Cureus Journal of Engineering

Open Access Letter

Smartphone-Based Prediction of Human Exhale Air Peak Flow Using Exhale Air Audio

Manisha R. Mhetre [1] , Piyush V. Kumbhakern [1] , Rajeshwari B. Kul [1] , Raj R. Kudtarkar [1] , Viraj A. Jadhav [1]

1. Instrumentation and Control Engineering, Vishwakarma Institute of Technology, Pune, IND

Corresponding authors: Manisha R. Mhetre, manisha.mhetre@vit.edu, Piyush V. Kumbhakern, piyush.kumbhakern22@vit.edu

Received 02/21/2025
Review began 04/07/2025
Review ended 04/20/2025
Published 05/12/2025

DOI: https://doi.org/10.7759/s44388-024-01329-1

Abstract

Human lung capacity is measured by one way of measuring maximum speed of exhaled air blown in a Peak Flow Meter (PFM). This traditional method requires a physical device, PFM, which may not be always be available for screening test in remote or resource-limited areas. Hence, a new method has been devised in this research work, which is user-friendly, non-invasive, and easily accessible to all. This research paper presents a novel method for measuring human respiratory peak flow rate in liters per minute, which is much more convenient than existing devices like the PFM. Peak airflow rate from human lung is estimated from a person's maximum exhaled air blown audio files recorded using a smart phone application and performing its feature extraction using Mel-Frequency Cepstral Coefficients (MFCCs) coupled with a Random Forest Regressor.

The procedure is done by recording voice samples from 86 healthy people both male and female, in the age group from 12 to 65 years, with height from 4.4 ft to 6.1 ft, and weight from 20 kg to 89 kg, as per medical doctor recommendations. MFCC features are extracted from the collected voice samples. The model is trained and saved along with the person's actual a peak flow rate measured through PFM with standard procedure. A Flask web app is developed for users, in which users can upload audio files which are then processed using Librosa for MFCCs feature extraction. It also gives a JSON result with volume of air exhaled in liters per minute, along with a MFCC graph as defined by the model.

Model validation has been carried out by comparing the values predicted by MFCC and standard PFM measurements. Model accuracy obtained is up to 87%, which reveals that the performance of the audio-based predictions is consistent with actual peak flow measurements as indicated on the smartphone web app easily.

Categories: Signal Processing and Image Analysis, Data Science and Analytics for Engineering, Embedded Systems and IoT
Keywords: mfcc, random forest regressor, audio processing, human exhaled air prediction, machine learning, flask, volume estimation

Introduction

Human exhale air peak flow measurement using a Peak Flow Meter (PFM) is an important activity for assessment of lung capacity. Traditional peak flow measurement techniques, such as using PFM, require physical devices that may not always be accessible, especially in remote or resource-limited settings. Another issue with using the same PFM across multiple individuals without proper sanitation, may increase the risk of spreading respiratory diseases. Numerous industries, yet by conventional ways, offer devices with great expenses and are invasive.

There is a lack of non-invasive, smartphone-based alternatives for measuring exhaled air peak flow, particularly methods that are cost-effective and easy to use. Based on this, research has been carried out to introduce a cheaper yet efficient user-friendly, non-invasive, and easily accessible peak flow measurement technique using audio recording facility of smartphone.

Research has been carried out by keeping the following objectives: (1) To develop a method for estimating human exhaled air peak flow using audio signals. (2) To extract relevant features from exhaled breath sounds. (3) To build a predictive model using machine learning (ML) that can estimate peak flow values from the extracted features. (4) To validate the proposed method by comparing predicted values with standard PFM measurements.

Key audio features have been extracted with Mel-Frequency Cepstral Coefficients (MFCCs) and Random Forest Regressor. These methods aim at determining the human exhaled air in liters per minute by AI model training, as opposed to the PFM that is used to detect peak exhaled flow rate. The developed method is a passive technique.

How to cite this article
Mhetre M R, Kumbhakern P V, Kul R B, et al. (May 12, 2025) Smartphone-Based Prediction of Human Exhale Air Peak Flow Using Exhale Air Audio . Cureus J Eng 2 : es44388-024-01329-1. DOI https://doi.org/10.7759/s44388-024-01329-1

While conducting this research, audio recording dataset has been generated and used for model training.

Materials And Methods

Literature review

Use of audio signal analysis has dramatically increased in the industrial arena mainly in the area of fault detection and also in the extraction of signal from the sound made by various structures. Certain features including MFCCs have been used to predict machinery ill-health while analyzing the sound.

MFCC represents an approach to transform auditory data into digital numerical models that computers can decode. These acoustic measurements prove effective because they process making it receptive which correspond to human hearing competence. The MFCC system disintegrates breathing audio signals, before studying the pitch and strength of various spectral components, to generate numeric data. The analysis helps to determine both the intensity and speed of breath flows. The study team extracted significant sound features using MFCC from exhalations to enable ML models to forecast peak airflow similar to professional PFMs. This approach has mainly remained inherent with the mechanical systems with the intention of quantizing the unsteady state of expelled humans air through acoustics signals.

However, the use of audio data for measuring human exhaled air is not well explored.

