, low-cost, and portable solution of yawning detection. **2).** We collect a dataset through earables from 10 participants, which contains yawning and various daily activities. **3).** We conduct a comprehensive feasibility study of yawning detection using earphone IMUs. Specifically, we employ a range of preprocessing methods and develop 5 neural networks along with 3 classical ML approaches based on a preliminary study on the IMU data, and evaluate yawning detection performance using these models on both the collected and public dataset. Our evaluation shows F_1 scores of up to 0.90 on the collected dataset and 0.71 on the public dataset, which validates the potential of earables for future fatigue detection methods.

2 RELATED WORK

Drowsiness detection. Existing studies for drowsiness detection mainly focus on three categories [25]: 1). Behavioural approaches, that typically utilize cameras with computer vision techniques to detect and extract drowsiness-related features such as yawning [13], facial posture [6, 29], and eye movement [14]. These methods are effective and non-invasive, but are susceptible to limitations such as significant privacy concerns, high system costs, limited service coverage, and variability in lighting conditions. 2). Physiological approaches, which detect drowsiness through sensing and analysing vital signs of individuals, like using a headpiece to detect EEG signals [5], or a chest harness to detect heart rate, breathing rate and other metrics [30]. Although these approaches have achieved high accuracies, their sensing modalities are invasive, leading to reduced user acceptance. 3). Vehicular approaches, which use data from the vehicle being driven to infer drowsiness; common features include the angle of the steering wheel [19] or a lane detection system [17]. These approaches are non-invasive, but are highly susceptible to external factors such as road conditions, weather and individual driving ability [16], and are limited to vehicle-based scenarios.

Yawning detection. Previous studies propose a broad set of approaches for yawning detection using vision-based techniques [4, 7, 21, 24, 28]. These studies commonly extract spatial and/or temporal features from captured human faces and behaviors using cameras through analysing the geometric, color and movement information manually [7, 28] or using learning techniques [4, 21, 24]. However, vision-based approaches are vulnerable to substantial privacy concerns, high system costs, restricted service coverage, and susceptibility to lighting conditions, *e.g.*, blocked line-of-sight. One study [27] recognizes yawning while driving using in-vehicle WiFi signals through monitoring the changes in the channel state information caused by yawns. However, this system is highly environment-dependent and requires dedicated infrastructure, *i.e.*, a signal transmitter and receiver. One wearable-based solution [9] proposes the use of a PPG sensor worn on the wrist for yawning detection while driving. While this approach is portable, the presence of various hand movements in daily life can interfere with the PPG signals and the proposed design is tailored specifically for the driving scenario, where the consistent hand movements are expected.

The most related work to ours is developed by Gashi *et al.* [12]. They conducted a study aiming to hierarchically classify 5 classes of head movements, including yawning, through a shallow classifier and a CNN model using IMUs integrated into eSense headphones [18]. The performance on yawning detection is subpar since this work focuses mainly on the broader classification of head movements, with two 5-class classifiers achieving F_1 scores of 30.63% and 30.24% respectively for the detection of yawns. This highlights an unexplored area in the field, revealing the need for further investigation into the feasibility of yawning detection using earphone-based IMUs.

3 PRELIMINARY

3.1 Preliminary Study

Before constructing any models, we first study the structure of both accelerometer and gyroscope data captured by the eSense IMU under various daily activities and yawning. This

