ESTIMATING PRESENCE OR ABSENCE OF SMOKING THROUGH BIO SIGNALS

AN INDUSTRY ORIENTED MINI REPORT

Submitted to

JAWAHARLAL NEHRU TECNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING(CSD)

Submitted By

GONE SRESHTA	**21UK1A6707**
KUSA NAVYA	**21UK1A6724**
CH PRUDHVI NATH REDDY	**21UK1A6719**
ANUMULA DEEXITH	**21UK1A6744**

Under the guidance of

N Rajesh

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

Affiliated to JNTUH, HYDERABAD

BOLLIKUNTA, WARANGAL (T.S) – 506005

DEPARTMENT OF

COMPUTER SCIENCE AND ENGINEERING (Data Science)

VAAGDEVI ENGINEERING COLLEGE(WARANGAL)

CERTIFICATE OF COMPLETION

INDUSTRY ORIENTED MINI PROJECT

This is to certify that the UG Project Phase-1 entitled " ESTIMATING PRESENCE OR ABSENCE OF SMOKING THROUGH BIO SIGNALS" is being submitted by GONE.SRESHTA (21UK1A6707), KUSA.NAVYA (21UK1A6724), CH.PRUDHVINATHREDDY (21UK1A6719), ANUMULA.DEEXITH(21UK1A6744) in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering to Jawaharlal Nehru Technological University Hyderabad during the academic year 2023-2024.

Project Guide	**HOD**
N Rajesh	**Dr. K. Sharmila Reddy**
(Assistant Professor)	(Professor)

External

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr. SYED MUSTHAK AHMED ,**Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this UG Project Phase-1 in the institute.

We extend our heartfelt thanks to **Dr.K.SHARMILA**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Project Phase-1.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited, for their constant supervision as well as for providing necessary information regarding the UG Project Phase-1 and for their support in completing the UG Project Phase-1. We express heartfelt thanks to the guide, **N.Rajesh,** Assistant professor, Department of CSE for his constant support and giving necessary guidance for completion of this UG Project Phase-1

Finally, we express our sincere thanks and gratitude to my family members, friends for their encouragement and outpouring their knowledge and experience throughout the project building

ABSTRACT

The detection of smoking behaviour through bio-signals has garnered significant attention due to its implications for public health monitoring and personalized healthcare. This paper provides a comprehensive review of existing research methodologies and technologies aimed at estimating the presence or absence of smoking using bio-signals.

Various bio-signals such as respiratory patterns, heart rate variability, skin conductance, and biochemical markers have been explored as indicators of smoking activity

Recent advancements in wearable sensor technologies and machine learning algorithms have enabled real-time and non-invasive monitoring of these bio-signals, enhancing the accuracy and reliability of smoking detection systems.

TABLE OF CONTENTS:-

1.INTRODUCTION

1.1 OVERVIEW

The main aim or goal of a machine learning project estimating the presence or absence of smoking through bio-signals is to develop an accurate and reliable system that can detect smoking behaviour using physiological signals.

The system would use machine learning algorithms to analyse bio-signals, such as heart rate, skin conductance, and respiratory rate, to determine if a person is smoking or not.

The system could be applied in various contexts, such as clinical settings or workplaces, to monitor smoking behaviour and help individuals quit smoking.

1.2.PURPOSE

The purpose of estimating the presence or absence of smoking through bio-signals serves several important objectives:

1. **Public Health Monitoring:** By accurately detecting smoking behaviour through bio-signals ,public health authorities can better understand patterns of smoking prevalence and trends. This data can inform targeted interventions and policies aimed at reducing smoking rates and improving public health outcomes.
2. **Personalized Healthcare:** For individuals, bio-signal-based smoking detection can provide real-time feedback on smoking habits. This can facilitate personalized healthcare interventions, such as smoking cessation programs tailored to an individual's specific smoking behaviors and triggers.
3. **Behavioural Research:** Bio-signal data related to smoking behaviour can be used in behavioural research to study factors influencing smoking initiation, cessation, and relapse. This can contribute to the development of more effective behavioural interventions and smoking cessation strategies.
4. **Validation of Self-Reported Data:** Bio-signal-based smoking detection can provide an objective measure of smoking behaviour, complementing self-reported data which may be subject to biases such as underreporting or social desirability bias.
5. **Early Detection of Health Risks:** Monitoring bio-signals associated with smoking can aid in the early detection of associated health risks, such as respiratory diseases or cardiovascular problems, allowing for timely medical intervention and management.

6. **Technological Advancements:** Research in this area drives advancements in wearable sensor technologies and machine learning algorithms, which can have broader applications beyond smoking detection, such as in remote patient monitoring and chronic disease management.

In summary, the purpose of estimating smoking presence or absence through bio-signals is to improve public health outcomes, support personalized healthcare, advance research in behavioural sciences, validate self-reported data, detect health risks early, and drive technological innovation in healthcare monitoring.

2.LITERATURE SURVEY

2.1 EXISTING PROBLEM

Several challenges exist when it comes to estimating the presence or absence of smoking through bio-signals:

1. **Variability in Bio-signals**: Bio-signals such as respiratory patterns, heart rate variability, and skin conductance can vary widely between individuals and even within the same individual under different conditions. This variability makes it challenging to establish universal thresholds or patterns that reliably indicate smoking behaviour.
2. **Ambiguity in Bio-signal Interpretation**: While certain bio-signals may change in response to smoking (e.g., increased heart rate, altered respiratory patterns), these changes can also be influenced by other factors such as physical activity, stress, or medical conditions. Distinguishing smoking-related changes from other causes of bio-signal variability is complex and requires sophisticated signal processing and interpretation techniques.
3. **Sensor Accuracy and Reliability**: The accuracy and reliability of wearable sensors used to capture bio-signals are crucial for robust smoking detection systems. Issues such as sensor drift, noise interference, and calibration errors can affect the quality of bio-signal data, leading to inaccurate estimations of smoking presence or absence.
4. **Ethical and Privacy Concerns**: Monitoring bio-signals for smoking detection raises ethical considerations regarding privacy, consent, and data security. Users may be concerned about the collection, storage, and potential misuse of sensitive health-related information obtained through bio-signal monitoring.
5. **Validation Against Gold Standards**: Establishing the validity of bio-signal-based smoking detection methods requires comparison against gold standard methods such as direct observation or biochemical testing (e.g., cotinine levels). Differences in sensitivity,

specificity, and reliability between bio-signal-based methods and established standards pose challenges for validation and real-world application.

6. **Individual Variability in Smoking Patterns**: Smoking behaviours can vary widely among individuals in terms of frequency, intensity, and smoking techniques (e.g., passive smoking, e-cigarettes). Bio-signal-based detection methods must account for these variations to accurately estimate smoking presence or absence across diverse populations.
7. **Real-World Implementation**: Translating bio-signal-based smoking detection from controlled research settings to real-world environments introduces additional challenges related to user compliance, system robustness, and environmental factors (e.g., ambient noise, movement).

2.2 PROPOSED SOLLUTION

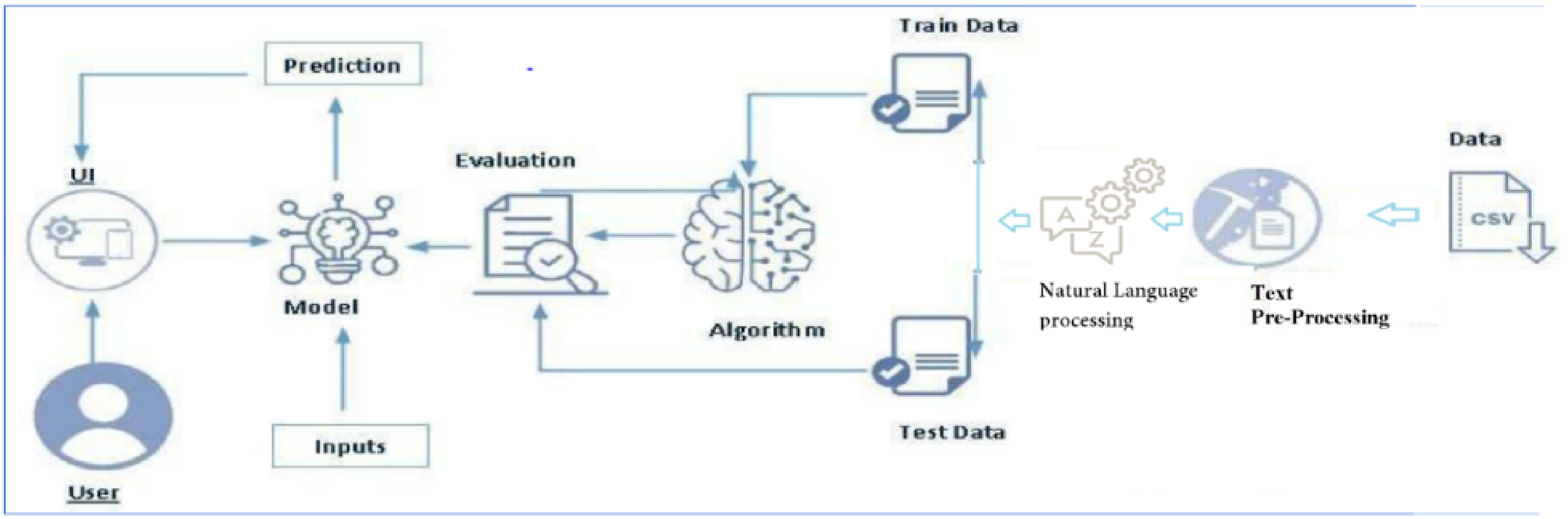
Proposing solutions for estimating the presence or absence of smoking through biosignals involves addressing several key challenges:

1. **Multimodal Bio-signal Integration**: Instead of relying on a single bio-signal, integrating multiple bio-signals (e.g., respiratory patterns, heart rate variability, skin conductance) can enhance the accuracy and reliability of smoking detection systems. Combining different modalities can mitigate the variability inherent in individual bio-signals and improve the robustness of smoking behaviour estimation.
2. **Advanced Signal Processing Techniques**: Developing and applying advanced signal processing algorithms tailored to bio-signal data from smoking-related behaviours is crucial. Techniques such as machine learning (e.g., deep learning, ensemble methods) can extract meaningful patterns and features from bio-signals, improving the discrimination between smoking and non-smoking states.
3. **Sensor Technology Advancements**: Investing in the development of high-quality, wearable sensors with improved accuracy, reliability, and real-time data transmission capabilities is essential. Sensors should be capable of capturing bio-signals accurately under various environmental conditions and user activities, minimizing noise and interference.

4. **Validation Against Gold Standards**: Conducting rigorous validation studies comparing bio-signal-based smoking detection methods against established gold standards (e.g., biochemical markers, self-reports) is critical. Validation ensures the accuracy, sensitivity, specificity, and reliability of bio-signal-based systems before deployment in real-world settings.
5. **Personalized Models and Adaptation**: Recognizing individual variability in smoking behaviours, personalized models based on longitudinal data and adaptive algorithms can improve the accuracy of smoking detection. These models can account for changes in bio-signals over time and adapt to individual smoking patterns and habits.
6. **Privacy and Ethical Considerations**: Addressing privacy concerns through transparent data handling practices, informed consent processes, and secure data storage and transmission protocols is essential. Upholding ethical guidelines ensures that bio-signal-based smoking detection respects user autonomy and confidentiality.
7. **Real-World Implementation and User Engagement**: Designing smoking detection systems that are user-friendly, non-intrusive, and integrate seamlessly into daily routines encourages user compliance and engagement. Considering real-world factors such as user acceptance, system usability, and environmental adaptability enhances the practical applicability of bio-signal-based smoking detection solutions.
8. **Collaborative Research and Stakeholder Engagement**: Foster collaboration among researchers, healthcare providers, policymakers, and technology developers to address interdisciplinary challenges and ensure the successful implementation of bio-signal-based smoking detection solutions. Engaging stakeholders early in the development process facilitates the adoption and scalability of innovative technologies in public health and clinical settings.

3.THEORITICAL ANALYSIS

3.1. BLOCK DIAGRAM



3.2. S

OFTWARE DESIGNING

The following is the Software required to complete this project:

- **Google-Colab**: Google-Colab will serve as the development and execution environment for your predictive modeling, data preprocessing, and model training tasks. It provides a cloud-based Jupyter Notebook environment with access to Python libraries and hardware acceleration.

- **Dataset (CSV File)**: The dataset in CSV format is essential for training and testing your predictive model. It should include historical air quality data, weather information, pollutant levels, and other relevant features.

- **Data Preprocessing Tools**: Python libraries like NumPy, Pandas, and Scikit-learn will be used to preprocess the dataset. This includes handling missing data, feature scaling, and data cleaning.

- **Feature Selection/Drop**: Feature selection or dropping unnecessary features from the dataset can be done using Scikit-learn or custom Python code to enhance the model's efficiency.

