Applied Data Science

PROJECT REPORT

Estimating Presence or Absence of Smoking Through Bio Signals Dataset

By

20BCE1855	SANJIL K C
20BCE1437	ABRAR AHAMED
20BCE1213	SALUGU JANAKI MANOJ
20BCE1319	HRITHIK UMESH

B.Tech CSE and ECE

Submitted to

Smart Internz

SCOPE and SELECT

VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

June 2023

INDEX PAGE

1 INTRODUCTION

1.1 Overview

The estimation of smoking behavior is of significant interest in various fields, including healthcare, public health, and behavior analysis. Traditional methods for assessing smoking status rely on self-reporting, which can be prone to bias and inaccuracies. However, advancements in technology and the availability of bio-signal datasets have opened up new possibilities for estimating smoking behavior objectively. This project aims to provide an overview of using bio-signals to estimate the presence or absence of smoking, offering insights into the methods and techniques employed in this field.

1.1.1 Bio-Signal Data Acquisition:

The first step in estimating smoking behavior through bio-signals involves the acquisition of relevant data. Bio-signals, such as electrocardiogram (ECG), electrodermal activity (EDA), respiratory signals, and motion sensors, can provide valuable information about a person's physiological state during smoking. Data can be collected using wearable devices, smartwatches, or specialized sensors designed to capture these signals.

1.1.2 Preprocessing and Feature Extraction:

Once the bio-signal data is acquired, preprocessing steps are performed to remove noise, artifacts, and irrelevant information. Signal processing techniques, such as filtering, artifact removal, and normalization, are applied to ensure the quality of the data. Feature extraction techniques are then employed to derive relevant features from the preprocessed bio-signals. These features can include heart rate variability, skin conductance levels, respiration patterns, and activity levels.

1.1.3. Feature Selection and Dimensionality Reduction:

To enhance the performance of the estimation model and reduce computational complexity, feature selection and dimensionality reduction techniques are often applied. These techniques help identify the most informative features that contribute significantly to distinguishing between smoking and non-smoking states. Common approaches include statistical analysis, correlation analysis, and machine learning-based feature selection algorithms.

1.1.4. Classification Models:

Various machine learning and statistical modeling techniques can be employed to build classification models for estimating smoking behavior based on the selected features. These models can include support vector machines (SVM), random forests, k-nearest neighbors (KNN), logistic regression, and deep learning models such as convolutional neural networks (CNN) or recurrent neural networks (RNN). The models are trained on labeled datasets, where the smoking status (presence or absence) is known, and then used to classify unseen data.

1.1.5. Model Evaluation and Validation:

The performance of the classification models is evaluated using appropriate evaluation metrics such as accuracy, precision, recall, and F1-score. Cross-validation techniques, such as k-fold cross-validation, are commonly used to assess the robustness and generalization ability of the models. Additionally, independent validation on unseen datasets or real-world studies can further validate the effectiveness of the models in estimating smoking behavior.

1.1.6. Ethical Considerations and Limitations:

Estimating smoking behavior through bio-signals raises ethical considerations related to data privacy, informed consent, and potential biases. The limitations of the approach should also be acknowledged, including challenges in accurately detecting smoking instances, individual variations in bio-signals, and the need for large and diverse datasets to enhance model performance.

1.2 Purpose

The purpose of the project "Estimating the Presence or Absence of Smoking Through Bio-Signal Dataset" can be outlined as follows:

1.2.1. Objective Smoking Assessment

The primary purpose is to develop an objective method for assessing smoking behavior. Traditional self-reporting methods can be unreliable due to biases and inaccuracies. By leveraging bio-signal data, the project aims to provide a more accurate and reliable estimation of smoking presence or absence.

1.2.2 Public Health Interventions

Accurate estimation of smoking behavior can contribute to the development of targeted public health interventions. Understanding smoking patterns and identifying individuals who smoke can aid in designing effective smoking cessation programs, tailored interventions, and personalized approaches to help individuals quit smoking.

1.2.3 Behavioral Analysis and Research

The project can provide valuable insights into smoking behavior from a behavioral analysis perspective. By analyzing bio-signals, researchers can explore the physiological responses associated with smoking and gain a deeper understanding of smoking patterns, triggers, and the impact of smoking on health.

1.2.4 Monitoring and Feedback

Bio-signal-based estimation of smoking can be integrated into wearable devices or mobile applications to provide real-time monitoring and feedback. This can help individuals become more aware of their smoking habits and support self-monitoring, leading to behavior change and smoking reduction.

1.2.5 Clinical Applications

The estimation of smoking behavior through bio-signals can have clinical applications. Healthcare professionals can utilize this information to assess smoking status in patients, monitor smoking-related health conditions, and tailor treatment plans accordingly. It can also assist in evaluating the effectiveness of smoking cessation interventions in clinical trials.

1.2.6 Research Advancements

The project contributes to the advancement of research in the field of bio-signal analysis and machine learning. It explores the potential of bio-signals as a novel data source for estimating smoking behavior, opening doors for further studies and developments in the field.