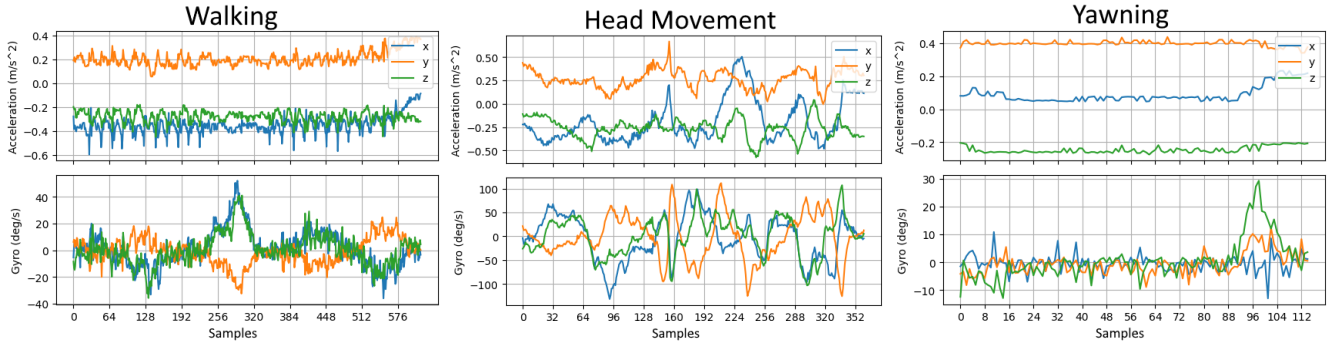

Figure 1: Time-domain plots illustrating the accelerometer and gyroscope data for various activities. All data was recorded at 32Hz.

understanding validates the feasibility of detecting yawning from IMUs on earables in a preliminary manner.

Figure 1 showcases the time-domain data while a user is walking, making head movements, and yawning. Distinct features can be observed in each of the time-domain signals. During walking, there are prominent spikes in the both axes with frequencies around 0.8Hz, with extended spikes in the gyroscope axes when turning. Head movements are characterized by erratic changes in both axes. Yawning has a unique pattern, with relatively stable readings leading up to a yawn, followed by a significant spike in the gyroscope z-axis during the exhale, before returning to a resting state. Given the presence and discernibility of these structures among various activities, it validates the feasibility of yawning detection from IMUs on earables. Moreover, a convolutional model is likely to be effective in detecting these patterns, similar to human perception. Additionally, the extended duration of a yawn and the notable changes in each axis over time suggest that an LSTM-based recurrent model would capture these temporal dynamics effectively.

3.2 Preprocessing

Data Transformation. Data is split into windows with a fixed length to ensure consistent input sizes for the networks. One of the following three types of input transformations can then be applied to these windows: **eIMU**: The data remains in its original format, with no further processing. Each window records the acceleration and gyroscope values for each timestep and axis, resulting in a matrix of shape $(s, 6)$, where s the number of samples in each processing window. **FFT**: The data is transformed into the frequency domain via a Fast Fourier Transform (FFT). Each window records the magnitude of the different frequency components present, resulting in a matrix of shape $(f, 6)$, where f the number of frequency components tracked. **Spectrogram**: Each window (plus some additional data padding to maintain the width) is transformed into a spectrogram, showing how the frequencies present in the window change over time. This is equivalent to splitting the window up into t equal sections, then applying an FFT transformation with f frequency components to each; thus, each window records the f frequency components for each of the t subsections, in a matrix of shape $(t, f, 6)$.

Filtering. Before splitting data into windows, we apply various filters to the data, primarily for noise reduction. In Section 5.4, we compare Butterworth low-pass and high-pass filters, as well as a moving average filter. Additionally, a scale normalization filter is employed to ensure the signal is scaled between -1 and 1, necessary for FFT transformation.

Equalisation. The final preprocessing step involves equalizing the number of positive and negative windows in each dataset since a balanced dataset is crucial for training a neural network to avoid biased learning and ensure fair testing. We observe an imbalance with fewer yawning windows compared to non-yawning windows. Therefore, random sampling (with removal) is performed on the non-yawning set to make the same number of windows as the yawning set.

4 YAWNING DETECTOR

4.1 Problem Statement

The yawning detection problem involves binary classification, where each input window is classified as either yawning or non-yawning. During model training/testing, the input to the model is a matrix consisting of m windows, with each window's size determined by the applied data transformation during preprocessing. The model output is an array of m floats ranging from 0 (indicating no yawn presence) to 1 (indicating full confidence in a yawn's presence), representing the network's predictions for each window. Additionally, models are trained using the binary cross-entropy loss function with the Adam optimizer. In total, 5 neural network models and 3 classical ML approaches are constructed and evaluated. The structure of each model is introduced as follows, ordered by

the type of data transformation listed in Section 3.2. Their performance will be illustrated in Section 5.4.

