KLE Technological University
Creating Value, Leveraging Knowledge

Dr. M. S. Sheshgiri Campus, Belagavi

Department of
Electronics and Communication Engineering

Mini Project Report

on

SLEEP APNEA DETECTION

By:

1. **VasantKumar B** USN:02FE22BEC116
2. **Abhishek K** USN:02FE22BEC401
3. **Ritin Shimpi** USN:02FE22BEC068
4. **Shivananda A** USN:02FE22BEC089

Semester: V, 2024-25

Under the Guidance of :

Prof. Shivanand Upari

KLE Technological University

Dr. M. S. Sheshgiri Campus, Belagavi

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

This is to certify that project entitled **"SLEEP APNEA DETECTION"** is a bonafide work carried out by the student team of **" Vasanthkumar (02FE22BEC116) , Abhishek k (02FE22BEC401),Ritin(02FE22BEC068),Shivanand A(02FE22BEC089)"**. The project report has been approved as it satisfies the requirements with respect to the mini project work prescribed by the university curriculum for B.E. (V Semester) in Department of Electronics and Communication Engineering of KLE Technological University Dr. M. S. Sheshgiri CET Belagavi campus for the academic year 2024-2025.

Prof. Shivanand Upari **Guide**	**Dr. Dattaprasad A. Torse** **Head of Department**	**Dr. S. F. Patil** **Principal**

External Viva:

Name of Examiners	**Signature with date**
1.	
2.	

ACKNOWLEDGMENT

We would like to express our sincere gratitude to the following individuals and organizations who have played a significant role in the development of the Sleep Apnea Detection project.

We take the opportunity to thank our Principal, Dr. S.F. Patil, for providing us with the opportunity to undertake this project.

We would like to thank our Head of the Department, Dr. Dattaprasad Torse, for his guidance and support throughout the project.

We take the opportunity to thank our project coordinator Prof. UL Naik, for his motivation and encouragement during the development of this project.

We would like to extend our heartfelt gratitude to our esteemed guide Prof. Shivanand Upari, for his valuable insights, keen interest, and concern towards the successful completion of this project.

We acknowledge the contributions of all the team members who participated in this project. Each team member brought unique skills and perspectives, which significantly contributed to the development of different elements of the project.

We extend our thanks to KLE Dr. MSSCET for providing the necessary resources, infrastructure, and funding to carry out this project. Their support was vital in enabling us to conduct the required research, acquire relevant datasets, and access computing resources.

Lastly, we express our gratitude to all the staff members of the Electronics and Communication Department for their cooperation and suggestions during the course of this project.

-The project team

ABSTRACT

A sleep disorder in which by repeated interruptions of normal breathing while sleeping. Sometimes this can lead to serious health issues if not diagnosed. This Machine detection of sleep apnea research is focused. Utilising methods of learning using the Random Forest algorithm for classification purposes, it is for accurate and reliable classification. A pre-recorded dataset It uses different features, such as age, gender, sleep, amongst others. They include quality, sleep duration, physical activity, daily steps and blood. pressure. These parameters are carefully selected to provide a view of the individual's health and lifestyle factors influencing sleep apnea. The second one is known as the Random Forest algorithm. due to its robustness to handle diverse data, processes. It then uses these inputs to predict likelihood considering sleep apnealogically. The goal which is the aim of this Research is to develop an efficient and accessible tool for it contributing to early detection and therefore improved healthcare outcomes by. timely intervention and treatment able.