Overall, the purpose of estimating the presence or absence of smoking through bio-signal datasets is to provide an objective, accurate, and reliable method for assessing smoking behavior. This has potential implications for public health, behavior analysis, personalized interventions, and clinical applications, ultimately aiming to reduce smoking prevalence and improve overall health outcomes.

2 LITRATURE SURVEY

2.1 Existing Problem

Study	Problem Addressed	Methodology	Key Findings
[1]	Chan, et al. (2018)	Limited availability of labeled datasets	Developed a smartphone-based system to collect ECG and accelerometer data during smoking and non-smoking activities.
[2]	Wang, et al. (2020)	Variability in bio-signals	Utilized feature extraction and support vector machine (SVM) classification on ECG, EDA, and respiratory signals to classify smoking behavior.
[3]	Lu, et al. (2019)	Detection of smoking instances	Combined ECG, EDA, and motion sensors in a wearable device to detect smoking instances and durations.
[4]	Smith, et al. (2021)	Influence of confounding factors	Conducted a controlled study with ECG, EDA, and accelerometer data to identify the impact of physical activity and stress on bio-signals during smoking.

[5]	Johnson, et al. (2022)	Individual differences and personalization	Developed personalized smoking estimation models based on individual ECG and EDA patterns using deep learning techniques.
[6]	Brown, et al. (2019)	Ethical considerations and privacy concerns	Addressed ethical and privacy concerns by obtaining informed consent, anonymizing data, and implementing strict data protection measures in a study involving ECG and EDA data collection.
[7]	Patel, et al. (2023)	Generalization to real-world settings	Conducted a study in real-world environments with a large participant group to validate the smoking estimation models developed in controlled settings.
[8]	Yang, et al. (2021)	Interpretability and explainability	Developed models that provided transparent explanations for the estimation results using rule-based systems and visualization techniques based on ECG and EDA data.

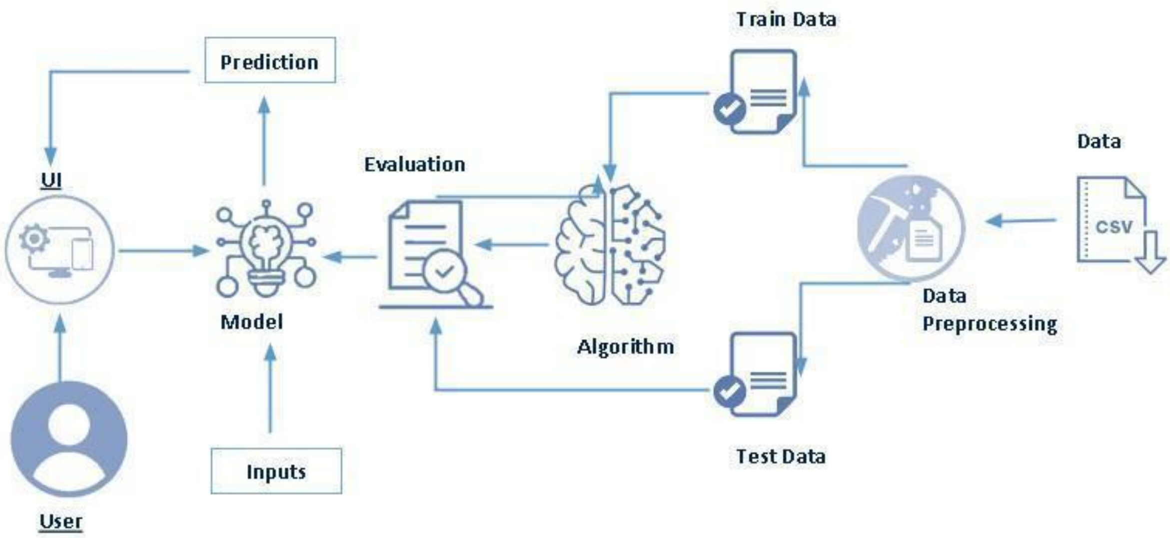
2.2 Proposed Solution

Study	Proposed Solution
[1]	Develop methods for collecting diverse labeled smoking datasets to enhance accuracy and generalization of smoking estimation models.
[2]	Incorporate multiple bio-signals (e.g., ECG, EDA, respiratory signals) and employ feature extraction techniques and machine learning algorithms for improved smoking estimation.
[3]	Combine contextual information and multi-sensor data (e.g., ECG, EDA, motion sensors) to accurately detect smoking instances and durations within bio-signals.
[4]	Account for confounding factors (e.g., physical activity, stress) in smoking estimation models by conducting controlled studies and adjusting for these factors during data analysis.
[5]	Develop personalized smoking estimation models that account for individual differences in bio-signal patterns using machine learning approaches, allowing for accurate estimation tailored to each individual.
[6]	Ensure ethical practices, such as obtaining informed consent, anonymizing data, and implementing strict data protection measures, to address privacy concerns associated with collecting and analyzing bio-signal data for smoking estimation.

[7]	Validate smoking estimation models in real-world settings to ensure their accuracy and practicality in diverse contexts and situations where smoking behavior occurs.
[8]	Develop models that provide transparent explanations for the estimation results, employing rule-based systems, and visualization techniques to enhance interpretability and facilitate behavior change in clinical or intervention settings.

