2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340643

Adolescent Asthma Monitoring: A Preliminary Study of Audio and Spirometry Modalities

Jeffrey A. Barahona[1], Katie Mills[2], Michelle Hernandez[2], Alper Bozkurt[1], Delesha Carpenter[3] and Edgar J. Lobaton[1]

***Abstract*— Asthma patients' sleep quality is correlated with how well their asthma symptoms are controlled. In this paper, deep learning techniques are explored to improve forecasting of forced expiratory volume in one second (FEV1) by using audio data from participants and test whether auditory sleep disturbances are correlated with poorer asthma outcomes. These are applied to a representative data set of FEV1 collected from a commercially available sprirometer and audio spectrograms collected overnight using a smartphone. A model for detecting nonverbal vocalizations including coughs, sneezes, sighs, snoring, throat clearing, sniffs, and breathing sounds was trained and used to capture nightly sleep disturbances. Our preliminary analysis found significant improvement in FEV1 forecasting when using overnight nonverbal vocalization detections as an additional feature for regression using XGBoost over using only spirometry data.**

***Clinical relevance*— This preliminary study establishes up to 30% improvement of FEV1 forecasting using features generated by deep learning techniques over only spirometry-based features.**

I. INTRODUCTION

Asthma is a chronic respiratory disease that affects the airways in the lungs. It is characterized by inflammation and narrowing of the airways which can make it difficult to breathe. Symptoms of asthma may include coughing, wheezing, shortness of breath, and chest tightness. These symptoms can range from mild to severe and may occur on a daily or intermittent basis. Asthma can be managed through a combination of medications, such as inhaled bronchodilators and corticosteroids [1], and lifestyle changes. Asthma attacks, also known as asthma exacerbations or flare-ups, can be triggered by a variety of factors, including exposure to allergens [2] (such as pollen, pet dander, or mold), exposure to irritants [3] (such as tobacco smoke or pollution), respiratory infections, and physical activity[4].

Asthma is a long-term condition that cannot be cured, but it can be controlled with proper treatment and management. The main goals of treatment are to control symptoms, prevent asthma attacks, and improve quality of life. Home-monitoring of physiological parameters including sleep quality, heart rate, respiratory rate, inhaler usage, and spirometry measurements has been shown to correlate with pediatrician based asthma assessment and control [5]. This indicates that wearable devices may be used as a complementary tool for monitoring asthma. Aside from direct monitoring, telemedicine solutions may play a significant role in improving adherence and enabling patients to achieve adequate awareness of and control over their own symptoms [6].

The aim of our research is to use physiological and environmental sensing modalities to draw inferences about changes in lung function and asthma exacerbations. By using spirometry measurements, we tracked forced expiratory volume in 1 second (FEV1), forced volume capacity (FVC), forced expiratory volume in 6 seconds (FEV6), Forced mid-expiratory flow (FEF 25-75), and the ratio FEV1/FVC. This paper highlights our preliminary assessment in using common commercial devices for monitoring asthma in adolescents using FEV1 for one month's worth of data using features extracted from mel-spectrograms.

This paper provides a preliminary analysis of this data and presents the feature selection for a machine learning model, Extreme Gradient Boosting (XGBoost) to forecast FEV1 and demonstrate improvement by measuring nightly sleep disturbances, successfully replicating the correlation levels reported earlier [7]. This promising result reinforces the use of machine learning on wearable devices to pave the way for using such continuous and quantitative monitoring tools to support asthma management and control.

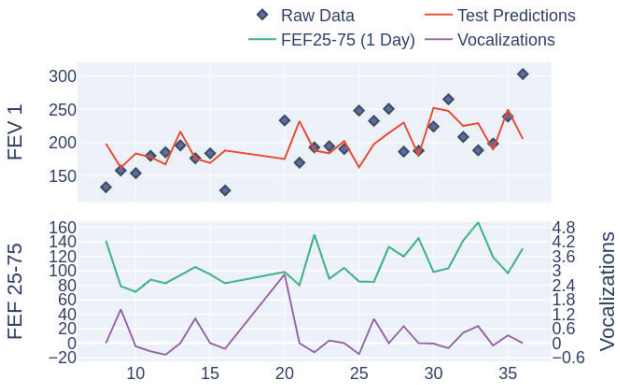

Fig. 1. Illustration of our prediction for FEV1 values using 1-day historic values (Top), and the FEF 25-75 and normalized vocalization values used as predictors (Bottom). We observe an improved on RMSE for the predictions of over 30% when including our deep-learning based vocalization detector.