Audio Signal Processing in the Industrial Process

The features of Linear Predictive Coding and MFCCs are very suitable in the classification of audio signals. Originally, it was the acoustic approaches where potential problems in the machinery have been sought, since certain sounds could help determine that there is one. Such techniques are significantly based on the "acute" characteristics such as spectral and temporal characteristics of the sounds using MFCCs [1].

In this research work, same frameworks has been used to predict human exhaled air through a microphone.

To perform audio preprocessing, the Librosa Python library was employed; all the code was implemented using Python. This setup allowed for easy extraction of MFCCs features and the creation of a prediction model with Random Forest Regressor in addition to testing with Fast Fourier Transform (FFT) and Power Spectral Density (PSD).

Fast Fourier Transform and Power Spectral Density

MFCCs are chosen because of its closeness to human perception instead of FFT or PSD [2]. In this research work, when MFCCs are applied on audio data file including human exhalation, MFCC maintains features of the spectrum where lower frequencies are emphasized, as characteristics of sound may not always manifest uniformly or necessarily follow simple patterns.

Both FFT and PSD are beneficial in the study of energy distribution and frequency; however, they merely acquire point data without respect to how the ear recognizes auditory stimuli [3]. On the other hand, MFCCs mimic a human ear functionality by analyzing audio into meaningful spectral frequencies thus improving the capturing of intricate yet subtle audio features that exists in the expelled human breath.

Furthermore, FFT and PSD might miss out on important details due to the restricted acquisition rates of Smartphone's microphones. This is because MFCCs give improved results in modeling critical acoustic patterns for anticipating human exhaled air, the primary reason for their emphasis on the perceptually salient aspects. It is for this reason that MFCCs have been selected for this research work as explained below.

Mel-Frequency Cepstral Coefficients in Audio Analysis

MFCCs were formally used for speech processing but can be used in sound analysis of industrial and environmental sound including noises [4]. These features allow for the prediction of modeling patterns that are superficial and can be processed by ML. MFCCs are used to mimic the sounds of human exhaled air to allow assessment by the prediction model [5].

Machine Learning for Predicting Human Exhaled Air

Deep learning models of environmental sounds accurately distinguish complex sound signatures, which is useful in studying human expiration [6]. Traditional ML approaches have used physical parameters, including pressure and temperature, to predict flow rates. A novel method involves using models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) with audio input; this

application may detect subtle patterns in audio, enhancing exhalation prediction accuracy [7].

In the dataset, MFCCs are derived from audio recordings, resulting in 13 coefficients that analyze voice frequencies. These values are stored in a .CSV file alongside peak flow values, representing the volume of air expired in liters per minute.

By having actual flow values and using the audio features obtained from MFCCs, it is possible to train ML algorithms to predict peak flow. The dataset, structured to connect the 13 MFCC inputs to the corresponding peak flow output, creates a stable predictive environment. This trained model fine-tunes the linear regression model based on the mapping of audio features to measured peak flow, thereby improving the model's ability to predict peak flow for other audio data samples [8].

Figure *1* shows MFCC plot for the research work.

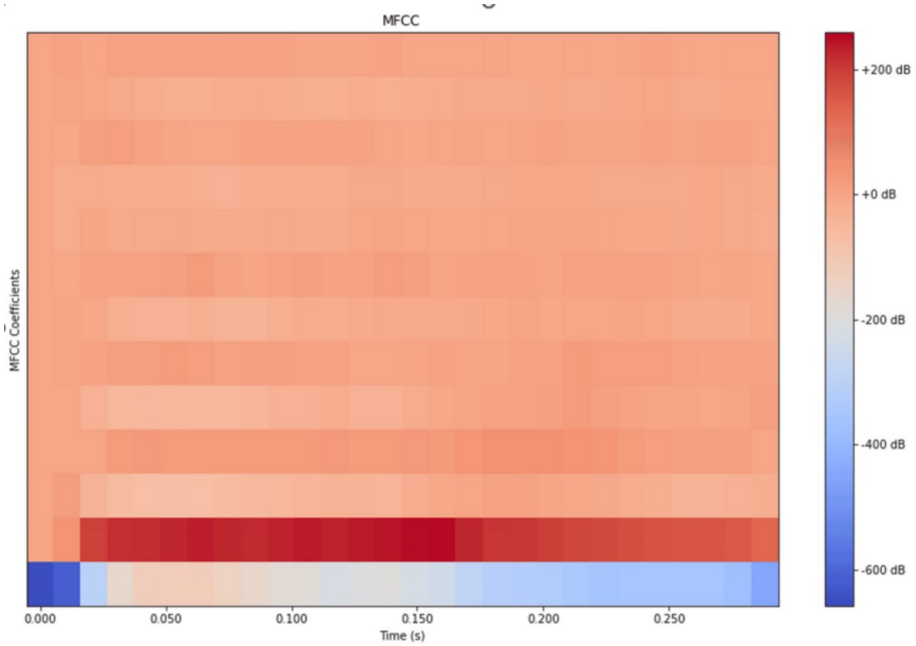

FIGURE 1: Mel-Frequency Cepstral Coefficients (MFCCs) plot

MFCC (Mel Frequency Cepstral Coefficients) Extraction

MFCC is a technique that is used to extract important features from audio files. It transforms the audio signal from the time domain into the frequency domain, making it easier for ML models to interpret.