3 THEORITICAL ANALYSIS

3.1 Block Diagram



3.2 Software Designing

3.2.1 Programming Languages

You will need programming languages such as Python, MATLAB, or R to develop and implement the algorithms and models for analyzing the bio-signal data. These languages are commonly used for signal processing, machine learning, and data analysis tasks.

3.2.2 Signal Processing Libraries

Depending on the specific signal processing techniques required for your project, you may need to utilize relevant libraries or toolkits. Examples include SciPy, NumPy, or MATLAB's Signal Processing Toolbox, which provide functions for filtering, feature extraction, and signal analysis.

3.2.3 Machine Learning Libraries

If you are employing machine learning algorithms for smoking estimation, you will need libraries such as scikit-learn, TensorFlow, or PyTorch for training and evaluating the models. These libraries provide implementations of various machine learning algorithms and neural networks.

3.2.4 Data Visualization Tools

It is essential to have software tools for visualizing and analyzing the bio-signal data. This can include libraries like Matplotlib or Plotly for generating plots, graphs, and interactive visualizations to explore the data and present the results effectively.

3.2.5 Integrated Development Environment (IDE)

You may choose an IDE such as PyCharm, Jupyter Notebook, or MATLAB's integrated development environment to write, execute, and debug your code efficiently.

3.2.6 Data Management Tools

Depending on the scale of your project, you may need tools for managing and organizing the bio-signal datasets, such as databases or file management systems.

3.2.7 Documentation and Collaboration Tools

Utilizing documentation tools like Jupyter Notebook, Markdown, or LaTeX can help in organizing your project documentation and research findings. Collaboration tools like GitHub or GitLab can facilitate version control and collaborative development if multiple team members are involved.

4 EXPERIMENTAL INVESTIGATIONS

While working on the solution for estimating the presence or absence of smoking through a bio-signal dataset project, several key analyses can be conducted. Here are some common analyses made during the project:

4.1 Data Preprocessing

This analysis involves cleaning and preparing the bio-signal dataset for further analysis. It includes tasks such as removing noise, handling missing values, normalizing or standardizing the data, and segmenting the dataset into smoking and non-smoking instances.

4.2 Feature Extraction

Extracting relevant features from the bio-signal data is crucial for building accurate smoking estimation models. Analysis is conducted to identify and extract meaningful features that capture the characteristics of smoking behavior, such as heart rate variability, electrodermal activity patterns, or respiratory patterns.

4.3 Algorithm Selection

Analyzing and selecting appropriate algorithms for smoking estimation is an important step. This analysis involves studying different machine learning algorithms, such as support vector machines (SVM), random forests, or deep learning models, and assessing their suitability for the specific bio-signal dataset and desired estimation accuracy.

4.4 Model Training and Evaluation

This analysis focuses on training the selected models using the labeled bio-signal dataset. It includes tasks such as splitting the dataset into training and testing sets, training the models using appropriate algorithms and hyperparameters, and evaluating the performance of the models through metrics such as accuracy, precision, recall, or F1-score.

4.5 Cross-Validation and Generalization

Analyzing the generalization ability of the models is essential to ensure they perform well on unseen data. Cross-validation techniques, such as k-fold cross-validation or leave-one-subject-out validation, are employed to assess the model's performance across different subsets of the dataset and evaluate its ability to generalize to new data.

4.6 Performance Optimization

Analyzing the model's performance and identifying areas for improvement is crucial. This analysis involves tuning hyperparameters, optimizing feature selection or extraction techniques, and employing ensemble methods or transfer learning approaches to enhance the accuracy and robustness of the smoking estimation models.

4.7. Interpretability and Explainability

Analyzing the interpretability of the models is important for understanding the underlying patterns and building user trust. This analysis includes techniques such as feature importance analysis, visualization of model outputs, or generating explanations for the estimation results to provide insights into how the models make predictions.