- **Model Training Tools**: Machine learning libraries such as Scikit-learn, TensorFlow, or PyTorch will be used to develop, train, and fine-tune the predictive model. Regression or classification models can be considered, depending on the nature of the AQI prediction task.

- **Model Accuracy Evaluation**: After model training, accuracy and performance evaluation tools, such as Scikit-learn metrics or custom validation scripts, will assess the model's predictive capabilities. You'll measure the model's ability to predict AQI categories based on historical data.

- **UI Based on Flask Environment**: Flask, a Python web framework, will be used to develop the user interface (UI) for the system. The Flask application will provide a user-friendly platform for users to input location data or view AQI predictions, health information, and recommended precautions.

- Google Colab will be the central hub for model development and training, while Flask will facilitate user interaction and data presentation. The dataset, along with data preprocessing, will ensure the quality of the training data, and feature selection will optimize the model. Finally, model accuracy evaluation will confirm the system's predictive capabilities, allowing users to rely on the AQI predictions and associated health information.

4.EXPERIMENTAL INVESTIGATION

In this project, we used Body Signal Of Smoking Dataset. This dataset is a csv file consisting of labelled data and having the following columns-

ID : index

gender

age : 5-years gap

height(cm)

weight(kg)

waist(cm) : Waist circumference length

eyesight(left)

eyesight(right)

hearing(left)

hearing(right)

systolic : Blood pressure

relaxation : Blood pressure

fasting blood sugar

Cholesterol : total

triglyceride

HDL : cholesterol type

LDL : cholesterol type

hemoglobin

Urine protein

serum creatinine

AST : glutamic oxaloacetic transaminase type

ALT : glutamic oxaloacetic transaminase type

Gtp : γ-GTP

oral : Oral Examination status

dental caries

tartar : tartar status

smoking

For the dataset we selected, it consists of more than the columns we want to predict it . So, we have chosen the feature drop it contains the columns that we are going to predict the AQI value.

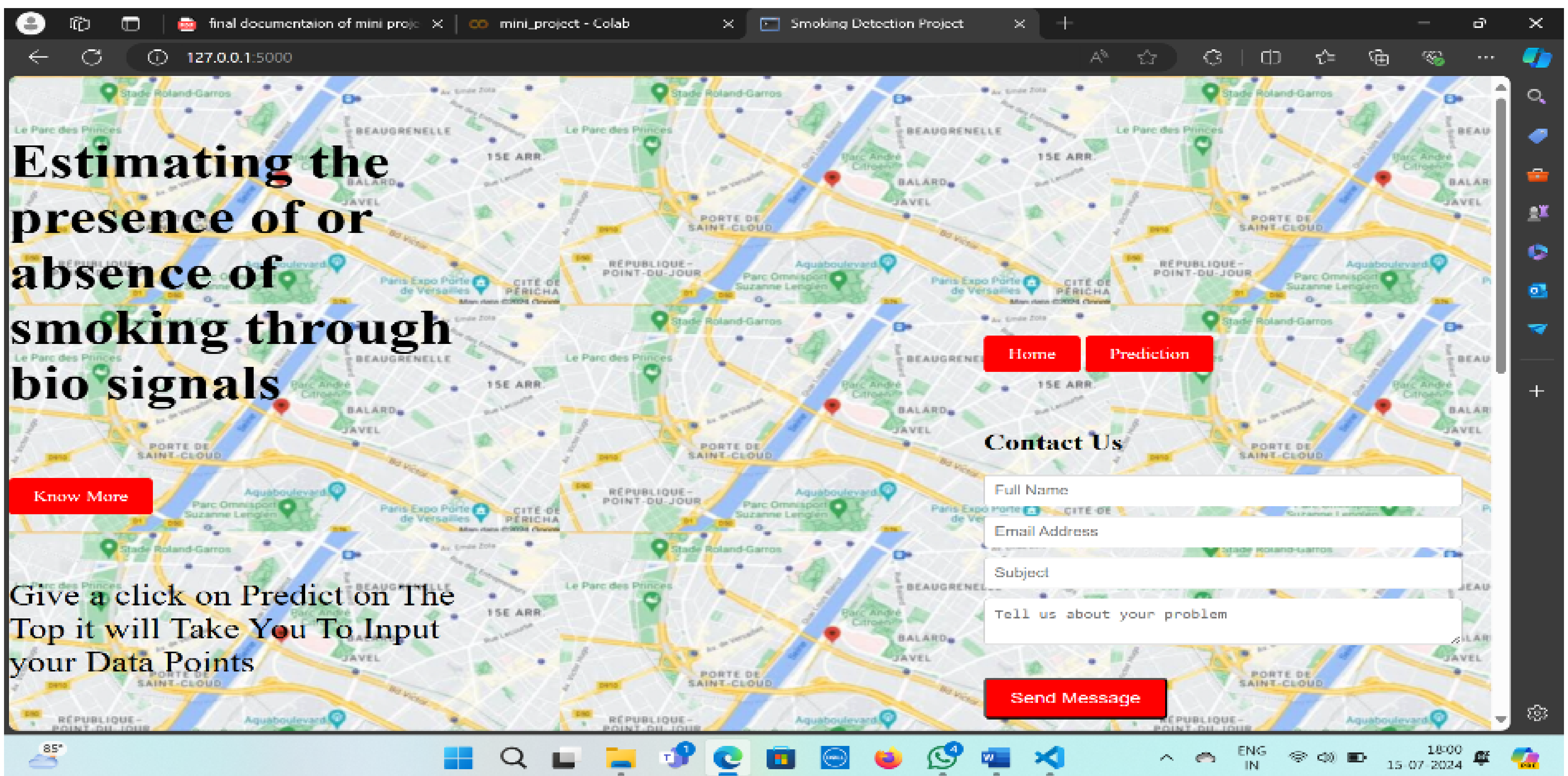
- Feature drop means it drops the columns that we don't want in our dataset.
- Feature drop = ['ID']

5. FLOW CHART

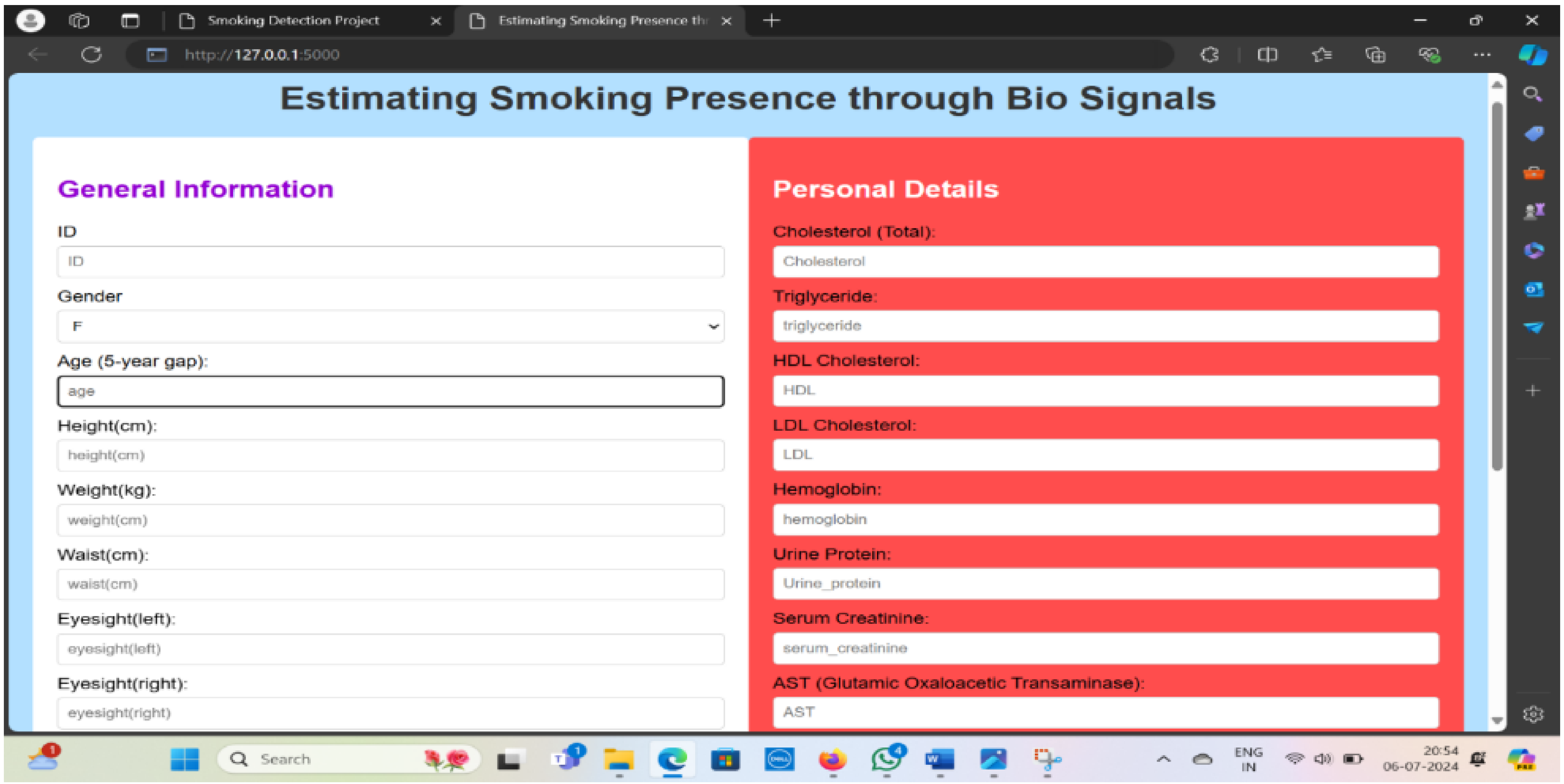
```
 ____________________________________________
            |           Start           |
|____________________________________________|
                        |
                        V
                 Acquire Bio Signals
                        |
                        V
                 Establish Baseline
                        |
                        V
                  Analyze Primary
                 Bio Signals (e.g.,
                heart rate, CO levels)
                        |
                 Yes    |    No
               ______   |   ______
               |    | |  |    |
             | Smoking|---+-->| End   |
            | Detected|     | (Absence) |
               |______|     |______|
                        |
                        V
                 Analyze Secondary
                Bio Signals (if needed)
                        |
                 Yes    |    No
               ______   |   ______
               |    | |  |    |
             | Smoking|---+-->| End   |
            | Detected|     | (Absence) |
               |______|     |______|
                        |
                        V
                   Confirmation
                        |
                 Yes    |    No
```

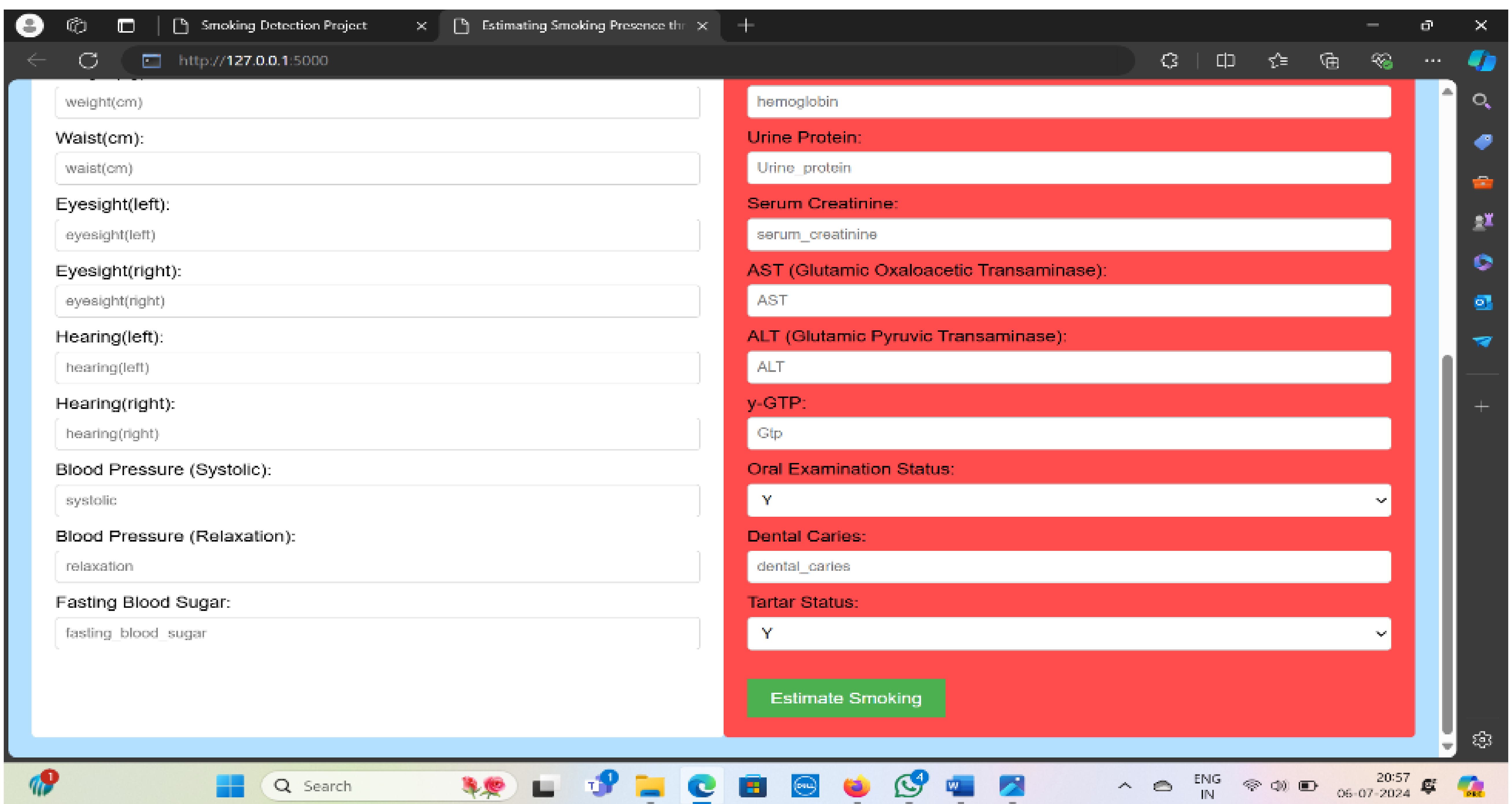
```
______   |   ______
|      |  |  |      |
| Smoking|---+-->| End   |
| Detected|      | (Absence) |
|______|     |______|
              |
              V
             End
```