4.2 eIMU Models

eIMU models take windows of eIMU data as input.

eIMU LSTM. The first model is an LSTM-based model, consisting of three LSTM layers, each with 128 units and a *tanh* activation function. These LSTM layers are followed by a global average pooling layer, which takes inputs from all three LSTM layers to perform average pooling. The resulting output is then passed through a single dense layer with one unit and a *sigmoid* activation function to get the final prediction. Additionally, there are three dropout layers, each applied after an LSTM layer, with a dropout rate of 0.3.

eIMU CNN. The second model is a CNN-based model. The input data is initially passed through a reshape layer, transforming the original $(w, 6)$ matrices into $(w, 1, 6)$ matrices – this extra dimension allows the utilization of 2D convolutional layers. Following the reshape layer, there are four convolutional layers, with each subsequent layer doubling the number of filters compared to the previous layer, and the first convolutional layer has 64 filters. The kernel size of each layer is (5×1), and padding is applied to ensure the output and input sizes of each layer remain equal. A *ReLU* activation function is used in each convolutional layer, followed by a max pooling layer with a pool size of (2×1), to reduce the number of parameters. A dropout layer with a dropout rate of 0.5 follows the last max pooling layer to mitigate overfitting, and a flatten layer is employed to reduce the number of dimensions in the resulting matrix. The network concludes with two dense layers: the first layer has 64 units and a *ReLU* activation function, while the second layer has 1 unit and a *sigmoid* activation function to provide the final prediction.

4.3 FFT Models

FFT models take windows of the FFT data as input. There are two FFT models: **FFT LSTM** and **FFT CNN**. Since both eIMU and FFT data transformations produce similar 2-dimensional data, the same model structure can be reused from the eIMU models. However, a slight difference arises in the FFT models where a scale normalization filter needs to be applied on the input data to ensure consistent FFT amplitudes despite sensor variations caused by *e.g.* variable head orientation.

4.4 Spectrogram Model

The spectrogram model takes windows of the spectrogram data as input. Since a spectrogram is naturally time-distributed, we only leverage CNN-based model for this input, **Spectrogram CNN**. The model consists of four convolutional layers, each employing 256 kernels with a kernel size (5×5). All four layers utilize a *ReLU* activation function and are followed by a max-pooling layer with a (2×2) pool size. The output of the final convolutional layer is then passed through a dropout layer with a dropout rate of 0.5, followed by a flatten layer to reduce the number of dimensions. Subsequently, the output is passed into two dense layers: the first layer comprises 64 units with a *ReLU* activation function, while the second layer consists of 1 unit employing a sigmoid activation function, to provide the final prediction.

4.5 Classical ML Approaches

Although neural networks were expected to be the most effective models due to their capability to consider time distribution, we also construct and evaluate three classical ML approaches, including K-nearest neighbors, support vector machine, and random forest.

5 IMPLEMENTATION AND EVALUATION

5.1 Data Collection

Datasets. A mobile app was developed to collect the eSense IMU data. Dataset A was collected with a 96Hz sampling rate by the app via a human study[1], whereby 10 participants were asked to wear the headphones and perform various tasks. The study was conducted in a quiet room, and the tasks performed were as follows:

- **Yawn:** Participants yawn (or mimic yawning[2]) 100 times.
- **Rest:** Participants sit comfortably for 5 minutes.
- **Walk:** Participants walk normally for 5 minutes.
- **Talk:** Participants read a provided text for 3 minutes.
- **Move Head:** Participants move their head 30 times in each direction.
- **Eat and Drink:** Participants eat from a small selection of food and drink water.
- **Expressions:** Participants smile, frown, and open their mouths repeatedly for 30 seconds each.

A second dataset, Dataset B, is a subset of the public HAFAR dataset [12]. The public dataset is collected from 21 participants wearing the eSense headphones and contains 32Hz IMU recordings of nodding, head shaking, talking, smiling and yawning activities. However, given the prevalence of non-yawning data, Dataset B contains all yawning samples from the public dataset mixed with an equal number of samples picked randomly from all other types. This aims to prevent class imbalance.

Data labelling. Data is labelled on a window-level, where 1 represents a yawn present in the window, and 0 otherwise. In Dataset A, these labels are input by the participant via a

[1]The experiment was approved by the Ethics Committee of our institution.
[2]During data collection, we attempted to replicate mimic yawnings as closely as possible to real yawnings.

button on the app acting as a marker indicating a yawn event took place around when the button was pressed. In Dataset B, activities are marked by type (*e.g. 'Yawn'*, *'Talk'*, *etc.*), and windows obtained from the recordings of these activities are labelled according to this marking, with any type other than *'Yawn'* being labelled 0.