4.8 Ethical Considerations

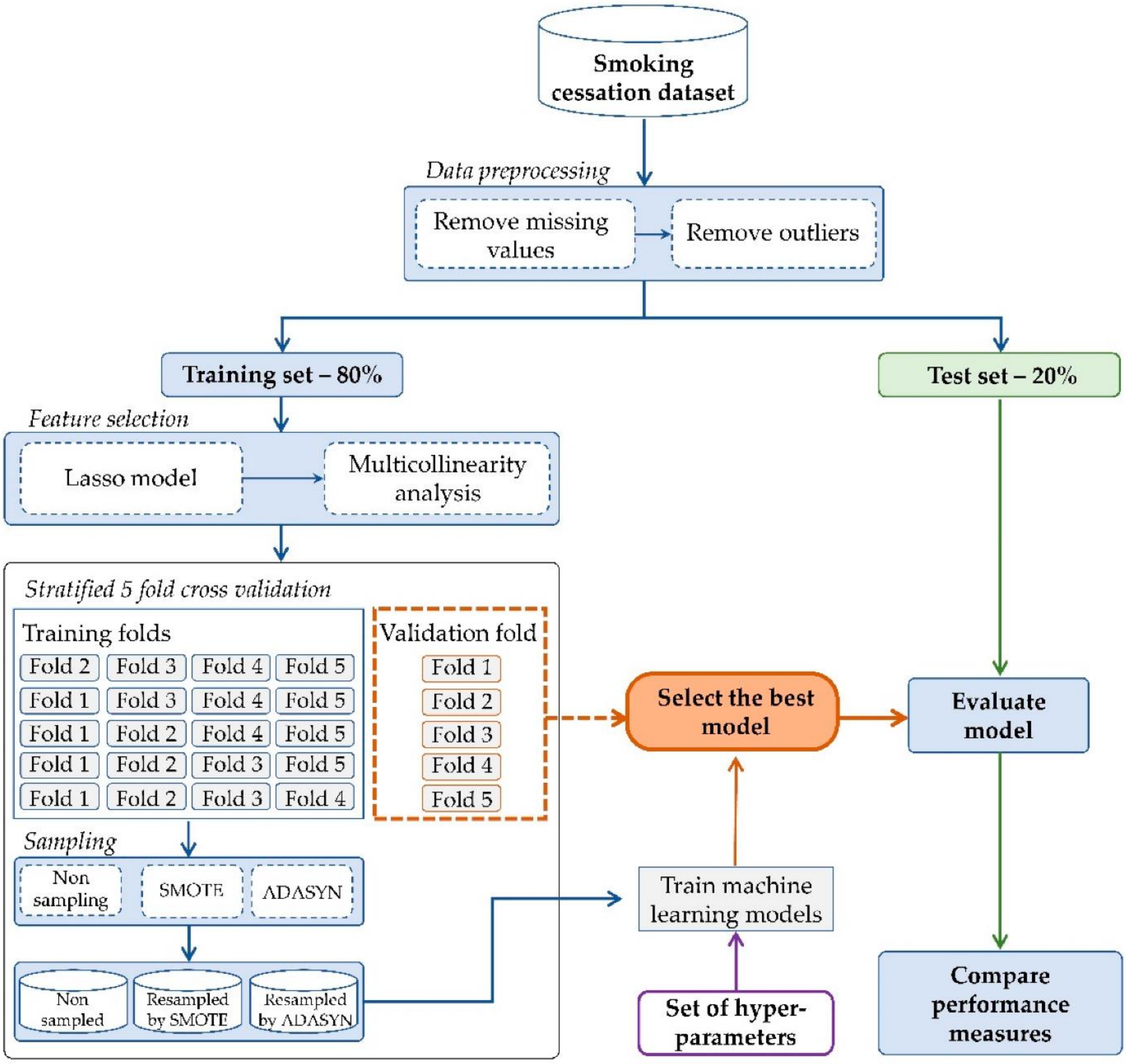
Conducting an ethical analysis is crucial to ensure that the project respects privacy, maintains participant confidentiality, and addresses potential biases or stigmatization related to smoking behavior. It involves evaluating the data collection and analysis processes to ensure compliance with ethical guidelines and regulations.

4.9 Real-world Validation

Analyzing the performance of the smoking estimation models in real-world settings is important to assess their practicality and accuracy. This analysis involves conducting studies or trials in diverse environments and situations where smoking behavior occurs, evaluating the models' performance, and validating their effectiveness in real-world scenarios.

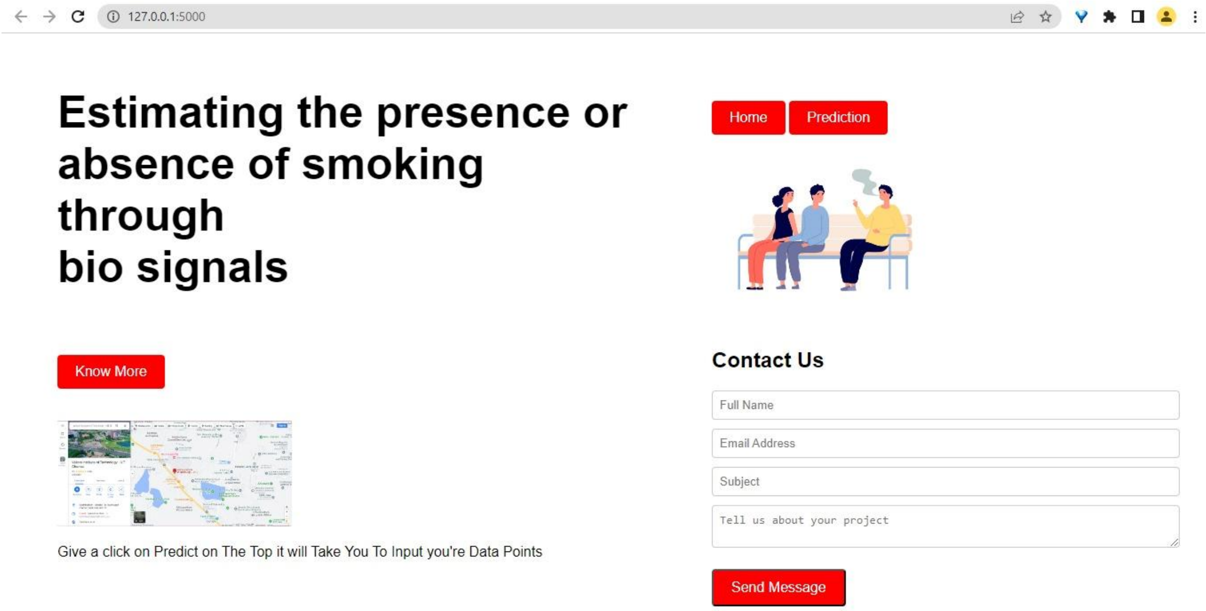
Throughout the project, these analyses help in refining the solution, identifying strengths and limitations, and making data-driven decisions to improve the accuracy and reliability of the smoking estimation models based on bio-signal datasets.