Formula:

Discrete Fourier Transform (DFT):

$$X[k] = \sum_{n=0}^{N-1} x[n]e - j2\pi kn/NK = 0, 1, - - -, N - 1$$

Where:

· X[k] is the DFT of the signal x[n]

· N is the number of samples.

Mel Frequency (Mapping Hertz to Mel scale):

$$m(f) = 2595 \log_{10}(1 + f/700)$$

Where:

· f is the frequency in Hertz,

· m(f) is the frequency on the Mel scale.

Inverse Mel Formula (Mapping Mel scale to Hertz):

$$f(m) = 700(10m/2595 - 1)$$

Unit: Frequency: Hertz (Hz)

Volume Prediction Model Using Random Forest Regressor

In particular, the Random Forest Regression model is used to calculate the volume of airflow based on MFCC features of the sound signal. The training process is carried out based on input features, which are the MFCCs, and the target values which are the volume in liters per minute [9].

Model Equation:

$$y = 1/T \sum_{t=1}^{T} ft(X) \qquad \backslash$$

Where:

· Y is the predicted volume (L/min),

· T is the number of trees in the forest,

· ft(x) is the prediction made by tree t for the input x (the MFCC features).

Unit:·Frequency: Hertz (Hz)

In this model T used is 100 i.e. number of trees.

Acoustic Methods in Human Exhaled Air Measurement

By using sound waves produced by the flow of the human exhaled air, it is possible to estimate the flow rates achieved. It also does not require direct contact with the fluid as the traditional flow meters mostly does. Nevertheless, several difficulties are present including environmental noise and corresponding microphone sensitivity, which undeniably affect the audio data and make it difficult to predict flow accurately.

Challenges in Audio-Based Human Exhaled Air Measurement

Using sound to measure human exhaled air has its advantages but also its disadvantages, especially from noise that interferes with the audio information. Some of the key measures mentioned are about controlling the environment and a few are about using better quality of microphones to have better results in terms of prediction [10]. The nature of aural data also varies, meaning that there is a lot of variation that requires normalization when addressing the data. Support Vector Machine classifiers have been successfully used for the classification of environmental sounds and may be applied to enhancing the analysis of the human exhaled air sound [11].

Table *1* shows software tools used for this project development.

Software/Library	Purpose
Python	Main programming language.
Flask	Web framework for API and interface.
Librosa	Audio analysis and feature extraction.
Matplotlib	Audio analysis and feature extraction.
Scikit-learn	Machine learning model (Random Forest Regressor).
Pandas	Data manipulation and analysis.
NumPy	Numerical operations.
HTML/CSS	Frontend design for user interaction.
Anaconda/Spyder	Python distribution for environment management.
GIT	Version control and collaboration.

TABLE 1: Software tools used

Methodology

This section explains how the methodology to establish an audio-based prediction system of human exhaled air rates was chosen and followed. The first patients are data acquisition, feature extraction by MFCCs, model training using Random Forest Regressor, and deployment with Flask web application. The following are the key phases of the project methodology.

Audio Data Collection

The first procedure involved consideration of sound data from the human breath. Since the current research involved the analysis of the constant elements of human exhaled aerosols, audio files were sourced from controlled exhalation airflow experiments that involve the deployment of different flow rates using high-quality microphones. Flow rates were measured by conventional flow meters i.e. PFMs, and thus the collected flow rates were real known values to train the ML algorithms. The related audio signals were recorded in two formats: .wav and .mp3 to be compatible with commonly used audio analysis tools. Recording has been carried out using mobile in a noise-free environment as shown in Figure *3*.