*This work was supported by National Science Foundation (NSF) under award IIS-1915599, IIS-2037328, and EEC-1160483 (ERC for ASSIST).

[1]Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695, USA; Corresponding author: edgar.lobaton@ncsu.edu

[2]Children's Research Institute, University of North Carolina at Chapel Hill, Chapel Hill NC 27599, USA

[3]Division of Pharmaceutical Outcomes and Policy, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27559, USA

II. METHODOLOGY

A. Data Collection

The participant data in this paper was collected as part of an on-going study under the NC State University Institutional Review Board (IRB) approved protocol 16598. In this study, we monitored adolescents that are between 11 years and 18 years old with poorly controlled or uncontrolled asthma using several wearable and portable devices. In this paper, we analyze representative data from one of the subjects. The participant was engaged in this study for 4 months, and this analysis was conducted on the first month's worth of data, comprised of over 160 hours of audio recorded at 16KHz and 30 days of spirometry measurements recorded daily. A more comprehensive analysis will be performed once the study is over and the data collection is completed.

The study used iOS device (Apple, Cupertino, CA, USA) for sleep acoustics data and the Spirobank Smart spirometer (Medical International Research (MIR), New Berlin, WI, USA) for FEV1. The iOS Device was an iPhone 8 which hosted commercially available and custom made apps for the wearable devices, and it was also used to record forced cough sounds, overnight audio and survey responses. The spirometer captured FEV1, FEV1/FVC, FEV6 and FEF 25-75 indexes. The relevant part of the protocol involved participants performing daily a spirometer test, recording a few instances of forced cough, and setting the phone for overnight recording of audio mel-spectrogram features (no raw audio due to privacy concerns). The spectrograms were continuously recorded throughout the night using a windows size of 2048 samples with a hop length of 512 samples.

B. Problem Statement

For this study, we determined the impact of features from overnight audio recordings on the prediction of lung function, specifically the FEV1 index. First, we determine features, x_t, to use by determining the significance and strength of correlations depending on the size of the lag between the forecasted FEV1 value, $\hat{y}_t$, and the point used as an input to the forecasting model, described effectively as $\hat{y}_t = g(x_{t-1}, x_{t-2})$.

A binary classification model, f, was trained to detect non-verbal vocalizations, returning 1 if a detection is made on the audio and 0 otherwise. Each night during the study, a set of audio spectrograms was recorded, denoted as S_i^t where t indicates the day in the study and i indicates a spectrogram in the set recorded that day. The number of detections for a given day can be described as $c_t = \sum_{i=1}^{N_t} f(S_i^t)$, where N_t denotes the number of samples recorded on a given day. The normalized number of detections for each night were used as features for the forecasting and can be described as

$$\bar{c}_t = \frac{c_t - \frac{1}{M}\sum_{k=1}^{M} c_k}{\sqrt{\frac{1}{M}\sum_{k=1}^{M}\left(c_k - \frac{1}{M}\sum_{j=1}^{M} c_j\right)^2}},$$

where M is the number of days in the study used to train the forecasting model.

The forecasting model with vocalization features included can be described as $\hat{y}_t = g(x_{t-1}, x_{t-2}, \bar{c}_{t-1})$. The RMSE error for each model on the forecasting tasks are calculated, and the percent improvement of the RMSE error of the vocalization-enhanced model over the forecasting model is reported in addition to a RMSE score normalized by the standard deviation of the FEV 1 values for a given day.

C. Datasets for Non-Verbal Vocalization

The datasets used are enumerated in Table I. All audio data was resampled to 16KHz, and the data was split in a 7:1:2 ratio between training, validation, and testing data while also ensuring that datasets that identified individual speakers did not share them across the subsets. Audio data from the participant was not used for training the model.