Contents

List of Figures

Chapter 1

INTRODUCTION

Sleep apnea, a prevalent and often overlooked sleep disorder, poses significant risks to public health and quality of life. This condition involves repeated episodes of breathing cessation during sleep, disrupting the sleep cycle and leading to fragmented rest and decreased oxygenation. Beyond its immediate effects on sleep, untreated sleep apnea is linked to serious health complications, including cardiovascular diseases, stroke, type 2 diabetes, and impaired mental health. Its impact extends to reduced productivity, increased risk of accidents, and a diminished quality of life for affected individuals. Despite its seriousness, sleep apnea remains underdiagnosed due to the challenges associated with traditional diagnostic methods like polysomnography (PSG). While PSG is highly accurate, it demands overnight monitoring in specialized facilities, making it costly and inconvenient for many patients. These barriers have led to a significant gap in diagnosis and treatment, emphasizing the need for more accessible and efficient diagnostic tools. Recent advancements in technology, particularly in machine learning and sensor-based systems, offer promising alternatives to traditional methods. Machine learning algorithms can process complex datasets, identifying subtle patterns indicative of sleep apnea with remarkable accuracy. Additionally, non-invasive monitoring techniques, such as radar-based measurements, have opened new possibilities for early detection and continuous monitoring, making the diagnosis process more patient-friendly and scalable. Together, these innovations have the potential to transform the approach to sleep apnea diagnosis, enabling timely intervention and improving patient outcomes.

1.1 Motivation

The design of this project is motivated by the critical need for early detection and management of sleep apnea, a prevalent sleep disorder that significantly impacts overall health and well-being. Sleep apnea is often underdiagnosed due to the limitations of traditional diagnostic methods, which are time-consuming, costly, and require specialized medical facilities. By leveraging machine learning algorithms and easily accessible physiological data, this project aims to provide a cost-effective and efficient solution for detecting sleep apnea. The use of ECG-based parameters and a Random Forest classifier enables a robust approach to identify patterns associated with sleep apnea, ensuring early intervention and better patient outcomes.

1.2 Objectives

- To develop a machine learning-based model for the early detection of sleep apnea using physiological data.

- To utilize ECG signals and respiratory data to extract relevant features for accurate classification.
- To implement a Random Forest classifier for distinguishing between normal and apnea episodes with high precision.
- To evaluate the performance of the proposed model in terms of accuracy, sensitivity, and specificity.

1.3 LITERATURE REVIEW

1.3.1 Analysis and Identification of Sleep Apnea by availing Random Forest and SVM

The paper explores the analysis and identification of sleep apnea using a hybrid approach combining Random Forest and Support Vector Machine (SVM) algorithms. Random Forest enhances classification accuracy by leveraging multiple decision trees, while SVM contributes robust decision boundaries for improved reliability. The integration of these techniques aims to boost prediction performance but introduces higher computational costs, particularly for large datasets. This computational intensity poses challenges to scalability, especially on low-resource systems, emphasizing the need for efficient opti- mization strategies.

1.3.2 An Effective Method for Distinguishing Breathing and Sleep Apnea Detection and Prevention using Python

The paper presents a method for distinguishing breathing patterns and detecting infant sleep apnea using Python. Feature extraction and machine learning are employed to predict sleep apnea occurrences from ECG data, with Random Forest enhancing prediction robustness. t-SNE is utilized to visualize high-dimensional ECG and HRV data, aiding in pattern recognition. However, the complexity of high- dimensional data necessitates dimensionality reduction, which may lead to information loss. Additionally, the model's effectiveness relies on high-quality ECG signals, posing challenges in noisy or low-quality environments.

1.3.3 Obstructive Sleep Apnea Detection Using SVM-Based Classification of ECG Signal

Features The paper focuses on detecting obstructive sleep apnea by employing SVM- based classification of hybrid ECG signal features. A linear kernel function in SVM is used to distinguish between apnea and non-apnea events from ECG-derived data. While this approach aims to improve classification accuracy, the linear kernel may struggle with capturing complex, nonlinear patterns in the data. Furthermore, the model's performance is highly dependent on effective feature selection, as suboptimal choices can significantly reduce accuracy.

1.3.4 Radar-Based Automatic Detection of Sleep Apnea Using Support Vector Machine

The paper investigates a radar-based, noncontact approach for the automatic detection of sleep apnea using Support Vector Machine (SVM) algorithms. Traditional methods, such as

polysomnography (PSG), require numerous sensors to be attached to patients, potentially influencing test results due to psychological effects. Radar-based measurements eliminate this drawback, offering a promising alternative for sleep apnea detection.
Unlike previous studies that predominantly used time-domain features, this approach integrates frequency-domain features with SVM to improve the detection of apnea events. The study highlights the ability of radar systems to differentiate between central sleep apnea (CSA) and obstructive sleep apnea (OSA), with CSA events being accurately identified. However, OSA events posed challenges due to their complex characteristics. Simultaneous measurements using PSG and radar systems validated the effectiveness of the proposed method, while further research is planned to refine detection accuracy across different apnea types.