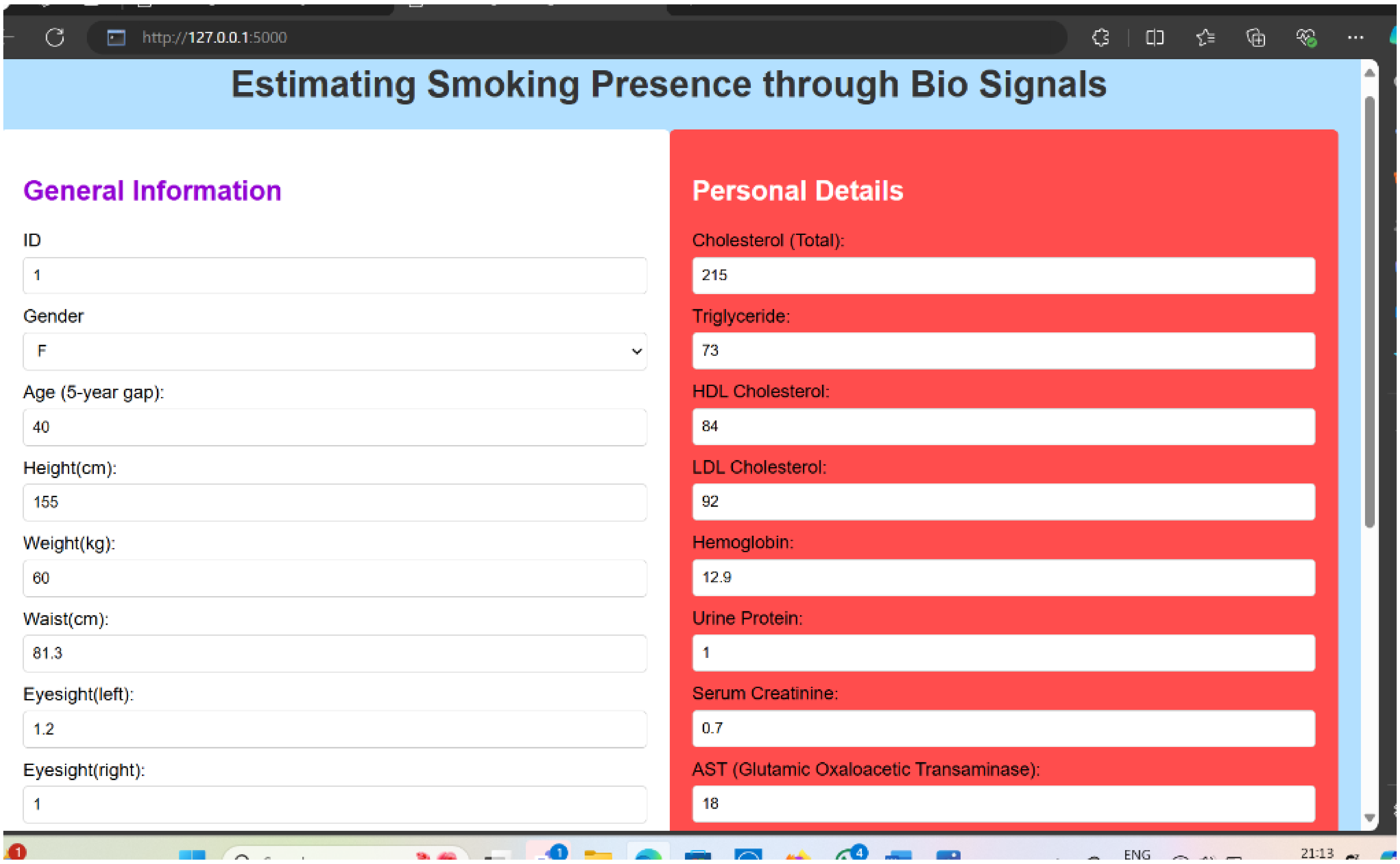
6.RESULT

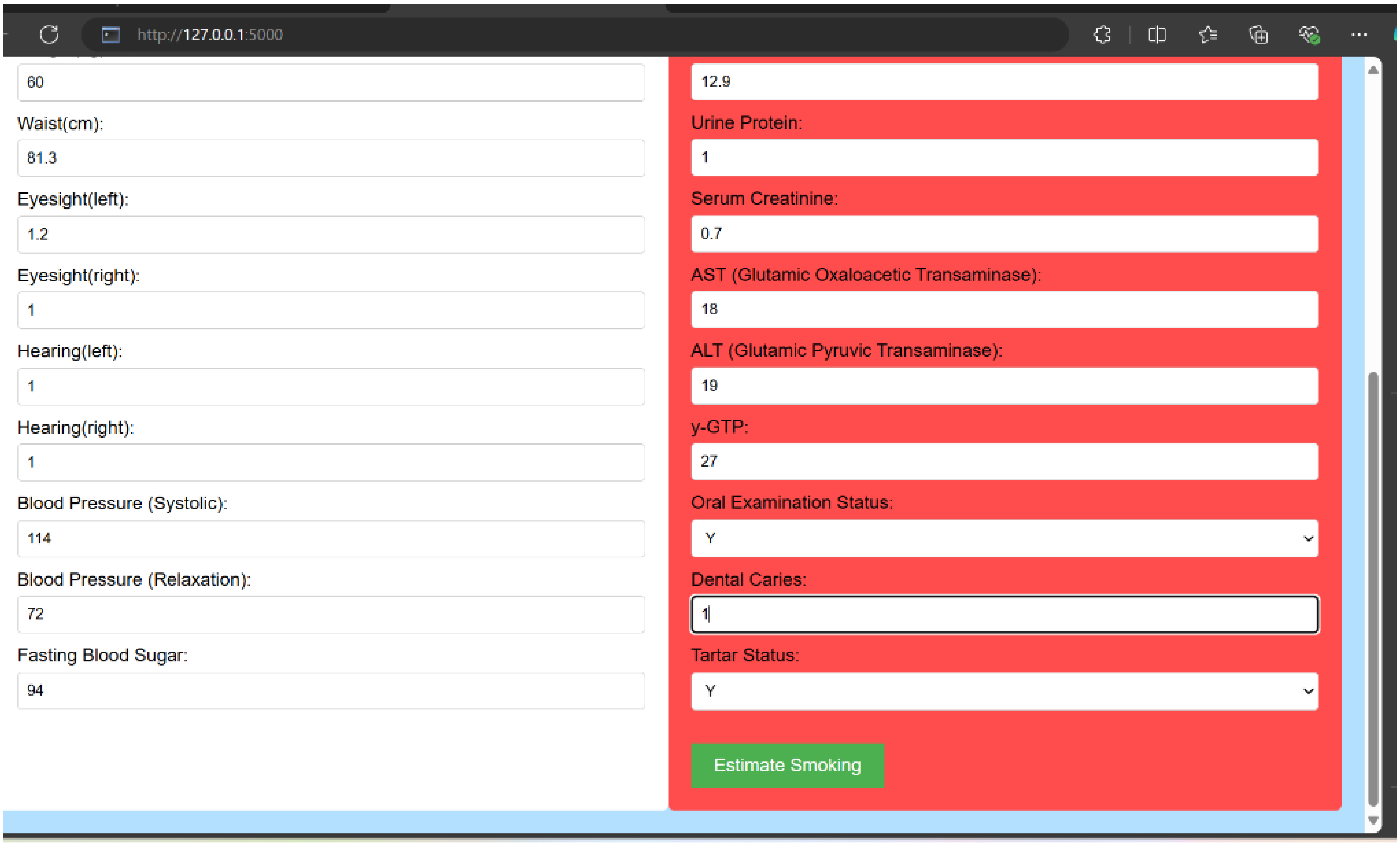


PREDICTIONS:

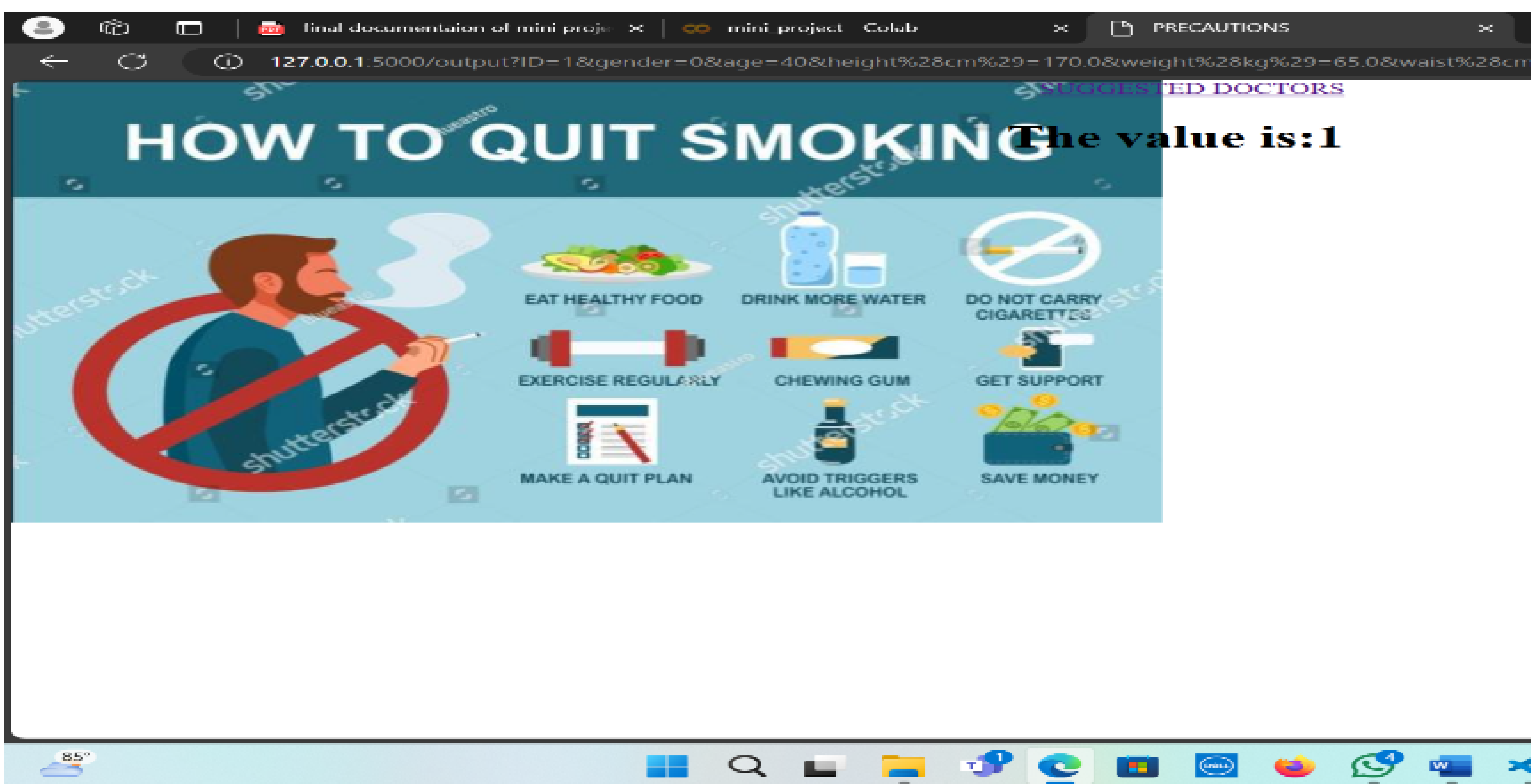

Smoking Detection Project
Estimating Smoking Presence th
http://127.0.0.1:5000
Estimating Smoking Presence through Bio Signals
General Information
ID
Gender
Age (5-year gap):
Height(cm):
Weight(kg):
Waist(cm):
Eyesight(left):
Eyesight(right):
Personal Details
Cholesterol (Total):
Triglyceride:
HDL Cholesterol:
LDL Cholesterol:
Hemoglobin:
Urine Protein:
Serum Creatinine:
AST (Glutamic Oxaloacetic Transaminase):