5.2 Training

For Dataset A, windows were split into an 80/20 train/test split over all participants. When training, the train split was further broken down into a validation set representing 10% of all data. Which training data was used in the validation set was changed each time (proportionally to the number of models being trained), simulating a cross-validation technique as to ensure the model was not overfit to the data it was trained on. For Dataset B, an approach simulating tailoring the dataset to specific individuals was taken. A set of users making up around 25% of all data was selected, then 80% of the data from these users was selected (20% of the total data) to be the test set. All remaining data formed the training and validation sets, as in Dataset A, after which the same cross-validation technique was applied.

5.3 Evaluation Metrics

The F_1 score was used as the primary metric for evaluating the performance of each model. This was chosen for being an all-round metric whereby a high F_1 indicates success in both correctly identifying yawns and correctly identifying non-yawns. Precision and recall are also given in the results.

5.4 Results

Results on neural network models. Figure 2 shows the results of the different models when trained on Dataset A. The models were optimised based on a comparison of the effectiveness of various filters, window size, separation, and FFT segment width and overlap. For the eIMU models, the optimal values used a 3-second window size with 1 second separation between each pair, with a smoothing filter zeroing the edges of the data to prevent windowing artefacts (such a filter is not applicable to the shorter, non-overlapping windows of Dataset B). For the FFT and spectrogram models, a 6-second window size with 1 second of additional segment overlap on either side with a 12Hz low-pass filter proved optimal. All models resulting from this perform well, with the eIMU LSTM model performing best on average with an F_1 score of 0.90. The model has a notably high average precision (0.94), meaning the false positive rate is low; however, a higher number of false negatives led to the recall being lower (0.86). This is also true of the FFT LSTM, having a much lower recall (0.69) than precision (0.81), though not clearly of other models, suggesting that there exist indicators of a yawn that our LSTM models are unable to extract. Dataset B also backs up this claim, with the eIMU LSTM reaching a precision of 0.83 with no filter but attaining only 0.61 recall, and the FFT LSTM comparing 0.60 precision to a (very variable) 0.5 recall under a moving average filter.

Figure 3 shows a comparison of results when different filters are applied to Dataset B. F_1 scores are universally lower than with Dataset A, likely a product of the inability to customize the models as greatly. In Dataset A, recordings were continuous throughout the activity, while Dataset B contains a multitude of short, discontinuous activities. With shorter recordings, window size is limited, itself limiting the FFT segment properties, and discontinuity prevents any overlap between windows. The eIMU models still perform well despite this, though the FFT and spectrogram models give somewhat poor results. Given that the optimal parameter search in Dataset A found the optimal window size for these models to be 6 seconds with 2 seconds of segment overlap, these results are best explained by the limitation imposed on the window size for these models.

Results on classical ML approaches. Regarding the classical ML approaches, the K-nearest neighbours approach achieved the best results on Dataset A, with an average F_1 score of 0.63, lower than the neural network approaches.

In summary, the results obtained are promising, with the most customized models reaching F_1 scores of up to 0.90 and 0.71 the two datasets respectively. All model types show promise, though the LSTM model over the raw data with a smoothing filter (where applicable) proved reliably consistent throughout. The FFT and spectrogram-based models were effective when given long, continuous data windows, but proved unreliable when the data was of a limited size. As expected, these neural network approaches proved better than single-point classical ML algorithms.

6 DISCUSSION

Yawning co-occurring with other activities. In our study, yawning is initially examined as a standalone activity. However, in real-world contexts, yawning sometimes coincides with activities like walking or head movements. Our primary objective in this study is to explore the feasibility of utilizing IMUs on earphones for detecting yawning. In future research, we aim to refine our model to ultimately discern the distinct characteristics of yawning, even when it co-occurs with other activities, by understanding the effect of interplay between yawning and other daily activities on the IMU data.