For robustness, a number of data augmentations were used during training but not during validation or testing. Adopting a similar strategy as the one presented by Xu et al. [16], we used time shifting, polarity inversion, pitch shifting, background noise augmentation, and mixup augmentation, each with 50% probability of being used, to increase the variability and number of training samples.

D. Training of Non-Verbal Vocalization Model

A model, developed in PyTorch, was trained to detect auditory night time disturbances in the form of coughs and other nonverbal vocalizations. The model used a dense convolutional layer to convert the single channel mel-spectrogram input to a 3 channel image for the purpose of being used as an input to a pre-trained image model, leveraging the performance of CV models trained on a much larger dataset than what is typically used for audio. The image model was used to extract features that were fed to a simple neural network composed of a linear layer with a rectified linear unit activation and another linear layer to perform the classification.

The pre-trained image model, the number of neurons used in the linear layer, and the learning rate was selected using the Hyperband based hyperparameter optimization pipeline provided by the Ray library. The hyperparameter ranges and options are listed in Table II. The final model used 2048 neurons on the linear layer, an EfficientNetV2L image pre-trained model, and an initial learning rate of 1e-4. In all trials, the Adam optimizer was used for training.

TABLE I
DATASETS FOR NON-VERBAL VOCALIZATION TRAINING.

Dataset	Purpose
Coughvid[8]	positive samples of cough
Flusense[9]	positive samples of cough
ESC 50[10]	positive and negative samples of nonverbal vocalizations
ESTI[11]	background audio augmentation
AIR[12]	room impulse response augmentation
DEMAND[13]	background audio augmentation
Musan[14]	negative samples of nonverbal vocalizations and background augmentation
VocalSound[15]	positive samples of nonverbal vocalizations

The hyperparameter search was performed on the training and validation subsets. After the model was selected, the best configuration was trained on the training and validation sets and then evaluated on the test set to determine its performance characteristics on unseen data. Afterwards, the model was trained on the entire dataset and used for detecting sleep disturbances on the study participant data.

The model selected by the hyperparameter pipeline was trained for 15 epochs. The test performances are: Accuracy (0.96), Precision (0.92), Recall (0.95) and F1 Score (0.93).

III. RESULTS AND DISCUSSION

First, we identified which variables to use for forecasting by performing a correlation analysis between the FEV1 values at time t; and the tested variable values 1 day or two days prior, and the vocalizations from the day prior. A Spearman correlation test was used with $\alpha < 0.1$ as the threshold to keep or exclude features. Table III enumerates the spirometry and vocalization features used, their Spearman coefficients and p-values. To perform the forecasting, we used a XGBoost model to regress on the spirometry data and the overnight disturbance detections and forecast future FEV1 values. Figure 1 illustrates the leave-one-out predictions obtained for the FEF 25-75 model with a 2-day history and vocalization.

Combinations of all spirometry features with and without added 1 day vocalization features were used to train XGBoost models using Leave-One-Out cross-Validation where a single days' worth of spirometry measurements are left out. Using Random Search, we found that an XGBoost Model with 1500 estimators, a maximum depth of 2, and a 0.625 subsampling ratio had the best performance across feature choices. The best performances were observed when considering 2 days of historical values. The top 5 performing XGBoost models using a 2-day history are shown in Figure 2. Overall, FEF 25-75 with vocalization features provides the best performance compared to all other combinations of spirometry and vocalization features. Figure 3 shows XGBoost performance using FEF 25-75 and vocalization features across cross-validation using different size folds. The cross-validation windows were generated by using a sliding window of the specified size to determine which days to use for testing (e.g., a fold size of 1 corresponds to a leave-one-out cross-validation). The downward trend in percent improvement was expected due to the limited amount of data (i.e., larger fold sizes reduced the amount of data available for training), and the temporal correlation observed (i.e., larger folds resulted in less correlated training and test sets). This also applied to the normalized RMSE score.