Smoking Detection Project
Estimating Smoking Presence th
http://127.0.0.1:5000
Waist(cm):
Eyesight(left):
Eyesight(right):
Hearing(left):
Hearing(right):
Blood Pressure (Systolic):
Blood Pressure (Relaxation):
Fasting Blood Sugar:
Urine Protein:
Serum Creatinine:
AST (Glutamic Oxaloacetic Transaminase):
ALT (Glutamic Pyruvic Transaminase):
y-GTP:
Oral Examination Status:
Dental Caries:
Tartar Status:
Estimate Smoking


http://127.0.0.1:5000
Estimating Smoking Presence through Bio Signals
General Information
ID
1
Gender
F
Age (5-year gap):
40
Height(cm):
155
Weight(kg):
60
Waist(cm):
81.3
Eyesight(left):
1.2
Eyesight(right):
1
Personal Details
Cholesterol (Total):
215
Triglyceride:
73
HDL Cholesterol:
84
LDL Cholesterol:
92
Hemoglobin:
12.9
Urine Protein:
1
Serum Creatinine:
0.7
AST (Glutamic Oxaloacetic Transaminase):
18
ENG
21:13



OUTPUT:



7.ADVANTAGES AND DISADVANTAGES

Estimating the presence or absence of smoking through biosignals has both advantages and disadvantages:

Advantages:

1. **Objective Measurement**: Biosignals provide an objective way to detect smoking behavior without relying on self-reporting, which can be biased or inaccurate due to social desirability or memory lapses.
2. **Real-Time Monitoring**: Biosignals can offer real-time detection of smoking episodes, allowing for immediate intervention or feedback.
3. **Accuracy**: By measuring physiological changes or biochemical markers associated with smoking (e.g., heart rate variability, carbon monoxide levels), the accuracy of detecting smoking behavior can be high.
4. **Quantitative Data**: Biosignals provide quantitative data that can be analyzed statistically, offering insights into smoking patterns, frequency, and intensity over time.
5. **Non-intrusive**: Monitoring through biosignals can be non-intrusive and discreet, reducing the likelihood of altering behavior due to awareness of being observed.
6. **Early Intervention**: Early detection of smoking behavior through biosignals can enable timely intervention and support for smoking cessation programs, potentially improving health outcomes.
7. **Research and Surveillance**: Biosignals can be used in research studies and public health surveillance to understand trends in smoking prevalence, effectiveness of tobacco control policies, and impacts on health.

Disadvantages:

1. **Cost and Accessibility**: The equipment and technology required to monitor biosignals can be costly, limiting widespread implementation, especially in resource-constrained settings.
2. **Interpretation Complexity**: Interpreting biosignal data requires specialized knowledge and expertise, which may not be readily available to all healthcare providers or users.
3. **Variability**: Individual variability in biosignals can make it challenging to establish universal thresholds or markers for detecting smoking behavior accurately.
4. **Privacy Concerns**: Continuous monitoring of biosignals raises privacy concerns, as it involves collecting sensitive health data that must be handled securely and ethically.
5. **Potential for False Positives or Negatives**: Despite high accuracy in many cases, biosignal-based detection of smoking may still yield false positives (indicating smoking when it did not occur) or false negatives (failing to detect smoking when it did occur).
6. **Ethical Considerations**: Monitoring biosignals for detecting behaviors like smoking raises ethical questions about consent, autonomy, and the potential for stigmatization or discrimination based on detected behaviors.

7. **Integration Challenges**: Integrating biosignal monitoring into daily life or existing healthcare systems can be challenging, requiring seamless integration with technology and workflows.

8.APPLICATIONS

Clinical Research and Trials: Biosignal monitoring can be used in clinical research to study the physiological effects of smoking, assess the efficacy of smoking cessation therapies, or evaluate the impact of smoking on health outcomes. It allows researchers to objectively measure smoking behavior and its associated health risks.

Health Monitoring and Management: Healthcare providers can use biosignals to monitor smoking behavior in patients as part of routine health assessments. This information can guide personalized health interventions, such as smoking cessation programs or adjustments in treatment plans for conditions exacerbated by smoking.

Public Health Surveillance: Biosignal-based monitoring can contribute to public health efforts by providing real-time data on smoking prevalence and trends within communities. This information helps in designing and evaluating tobacco control policies and interventions aimed at reducing smoking rates.

Behavioral Studies: Researchers studying behavioral patterns related to smoking can utilize biosignals to analyze smoking habits, triggers, and cessation strategies. This includes understanding the physiological responses associated with cravings and withdrawal symptoms.

Occupational Health and Safety: In workplaces where smoking is prohibited or restricted, biosignal monitoring can help enforce policies by detecting unauthorized smoking breaks or exposure to secondhand smoke among employees.

Personal Health Devices: Wearable biosensors integrated into consumer devices (e.g., fitness trackers, smartwatches) can provide individuals with real-time feedback on their smoking behavior. This can empower users to monitor and manage their smoking habits independently.

Forensic Investigations: Biosignal data can be used in forensic investigations to verify claims related to smoking behavior in legal cases, such as insurance claims or workplace disputes.

Psychological and Addiction Studies: Biosignals can contribute to understanding the psychological and physiological aspects of nicotine addiction, including its impact on stress responses, mood regulation, and cognitive functions.

Sports Medicine: Monitoring biosignals related to smoking can be relevant in sports medicine to assess the impact of smoking on athletic performance, recovery, and overall health of athletes.

9.CONCLUSION

Estimating the presence or absence of smoking through biosignals represents a promising approach with diverse applications across healthcare, research, public health, and beyond. By leveraging physiological measurements and biochemical markers associated with smoking, this method offers several advantages:

1. **Objective and Reliable Measurement**: Biosignals provide an objective means to detect smoking behavior, overcoming the limitations of self-reported data which can be biased or unreliable.
2. **Real-Time Monitoring**: The ability to monitor smoking behavior in real-time enables prompt intervention and support, crucial for smoking cessation efforts and health management.
3. **Enhanced Research and Surveillance**: Biosignal data contributes valuable insights into smoking prevalence, trends, and health impacts, supporting evidence-based policy-making and public health initiatives.

However, challenges such as cost, interpretational complexity, privacy concerns, and ethical considerations must be carefully addressed. Despite these challenges, the continued advancement of biosensor technology and integration with everyday devices holds great potential for improving both individual health outcomes and population-wide efforts to reduce smoking rates.

10.FUTURE SCOPE

Data driven by advancements in technology, healthcare integration, and societal needs. Here are several potential future directions:

1. **Advanced Sensor Technology**: Continued development in wearable biosensors and miniaturized devices will improve the accuracy, comfort, and accessibility of biosignal monitoring for detecting smoking behavior. This could include sensors capable of detecting specific biomarkers associated with smoking in real-time.
2. **Integration with AI and Machine Learning**: AI algorithms can analyze complex biosignal data patterns to enhance the accuracy of smoking detection. Machine learning models can learn from diverse datasets to improve predictive capabilities and adapt to individual variations in biosignals.
3. **Personalized Health Interventions**: Biosignal data can inform personalized smoking cessation programs tailored to individual behaviors, triggers, and responses. This could include real-time feedback and interventions delivered through mobile health applications or smart devices.
4. **Telehealth and Remote Monitoring**: Remote biosignal monitoring will enable healthcare providers to monitor smoking behavior and provide support remotely, expanding access to care and improving continuity of management.

5. **Population Health Surveillance**: Biosignal-based monitoring can contribute to comprehensive public health surveillance systems, providing timely data on smoking prevalence, trends, and associated health outcomes at both local and global levels.
6. **Biofeedback and Behavioral Therapy**: Utilizing biosignals for biofeedback mechanisms can aid in behavioral therapies aimed at managing cravings, stress responses, and withdrawal symptoms associated with smoking cessation.
7. **Integration into Smart Environments**: Biosensors embedded in smart environments (e.g., homes, workplaces) could automatically detect and respond to smoking behavior, supporting smoke-free policies and promoting healthier environments.
8. **Forensic and Legal Applications**: Biosignal data can be utilized in forensic investigations, insurance claims, and legal contexts to verify smoking behavior, contributing to evidence-based decision-making.
9. **Ethical and Privacy Considerations**: Future developments will need to address ethical considerations regarding consent, data security, and the responsible use of biosignal data in monitoring and intervention.

11.BIBILIOGRAPHY

1. Benowitz, Neal L. "Biomarkers of environmental tobacco smoke exposure." *Environmental Health Perspectives* 107.Suppl 2 (1999): 349-355.
2. Ferguson, Steven G., et al. "Measuring acute effects of cigarette smoking on behavioral tasks: A new approach for immediate biosignal tracking." *Addictive Behaviors* 86 (2018): 90-96.
3. Giovino, Gary A., et al. "Epidemiology of tobacco use in the United States." *Epidemiologic reviews* 15 (1993): 195-211.
4. Hossain, Md Zakir, et al. "Wearable sensors for remote health monitoring: A review." *Sensors* 19.21 (2019): 1-32.
5. McClure, Erin A., et al. "A review of wearable sensors and systems with application in rehabilitation." *Journal of Neuroengineering and Rehabilitation* 14.1 (2017): 1-17.
6. Mukasa, Oscar, et al. "Real-time detection of smoking activities with wrist-worn inertial measurement unit and bio-impedance sensor." *IEEE Transactions on Biomedical Engineering* 66.2 (2018): 273-281.
7. Shih, Ping-Tung, et al. "Tobacco smoke exposure and levels of urinary metals in the U.S. youth and adult population: The National Health and Nutrition Examination Survey (NHANES) 1999-2004." *International Journal of Environmental Research and Public Health* 7.6 (2010): 2472-2485.

8. Stokes, Andrew, et al. "Exposure measurement error in studies of environmental tobacco smoke exposure and lung cancer risk." *American Journal of Epidemiology* 170.6 (2009): 746-751.
9. Wang, Ping, et al. "Biomarkers of tobacco exposure: summary of an FDA-sponsored public workshop." *Cancer Epidemiology and Prevention Biomarkers* 15.5 (2006): 1554-1555.
10. West, Robert. "Assessment of dependence and motivation to stop smoking." *British Journal of Addiction* 86.9 (1991): 1119-1127.

These references cover various aspects of biosignal-based smoking detection, including physiological biomarkers, wearable sensors, epidemiological studies, and public health implications. Adjust the format according to the citation style (e.g., APA, MLA) you're using for your bibliography.

12.APPENDIX

Model building :

1)Dataset

2)Google colab and VS code Application Building

1. HTML file (Index file, Predict file)
2. CSS file
3. Models in pickle format

SOURCE CODE:

HOME.HTML

```
 <!DOCTYPE html>
<html lang="en">
<head>
   <meta charset="UTF-8">
   <meta name="viewport" content="width=device-width, initial-scale=1.0">
   <title>Smoking Detection Project</title>
   <style>
   body{
      margin: 0;
      padding: 0;
      color: black;
      background-image: url("../static/map.jpg");
      height: 1500px;
      width: 750px;
  }
   .container {
      max-width: 1200px;
      margin: 0 auto;
      padding: 20px;
      display: flex;
      flex-wrap: wrap;
      }
   .left-column {
      position: relative;
      right: 20px;
      bottom: 10px;
      flex-basis: 55%;
      padding-right: 40px;
      }
   .right-column {
      position: absolute;
      left: 790px;
```

```
  bottom: 12px;
  flex-basis: 50%;
  padding-left: 10px;
  }
 h3 {
  color: #ff0000;
  margin-top: 0;
  }
.button {
  display: inline-block;
  padding: 10px 20px;
  font-size: 16px;
  color: #ffffff;
  background-color: #ff0000;
  text-decoration: none;
  border-radius: 4px;
  margin-top: 20px;

  }
.contact-form {
  margin-top: 40px;
}
 .contact-form input[type="text"],
.contact-form textarea {
  width: 90%;
  padding: 8px;
  margin-bottom: 10px;
  border: 1px solid #ccc;
  border-radius: 4px;
 }
.contact-form.button {
  margin-top: 10px;
```

```
  }
  </style>
  </head>
 <body>
     <div class="container">
     <div class="left-column">
        <p style="font-size: 50px;"><b>Estimating the presence of       or<br>absence of
smoking through<br>bio signals</b></p>
        <a href="#" class="button">Know More</a>
<br><br><br>
<p style="font-size: 30px;"> Give a click on Predict on The Top it will Take You To Input your
Data Points</p>
 </div>
<div class="right-column">
        <a href="/" class="button">Home</a>
        <a href="/predict" class="button">Prediction</a>
        <br><br>
 <div class="contact-form">
     <h2>Contact Us</h2>
     <form action="https://formsubmit.co/0faf1458d3cf0e276f4f899db62fe0bb"
method="post">
        <input type="text" placeholder="Full Name" name="fullname" required>
        <input type="text" placeholder="Email Address" name="email" required>
        <input type="text" placeholder="Subject" name="Subject" required>
        <textarea placeholder="Tell us about your problem" name="problem"></textarea>
        <button type="submit" class="button">Send Message</button>
     </form>
</div>
</div>
</div>
</body>
</html>
```