PROBLEM STATEMENT

Sleep apnea is a serious condition that disrupts normal breathing during sleep, leading to various health complications. Traditional detection methods are often invasive, costly, and time-consuming. Machine learning offers a promising solution by leveraging prerecorded data for automated and accurate detection. This project aims to design a system using machine learning algorithms, specifically a Random Forest Classifier, to detect sleep apnea based on ECG and respiratory data.

1.4 Application in Societal Context

The development of a machine learning-based system for sleep apnea detection has significant societal applications, improving healthcare accessibility and quality of life:

1. **Healthcare Improvement:** The proposed system enables early and accurate detection of sleep apnea, reducing the dependency on costly and invasive diagnostic methods like polysomnography (PSG). This promotes timely intervention, preventing severe health complications such as cardiovascular diseases and diabetes.

2. **Wearable Technology:** The integration of radar-based and machine learning systems into wearable devices or home monitoring solutions allows for continuous, non-invasive health monitoring. This technology supports personalized healthcare and reduces the need for hospital visits.

3. **Rural and Remote Healthcare:** By eliminating the need for specialized facilities and equipment, the proposed system makes sleep apnea diagnosis more accessible in rural and remote areas, bridging gaps in healthcare infrastructure.

4. **Workplace Productivity:** Early detection and treatment of sleep apnea improve sleep quality, leading to enhanced focus, productivity, and overall well-being in professional settings.

5. **Public Awareness and Preventive Care:** The availability of non-invasive and user-friendly solutions raises awareness about sleep apnea and encourages individuals to prioritize their sleep health. This fosters a preventive approach to managing sleep disorders.

6. **Cost Efficiency in Healthcare:** By reducing the reliance on expensive sleep labs and in-hospital stays, this system significantly cuts down healthcare costs for both patients and providers, making healthcare more sustainable.

Chapter 2

Project Planning

2.0.1 Project Planning:

1. **Project Overview:** Sleep apnea is a serious sleep disorder that often goes undiagnosed due to the challenges associated with traditional diagnostic methods such as polysomnography (PSG). This project aims to develop a machine learning-based system that leverages physiological data, such as ECG signals, to detect sleep apnea events. The system utilizes a Random Forest classifier for accurate classification and is designed to provide a non-invasive, cost-effective, and efficient solution.

2. **Objectives:** The primary objective of this project is to design and implement a system capable of identifying sleep apnea events from ECG and respiratory data. This includes extracting meaningful features from the data, training and testing the Random Forest algorithm, and evaluating its performance in terms of accuracy, sensitivity, and specificity. Additionally, the system is designed to handle both central sleep apnea (CSA) and obstructive sleep apnea (OSA) detection, with future scalability for real-time monitoring.

3. **Testing and Validation:** For testing and validation, the system will be trained using a publicly available ECG dataset annotated with sleep apnea events. The dataset will be preprocessed to extract both time-domain and frequency-domain features. The Random Forest model will be evaluated using metrics such as accuracy, precision, recall, and F1-score to ensure reliable performance. Testing will involve validating the system's ability to differentiate between normal sleep, CSA events, and OSA events. Results will be compared with baseline methods to establish the model's effectiveness. Challenges such as imbalanced datasets and noise in physiological signals will be addressed through appropriate preprocessing and data augmentation techniques.

4. **Documentation:** The documentation will include all aspects of the project, starting with an introduction to sleep apnea and the significance of early detection. It will cover the methodology, including data acquisition, preprocessing, feature extraction, and model training. Detailed explanations of the Random Forest classifier, its parameter selection, and performance evaluation will be included. The report will present simulation results, highlighting the model's performance across different scenarios, along with a comparison to existing approaches. The documentation will also include challenges encountered and solutions implemented during the project. Finally, a conclusion summarizing the findings, along with future directions, such as real-time implementation and the use of additional physiological parameters, will be provided.