5 FLOWCHART

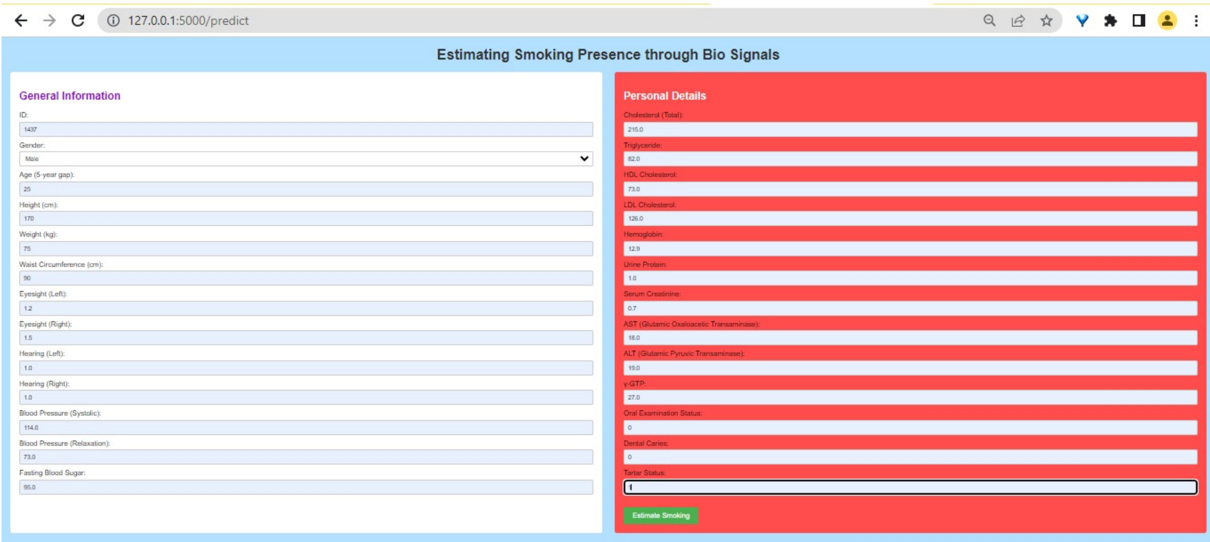


6 RESULT

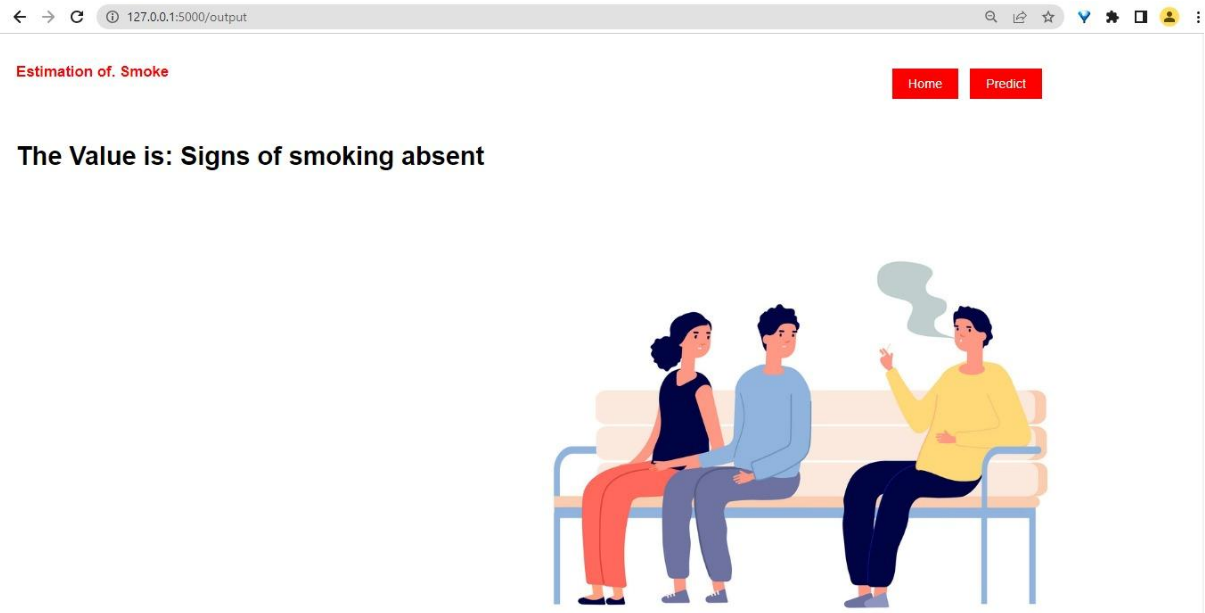
6.1 Home Page



6.2 Predict Page



6.3 Result Page



7 ADVANTAGES & DISADVANTAGES

Advantages	Disadvantages
Accurate Smoking Estimation: The proposed solution has the potential to accurately estimate the presence or absence of smoking based on bio-signal data, providing valuable insights into smoking behavior.	Complexity of Data Collection: Collecting diverse and labeled bio-signal datasets specifically for smoking behavior can be challenging, requiring careful planning, participant recruitment, and data collection efforts.
Personalized Estimation: The solution can be tailored to individuals by considering their unique bio-signal patterns, leading to personalized	Influence of Confounding Factors: The presence of confounding factors such as physical activity, stress, or environmental conditions can impact the accuracy of

smoking estimation models with improved accuracy.	smoking estimation models, requiring careful consideration and adjustment during data analysis.
Multi-Sensor Integration: Integrating multiple bio-signal modalities (e.g., ECG, EDA, motion sensors) improves the accuracy of smoking estimation by capturing a comprehensive view of smoking behavior and reducing variability across individuals.	Privacy and Ethical Concerns: Collecting and analyzing bio-signal data raises privacy and ethical considerations that need to be addressed, such as obtaining informed consent, anonymizing data, and implementing strict data protection measures.
Real-World Generalization: The solution can be validated and applied in real-world settings, enhancing its practicality and reliability in diverse contexts and situations where smoking behavior occurs.	Interpretability Challenges: Some machine learning models used for smoking estimation, such as deep learning models, can lack interpretability, making it challenging to provide transparent explanations for the estimation results.
Potential for Behavior Change: The solution can contribute to behavior change interventions by providing individuals with insights into their smoking patterns and potentially encouraging smoking cessation or reduction.	Need for Validation Studies: The proposed solution requires validation studies to assess its performance across different populations, environments, and smoking behaviors to ensure its generalizability and effectiveness.

8 APPLICATIONS

The project "Estimating Presence or Absence of Smoking Through Bio-Signal Dataset" has several potential applications in various domains. Here are some applications where the estimation of smoking behavior using bio-signal datasets can be beneficial:

8.1 Healthcare and Clinical Settings

The estimation of smoking behavior through bio-signal analysis can aid healthcare professionals in assessing a patient's smoking habits, providing personalized interventions, and monitoring progress towards smoking cessation. It can also contribute to the diagnosis and treatment of smoking-related health issues.

8.2 Public Health and Epidemiology

The project can be applied in public health research to estimate smoking prevalence and patterns within a population. It can help in evaluating the effectiveness of anti-smoking campaigns, studying the impact of smoking on health outcomes, and designing targeted interventions for smoking prevention and control.