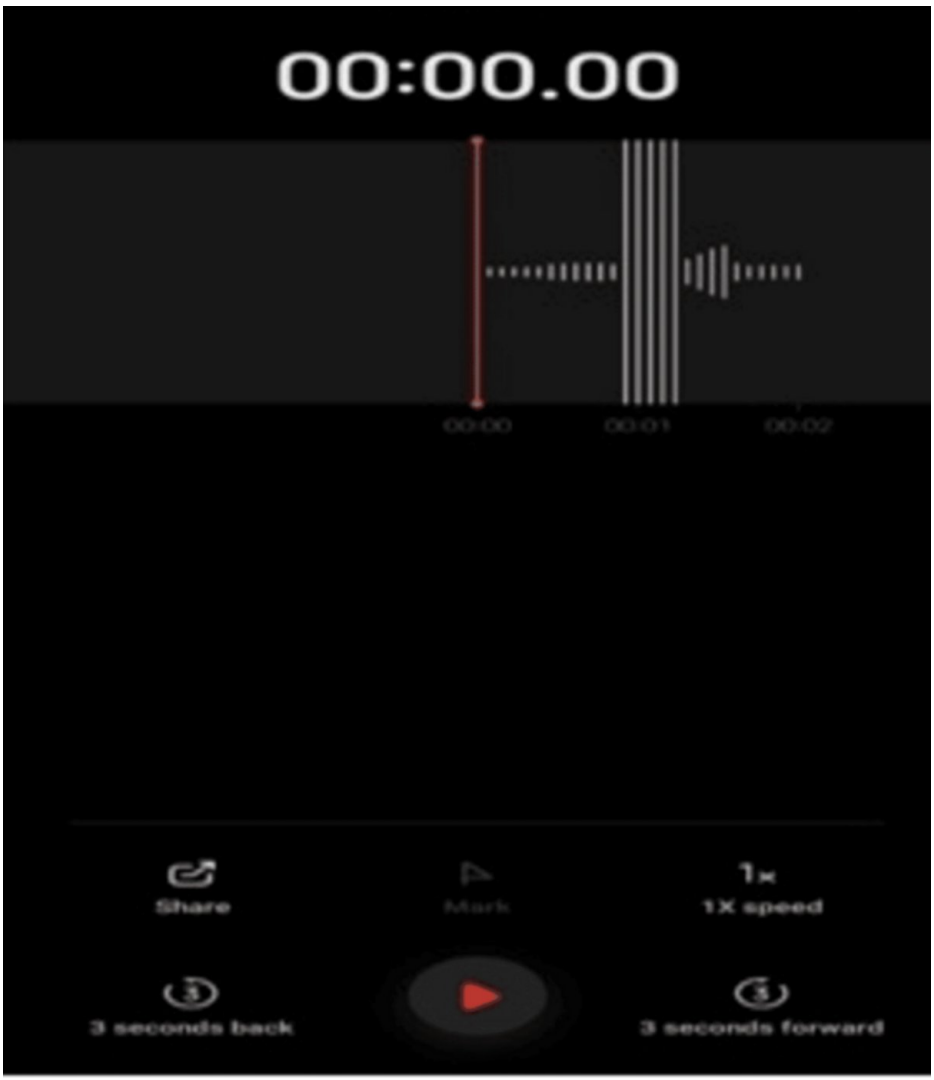

FIGURE 2: Recorder app

Figure *2* shows the app used for human exhaled air sound recording purpose.

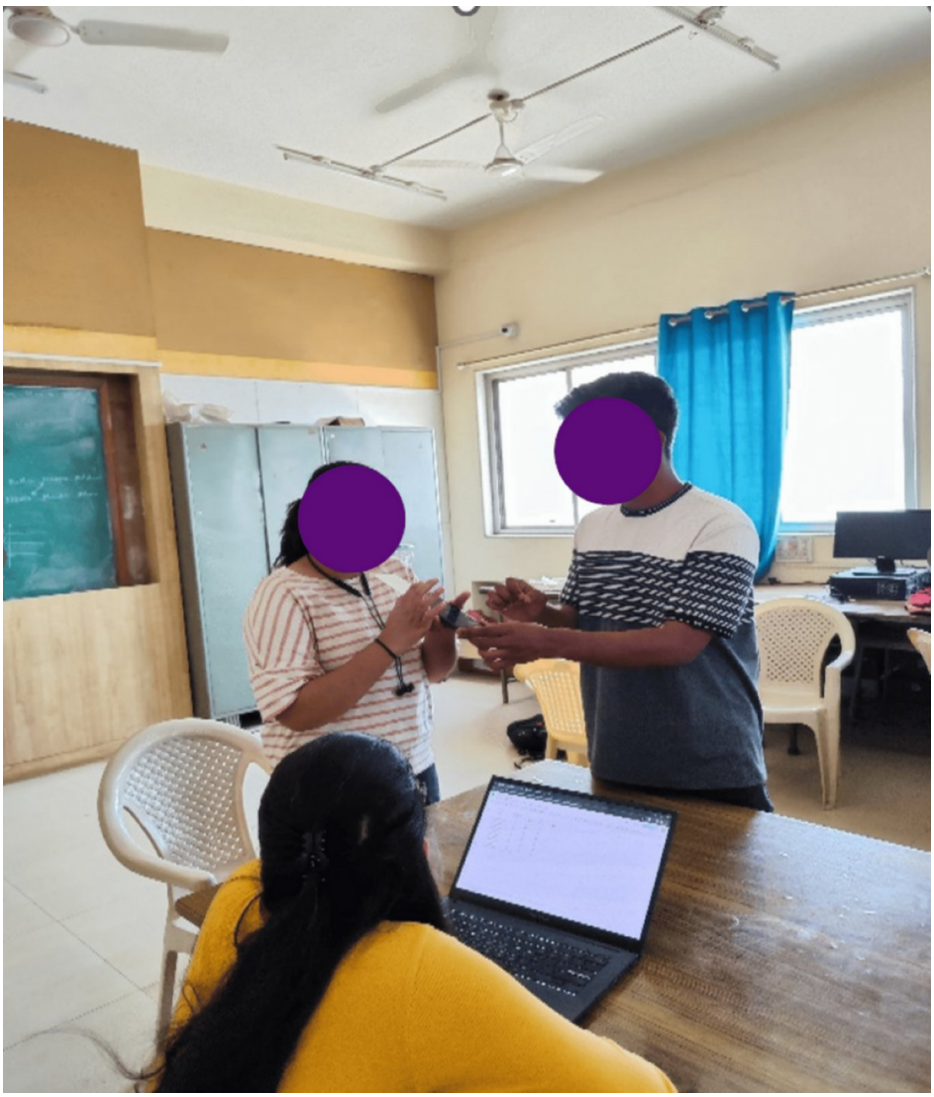

FIGURE 3: Human exhaled air sound recording

Figure 3 shows actual recording setup and procedure.

Duration of Audio

The duration of the audio is calculated based on the number of samples and the sampling rate.