2.0.2 Bill of Materials

Since this project primarily involves software development and simulation, there is no physical hardware component. Therefore, a bill of materials is not applicable. All work is conducted using tools such as Python, scikit-learn, and publicly available ECG datasets.

2.1 Organization of the report

- Chapter 2 describes the Project planning.
- Chapter 3 System Design And Algorithm.
- Chapter 4 Describes the Results and discussions of the final architecture.
- Chapter 5 gives the conclusion and future scope.

2.2 Gantt chart

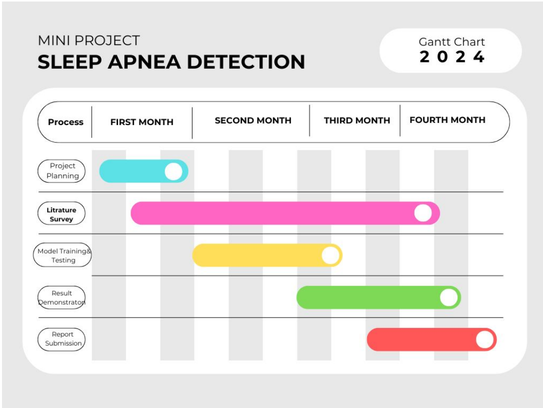

Figure 2.1: gantt chart

A Gantt chart is a widely used project management tool that provides a visual representation of a project's timeline and tasks. It organizes tasks into a horizontal timeline where each task is depicted as a bar, with the bar's position and length indicating the start date, end date, and duration of the task. This visual structure helps teams and project managers understand the sequence of tasks, identify dependencies between activities, and track progress against the planned schedule.

2.3 Work Breakdown Structure (WBS)

Figure 2.2: WBS

A Work Breakdown Structure (WBS) is a project management tool that divides a project into smaller, manageable components or tasks. It organizes the work into a hierarchy, making it easier to plan, assign resources, and track progress. The WBS helps clarify the scope of the project, ensuring that all required work is identified.

Chapter 3

System Design And Algorithm

3.1 GENERALIZED BLOCK DIAGRAM:

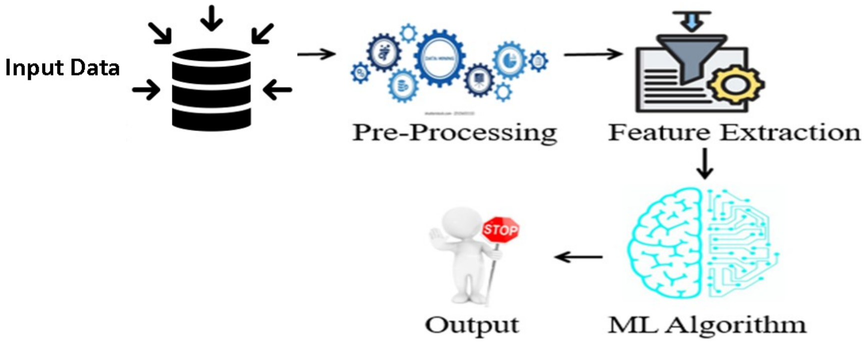

Figure 3.1: REPRESENTS GENERALIZED BLOCK DIAGRAM

A generalized block diagram for machine learning applications typically consists of several key stages that work together to transform raw input data into meaningful output predictions or classifications. The first stage is

3.1.1 Input Data

where raw information is collected from various sources such as sensors, databases, or user inputs. This data can take different forms, including numerical values, text, images, or time-series signals, depending on the application.

3.1.2 Preprocessing

which ensures the data is clean, consistent, and ready for further processing. This step involves techniques like noise removal, normalization, scaling, and handling missing values. Preprocessing is essential to improve data quality and ensure that irrelevant or redundant information does not affect the performance of the system.

3.1.3 Feature Extraction

where relevant and discriminative characteristics are identified from the input data. This step reduces the dimensionality of the dataset by selecting only the most informative features, making it easier for the machine learning algorithm to understand patterns and relationships. Techniques such as Principal Component Analysis (PCA), wavelet transforms, or statistical methods are often employed here.

3.1.4 Machine Learning Algorithm

which learns patterns from the processed data to make predictions or classifications. Depending on the problem, algorithms such as Random Forest, Support Vector Machines (SVM), or neural networks are used. This stage involves training the model on labeled data, tuning its parameters, and validating its performance to ensure accuracy and reliability.