To assess of the impact of the variability of the vocalization features, multiple iterations of the non-verbal vocalization model were trained using different combinations of data augmentations and used to generate vocalization features for the XGBoost model. Section II-C lists the different type of data augmentation. Figure 4 depicts the distribution of the percent improvement of using vocalization features in addition to spirometry features compared to just using spirometry features for various fold sizes. We observed that vocalization features improve forecasting model performance

TABLE II
HYPERPARAMETER RANGES AND OPTIONS

Hyperparameter	Range/Options
Image Model	DenseNet, ResNet101, EfficientNet B4, EfficientNet B7, EfficientNet V2M, EfficientNet V2L
Linear Layer Size	1024-8096
Learning Rate	10^{-5}-10^{-2}

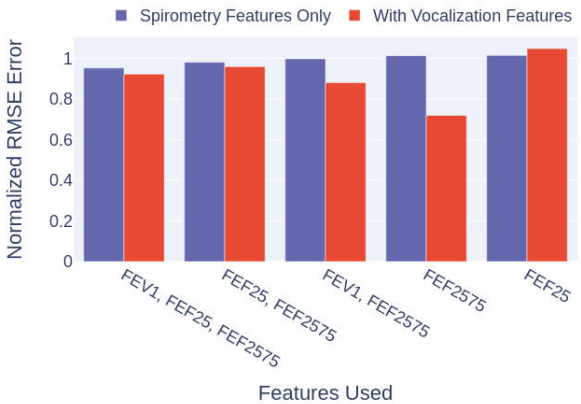

Fig. 2. Top Performing Forecasts Using Two Days of Features.

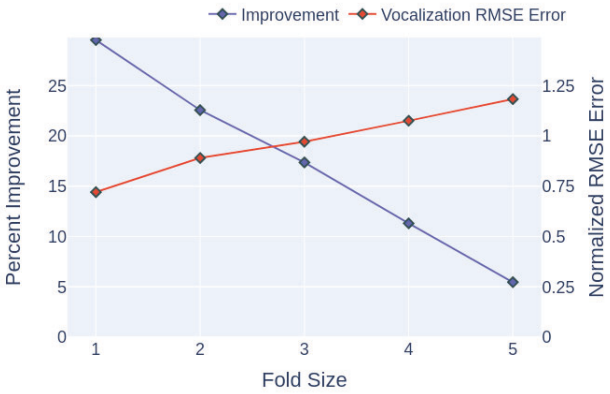

Fig. 3. Improvement of model using FEF 25-75 with 2-Day history when using vocalization features for various cross-validation fold sizes (blue). The normalized RMSE of the model with vocalization is also shown (red).

TABLE III
SPEARMAN VALUES FOR SPIROMETRY PARAMETERS

Feature	1 Day		2 Days	
	ρ	p	ρ	SP
FEV1	0.541	$\mathbf{1.26e-4}$	0.357	**0.028**
FEV6	0.099	0.517	0.131	0.427
FVC	0.142	0.353	0.160	0.330
FEF 25	0.598	$\mathbf{1.41e-5}$	0.403	**0.010**
FEF 75	0.331	**0.027**	0.169	0.303
FEF 25-75	0.538	$\mathbf{1.37e-4}$	0.333	**0.038**
Vocalizations	-0.690	**0.059**	-0.011	0.972

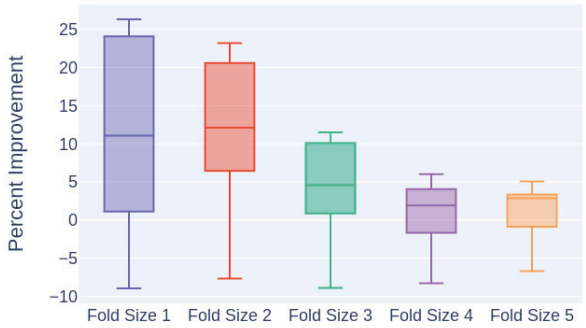

Fig. 4. Aggregated cross-validation performances for various window Sizes. Percentage of improvements of the model with vocalization are shown on the y-axis. This shows the variability of the impact of the model.

overall.

IV. DISCUSSION AND CONCLUSION

We demonstrate improvement on FEV1 forecasting over our baseline methods using features generated by deep learning techniques. Future work will incorporate additional modalities from data recorded in the study and report results on more participants. We also aim to demonstrate model reliability by incorporating interpretability through uncertainty modeling and out of distribution detection.