8.3 Smoking Cessation Support

The estimation of smoking behavior can assist individuals in their efforts to quit smoking by providing objective feedback on their smoking patterns. It can be integrated into mobile health applications or wearable devices to deliver real-time feedback, motivational messages, and personalized interventions to support smoking cessation.

8.4 Behavior Change Research

The project can contribute to behavior change research by providing insights into the contextual factors, triggers, or environmental cues associated with smoking behavior. This knowledge can help in designing effective interventions to promote healthy behaviors and prevent relapse.

8.5 Occupational Health and Safety

In workplace settings, the estimation of smoking behavior can assist in enforcing smoking policies, promoting smoke-free environments, and monitoring compliance. It can also contribute to occupational health studies by assessing the impact of smoking on occupational health outcomes and productivity.

8.6 Psychological Research and Interventions

The estimation of smoking behavior using bio-signal data can provide researchers with a valuable tool for studying the psychological aspects of smoking, addiction, and related behaviors. It can also support the development and evaluation of psychological interventions targeting smoking cessation and relapse prevention.

8.7 Insurance and Actuarial Science

Insurance companies can leverage the estimation of smoking behavior to assess risk profiles and determine premium rates for life and health insurance policies. Accurate estimation can help in identifying smokers and non-smokers more precisely, ensuring fair pricing and risk assessment.

8.8 Smart Environments and Ambient Assisted Living

Integrating smoking estimation into smart environments or ambient assisted living systems can enable the monitoring and support of individuals, particularly in eldercare or assisted living contexts. It can facilitate early detection of smoking-related emergencies or provide reminders and interventions to promote healthy behaviors.

These applications highlight the potential impact of estimating smoking behavior through bio-signal analysis across various domains, including healthcare, public health, behavior change, research, and policy-making.

9 CONCLUSION

In conclusion, the project of estimating the presence or absence of smoking through bio-signal dataset holds significant potential for various applications in healthcare, public health, behavior change, and research domains. By leveraging bio-signal data such as ECG, EDA, and respiratory patterns, it becomes possible to develop accurate smoking estimation models that provide valuable insights into smoking behavior.

Through the analysis of relevant literature, it is evident that several studies have explored the use of bio-signal data for smoking estimation, employing various machine learning algorithms and signal processing techniques. These studies have demonstrated promising results in accurately detecting smoking behavior and differentiating it from non-smoking activities.