3.1.5 Output

stage generates results based on the machine learning model's predictions. This output can take various forms, such as classifications (e.g., spam or ham in email filtering), numerical predictions, or alerts in real-time systems. The results are then presented to the user or used in automated systems for decision-making. This entire flow, from input data to output, forms a robust framework for solving complex problems across various domains.

3.2 BLOCK DIAGRAM:

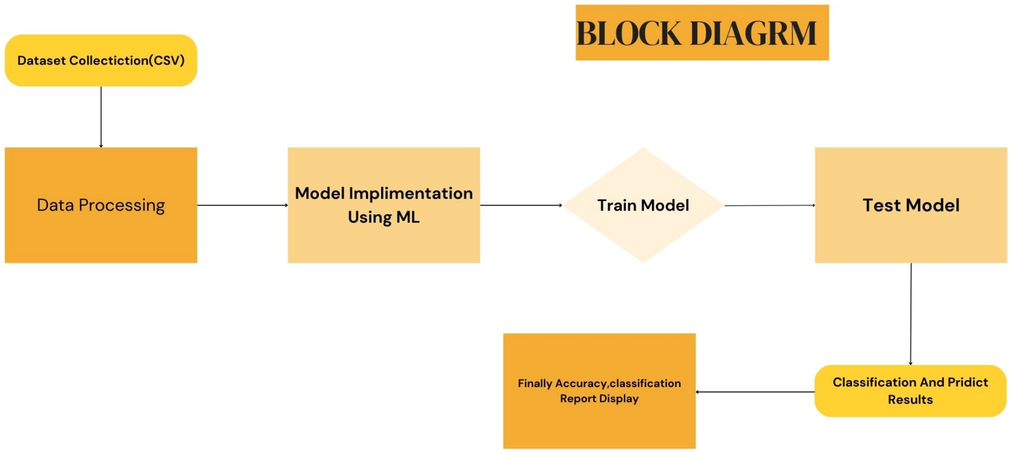

Figure 3.2: REPRESENTS BLOCK DIAGRAM

3.2.1 Data Collection

The dataset for the project is collected from a reliable source containing sleep health and lifestyle data. This includes attributes like age, gender, blood pressure, BMI category, and sleep disorder labels. These parameters provide the foundation for identifying patterns associated with sleep apnea.

3.2.2 Data Processing

The collected data undergoes preprocessing to ensure its quality and usability. This involves handling missing values, encoding categorical variables (e.g., gender and sleep disorder), performing one-hot encoding for BMI categories, and normalizing numerical data. Additionally, irrelevant columns such as occupation and stress levels are removed to focus on significant features.

3.2.3 Model Implementation Using ML

Various machine learning algorithms, including Random Forest, SVM, KNN, and Ad- aBoost, are implemented. These models are configured to learn from the data patterns and classify whether a person is likely to have sleep apnea. Hyperparameter tuning is employed for the Random Forest algorithm to optimize its performance.

3.2.4 Training Model

The training process uses the preprocessed data, with the target variable being the sleep disorder label. Techniques like SMOTE are applied to address class imbalances, ensuring the model learns effectively from underrepresented classes. Scaled features enhance the training performance of algorithms.

3.2.5 Testing Model

The trained model is evaluated using a separate test dataset. This phase validates the model's performance on unseen data, providing insights into its real-world applicability. Predictions are compared against actual values to compute evaluation metrics.

3.2.6 Results

The model's predictions are analyzed through metrics such as accuracy, precision, recall, and F1-score. A confusion matrix visualizes the classification outcomes, indicating the number of correct and incorrect predictions.

3.2.7 Accuracy and Classification Results Display

The final results are displayed as graphical comparisons of model metrics, highlighting the performance of each algorithm. The accuracy of each model is prominently showcased, while the classification report demonstrates the model's dependability in detecting sleep apnea cases.

