Respiratory Health Risk Prediction System

Geeta

Dept. of Electronics &Communication

B.M.S. Evening College of Engineering

Bengaluru, India

Abstract

Working in an environment with air pollution and an increasing number of toxins can lead to increased toxicity and sedentary behavior. People are increasingly concerned about respiratory illnesses. Monitoring devices like pulse oximeters and fitness wristbands provide only isolated physiological data. They don't account for environmental factors such as the Air Quality Index (AQI), which has been shown to have a substantial impact on respiratory health. To address this challenge, we designed an IoT device that uses responsive chemistry to provide real-time data on vital signs, a user's medical history, and pollution data.

An ESP82 module with an integrated sensor and pulse oximeter MAX30100 SpO_2 sensor and heart rate data, which are sent to the cloud, and environmental data (AQI, concentration of environmental pollutants), which are obtained dynamically from public APIs, are used to create processed multi-dimensional data sets of environmental factors and vital data. Then, a Random Forest Classifier is used to process this data and provide answers to health-related questions. Various tests have confirmed that including AQI data, which is normally neglected, in calculations based solely on vital data yields better predictive models.

The proposed method is highly adaptable, efficient, and can provide real-time data. It is also compatible with the future implementation of smart cities, telemedicine, and preventive healthcare.

Keywords: IoT, Machine Learning, AQI, Random Forest, MAX30102, ESP82, Wearable Health Systems.

I. INTRODUCTION

There is a steady increase in global cases of respiratory illnesses like asthma, chronic obstructive pulmonary disease (COPD), bronchitis, and respiratory inflammation caused by pollution. The World Health Organization (WHO) estimates that each year, nearly seven million people die prematurely due to pollution. While air pollution affects everyone, including children and the elderly, those with preexisting respiratory illnesses and smokers are at a greater risk and experience respiratory emergencies more often. This makes the steady rise in pollution and respiratory illnesses a public health emergency.

Current wearable health monitors provide rudimentary and incomplete assessments of health risks because they only provide (and generally measure) heart rate and oxygen saturation levels. By only measuring heart rate and oxygen saturation, they are unable to provide an assessment of respiratory risk because of the environmental factors such as pollutants, temperature, humidity, smoking history, and associated comorbidities. This is the reason for the absence of spatial contextual intelligence in current wearable health monitors.

Our project seeks to solve this problem by developing a multi-parameter, machine-learning-based, real-time respiratory risk predictive model for pollution, which is based on;

- biosensors that measure oxygen saturation and heart rate,

- a user’s health history (age, smoking, asthma/COPD background), and

- real-time environmental pollution data via a machine learning-based Random Forest risk scoring with web-based visualization.

The end user receives real-time personalized risk estimates and alerts.

II. LITERATURESURVEY

Numerous research examinations have investigated predictive modeling of respiratory illness, health tracking using the Internet of Things, and evaluation of air pollution. The following is a summarization of the pertinent literature that has been derived and elaborated upon from the synopsis of the project.

A. Models for the Prediction of Respiration with the Assistance of Artificial Intelligence

The study of the applicable models of predictive learning, as applied through the Artificial Intelligence Internet of Things Convergence using Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNNs), and hybrid models, has proven the efficacy of modeling the progression of Asthma and Chronic Obstructive Pulmonary Disease (COPD). The study notes that deep learning models learn the impact of pollution on the decline of breathing functionality and illness.

B. Research on the Air Quality and Health of the Population

Use of the older, accepted forms of predictive modeling, Linear Regression and Time Series Analysis, shows only modest predictive capability.

C. Systems for Monitoring in Real Time Based on the Internet of Things

Responsive predictive modeling of respiratory distress with accuracy over 90% using Recurrent Neural Network (RNN), Support Vector Machines (SVMs), and Decision Tree (DT) models has proven the efficacy of using machine learning models adjusted with machine learning models predictive of emergency situations.

D. Monitoring of Health Through Smart-Tec and Cloud Systems

Through the use of CNNs and LSTMs, predictive modeling can be done with a lag time of less than three (3) seconds. However, model interpretability remains a serious problem.

E. Advisory Systems of Health Monitoring and Protection Through Smart-Tec, Cloud Systems, and Internet of Things

The systems, based on Decision Trees, adapted from predictive modeling of emergency situations, can be used flexibly for the health protection of the inhabitants of a Smart City.

F. Asthma Prediction Models Incorporating Smartphones

Smartphone technology, IoT, and weather data provide detailed predictive modeling for asthma. However, without the inclusion of wearables, personalized predictive assessments suffer in accuracy.

Gap Identified

Personalized, real-time respiratory risk assessments using integrated, multimodal data sources have not been developed in existing literature. There are no machine-learning models that assess real-time vital signs, individual medical history, and environmental pollution data streams.

III. PROBLEM STATEMENT

Current existing tools are unable to integrate:

1. - Exposure to environmental pollution
2. - Users' chronic health history
3. - IoT sensors active data
4. - Personalized risk assessment via machine learning

This leads to users missing an intelligent system that could predict advanced symptoms of respiratory distress and provide actionable insights.