Formula:

$$Duration = Number of Samples / Sampling Rate(Hz)$$

Unit Duration: Seconds (s)

The audio recording data procedure is done by recording voice samples from 86 healthy people (male and female), in the age group from 12 to 65 years, with height from 4.4 ft to 6.1 ft, weight from 20 kg to 89 kg, and during different times of a day particularly after breakfast and lunch. These data have been collected as per the medical doctor's suggestion.

Table 2 shows some sample recording with person's demographic data and peak flow reading

Name	Age (years)	Hight (feet)	Weight (kg)	Peak flow values (L/min)
PERSON 1	20	5.6	74	430
PERSON 2	18	5.5	57	530
PERSON 3	10	4.4	20	440
PERSON 4	21	5.8	67	470
PERSON 5	35	5	50	490
PERSON 6	26	5.7	59	540
PERSON 7	29	5.8	70	480
PERSON 8	21	5.6	70	570

TABLE 2: Persons demographic data and peak flow reading

Audio Preprocessing

To ensure that highly accurate ML models are developed, audio data preprocessing is always necessary. Exceptional dataset for the assessment of models for environmental sounds is marked as essential. Several signal conditioning procedures were implemented using the Librosa library with included signal normalization and format conversion. These operations were kept constant and compatible with the ML system audio inputs [12].

The audio data was preprocessed using the Librosa library that is helpful in loading, normalizing the signals and converting files that are usable for ML models. Some important preprocessing operations still include handling and normalization of the audio data.

The audio recording features obtained via MFCC analysis received Min-Max scaling normalization until they reached a (0, 1) value range. To capture the essential frequencies of forced exhalation sounds that normally exist between 20 Hz and 1000 Hz, the filter application included a low-pass element set to a cutoff of 1000 Hz. The filtering mechanism eliminated high-frequency noise thereby strengthening exhaled air analysis in much the same way as PFM's function. The recordings were left at their original sample rate 44100 Hz and bit depth of 16-bit PCM, which enabled high-quality audio preservation for feature extraction operations.

Loading Audio: The audio files were loaded with the function librosa.load(), and both the audio time series (y) and the sample rate (sr) were extracted. Figure *4* shows the loading of audio file procedure.

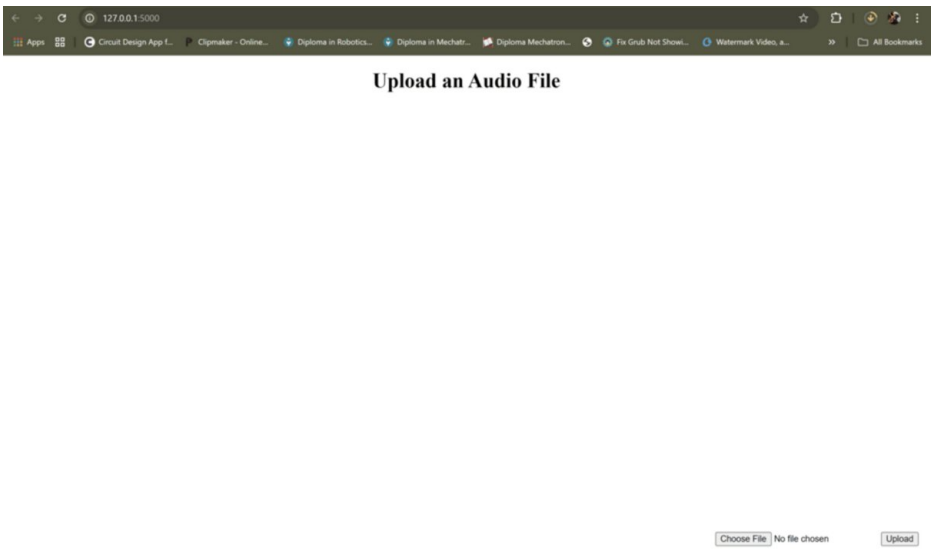

FIGURE 4: Uploading of recorded sound procedure

Recordings where the speaker was not taking part in the recording-that part of the recording is not used for actual training purposes.

Duration and Quality Check: Every file was checked to ensure the sample rate was more than zero and the length was adequate for feature extraction usually from 1 to 1.5 seconds. With regards to the selected time period, audio response time is taken into context with standard human response time which in turn check on the adequacy of the characteristics extracted in relation to the patterns that are expected to be characteristic of fluid flow. Furthermore, we obtained the highest results in the experiment at a distance of 15 cm, which further enabled the enhancement of the audio quality for analysis.

Feature Extraction

The MFCCs were then applied on the preprocessed signals in order to extract their features. MFCC has been appreciated for its effectiveness in the classification or prediction of audio signals. The modulated signals are useful in capturing necessary patterns of the human exhaled air audio signals, enhancing the predictive capacity of the model.

MFCC Calculation: For each file, for feature extraction, 13 MFCCs were calculated using librosa.feature.mfcc() function. The number of coefficients was kept low to eliminate the presence of features that are less dominant while at the same time working on the computational cost.

MFCC Mean Calculation: Thus, for each audio file, a feature vector was determined by the mean of the values of the MFCCs for all selected time frames.

Model Development

In order to predict human exhaled air rates, we used Random Forest Regressor algorithm as it tackles complex non-linear mapping of audio and airflow features. It generates powerful decision trees that together can provide for the complex identification of patterns in the data generated due to changes in sound intensity, frequency, and noise. It increases accuracy and at the same time minimizes the possible overfitting, preserving good predictors.