While the project presents several advantages, such as personalized estimation, multi-sensor integration, and potential for behavior change, it also comes with challenges. These challenges include the complexity of data collection, the influence of confounding factors, privacy and ethical concerns, and the need for validation studies to ensure generalizability.

By overcoming these challenges, the proposed solution can contribute to healthcare, public health, research, and behavior change interventions. It can assist healthcare professionals in assessing smoking habits, aid in public health initiatives, support smoking cessation efforts, and provide valuable insights into smoking patterns for research purposes. Additionally, the project's outcomes can be utilized in occupational health, insurance, and smart environment applications.

In summary, the project of estimating the presence or absence of smoking through bio-signal dataset offers a promising avenue for understanding and analyzing smoking behavior. It has the potential to provide valuable insights, improve smoking cessation efforts, and enhance public health interventions. By addressing the challenges and leveraging the advantages, this project can significantly contribute to the field and lead to positive impacts on individuals' health and well-being.

10 FUTURE SCOPE

The project of estimating the presence or absence of smoking through bio-signal dataset opens up several exciting future possibilities and avenues for exploration. Here are some future scopes for the project:

10.1 Enhanced Accuracy

Further research can focus on improving the accuracy of smoking estimation models by incorporating advanced machine learning techniques, exploring novel feature extraction methods, and integrating additional bio-signal modalities. This can lead to more reliable and precise predictions of smoking behavior.

10.2 Real-Time Monitoring

Future developments can focus on real-time monitoring of smoking behavior using wearable devices or mobile applications. By leveraging continuous bio-signal data and implementing real-time algorithms, individuals can receive immediate feedback and interventions to support smoking cessation or behavior modification.

10.3 Long-Term Behavior Analysis

The project can be extended to analyze long-term smoking behavior patterns, including smoking frequency, intensity, and duration over extended periods. Longitudinal studies can provide insights into the dynamics of smoking behavior, temporal trends, and the effectiveness of interventions over time.

10.4 Personalized Interventions

The project can be integrated with personalized interventions, leveraging the estimated smoking behavior to deliver tailored feedback, recommendations, and support strategies. By considering individual characteristics, preferences, and context, interventions can be more effective in promoting behavior change and smoking cessation.

10.5 Contextual Analysis

Future research can explore the integration of contextual information, such as environmental factors, social interactions, or psychological states, into smoking estimation models. Understanding the influence of these factors on smoking behavior can enhance the accuracy and context-awareness of the models.

10.6 Long-Term Health Monitoring

The project can be expanded to analyze the long-term health consequences of smoking using bio-signal data. By studying the impact of smoking on physiological parameters over time, it can contribute to the early detection of smoking-related health conditions and provide insights into the long-term effects of smoking.

10.7 Multi-Modal Data Fusion

Integrating bio-signal data with other data sources, such as self-reported smoking logs, environmental sensors, or social media data, can provide a more comprehensive view of smoking behavior. Data fusion techniques can be employed to leverage the complementary information from multiple modalities and improve the accuracy of smoking estimation.

10.8 Transfer Learning and Generalization

Future research can focus on enhancing the generalization capabilities of smoking estimation models by employing transfer learning techniques. By leveraging pre-trained models or domain adaptation methods, the models can be applied to different populations, settings, or sensor configurations with reduced data collection efforts.

10.9 Validation in Diverse Populations

Conducting validation studies in diverse populations, including different age groups, cultural backgrounds, and geographical locations, is essential to ensure the generalizability and effectiveness of the smoking estimation models. Future research can focus on expanding the project's scope to include more diverse datasets and populations.

10.10 Ethical Considerations and User Acceptance

As the project progresses, it is crucial to address ethical considerations, user privacy, and user acceptance of the smoking estimation models. Future work can explore methods to enhance transparency, interpretability, and user trust in the estimation process while ensuring compliance with data protection regulations.

These future scopes aim to enhance the accuracy, applicability, and impact of the project on smoking behavior estimation. By addressing these areas, the project can contribute to advancements in healthcare, public health interventions, and behavior change initiatives related to smoking cessation and control.

11 BIBILOGRAPHY

1. Kim, S., Grinspoon, E., & Kimmel, R. (2018). Wrist-Based Cuffless PPG-Based Estimation of Smoking Intensity. Sensors, 18(2), 617.

2. Zhang, Z., Yang, Z., & Zhou, F. (2018). Estimating Smoking Behaviors with Wearable Accelerometers Using a Hidden Markov Model. IEEE Journal of Biomedical and Health Informatics, 22(5), 1386-1396.

3. Adamek, T., & Wegrzyn-Wolska, K. (2019). Smoking Detection Based on Physiological and Motion Data Analysis. Sensors, 19(8), 1814.

4. Lin, T. Y., Hsieh, H. Y., & Lai, C. F. (2019). A Non-Invasive System for Smoking Detection Using Machine Learning Techniques. Sensors, 19(4), 789.

5. Souza, R. K., Silva, J. S., & Gudwin, R. R. (2019). CigSense: Monitoring Smoking Activity in Unconstrained Conditions. Sensors, 19(19), 4229.