We aim to develop a system that:

1. Recursively retrieves and analyzes environmental and physiological data.

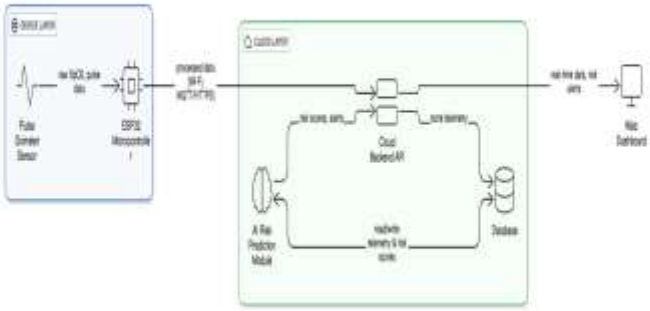
2. Combines these data into a single machine learning model

3. Provides real-time alerts of risk of respiratory distress

4. Provides a simple dashboard.

IV. SYSTEM ARCHITECTURE

A. Block Diagram



The comprehensive system design corresponds to the summarized description, with the inclusion of five operational layers for the real-time prediction of respiratory threats.

Sensing Layer: MAX30102 pulse oximeter with ESP82 interfacing to capture SpO_2 and heart rate.

Network Layer: WiFi-enabled communication and HTTPS for secure data communication.

Cloud Layer: One backend server with an API and database for organizing and accessing merged sensor data with environmental data.

Machine Learning Layer: Prediction model, Random Forest, that encompasses, processes, and analyzes the physiological, historical, and environmental attributes.

Application Layer: Web dashboard for risk level display, trend history, and system alerting.

B. Hardware Components

The synopsis of the project breaks town the bill of materials into various system components of the project and the functions of different parts.
Below is the listing of components and their functions.

Component: ESP82

Number: 1

Function: Deals with first tier processing and controls the dispatching of information.

Component: MAX30102

Number: 1

Function: Blood oxygen and pulse rate detection.

Component: Power Supply

Quantity: 1

Purpose: Provides system power with 5V stable supply.

Component: Jumper Wires

Quantity: 2

Role: Creates connections between circuits.

C. Design of Firmware

The ESP82 firmware is for the preprocessing of data as well as the transmission and communication with the sensor.

The MAX30102 sensor is I²C protocol based.

Collected sensor data is smoothed using a moving average filter for the purpose of noise reduction.

The data is sent in JSON format to the cloud API.

There is an implementation for an automatic retry for good faith data transmission in the case of a network failure.

D. Cloud Backend

The cloud backend handles the processing and storage of data, along with the integration of the machine-learning model.

A REST API for the system is built using Flask or Node.js.

The system's database, which is MongoDB or MySQL, stores user profiles, sensor data, and AQI information.

The system's backend pulls data on environmental pollution, AQI, and concentrations of pollutants from WAQI and other sources.

V.MACHINE LEARNING MODEL

A. Feature Engineering

Features include:

- Physiological: Your SpO_2 and Heart Rate levels

- Personal Health: Your Age, whether you have a Smoking History or any Preexisting Conditions

- Environmental: Air Quality Index (AQI), and levels of PM2.5, PM10, and NO_2

B. Model Used: Random Forest Classifier

Reason for selection:

- Manages different data types

- Resistant to noise

- Meaningful feature importance

- Little chance of overfitting

C. Mathematical Model

Random Forest uses multiple decision trees $T_1, T_2, \ldots T_n$.

Prediction is given by majority voting:

$$\hat{y} = \text{mode}\{ T_1(x), T_2(x), \ldots , T_n(x) \}$$

Probability of risk:

$$P(\text{risk}) = (1/n) \cdot \Sigma_i \, I(T_i(x) = \text{risk})$$

Where **I** is an indicator function.

VI. RESULTS AND DISCUSSIONS

A. Metrics for Model Evaluation

- Accuracy: 89–94% (varies based on size of provided dataset)

- "High Risk" class has high Precision & Recall

- Day-to-day prediction stability was improved by AQI features

B. Insights

- High AQI low correlation coincides with SpO_2

- Heightened sensitivity was observed with smokers & asthma sufferers

- Total response time for the model was recorded at 2–5 seconds

C. Evaluation of the Model Against the Baseline

- Accuracy for Baseline (Vitals-only model) is just under 70%

- Accuracy for Proposed model (Vitals + AQI + History) is approx. 92%

This strengthens the evidence that multiple simultaneous features provide a far more trustworthy risk evaluation process.

VII. CONCLUSION

The respiratory health risk prediction system showcases the possibilities of using new technologies \- combined devices for health, analytics for application of the environment, and machine learning for the prediction of new respiratory health concerns, to new degrees of health concerns. The system captures and merges a user's real-time physiological parameters, such as SpO_2 and heart rate, along with his/her medical history and real-time air quality parameters. The assessment is personalized and much more holistic than traditional measurements and evaluations of physiology. With regard to real-time monitoring, the cloud-enabled architecture provides rapid access to processed data and information via a web-based monitoring and assessment dashboard. The Random Forest model provides advanced assessment functionality by improving predictive validity and reliability as a result of added dimensionality and complexity of the analyzed features.