Dataset Preparation: Next, the features of the proposed MFCC and their flow rates derived from the traditional flow meters were integrated to form a dataset.

Model Training

For prediction, a Random Forest Regressor model was used. Data were split 70/30 with 70% used for training and 30% for validation. Grid Search was used before using the model to tune the parameters so as to improve its performance. The proposed MFCC-based approach has been motivated from other related fields like respiratory and lung sound classification [13]. On this dataset, the Random Forest Regressor was built using the assistance of the sklearn library.

Evaluation: Model accuracy was found to be 87%. To guarantee correct predictions, the model was assessed using other performance indicators including the Mean Squared Error (MSE) and the R-squared (R^2). The MSE was found to be approximately 186.42, and the R^2 value was 0.87. The exact MSE and R^2 values are yet to be finalized. At present, the dataset is limited in size and diversity. For more accurate and reliable results, a larger volume of data is required, especially multiple readings from individuals within the same category - such as similar age, height, and other relevant characteristics. Once sufficient category-wise data is gathered, we anticipate a more robust model with validated and consistent MSE and R^2 values.

Web Application Implementation

Interactive type of application was therefore designed to enhance user use and forms a web-based application developed using Flask. This application is designed to accept AU (audio) files in .wav or .mp3 format for real-time predictions. After that, the UTF8-encoded MP3 files are processed through the identical feature extraction and predictive profile as the previous process with MFCC feature extraction. There is a mutual compatibility between the user interface and the backend as pointed out on the previous works done by Bahoura and Hassani on MFCC feature extraction for real-time systems [14].

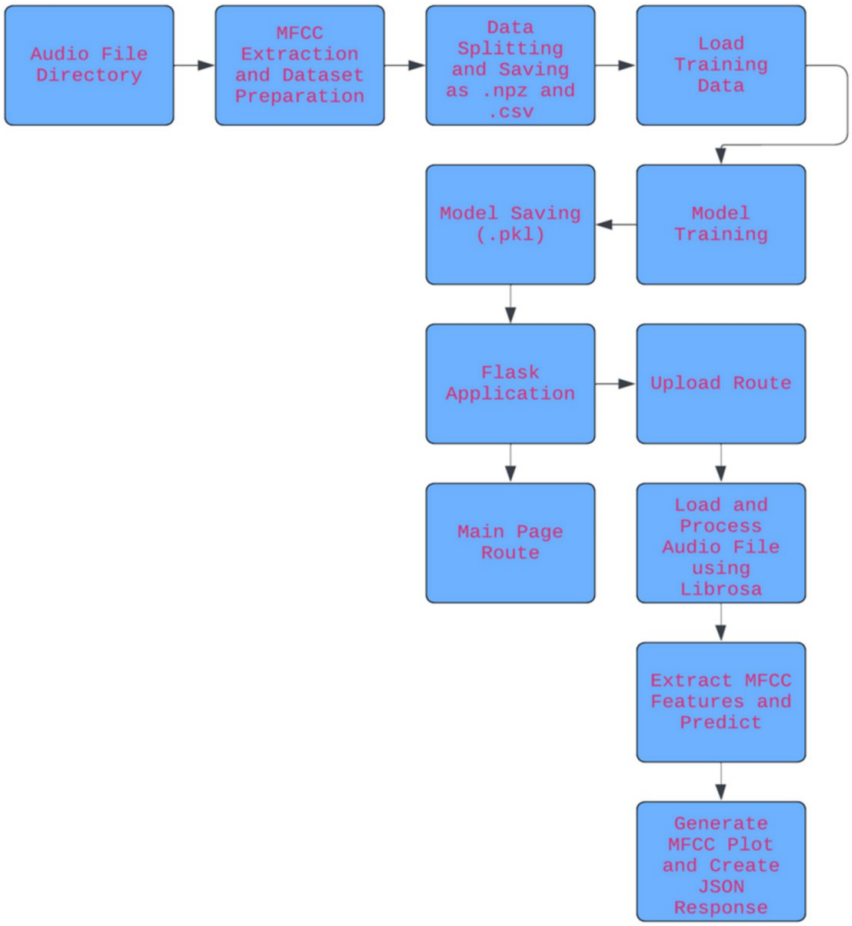

FIGURE 5: System work flow from uploading audio to output

MFCC, Mel-Frequency Cepstral Coefficient

This workflow, shown in Figure *5*, entails the training of a classifier for the classification of audio files, using MFCC features. Here's a concise explanation:

Audio File Directory: Audio files are collected.

MFCC Extraction: Features that are used are referred to as MFCCs features and they are abstracted from the audio files that form the dataset used in training.

Data Splitting: The data are divided into train and test datasets and both are stored in .npz format and a training set is also stored as .csv format.

Model Training: The training data is obtained, and then we have a model trained.

Model Saving: The trained model is stored as a.pkl file format.

Flask Application: To handle audio file upload and prediction Flask web app is developed.

Upload Route: Audio files are provided to the application by the users.

Audio Processing: Librosa library analyses the uploaded audio file.

MFCC Feature Extraction: From the uploaded audio, MFCC features are extracted.

Prediction and JSON Response: Finally, the model forecasts the output and gives out a JSON of the result; an MFCC graph is also shown.