Unbalanced dataset. In line with our primary objective, both datasets were balanced in order to most accurately test the classification abilities of the models. Indeed, in a production setting, a yawning event is significantly less likely to occur than a non-yawning event. If the models defined

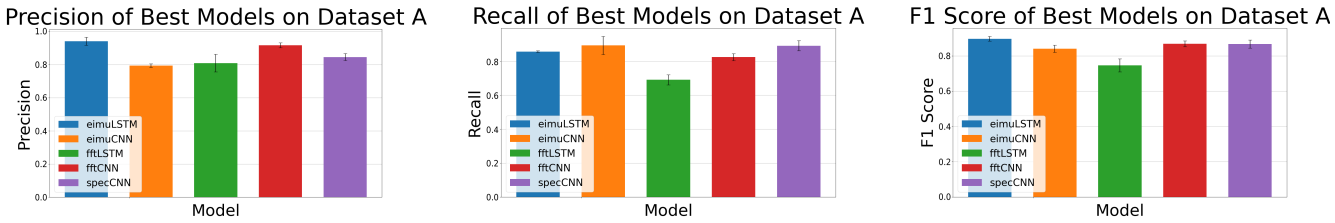

Figure 2: A comparison of the 5 deep learning models with their optimal parameters on Dataset A.

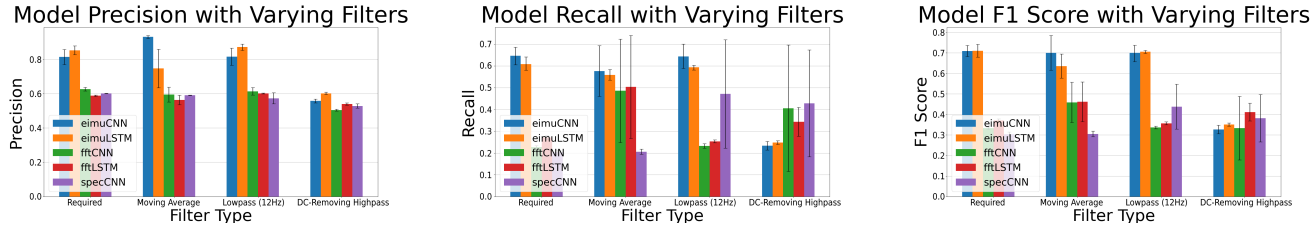

Figure 3: A comparison of the 5 deep learning models with various filters on Dataset B.

here were used, the class imbalance between training and deployment would hinder the generalisability onto new data; in such a case, the dataset should be only be balanced as far as the proportion of yawning to non-yawning events in real-world scenarios.

Model generalizability. To ensure our model's generalizability, we employ cross-validation, a proven method to evaluate performance on unseen data and mitigate potential overfitting. While this has strengthened our model's generalization capability, we acknowledge the need for continuous improvement. Therefore, we intend to explore and implement more methodologies in the future to further enhance the generalization capability.

Long-term usage. In order to be truly effective in real-world contexts, the system must be able to work for extended periods of time. On the one hand, the models process IMU data in short intervals for each detection, specifically 3, 3, and 6 seconds for the eIMU, FFT, and spectrogram models, respectively. This setting inherently minimizes the potential for long-term drifting issue of IMU data. Since the drifting issue primarily manifest over extended periods, the use of such short data windows makes the drift negligible. On the other hand, given the portable context, reserving power where possible is of high importance, and applying the computed models continuously would be a heavy drain on battery life. As such, a lightweight window detector with a low false-negative rate could be employed to first detect the possibility of a yawn within a window, before being passed to the computed models for a more accurate analysis. We plan to address this power reservation design in our future work.

7 CONCLUSION AND FUTURE WORK

In conclusion, this paper proposes a privacy-resistant, low-cost, portable solution for yawning detection by leveraging earphones equipped with IMUs. We collect a dataset from 10 participants using the eSense headphones, containing yawning and various daily activities. A range of preprocessing methods are employed and 5 neural networks along with 3 classical ML approaches are developed based on a preliminary study on earphone IMU data. We evaluate yawning detection performance using these models on both the collected and one public dataset. The results demonstrate promising performance, with F_1 scores of up to 0.90 on the collected dataset and 0.71 on the public dataset, which indeed imply the potential success of yawning detection from earables. Further research can build upon this solution to advance fatigue detection methods.

Regarding future work, an interesting addition would be to train the models with other types of data. The eSense headphones come with a microphone which was not used here due to stricter ethical guidelines surrounding user data; however, the microphone could detect the sound of a yawn as a third feature to train with, giving more data to train on and thus generating a more robust model. Other types of data that could be collected from an earphone include heart rate, blood oxygen levels and temperature, all helping create comprehensive model for yawning detection.