Privacy remains a major concern for participants, and minors justifiably receive additional protections in medical studies. For this reason, we chose to only record spectrograms of the overnight audio, and this introduces its own challenges. This makes it difficult to verify how well the models perform on each participant, and, while forced cough data can be helpful for this assessment, forced coughs are not truly representative of the sound of spontaneous coughs and their context of occurrence.

REFERENCES

[1] D. M. Sobieraj, W. L. Baker, E. Nguyen, *et al.*, "Association of inhaled corticosteroids and long-acting muscarinic antagonists with asthma control in patients with uncontrolled, persistent asthma: A systematic review and meta-analysis," *JAMA*, vol. 319, no. 14, pp. 1473–1484, Apr. 10, 2018.

[2] S. R. Del Giacco, A. Bakirtas, E. Bel, *et al.*, "Allergy in severe asthma," *Allergy*, vol. 72, no. 2, pp. 207–220, Feb. 2017.

[3] O. Fuchs, T. Bahmer, K. F. Rabe, and E. von Mutius, "Asthma transition from childhood into adulthood," *The Lancet. Respiratory Medicine*, vol. 5, no. 3, pp. 224–234, Mar. 2017.

[4] L.-P. Boulet and P. M. O'Byrne, "Asthma and exercise-induced bronchoconstriction in athletes," *The New England Journal of Medicine*, vol. 372, no. 7, pp. 641–648, Feb. 12, 2015.

[5] M. R. van der Kamp, E. C. Klaver, B. J. Thio, *et al.*, "WEARCON: Wearable home monitoring in children with asthma reveals a strong association with hospital based assessment of asthma control," *BMC Medical Informatics and Decision Making*, vol. 20, p. 192, Aug. 14, 2020.

[6] B. Davies, P. Kenia, P. Nagakumar, and A. Gupta, "Paediatric and adolescent asthma: A narrative review of telemedicine and emerging technologies for the post-COVID-19 era," *Clinical & Experimental Allergy*, vol. 51, no. 3, pp. 393–401, 2021.

[7] F. S. Luyster, M. Teodorescu, E. Bleecker, *et al.*, "Sleep quality and asthma control and quality of life in non-severe and severe asthma," *Sleep & breathing = Schlaf & Atmung*, vol. 16, no. 4, pp. 1129–1137, Dec. 2012.

[8] L. Orlandic, T. Teijeiro, and D. Atienza, "The COUGHVID crowdsourcing dataset: A corpus for the study of large-scale cough analysis algorithms," *Scientific Data*, vol. 8, no. 1, p. 156, Dec. 2021.

[9] F. Al Hossain, A. A. Lover, G. A. Corey, N. G. Reich, and T. Rahman, "FluSense: A contactless syndromic surveillance platform for influenza-like illness in hospital waiting areas," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 1, 1:1–1:28, Mar. 18, 2020.

[10] K. J. Piczak, "ESC: Dataset for environmental sound classification," in *Proceedings of the 23rd ACM international conference on Multimedia*, ser. MM '15, New York, NY, USA: Association for Computing Machinery, Oct. 13, 2015, pp. 1015–1018.

[11] *Speech processing, transmission and quality aspects (STQ); speech quality performance in the presence of background noise;* 2008.

[12] M. Jeub, M. Schafer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *2009 16th International Conference on Digital Signal Processing*, ISSN: 2165-3577, Jul. 2009, pp. 1–5.

[13] J. Thiemann, N. Ito, and E. Vincent, *DEMAND: A collection of multi-channel recordings of acoustic noise in diverse environments*, Type: dataset, Jun. 9, 2013.

[14] D. Snyder, G. Chen, and D. Povey, *MUSAN: A music, speech, and noise corpus*, Oct. 28, 2015.

[15] Y. Gong, J. Yu, and J. Glass, "Vocalsound: A dataset for improving human vocal sounds recognition," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 23, 2022, pp. 151–155.

[16] X. Xu, E. Nemati, K. Vatanparvar, *et al.*, "Listen2cough: Leveraging end-to-end deep learning cough detection model to enhance lung health assessment using passively sensed audio," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 1, pp. 1–22, Mar. 19, 2021.