INDEX.HTML

```
<!DOCTYPE html>
<html>
<head>
  <title>Estimating Smoking Presence through Bio Signals</title>
  <style>
  body {
    font-family: Arial, sans-serif;
    margin: 20px;
    background-color: #b3e0ff;
  }
  h1 {
    color: #333;
  }
  p {
    margin-bottom: 10px;
  }
  form {
    margin-top: 20px;
    display: flex;

    justify-content: space-between;
  }
  .general-info {
    width: 48%;
    background-color: white;
    padding: 20px;
    border-radius: 5px;
  }
```

```
.personal-details {
   width: 48%;
   background-color: #ff4d4d;
   padding: 20px;
   border-radius: 5px;

}

label {
   display: block;
   margin-bottom: 5px;
}

input[type="text"], select {
   width: 100%;
   padding: 8px;
   margin-bottom: 10px;
   border: 1px solid #ccc;
   border-radius: 4px;
   box-sizing: border-box;
}
button {
   padding: 10px 20px;
   font-size: 16px;
   background-color: #4CAF50;
   color: #fff;
   border: none;
   cursor: pointer;
}

#result {
   margin-top: 20px;
```

```
    }

    table {
      border-collapse: collapse;
      width: 100%;
    }

    th, td {
      text-align: left;
      padding: 8px;
    }

    th {
      background-color: #4CAF50;
      color: white;
    }

    </style>

</head>
<body>
    <h1 style="text-align:center">Estimating Smoking Presence through Bio Signals</h1>
    <form action="/output", methods="post">
    <div class="general-info">
    <h2 style="color: #9400d3">General Information</h2>
    <label for="ID">ID</label>
    <input type="text" name="ID" placeholder="ID">
    <label for="gender">Gender</label>
    <select id="gender" name="gender" required>
      <option value="0">F</option>
      <option value="1">M</option>
```

```
</select>
<label for="age">Age (5-year gap): </label>
<input type="text" name="age" placeholder="age">
<label for="height(cm)">Height(cm): </label>
<input type="text" name="height(cm)" placeholder="height(cm)">
<label for="weight(kg)">Weight(kg):</label>
<input type="text" name="weight(kg)" placeholder="weight(cm)">
<label for="waist(cm)">Waist(cm): </label>
<input type="text" name="waist(cm)" placeholder="waist(cm)">
<label for="eyesight(left)">Eyesight(left):</label>
<input type="text" name="eyesight(left)" placeholder="eyesight(left)">
<label for="eyesight(right)">Eyesight(right): </label>
<input type="text" name="eyesight(right)" placeholder="eyesight(right)">
<label for="hearing(left)">Hearing(left): </label>
<input type="text" name="hearing(left)" placeholder="hearing(left)">
<label for="hearing(right)">Hearing(right): </label>
<input type="text" name="hearing(right)" placeholder="hearing(right)">
<label for="systolic">Blood Pressure (Systolic): </label>
<input type="text" name="systolic" placeholder="systolic">
<label for="relaxation">Blood Pressure (Relaxation): </label>
<input type="text" name="relaxation" placeholder="relaxation">
<label for="fasting_blood_sugar">Fasting Blood Sugar: </label>
<input type="text" name="fasting_blood_sugar" placeholder="fasting_blood_sugar">
</div>
<div class="personal-details">

  <h2 style="color: white">Personal Details</h2>
  <label for="Cholesterol">Cholesterol (Total): </label>
  <input type="text" name="Cholesterol" placeholder="Cholesterol">
  <label for="triglyceride">Triglyceride: </label>
  <input type="text" name="triglyceride" placeholder="triglyceride">
  <label for="HDL">HDL Cholesterol:</label>
```

```html
    <input type="text" name="HDL" placeholder="HDL">
    <label for="LDL">LDL Cholesterol: </label>
    <input type="text" name="LDL" placeholder="LDL">
    <label for="hemoglobin">Hemoglobin:</label>
    <input type="text" name="hemoglobin" placeholder="hemoglobin">
    <label for="Urine_protein">Urine Protein: </label>
    <input type="text" name="Urine_protein" placeholder="Urine_protein">
    <label for="serum_creatinine">Serum Creatinine:</label>
    <input type="text" name="serum_creatinine" placeholder="serum_creatinine">
    <label for="AST">AST (Glutamic Oxaloacetic Transaminase):</label>
    <input type="text" name="AST" placeholder="AST">
    <label for="ALT">ALT (Glutamic Pyruvic Transaminase): </label>
    <input type="text" name="ALT" placeholder="ALT">
    <label for="Gtp">y-GTP:</label>
    <input type="text" name="Gtp" placeholder="Gtp">
    <label for="oral">Oral Examination Status: </label>
    <select id="oral" name="oral" required>
      <option value="1">Y</option>
      <option value="0">N</option>
    </select>
    <label for="dental_caries">Dental Caries: </label>
    <input type="text" id="dental_caries" placeholder="dental_caries">
    <label for="tartar">Tartar Status:</label>
    <select id="tartar" name="tartar" required>
      <option value="0">Y</option>
    </select>
    <br><br>
  <!--    <input type="text" id="smoking" name="smoking" readonly hidden>-->
    <button type="submit">Estimate Smoking</button>
  </div>
  </form>
</body>
```

```
  </html>
```

SMOKE_PREDICT.HTML

```
<!DOCTYPE html>
<html>
<head>
    <title>Smoking Estimation Result</title>
    <style>
    body {
        font-family: Arial, sans-serif;
        margin: 0;
        color: white;
        background-image: url("../static/smoke.jpg");
        background-repeat: no-repeat;
        height: 1500px;
        width: 750px;
    }
    h1 {
      color: black;
      text-align: right;
    }
    h3 {
        color: red;
    }
    .container {
        padding: 20px;
    }
    .top-right {
        position: absolute;
        top: 10px;
        right: 200px;
    }
```

```
    .top-img {
        position: absolute;
        top: 10px;
        right: 1px;
    }

    .button {
        background-color: red;
        color: white;
        padding: 10px 20px;
        margin-left: 10px;
        text-decoration: none;
        display: inline-block;
    }
    .button:hover {
        background-color: #800000;
    }
    img {
        position: relative;
        display: block;
        margin-top: 40px;
        max-width: 80%;
        top: 100px;
        left: 100px;
        height: 80%;
    }

    </style>

</head>
<body>
    <div class="container">
```

```
        <div class="top-right">
          <br><br>
          <a href="/" class="button">Home</a>
          <a href="/predict" class="button.">Prediction</a>
          <br><br>

        </div>

        <div class="top-img">
          <br><br><br><br>
        </div>

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

    </div>

</body>
</html>
```

PRECAUTIONS.HTML

```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
```

```
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>PRECAUTIONS</title>
    <style>
        h1{
            text-align:right;
        }
        body{
            margin: 0;
            padding: 0;
            color: black;
            text-align:right;
            background-image: url("../static/safety.jpg");
            place-items: center;
            background-repeat: no-repeat;
            height: 750px;
            width: 750px;

        }
    </style>
</head>
<body>
    <a href="https://www.askapollo.com/diseases/quitting-smoking" >SUGGESTED
DOCTORS</a>
    <h1>The value is:{{pred}}</h1>

</body>
</html>
```

APP.PY

```
import flask
from flask import Flask,render_template,request,url_for
import pickle
```

```
import numpy as np
import pandas as pd
from sklearn.preprocessing import StandardScaler,LabelEncoder
from sklearn import preprocessing

app=Flask(_name_)
@app.route('/')
def Hello_world():
   return render_template('home.html')
@app.route('/predict')
def home():
   return render_template('index.html')
@app.route('/output',methods=['POST','GET'])
def predict():
   if request.method=="POST":
      ID=request.form['ID']
      Gender=request.form['gender']
      Age=request.form['age']
      Height=request.form['height(cm)']
      Weight=request.form['weight(Kg)']
      Waist=request.form['waist(cm)']
      Eyesight_left=request.form['eyesight(left)']
      Eyesight_right=request.form['eyesight(right)']
      Hearing_left=request.form['hearing(left)']
      Hearing_right=request.form['hearing(right)']
      systolic=request.form['systolic']
      relaxation=request.form['relaxation']
      fasting_blood_sugar=request.form['fasting_blood_sugar']
      Cholesterol=request.form['Cholesterol']
      triglyceride=request.form['triglyceride']
      HDL=request.form['HDL']
      LDL=request.form['LDL']
```

```
    Hemoglobin=request.form['hemoglobin']
    urine_protein=request.form['Urine_protein']
    serum_creatinine=request.form['serum_creatinine']
    AST=request.form['AST']
    ALT=request.form['ALT']
    Gtp=request.form['Gtp']
    oral=request.form['oral']
    dental_caries=request.form['dental_caries']
    tartar=request.form['tartar']

input=[[float(ID),float(Gender),float(Age),float(Height),float(Weight),float(Waist),float(Eyesight_left),float(Eyesight_right),float(Hearing_left),

float(Hearing_right),float(systolic),float(relaxation),float(fasting_blood_sugar),float(Cholesterol),float(triglyceride),float(HDL),float(LDL),

float(Hemoglobin),float(urine_protein),float(serum_creatinine),float(AST),float(ALT),float(Gtp),float(oral),float(dental_caries),float(tartar)]]
    #Handling Missing Values

    model=pickle.load(open('smoke_model.pkl','rb'))
    names=[['ID','gender','age','height(cm)','weight(kg)','waist(cm)','eyesight(left)','eyesight(right)',
        'hearing(left)','hearing(right)','systolic','relaxation','fasting_blood_sugar','Cholesterol',
        'triglyceride','HDL','LDL','hemoglobin','Urine_protein','serum creatinine','AST','ALT',
        'Gtp','oral','dental_caries','tartar']]
    final=[np.array(input)]
    data=pd.DataFrame(final,columns=names)
    scale=StandardScaler()
    data=scale.fit_transform(final)
    label_encoder = preprocessing.LabelEncoder()
    # Encode labels in column 'gender'.
    data['gender']= label_encoder.fit_transform(data['gender'])
    data['gender'].unique()
```

```
        data['oral']= label_encoder.fit_transform(data['oral'])
        data['oral'].unique()
        data['tartar']= label_encoder.fit_transform(data['tartar'])
        data['tartar'].unique()
        prediction=model.predict(data)[0]
        if prediction==0 :
            return render_template('/smoke_predict.html',pred="0")
    #rendering
    return render_template('precautions.html',pred="1")
if _name=='main_':
    app.run(debug=True)
```

CODE SNIPPETS

MODEL BUILDING

mini_project

File Edit View Insert Runtime Tools Help Last saved at July 14

+ Code + Text Connect Gemini

```
import pandas as pd
import numpy as np
import seaborn as sns
```

```
[ ] df=pd.read_csv("/content/smoking.csv")
```

```
df.shape
```

```
(55692, 27)
```



```
df.shape
```

```
(55692, 27)
```



```
] df
```

	ID	gender	age	height(cm)	weight(kg)	waist(cm)	eyesight(left)	eyesight(right)	hearing(left)	hearing(right)	...	hemoglobin	p
0	0	F	40	155	60	81.3	1.2	1.0	1.0	1.0	...	12.9	
1	1	F	40	160	60	81.0	0.8	0.6	1.0	1.0	...	12.7	
2	2	M	55	170	60	80.0	0.8	0.8	1.0	1.0	...	15.8	
3	3	M	40	165	70	88.0	1.5	1.5	1.0	1.0	...	14.7	
4	4	F	40	155	60	86.0	1.0	1.0	1.0	1.0	...	12.5	
...	...	...	...	...	...	...	...	...	...	...	...	...	
55687	55676	F	40	170	65	75.0	0.9	0.9	1.0	1.0	...	12.3	
55688	55681	F	45	160	50	70.0	1.2	1.2	1.0	1.0	...	14.0	
55689	55683	F	55	160	50	68.5	1.0	1.2	1.0	1.0	...	12.4	
55690	55684	M	60	165	60	78.0	0.8	1.0	1.0	1.0	...	14.4	
55691	55691	M	55	160	65	85.0	0.9	0.7	1.0	1.0	...	15.0	