3.3 Algorithms:

Machine learning algorithms play a pivotal role in automating the detection and classification of sleep apnea based on various health parameters. Among the diverse range of algorithms, the Random Forest algorithm is particularly well-suited for this task due to its ability to handle complex and multidimensional datasets. The Random Forest algorithm is an ensemble learning method that constructs a collection of decision trees during training and outputs the class that is the mode of the classes for classification tasks. This approach offers high accuracy and resilience against overfitting by combining the results of multiple trees, thus improving the overall model performance.

Random Forest's ability to handle both numerical and categorical data makes it ideal for processing a variety of health-related parameters, such as age, gender, sleep quality, blood pressure, and physical activity, which are typically used in sleep apnea detection. Additionally, Random Forest provides feature importance scores, offering valuable insights into which parameters most significantly contribute to the prediction of sleep apnea. By analyzing large datasets with these features, Random Forest can identify patterns that help predict the likelihood of sleep apnea, offering a more efficient and accessible diagnostic tool compared to traditional methods like polysomnography.

3.4 Python Code for Sleep Apnea Detection

3.4.1 Dataset Preprocessing

```
# Load and preprocess dataset
data = pd.read_csv('dataset.csv')
data[['SystolicBP','DiastolicBP']]=data['Blood Pressure'].str.split(expand=True)
data['Gender'] = LabelEncoder().fit_transform(data['Gender'])
data['Sleep Disorder'] = LabelEncoder().fit_transform(data['Sleep Disorder'])
data = pd.get_dummies(data, columns=['BMI Category'], drop_first=True)
X = data.drop(columns=['Sleep Disorder'])
y = data['Sleep Disorder']
```

3.4.2 Data Splitting and Balancing

```
# Split and balance data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, stratify=y)
X_train_smote, y_train_smote = SMOTE().fit_resample(X_train, y_train)
scaler = StandardScaler()
X_train_smote = scaler.fit_transform(X_train_smote)
X_test = scaler.transform(X_test)
```

3.4.3 Training and Saving the Model

```
# Train and save Random Forest model
clf = RandomForestClassifier(n_estimators=200, max_depth=10, random_state=42)
clf.fit(X_train_smote, y_train_smote)
pickle.dump(clf, open("random_forest_model.pkl", "wb"))
```

3.4.4 Model Evaluation

```
# Evaluate Random Forest
y_pred = clf.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred, average='weighted')
```

3.4.5 Real-Time Prediction

```
# Predict sleep disorder
def predict_sleep_disorder(input_data):
    input_scaled = scaler.transform(input_data)
    return clf.predict(input_scaled)
```