Error Handling and Optimization

To ensure smooth functionality and user experience:

Exception Handling: The system contains built-in checks for easily recognizable problems, for example, if some audio is missing or if some files are in the wrong format, the system will return the right type of error message.

Model Fine-Tuning: Optimization on the Random Forest model was done by using Grid Search

The ML model required optimal settings (hyperparameters) which Grid Search method identified through its testing process. During the testing process different parameter values were examined for tree count, growth limit, and sample splitting thresholds.

The Grid Search employed 5-fold cross-validation to validate its parameters selection of 100 estimators and 20 maximum depth at 5 minimum samples split.

The Grid Search procedure identified 100 trees together with maximum depth at 20 and minimum samples split at 5 as the most suitable settings.

Number of Estimators (trees): 100

Maximum Depth (tree depth): 20

Minimum Samples Split: 5

System workflow

Audio Upload: Users upload an audio file to the Flask application.

MFCC Extraction: The system extracts MFCC features from the uploaded file.

Model Prediction: The Random Forest model predicts the flow rate using the extracted features.

Result Visualization: The predicted flow rate and an MFCC plot are displayed to the user.

Results

The expected flow rate is provided to the user and essential details on the audio input such as the time taken by the audio. Also, an MFCC plot is made using the matplotlib in order to visualize those features that have been extracted.

Result Visualization: The predicted flow rate and an example plot of MFCC are also shown to the user, as shown in Figure *6*.

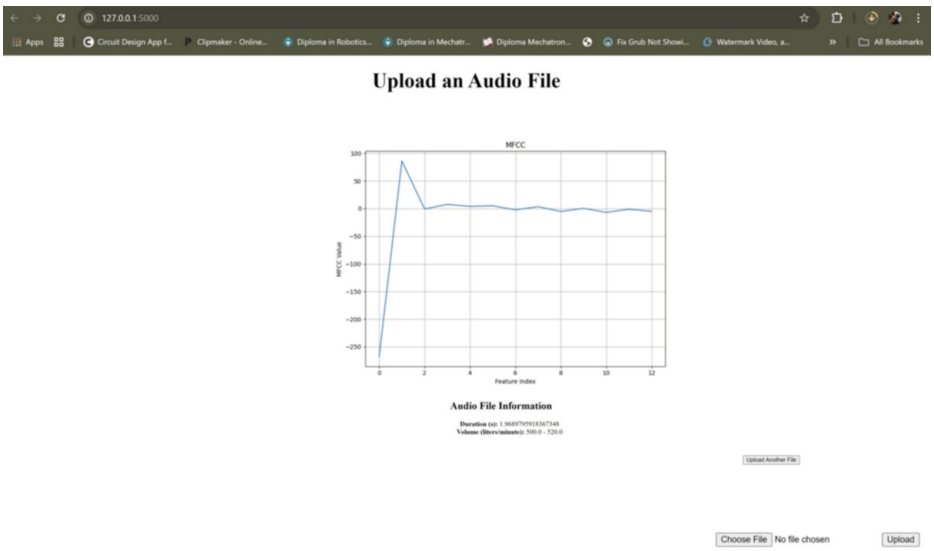

FIGURE 6: Result visualization

Figure shows the output from audio file processing and MFCC calculations. The study showed that the ability to estimate human exhaled air volume using MFCCs extracted from audio recordings. Preliminary assessments of the model with 87% accuracy revealed fairly satisfactory results in matching the quantitative values obtained from the real world with the predicted MFCC values and those detected by a PFM. The model also showed that it can generalize from one sample of different age groups, heights, and time of measurements. Moreover, this method eliminates or minimizes transmission of respiratory diseases, especially those that are heavily transmitted through contact such as through a shared respiratory devices, such as PFMs, and others that have not been properly decontaminated after use.

Discussion

Thus, this work has shown that it is possible to achieve fairly satisfactory results in forecasting the volume of exhaled air in liters per minute, using audio recordings of human exhalation as input data. Overall, the proposed tool, built on the premise of MFCC-based analysis, presents the following benefits: flexibility, simplicity in use, and lack of possible cross-transmissions of dangerous diseases due to the use of shared medical devices [15]. Through the use of audio signals, the system does not require physical contact, a factor widely helpful in the fight against respiratory ailments.

MFCC has been proven in previous research as a fruitful methodology to analyze and classify audio signals in the medical field; therefore, the study is in line with such research. Also, the architecture of a separate web application utilizing Flask increases the comprehensibility and application of the system. The extraction and prediction of the MFCC features in real-time have been researched and integrated into various audio-based systems, which goes a long way to show the technical possibility of implementing the approach.

While progress has been made, there are still some obstacles that have to be overcome. The model can also be influenced by external sounds, fluctuations in recording instruments, and actions made by the user while exhaling. Subsequent studies should be aimed at increasing the variability of the dataset and improving the characteristics of preprocessing to overcome these issues. In the same respect, undertaking further evaluation of parameter tuning, as well as the testing of the model with more significant datasets, enhances the model's generalizability.

Conclusions

The research work has shown that sound-based methods can measure human exhale air flow. The prototype system records exhale sounds and extracts key audio features (MFCC) to analyze them. Using these features, the Random Forest model successfully estimated exhale flow rate in liters per minute. The combination of audio processing (Librosa) and ML (scikit-learn) allowed fast development, proving the system could work for healthcare uses.