55692 rows × 27 columns

```
df.describe()
```

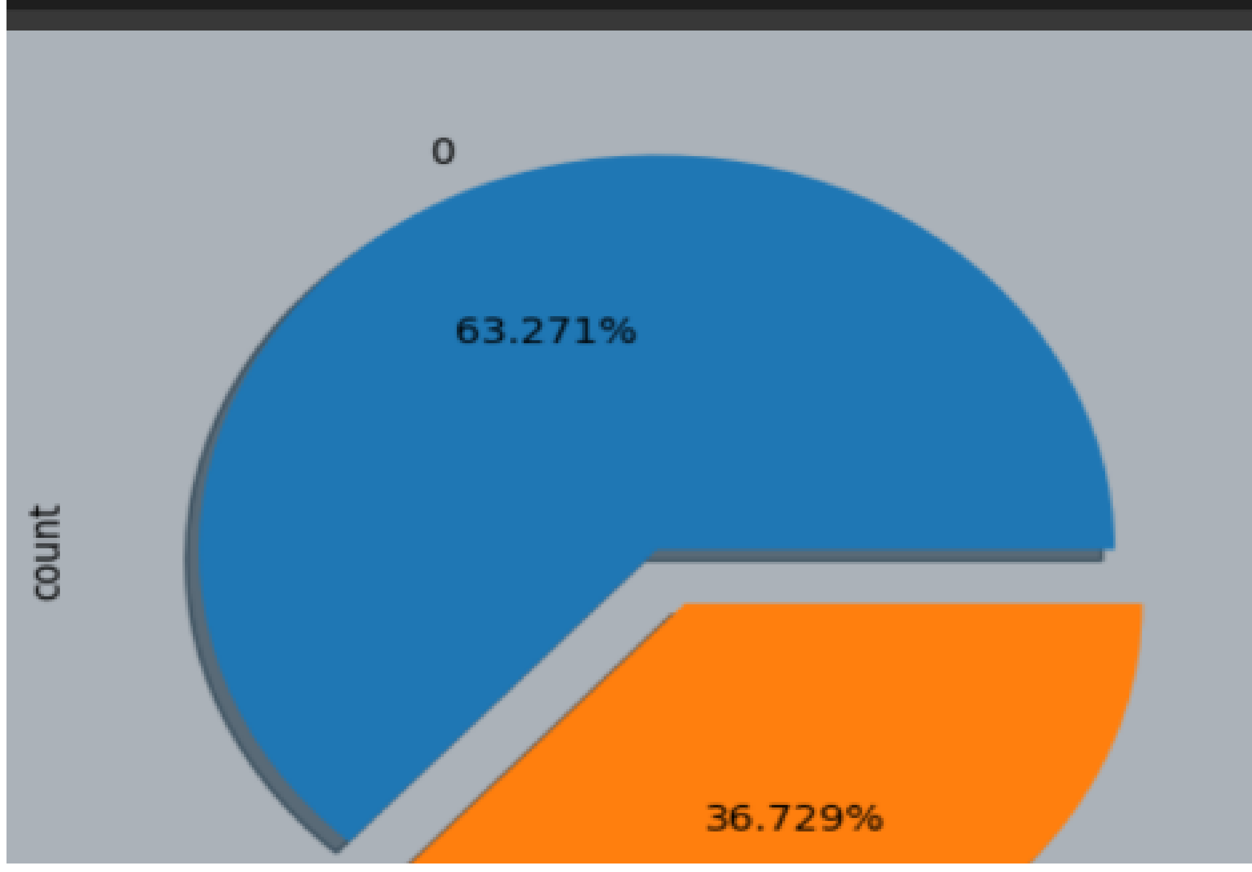
	ID	gender	age	height(cm)	weight(kg)	waist(cm)	eyesight(left)	eyesight(right)	hearing(left)	hearing(
count	55692.000000	55692.000000	55692.000000	55692.000000	55692.000000	55692.000000	55692.000000	55692.000000	55692.000000	55692.
mean	27845.500000	0.635657	44.182917	164.649321	65.864936	82.046418	1.012623	1.007443	1.025587	1.
std	16077.039933	0.481250	12.071418	9.194597	12.820306	9.274223	0.486873	0.485964	0.157902	0.
min	0.000000	0.000000	20.000000	130.000000	30.000000	51.000000	0.100000	0.100000	1.000000	1.
25%	13922.750000	0.000000	40.000000	160.000000	55.000000	76.000000	0.800000	0.800000	1.000000	1.
50%	27845.500000	1.000000	40.000000	165.000000	65.000000	82.000000	1.000000	1.000000	1.000000	1.
75%	41768.250000	1.000000	55.000000	170.000000	75.000000	88.000000	1.200000	1.200000	1.000000	1.
max	55691.000000	1.000000	85.000000	190.000000	135.000000	129.000000	9.900000	9.900000	2.000000	2.

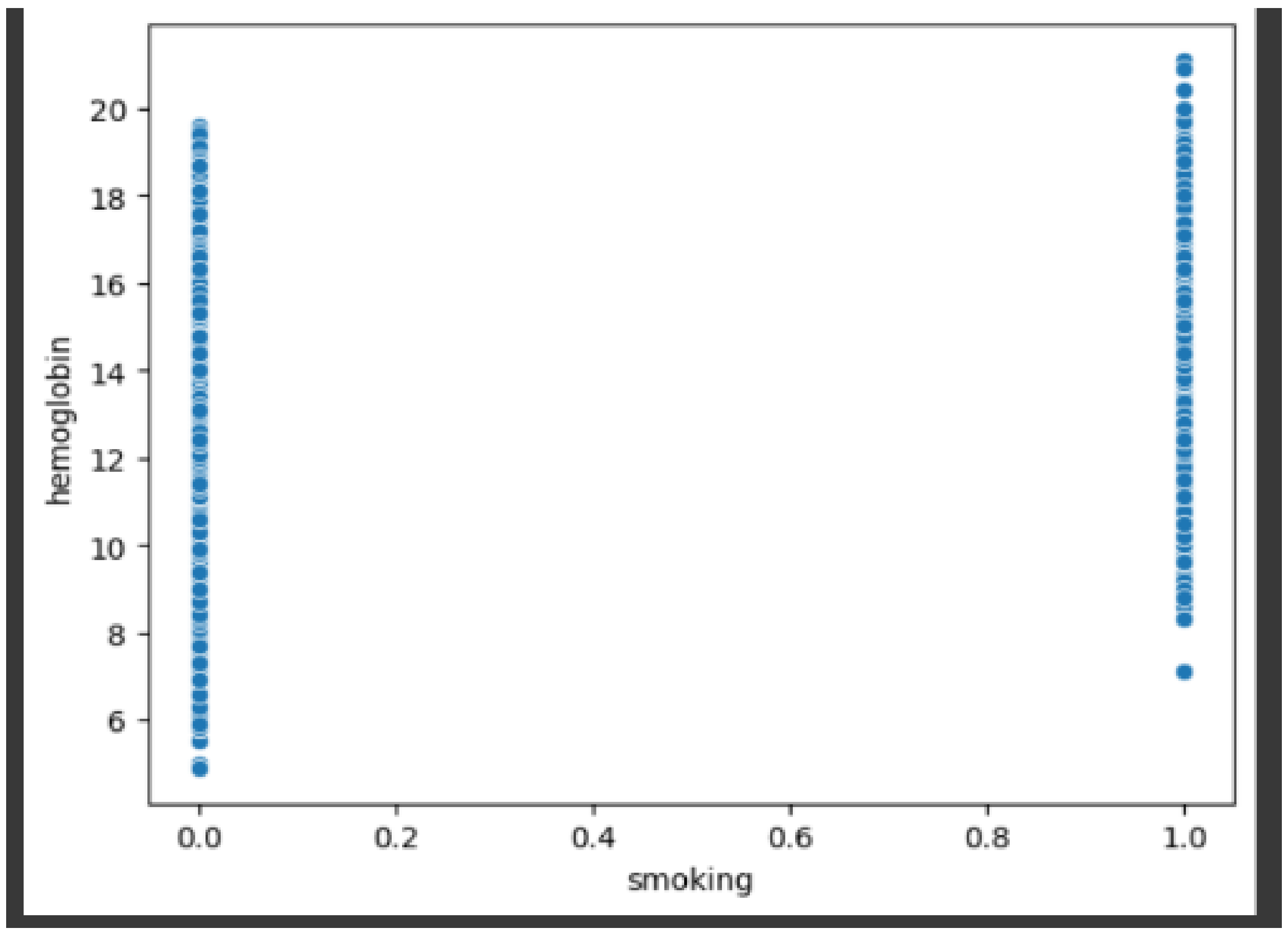
8 rows × 27 columns

hearing(left)	hearing(right)	...	hemoglobin	Urine_protein	serum_creatinine	AST	ALT	Gtp	oral	dental_caries
55692.000000	55692.000000	...	55692.000000	55692.000000	55692.000000	55692.000000	55692.000000	55692.000000	55692.0	55692.000000
1.025587	1.026144	...	14.622592	1.087212	0.885738	26.182935	27.036037	39.952201	0.0	0.213334
0.157902	0.159564	...	1.564498	0.404882	0.221524	19.355460	30.947853	50.290539	0.0	0.409665
1.000000	1.000000	...	4.900000	1.000000	0.100000	6.000000	1.000000	1.000000	0.0	0.000000
1.000000	1.000000	...	13.600000	1.000000	0.800000	19.000000	15.000000	17.000000	0.0	0.000000
1.000000	1.000000	...	14.800000	1.000000	0.900000	23.000000	21.000000	25.000000	0.0	0.000000
1.000000	1.000000	...	15.800000	1.000000	1.000000	28.000000	31.000000	43.000000	0.0	0.000000
2.000000	2.000000	...	21.100000	6.000000	11.600000	1311.000000	2914.000000	999.000000	0.0	1.000000

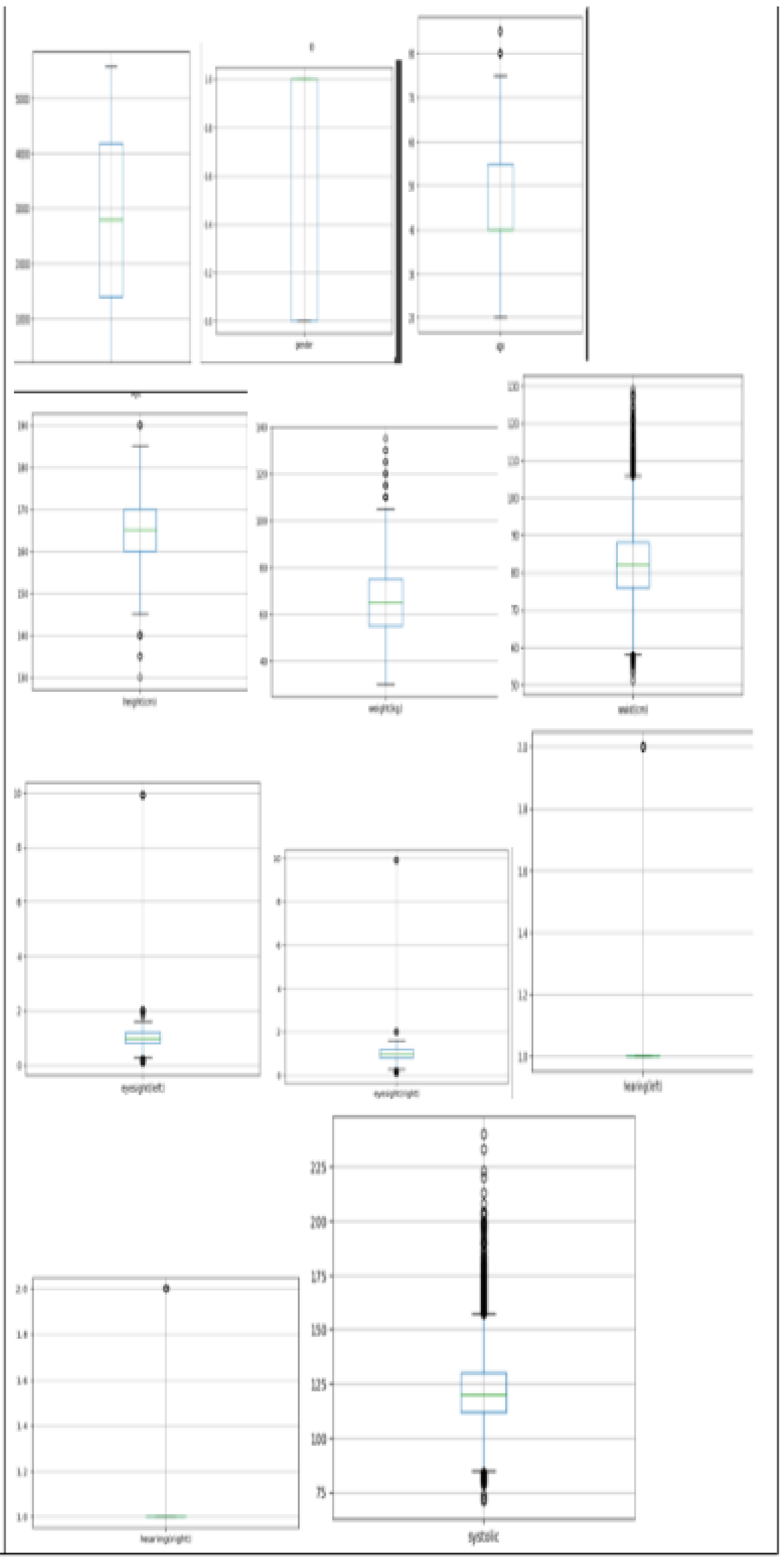
dental_caries	tartar	smoking
55692.000000	55692.000000	55692.000000
0.213334	0.555556	0.367288
0.409665	0.496908	0.482070
0.000000	0.000000	0.000000
0.000000	0.000000	0.000000
0.000000	1.000000	0.000000
0.000000	1.000000	1.000000
1.000000	1.000000	1.000000