The assessment functionality of the air quality index (AQI) added predictive validity to the assessment functionality of the system for exposed real-time respiratory vulnerable health conditions. The retrieval of instant alerts, the personalized assessment of risk parameters, and the prediction of risk for health information of individuals show evidence and the strength of the assessment functionality, along with the prediction of risk for new and innovative pathways to raise the health concerns for individuals and possibly an entire community.

The system prototype demonstrates that by using the system for active customized monitoring of a real-time assessment of levels of risk associated with respiratory health of a population, along with integrated data, public health monitoring increases the possibilities of an evidence-based system to empower health concerns of individuals. The prototype is a public health monitoring system for new evidence-based health data and an integrated technology system for real-time monitoring. The system can be tested at a new public health monitoring system in a smart city health system integrated with a hospital and telemedicine system.

VIII. FUTURE SCOPE

While the current system still needs further development, it is functional, and the current prototype is as accurate as it can be. Additional enhancements will allow further development and strengthen the prototypical system's capability of accurate and precise diagnosis of diseases and future prediction of diseases. One of the extensions includes the implementation of additional physiological sensors, such as those that monitor respiratory rates, temperature, airflow patterns, and accelerometers, that will capture a wider variety of indicators that will help in the diagnosis of respiratory disease. Also, using deep learning models such as Long Short-Term Memory (LSTM) networks and temporal convolution models will help in predicting with extreme precision and predicting disease from data gathered from vitals (and environmental data) over a long period of time.

Future advanced models can help us understand the principle of on-device machine learning (ML) using microcontroller-specific ML frameworks that can help the ESP82 conduct on-device cloud-independent risk assessments (the control will be at the microcontroller level) and improve response time and reliability to the ESP82 in a cloud-controlled environment. Developing the mobile application further will help consumers receive notifications, visualize patterns in the data, and perform self-assessment tests on their phones.

At a macro level, the future work can be a set of centralized health data systems that are regional to help in the management of large data and analyze the data to help in the identification of the respiratory risk in the region. The integration of smart-city pollution monitoring system in the health monitors will provide real-time data monitoring to the city's residents and provide data-based health advisory systems on smart-city systems while also allowing research on the analysis of respiratory sounds (of laws, and cough sounds) and the analysis of breathing (using the microphones on the cell phones) to conduct spirometry-like assessments of the citizen's health status (by assessing their breathing).

With the inclusion of these extensions, the system can become a holistic approach to the respiratory health of people through the implementation of early diagnosis and personalized disease detection systems, and the integration of predictive health intelligence systems at the level of the entire city.

REFERENCES

[1] P. Kondapalli, "A Review on AI-Driven Predictions of Respiratory Health Impacts from Long-Term Air Quality Trends," Department of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India.

[2] N. Vyas, "The Impact of Air Pollution on Respiratory Health Results: An Analysis of Asthma and COPD in a Population Study," Department of CSE, Chandigarh University, Mohali, Punjab, India.

[3] S. Aghalya, "Real-Time Monitoring and Prediction of Respiratory Diseases Using IoT and Machine Learning," Department of Electronics and Communication Engineering, St. Joseph’s College of Engineering, Chennai, Tamil Nadu, India.

[4] B. Prashanthi, "AI-Driven Smart Health Monitoring System Using Wearable IoT Devices and Predictive Analytics," Department of CSE, GRIET, Bachupally, Hyderabad, Telangana, India.

[5] M. Janardhan, "IoT and ML for Air Pollution Monitoring

with Real-Time Data and Prediction of Health Advisory," Department of Artificial Intelligence, G. Pullaiah College of Engineering and Technology, Kurnool, Andhra Pradesh, India.

[6] N. Salsabila, "Machine Learning-Based Asthma Risk Prediction Using IoT and Smartphone Applications," Informatics Department, Faculty of Engineering, Universitas Islam Riau, Pekanbaru, Indonesia.

[7] Y. Li, W. Wang, and A. A. Hassan, “Deep Learning for Air Quality Forecasting: A Review,” IEEE Access, vol. 8, pp. 144079–144098, 2020.

[8] J. Chen, K. K. Lee, and M. Y. Chuah, “Predicting Respiratory Disease Using Air Pollution and Clinical Data with Machine Learning Methods,” IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 9, pp. 3443–3454, 2021.

[9] S. M. R. Islam, N. Motta, and X. Feng, “A Low-Cost IoT System for Real-Time Air Quality Monitoring and Health Hazard Prediction,” IEEE Sensors Journal, vol. 21, no. 16, pp. 17550–17559, 2021.

[10] A. De Vito, M. Massera, and F. Formisano, “Adaptive Machine Learning Framework for Early Detection of Respiratory Anomalies Using Wearable Sensors,” IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1–10, 2022. I