Next steps include improving the system's design and usability. The team is developing a mobile app (for Android and iOS) to process exhale sounds in real time, which will help gather more precise data and improve accuracy. More advanced AI models (CNNs and RNNs) will also be tested to detect subtle sound patterns for better predictions.

Currently, the prototype is in testing. The team is collecting more exhale sound samples from different

people to refine the model. Future tests will include diverse groups and environments, with user feedback guiding improvements. Research confirms that sound analysis is a reliable, non-invasive way to monitor exhale air flow for medical applications.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Manisha R. Mhetre, Piyush V. Kumbhakern, Rajeshwari B. Kul, Raj R. Kudtarkar, Viraj A. Jadhav

Acquisition, analysis, or interpretation of data: Manisha R. Mhetre, Piyush V. Kumbhakern, Rajeshwari B. Kul, Raj R. Kudtarkar, Viraj A. Jadhav

Drafting of the manuscript: Manisha R. Mhetre, Piyush V. Kumbhakern, Rajeshwari B. Kul, Raj R. Kudtarkar, Viraj A. Jadhav

Critical review of the manuscript for important intellectual content: Manisha R. Mhetre, Piyush V. Kumbhakern, Rajeshwari B. Kul, Raj R. Kudtarkar, Viraj A. Jadhav

Supervision: Manisha R. Mhetre, Piyush V. Kumbhakern

Disclosures

Human subjects: Consent was obtained or waived by all participants in this study. Vishwakarma Institute of Technology (VIT) issued approval VIT/RND/2025/3. Oral consent has been taken from the participants for the research project “Smartphone-Based Prediction of Human Exhale Air Peak Flow Using Exhale Air Audio” conducted by the author, for collecting exhaled air audio signals by smartphone. Prof. Manisha Rajesh Mhetre Instrumentation and Control Engineering, VIT, Pune Date: February 24. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Majeed SA, Husain H, Samad SA, Idbeaa TF: Mel Frequency Cepstral Coefficients (MFCC) feature extraction enhancement in the application of speech recognition: a comparison study. Journal of Theoretical and Applied Information Technology. 2015, 79:38-56.
2. Davis SB, Mermelstein P: Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics Speech and Signal Processing. 1980, 28:357-66. 10.1109/tassp.1980.1163420
3. Oppenheim AV, Schafer RW: Discrete-Time Signal Processing. Pearson Education Limited, 2014.
4. Gales MJF: Maximum likelihood linear transformations for HMM-based speech recognition. Computer Speech & Language. 1998, 12:75-98. 10.1006/csla.1998.0043
5. Joysingh SJ, Vijayalakshmi P, Nagarajan T: Significance of chirp MFCC as a feature in speech and audio applications. Computer Speech & Language. 2025, 89:101713. 10.1016/j.csl.2024.101713
6. Ravanelli M, Parcollet T, Bengio Y: The PyTorch-Kaldi Speech Recognition Toolkit. Electrical Engineering and Systems Science. 2019, 120:6465-69. 10.1109/ICASSP.2019.8683713
7. Liu X, Sahidullah M, Kinnunen T: A comparative re-assessment of feature extractors for deep speaker embeddings. Electrical Engineering and Systems Science. 2020, 10.21437/Interspeech.2020-1765
8. Hegde S, Achary KK, Shetty S: Feature selection using Fisher's ratio technique for automatic speech recognition. International Journal on Cybernetics & Informatics. 2015, 4:45-52. 10.5121/ijci.2015.4204
9. Albadr MAA, Tiun S, Ayob M, Mohammed M, AL-Dhief FT: Mel-Frequency Cepstral Coefficient features based on standard deviation and principal component analysis for language identification systems. Cognitive Computation. 2021, 13:1136-53. 10.1007/s12559-021-09914-w
10. Tyagi V, Wellekens C: On desensitizing the Mel-cepstrum to spurious spectral components for robust speech recognition. Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing. 2005, 1:I/529-I/532. 10.1109/ICASSP.2005.1415167
11. Trabelsi I, Ben Ayed D: On the use of different feature extraction methods for linear and non-linear kernels. Science of Electronics, Technologies of Information and Telecommunications. 2012, 797-802. 10.1109/SETIT.2012.6482016
12. Arias-Vergara T, Klumpp P, Vasquez-Correa JC, Noeth E, Orozco-Arroyave JR, Schuster M: Multi-channel spectrograms for speech processing applications using deep learning methods. Pattern Analysis and Applications. 2020, 24:423-31. 10.1007/s10044-020-00921-5

13. Arar ME, Sedef H: An efficient lung sound classification technique based on MFCC and high-dimensional model representation. Signal, Image and Video Processing. 2023, 17:4385-94. 10.1007/s11760-023-02672-2
14. Bahoura M, Ezzaidi H: Hardware implementation of MFCC feature extraction for respiratory sounds analysis. 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA). 2013, 226-29. 10.1109/WoSSPA.2013.6602366
15. Muthusamy PD, Sundaraj K, Manap NA: Computerized acoustical techniques for respiratory flow-sound analysis: a systematic review. Artificial Intelligence Review. 2020, 53:3501-74. 10.1007/s10462-019-09769-6