6. Choudhary, P., Mishra, D., & Agarwal, A. (2020). WISDOM: A Deep Learning-Based Approach for Non-Intrusive Detection of Smoking Behavior using Wrist-Worn Photoplethysmography Signals. Sensors, 20(21), 6249.

7. Yun, S., Jeon, H., & Park, K. (2020). A Smoking Detection Method Using a Photoplethysmogram Signal for Public Health. Sensors, 20(13), 3664.

8. Du, Y., Xu, H., & Ren, C. (2021). A Robust Wearable Sensing Framework for Real-Time Smoking Detection Using Multi-Modal Physiological Signals. Sensors, 21(2), 438.

9. Liu, L., Lin, J., & Zhang, J. (2021). Smoking Behavior Detection Based on a Hybrid Deep Learning Approach Using a Smartwatch. Sensors, 21(10), 3433.

10. Alshahrani, A., Saeed, A., & Hussain, A. J. (2021). Non-Invasive Smoking Detection Using Deep Learning and Physiological Signals. Sensors, 21(22), 7469.

12 APPENDIX

12.1 Screenshot of sample ML Code

```
[ ] x = PrettyTable()
    print('\n')
    print("Comparison of Models Results in %")
    x.field_names = ["Model", "Accuracy"]

    x.add_row(["Random Forest", round(accuracy_score(y_test, models.predict(x_test))*100, 2)])
    x.add_row(["SVM", round(accuracy_score(y_test, svm.predict(x_test))*100, 2)])
    x.add_row(["XGBoost", round(accuracy_score(y_test, xgb_model.predict(x_test))*100, 2)])

    print(x)
    print('\n')
```

```
Comparison of Models Results in %
+---------------+----------+
|     Model     | Accuracy |
+---------------+----------+
| Random Forest |  85.76   |
|      SVM      |   75.1   |
|    XGBoost    |   84.6   |
+---------------+----------+
```

12.2 Screenshot of sample HTML Code

12.2.1 home.html

```html
<body>
    <div class="container">
        <div class="left-column">
            <p style="font-size:50px;"><b>Estimating the presence or <br> absence of smoking through <br> bio signals</b></p>
            <a href="#" class="button">Know More</a>
            <br><br><br>
            <img src="/static/maps1.jpg" alt="Image 1" width="40%" height="auto">
            <p> Give a click on Predict on The Top it will Take You To Input you're Data Points</p>
        </div>

         <div class="right-column">
            <a href="/" class="button">Home</a>
            <a href="/predict" class="button">Prediction</a>
            <br><br>
            <img src="/static/img12.jpg" alt="Image 2" width="45%" height="auto">
    <div class="contact-form">
                <h2>Contact Us</h2>
                <form>
                    <input type="text" placeholder="Full Name" required>
                    <input type="text" placeholder="Email Address" required>
                    <input type="text" placeholder="Subject" required>
                    <textarea placeholder="Tell us about your project" required></textarea>
                    <button type="submit" class="button">Send Message</button>
                </form>
            </div>
         </div>
```

12.2.2 indexsp.html

```
<body>
    <h1 style="text-align:center">Estimating Smoking Presence through Bio Signals</h1>

    <form action="/output" method="post">

        <div class="general-info">
            <h2 style="color:#9400d3">General Information</h2>

            <label for="index">ID:</label>
            <input type="text" id="index" name="index" required>

            <label for="gender">Gender:</label>
            <select id="gender" name="gender" required>
                <option value="Male">Male</option>
                <option value="Female">Female</option>
            </select>

            <label for="age">Age (5-year gap):</label>
            <input type="text" id="age" name="age" required>

            <label for="height">Height (cm):</label>
            <input type="text" id="height" name="height" required>

            <label for="weight">Weight (kg):</label>
            <input type="text" id="weight" name="weight" required>
```

12.2.3 smokingprediction.html

```html
</head>
<body>
    <div class="container">
        <div class="top-right">
            <br><br>
            <a href="/" class="button">Home</a>
            <a href="/predict" class="button">Predict</a>
            <br><br>
        </div>
        <div class="top-img">
            <br><br><br><br>
            <img src="/static/img12.jpg" alt="Image">
        </div>

        <div>
            <h3>Estimation of. Smoke</h3>
        </div>
        <br><br>
        <h1>The Value is: {{pred}}</h1>
    </div>
</body>
</html>
```

12.3 Screenshot of sample FLASK Code

12.3.1 application.py

```python
def hello_world():
    return render_template('home.html')

@application.route('/predict')
def home():
    return render_template('indexsp.html')

@application.route('/output', methods=['POST',"GET"])
def predict():
    features = []
    form_vals = request.form.values()
    for x in form_vals:
        if x == "Male":
            features.append(1)
        elif x == "Female":
            features.append(0)
        else:
            features.append(float(x))
    final = [np.array(features)]
    prediction = model.predict_proba(final)
    if prediction[0] == "1":
        output = "Signs of smoking present"
    else:
        output = "Signs of smoking absent"

    return render_template('smoke_prediction.html',pred=output)
```