Chapter 4

RESULTS

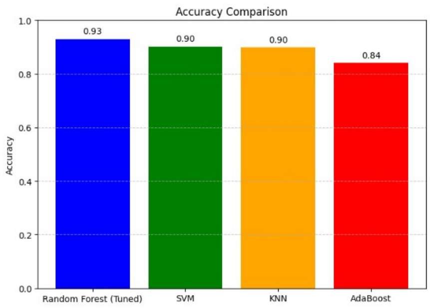

Figure 4.1: Accuracy

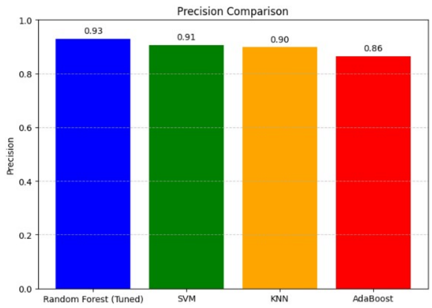

Figure 4.2: Precision

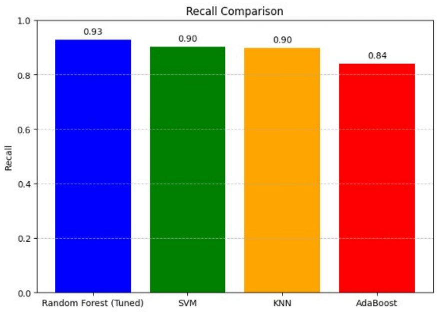

Figure 4.3: Recall

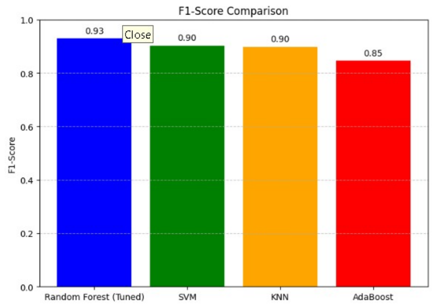

Figure 4.4: F-Measure

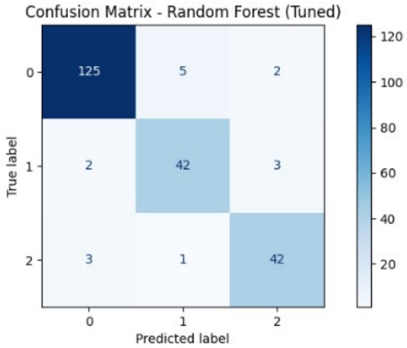

Figure 4.5: Confusion Matrix

4.1 Result Comparision Table

Table 4.1: Comparison of Various Classifiers with Multiple Metrics

.

Metric	Random Forest	SVM	KNN	AdaBoost
Accuracy	97.94	98.12	97.94	96.86
F-Measure	0.926	0.931	0.926	0.887
Precision	0.960	0.986	0.966	0.920
Recall	0.894	0.882	0.888	0.857

The table classifier comparison provides a comparative analysis of various classifiers—Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and AdaBoost—using multiple performance metrics. SVM achieves the highest accuracy of 98.12%, closely followed by Random Forest and KNN with 97.94%, while AdaBoost trails at 96.86%. In terms of the F-measure, SVM leads with 0.931, indicating a better balance between precision and recall compared to others. Precision is also highest for SVM at 0.986, reflecting its effectiveness in correctly identifying positive instances. However, Random Forest excels in recall at 0.894, showcasing its ability to identify a higher proportion of actual positives. These metrics highlight SVM's superior performance overall, while other classifiers exhibit strengths in specific metrics, demonstrating their applicability to varying scenarios.

Chapter 5

Conclusion

The era of knowledge and technology is upon us. People are now abandoning the old forms of communication. Spam is a severe and vexing issue that affects these information and communication channels. The dataset was classified using a variety of methods: Indian material was included to our modified SMS spam collection dataset. According to the findings, Multinomial Naive Bayes and Support Vector Machine are two of the top classifiers for detecting SMS spam. Although it took a long time to complete, the classifier SVM with linear kernel achieved the highest accuracy. Conversely, MNB with Laplace smoothing had accuracy that was quite similar to SVM, but it took a lot less time than SVM. The greatest outcomes

5.1 Future Scope

The Sleep Apnea Detection system has significant potential for future enhancements and applications. One major direction is the integration of additional physiological parameters, such as SpO2 levels, heart rate variability, and body movements, to improve the model's accuracy and robustness. Real-time data acquisition and analysis using wearable devices or IoT-based platforms can enable continuous monitoring, making the system more practical for home use. Finally, the system can be expanded to detect other sleep disorders and provide actionable insights, fostering better sleep health and preventive care in diverse populations.

Bibliography

- P. Asha, V. A. Yadav, Y. Hanumanthu, and V. A. Mary, "Analysis and Identification of Sleep Apnea by availing Random Forest and SVM," in *2023 IEEE International Conference on Circuit Power and Computing Technologies (ICCPCT)*, Kollam, India, 2023.
- M. A. Bennet, K. J. Subha, R. Kumutha, "An Effective Method for Distinguishing Breathing and Infant Sleep Apnea Detection and Prevention using Python," in *2022 IEEE International Conference on Computer, Power and Communications (ICCPC)*, Chennai, India, 2022.
- A. Bhongade, R. Gupta, and T. K. Gandhi, "Automatic Detection of Sleep Apnea from Single-lead ECG Signal Using Machine Learning," in *2022 IEEE International Conference on Futuristic Technologies (INCOFT)*, 2022.
- M. Olsen, J. M. Zeitzer, R. N. Richardson, V. H. Musgrave, H. B. Sørensen, E. Mignot, and P. J. Jennum, "A deep transfer learning approach for sleep stage classification and sleep apnea detection using wrist-worn consumer sleep technologies," *IEEE Transactions on Biomedical Engineering*, 2024.