ACKNOWLEDGMENTS

This work was supported by European Research Council (ERC) project 833296 (EAR). We would like to express our gratitude to the Nokia Bell Labs for the provision of the

eSense device, and to Silvia Santini, Lidia Alecci and Shkurta Gashi of the Università della Svizzera italiana for generously sharing the public dataset with us.

REFERENCES

[1] [n. d.]. Apple Airpods. https://www.ifixit.com/Teardown/AirPods+2+Teardown/121471.

[2] [n. d.]. Google Pixel Buds. https://medium.com/@justlv/google-pixel-buds-teardown-396183cbbc18.

[3] [n. d.]. Samsung Galaxy Buds. https://root-nation.com/audio-en/headphones-en/en-samsung-galaxy-buds-review.

[4] Belhassen Akrout and Walid Mahdi. 2016. Yawning detection by the analysis of variational descriptor for monitoring driver drowsiness. In *2016 International Image Processing, Applications and Systems (IPAS)*. 1–5. https://doi.org/10.1109/IPAS.2016.7880127

[5] Hamzah S. AlZu'bi, Waleed Al-Nuaimy, and Nayel S. Al-Zubi. 2013. EEG-based Driver Fatigue Detection. *2013 Sixth International Conference on Developments in eSystems Engineering* (2013), 111–114.

[6] Mohammad Amin Assari and Mohammad Rahmati. 2011. Driver drowsiness detection using face expression recognition. In *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. 337–341. https://doi.org/10.1109/ICSIPA.2011.6144162

[7] Jing Bai, Wentao Yu, Zhu Xiao, Vincent Havyarimana, Amelia C Regan, Hongbo Jiang, and Licheng Jiao. 2021. Two-stream spatial–temporal graph convolutional networks for driver drowsiness detection. *IEEE Transactions on Cybernetics* 52, 12 (2021), 13821–13833.

[8] Michael H. Bonnet. 1985. Effect of Sleep Disruption on Sleep, Performance, and Mood. *Sleep* 8, 1 (03 1985), 11–19. https://doi.org/10.1093/sleep/8.1.11 arXiv:https://academic.oup.com/sleep/article-pdf/8/1/11/13678498/080102.pdf

[9] Yetong Cao, Fan Li, Xiaochen Liu, Song Yang, and Yu Wang. 2023. Towards reliable driver drowsiness detection leveraging wearables. *ACM Transactions on Sensor Networks* 19, 2 (2023), 1–23.

[10] G Daquin, J Micallef, and O Blin. 2001. Yawning. *Sleep Medicine Reviews* 5, 4 (2001), 299–312.

[11] Christiaan Jacob Doelman and Johannes Adriaan Rijken. 2022. Yawning and airway physiology: a scoping review and novel hypothesis. *Sleep and Breathing* 26, 4 (01 Dec 2022), 1561–1572. https://doi.org/10.1007/s11325-022-02565-7

[12] Shkurta Gashi, Aaqib Saeed, Alessandra Vicini, Elena Di Lascio, and Silvia Santini. 2021. Hierarchical Classification and Transfer Learning to Recognize Head Gestures and Facial Expressions Using Earbuds. In *Proceedings of the 2021 International Conference on Multimodal Interaction* (Montréal, QC, Canada) *(ICMI '21)*. Association for Computing Machinery, New York, NY, USA, 168–176. https://doi.org/10.1145/3462244.3479921

[13] Behnoosh Hariri, Shabnam Abtahi, Shervin Shirmohammadi, and Luc Martel. 2011. Demo: Vision based smart in-car camera system for driver yawning detection. In *2011 Fifth ACM/IEEE International Conference on Distributed Smart Cameras*. 1–2. https://doi.org/10.1109/ICDSC.2011.6042952

[14] Wen-Bing Horng, Chih-Yuan Chen, Yi Chang, and Chun-Hai Fan. 2004. Driver fatigue detection based on eye tracking and dynamic template matching. In *IEEE International Conference on Networking, Sensing and Control, 2004*, Vol. 1. 7–12. https://doi.org/10.1109/ICNSC.2004.1297400

[15] Rui Huang, Yan Wang, Zijian Li, Zeyu Lei, and Yufan Xu. 2020. RF-DCM: multi-granularity deep convolutional model based on feature recalibration and fusion for driver fatigue detection. *IEEE Transactions on Intelligent Transportation Systems* 23, 1 (2020), 630–640.