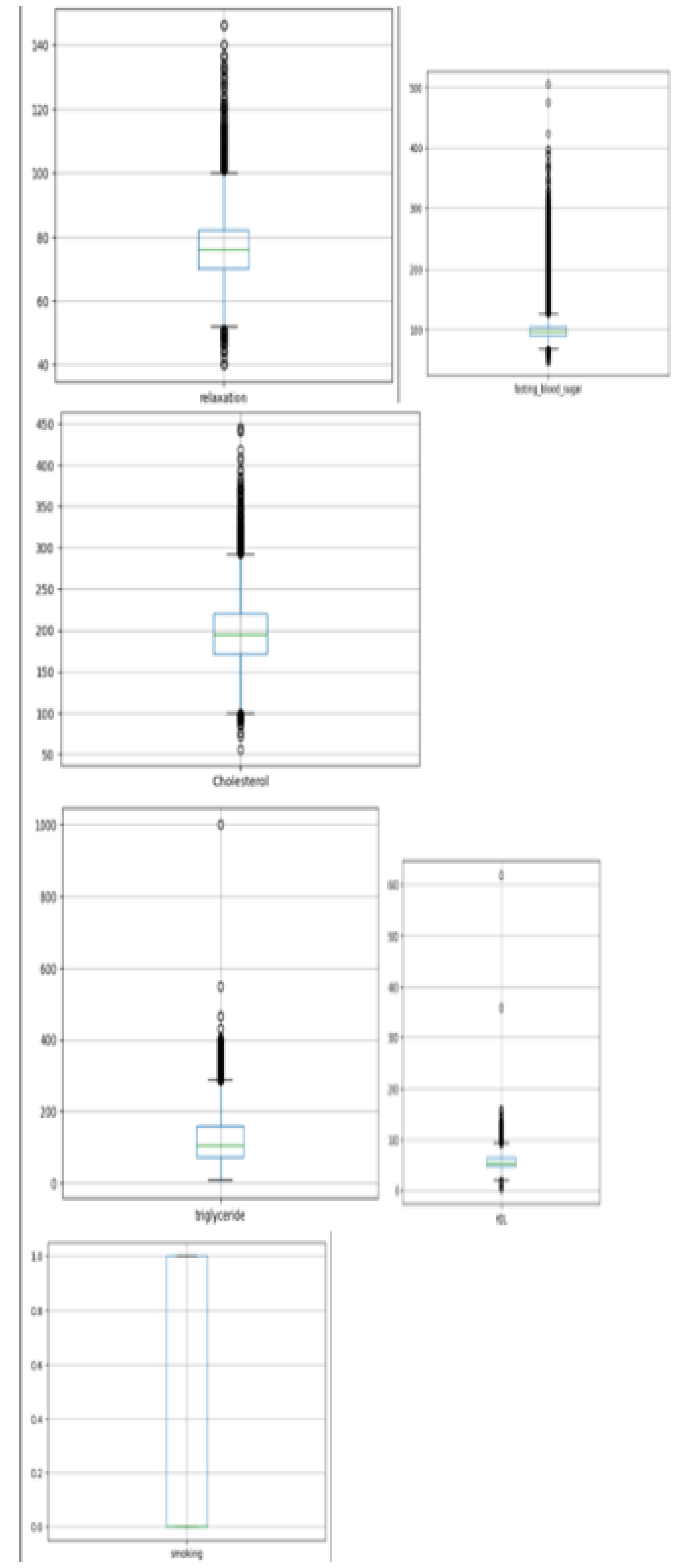
```
#Data Visualization
#Unnivariate Analysis
plt.figure(figsize=[5,5],clear=True,facecolor="#ABB2B9")
df["smoking"].value_counts().plot.pie(explode=[0,0.15],autopct="%1.3f%%",shadow=True);
```




hemoglobin
smoking
20
18
16
14
12
10
8
6
0.0
0.2
0.4
0.6
0.8
1.0

utliers and
nomalies





Loading Data	(see tables below)
Handling Missing Data	(see list below)

Loading Data

	ID	gender	age	height(cm)	weight(kg)	waist(cm)	eyesight(left)	eyesight(right)	hearing(left)	hearing(right)	...	hemoglobin
0	0	F	40	155	60	81.3	1.2	1.0	1.0	1.0	...	12.9
1	1	F	40	160	60	81.0	0.8	0.6	1.0	1.0	...	12.7
2	2	M	55	170	60	80.0	0.8	0.8	1.0	1.0	...	15.8
3	3	M	40	165	70	88.0	1.5	1.5	1.0	1.0	...	14.7
4	4	F	40	155	60	86.0	1.0	1.0	1.0	1.0	...	12.5
...	...	...	...	...	...	...	...	...	...	...	...	...
55687	55676	F	40	170	65	75.0	0.9	0.9	1.0	1.0	...	12.3
55688	55681	F	45	160	50	70.0	1.2	1.2	1.0	1.0	...	14.0
55689	55683	F	55	160	50	68.5	1.0	1.2	1.0	1.0	...	12.4
55690	55684	M	60	165	60	78.0	0.8	1.0	1.0	1.0	...	14.4
55691	55691	M	55	160	65	85.0	0.9	0.7	1.0	1.0	...	15.0

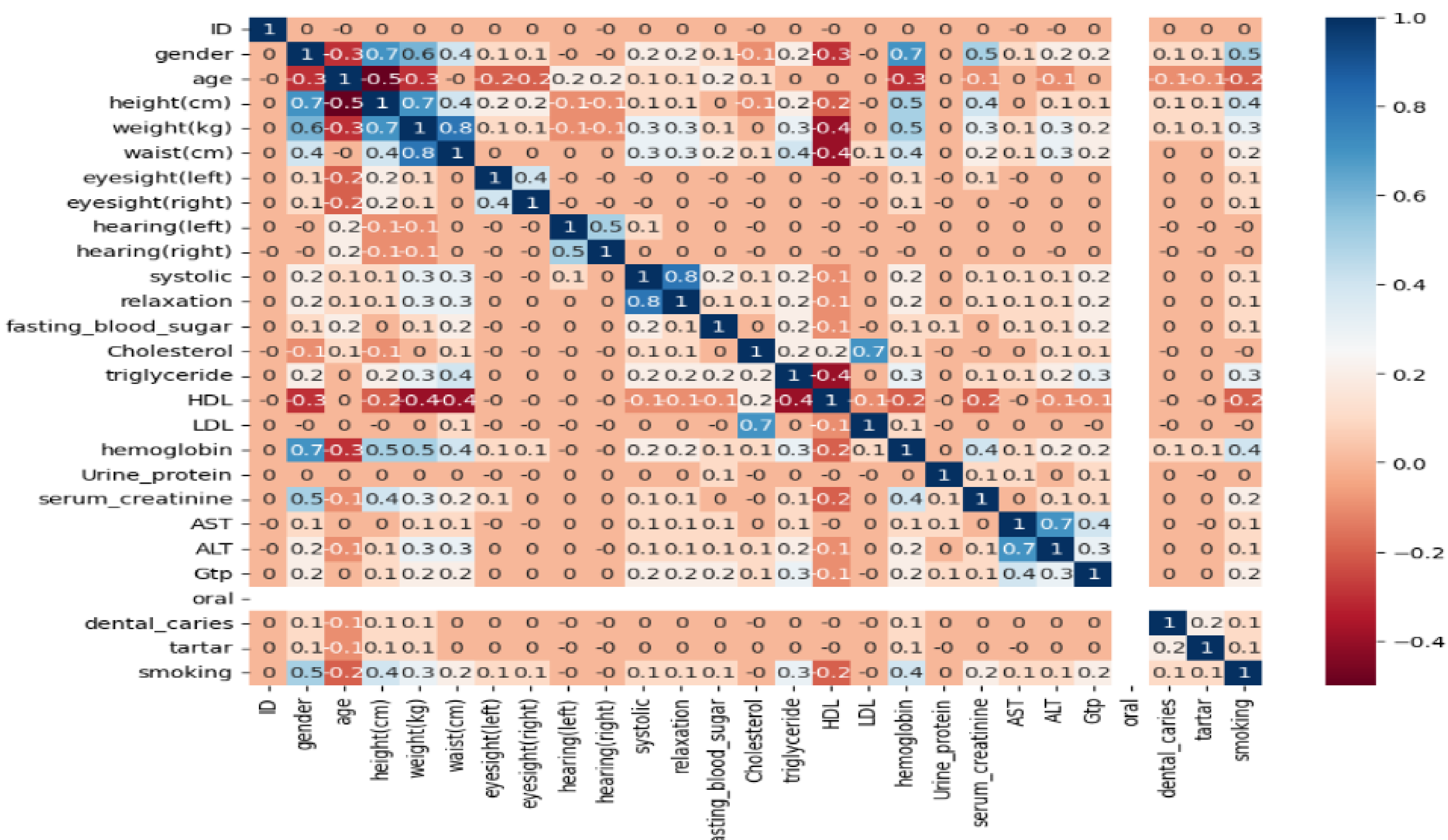
55692 rows × 27 columns

hemoglobin	Urine protein	serum creatinine	AST	ALT	Gtp	oral	dental caries	tartar	smoking
12.9	1.0	0.7	18.0	19.0	27.0	Y	0	Y	0
12.7	1.0	0.6	22.0	19.0	18.0	Y	0	Y	0
15.8	1.0	1.0	21.0	16.0	22.0	Y	0	N	1
14.7	1.0	1.0	19.0	26.0	18.0	Y	0	Y	0
12.5	1.0	0.6	16.0	14.0	22.0	Y	0	N	0
...	...	...	...	...	...	...	...	...	...
12.3	1.0	0.6	14.0	7.0	10.0	Y	1	Y	0
14.0	1.0	0.9	20.0	12.0	14.0	Y	0	Y	0
12.4	1.0	0.5	17.0	11.0	12.0	Y	0	N	0
14.4	1.0	0.7	20.0	19.0	18.0	Y	0	N	0
15.0	1.0	0.8	26.0	29.0	41.0	Y	0	Y	1