[16] Michael Ingre, Torbjörn Åkerstedt, Björn Peters, Anna Anund, and Göran Kecklund. 2006. Subjective sleepiness, simulated driving performance and blink duration: examining individual differences. *Journal of Sleep Research* 15, 1 (2006), 47–53. https://doi.org/10.1111/j.1365-2869.2006.00504.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2869.2006.00504.x

[17] Yashika Katyal, Suhas V. Alur, and Shipra Dwivedi. 2014. Safe driving by detecting lane discipline and driver drowsiness. *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies* (2014), 1008–1012.

[18] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Alessandro Montanari. 2018. Earables for Personal scale Behaviour Analytics. *IEEE Pervasive Computing* 17, 3 (2018).

[19] Zuojin Li, Shengbo Eben Li, Renjie Li, Bo Cheng, and Jinliang Shi. 2017. Online Detection of Driver Fatigue Using Steering Wheel Angles for Real Driving Conditions. *Sensors* 17, 3 (2017). https://doi.org/10.3390/s17030495

[20] Yang Liu, Zhenjiang Li, Zhidan Liu, and Kaishun Wu. 2019. Real-time arm skeleton tracking and gesture inference tolerant to missing wearable sensors. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*. 287–299.

[21] Yansha Lu, Chunsheng Liu, Faliang Chang, Hui Liu, and Hengqiang Huan. 2023. JHPFA-Net: Joint Head Pose and Facial Action Network for Driver Yawning Detection Across Arbitrary Poses in Videos. *IEEE Transactions on Intelligent Transportation Systems* (2023).

[22] Chulhong Min, Akhil Mathur, and Fahim Kawsar. June 2018. Exploring Audio and Kinetic Sensing on Earable Devices. In *WearSys 2018*. The 16th ACM Conference on Mobile Systems, Applications, and Services (MobiSys 2018).

[23] NHTSA. 2017. Drowsy Driving. https://www.nhtsa.gov/risky-driving/drowsy-driving. Accessed: 2023-23-03.

[24] Mona Omidyeganeh, Shervin Shirmohammadi, Shabnam Abtahi, Aasim Khurshid, Muhammad Farhan, Jacob Scharcanski, Behnoosh Hariri, Daniel Laroche, and Luc Martel. 2016. Yawning detection using embedded smart cameras. *IEEE Transactions on Instrumentation and Measurement* 65, 3 (2016), 570–582.

[25] Muhammad Ramzan, Hikmat Ullah Khan, Shahid Mahmood Awan, Amina Ismail, Mahwish Ilyas, and Ahsan Mahmood. 2019. A Survey on State-of-the-Art Drowsiness Detection Techniques. *IEEE Access* 7 (2019), 61904–61919. https://doi.org/10.1109/ACCESS.2019.2914373

[26] Judith A. Ricci, Elsbeth Chee, Amy L. Lorandeau, and Jan Berger. 2007. Fatigue in the U.S. Workforce: Prevalence and Implications for Lost Productive Work Time. *Journal of Occupational and Environmental Medicine* 49, 1 (2007), 1–10. http://www.jstor.org/stable/44997095

[27] Hashim Saeed, Tabish Saeed, Muhammad Tahir, and Momin Uppal. 2018. Risky driving behavior detection using in-vehicle WiFi signals. In *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*. IEEE, 1–5.

[28] Mandalapu Saradadevi and Preeti Bajaj. 2008. Driver fatigue detection using mouth and yawning analysis. *International journal of Computer science and network security* 8, 6 (2008), 183–188.

[29] Ines Teyeb, Olfa Jemai, Mourad Zaied, and Chokri Ben Amar. 2014. A novel approach for drowsy driver detection using head posture estimation and eyes recognition system based on wavelet network. In *IISA 2014, The 5th International Conference on Information, Intelligence, Systems and Applications*. 379–384. https://doi.org/10.1109/IISA.2014.6878809

[30] Brandy Warwick, Nicholas Symons, Xiao Chen, and Kaiqi Xiong. 2015. Detecting Driver Drowsiness Using Wireless Wearables. *2015 IEEE 12th International Conference on Mobile Ad Hoc and Sensor Systems* (2015), 585–588.