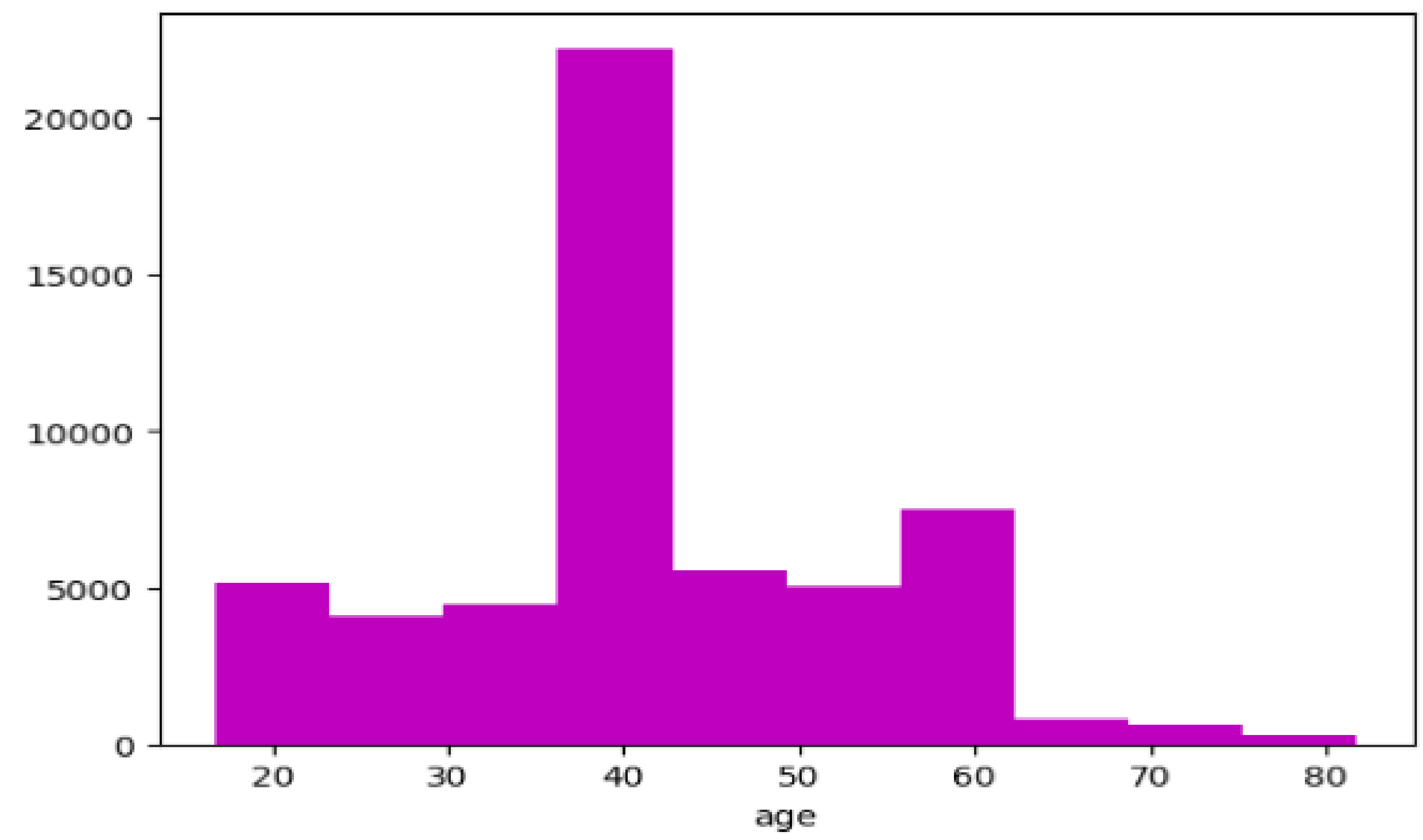
Handling Missing Data

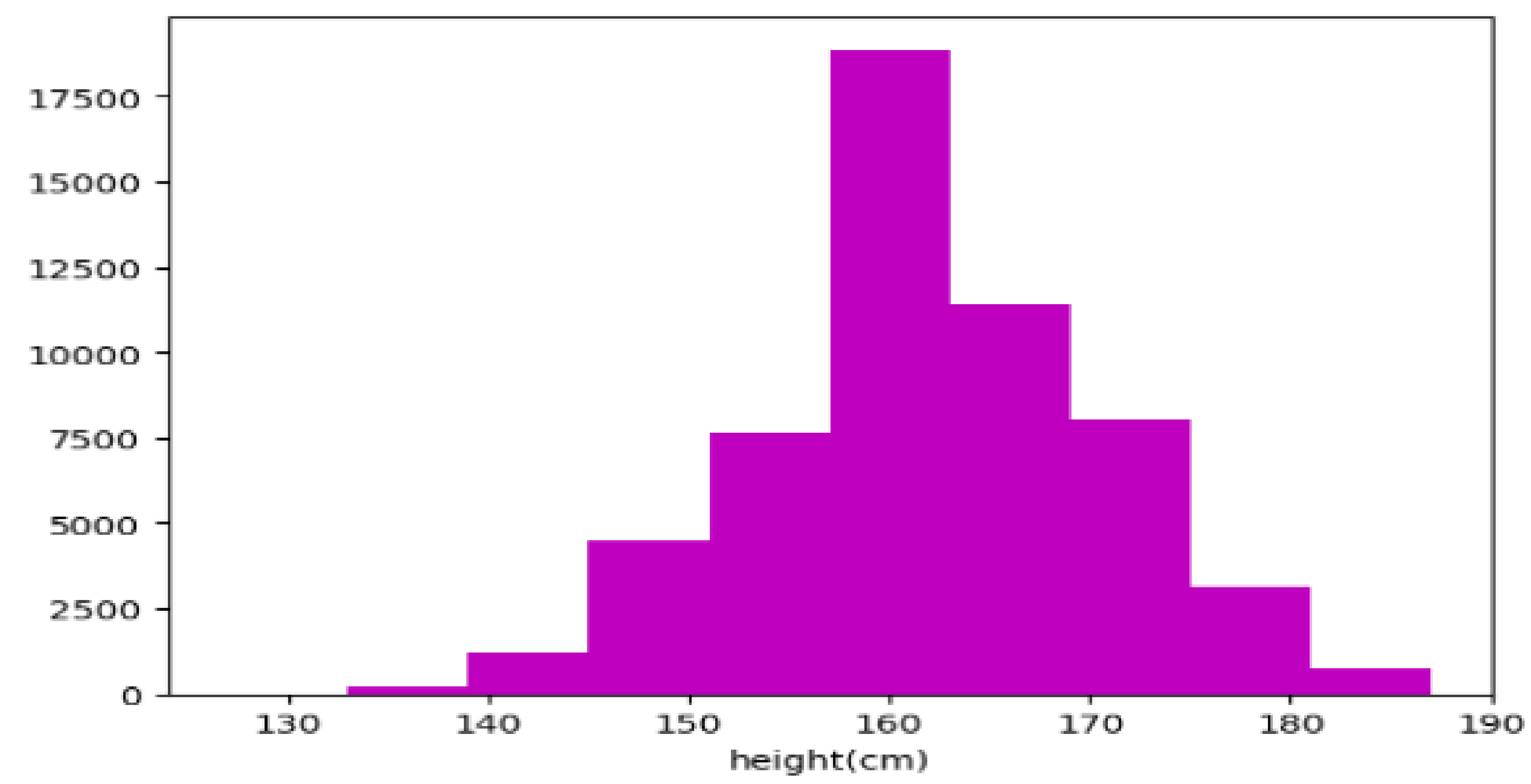
```
ID                     0
gender                 0
age                    0
height(cm)             0
weight(kg)             0
waist(cm)              0
eyesight(left)         0
eyesight(right)        0
hearing(left)          0
hearing(right)         0
systolic               0
relaxation             0
fasting_blood_sugar    0
Cholesterol            0
triglyceride           0
HDL                    0
LDL                    0
hemoglobin             0
Urine_protein          0
serum_creatinine       0
AST                    0
ALT                    0
Gtp                    0
oral                   0
dental_caries          0
tartar                 0
smoking                0
dtype: int64
```

HEAT MAP:



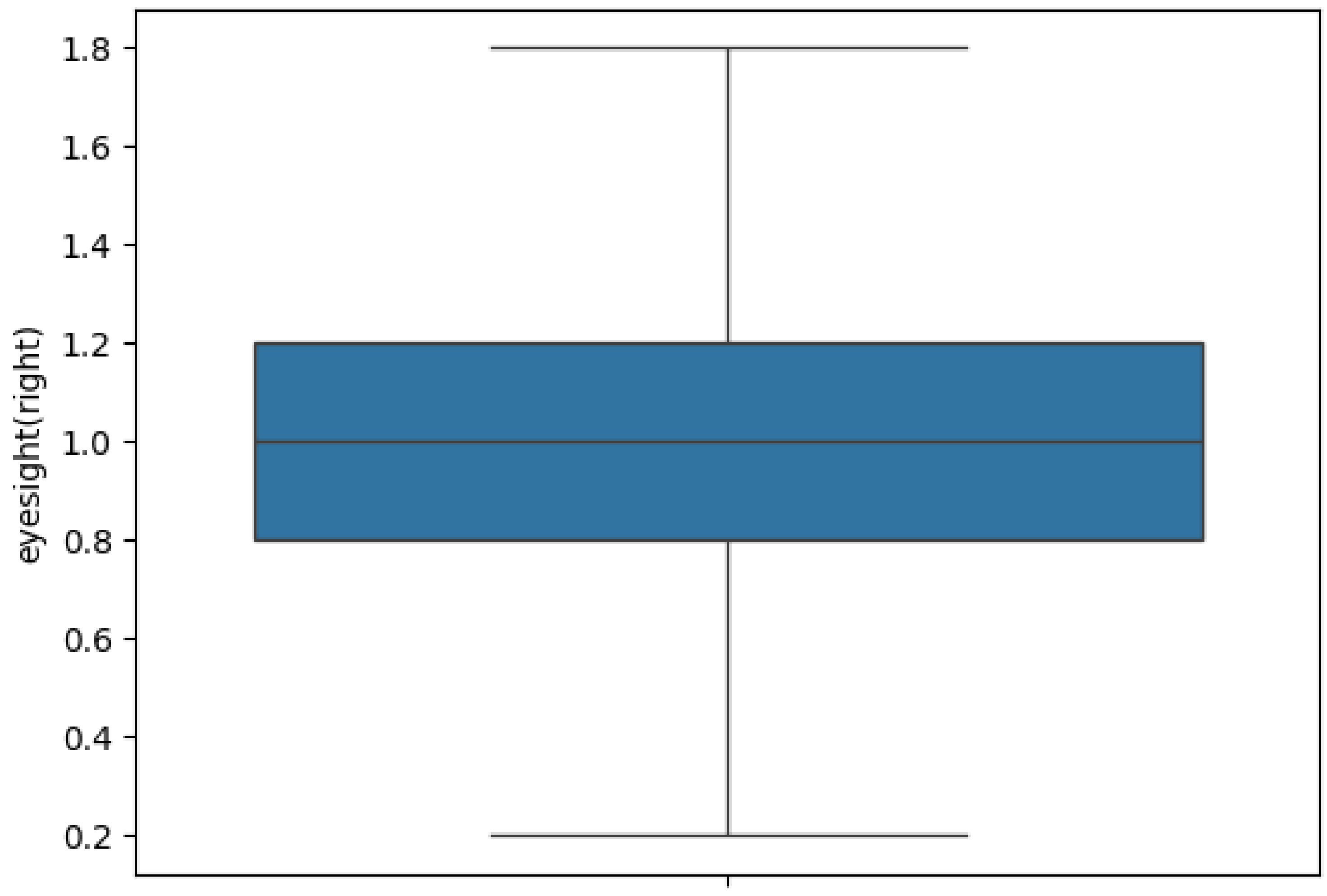
HISTOGRAM:





RECTIFYING OYTLIERS:

```
[ ] IQR=df['eyesight(right)'].quantile(0.75)-df['eyesight(right)'].quantile(0.25)
    IQR

    0.3999999999999999

[ ] lowerBound=df['eyesight(right)'].quantile(0.25)-(1.5*IQR)
    lowerBound

    0.20000000000000018

    upperBound=df['eyesight(right)'].quantile(0.75)+(1.5*IQR)
    upperBound

    1.7999999999999998

[ ] df['eyesight(right)']=np.where(df['eyesight(right)']>upperBound,upperBound,df['eyesight(right)'])
    df['eyesight(right)']=np.where(df['eyesight(right)']<lowerBound,lowerBound,df['eyesight(right)'])

[ ] sns.boxplot(df['eyesight(right)'])
```



DATA PREPROCESSING

```
#Seperating Data For Training and Testing
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
```

```
x=df.drop(['smoking'],axis=1)
y=df['smoking']
```

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.1,random_state=1)
```

```
from sklearn.preprocessing import MinMaxScaler,RobustScaler,StandardScaler
```

```
scaler=MinMaxScaler()
scaler.fit(x_train)
x_train=scaler.transform(x_train)
x_test=scaler.transform(x_test)
```

```
sc=StandardScaler()
sc
```

```
StandardScaler
StandardScaler()
```

```
clf=LogisticRegression()
clf
```

```
LogisticRegression
LogisticRegression()
```

```
clf.fit(x_train,y_train)
```

```
LogisticRegression
LogisticRegression()
```

```
pred=clf.predict(x_test)
pred
```

```
array([1, 0, 1, ..., 1, 0, 1])
```

```
y_test
```

```
13294    1
48549    1
21381    0
18058    0
32466    0
```

```
[ ] clf.score(x_test,y_test)*100

76.31956912028726

[ ] Start coding or generate with AI.

[ ] clf.predict([[1,1,40,155,60,81.3,1.2,1,1,1,114,73,94,215,82,73,126,12.9,1,0.7,18,19,27,1,0,1]])

array([1])

from sklearn.metrics import confusion_matrix
confmat = confusion_matrix(y_test, pred)

# Print the confusion matrix
print(confmat)

[[2790  717]
 [ 602 1461]]
```

HYPER PARAMETER TUNING:

```
from sklearn.model_selection import GridSearchCV

[ ] param_grid = [
        {'penalty' : ['l1', 'l2', 'elasticnet', 'none'],
        'C' : np.logspace(-4, 4, 20),
        'solver' : ['lbfgs','newton-cg','liblinear','sag','saga'],
        'max_iter' : [100, 1000,2500, 5000]
        }
    ]

[ ] gs = GridSearchCV(clf, param_grid = param_grid, cv = 3, verbose=True, n_jobs=-1)

[ ] best_clf=clf.fit(x,y)

[ ] print("Accuracy - :",{best_clf.score(x,y)})

Accuracy - : {0.7519033254327372}
```

LOADING FILE INTO PICKLE MODULE:

```
[ ] import pickle

[ ] filename = 'smoke_model.pkl'

[ ] pickle.dump(clf,open(filename,'wb'))

[ ] pickle.load(open('smoke_model.pkl','rb'))
```