

Making Sleep Disorder Classification Using Optimized Machine Learning Models

*A Project Report submitted in the partial
fulfillment of the Requirements for the
award of the degree*

BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING

Submitted by

M. Dileep Kumar	(21471A05A2)
D. Abhishikth	(21471A0571)
Sk. Alimoon	(21471A05C5)

Under the esteemed guidance of

Y.Chandana, M.Tech.,

Assistant Professor



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

**NARASARAOPETA ENGINEERING COLLEGE: NARASAROPET
(AUTONOMOUS)**

Accredited by NAAC with A+ Grade and NBA under
Tyre -1 NIRF rank in the band of 201-300 and an
ISO 9001:2015 Certified

Approved by AICTE, New Delhi, Permanently Affiliated to JNTUK, Kakinada
KOTAPPAKONDA ROAD, YALAMANDA VILLAGE,
NARASARAOPET- 522601

2024-2025

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)
DEPARTMENT OF COMPUTER SCIENCE AND
ENGINEERING



CERTIFICATE

This is to certify that the project that is entitled with the name **“Making Sleep Disorder Classification Using Optimized Machine Learning Models ”** is a bonafide work done by the team **M. Dileep Kumar (21471A05A2), D. Abhishikth (21471A0571), Sk. Alimoon (21471A05C5)** in partial fulfillment of the requirements for the award of the degree of **BACHELOR OF TECHNOLOGY** in the Department of **COMPUTER SCIENCE AND ENGINEERING** during 2024-2025.

PROJECT GUIDE

Y.Chandana, M.Tech.
Asst. Professor

PROJECT CO-ORDINATOR

Dr. M. Sireesha, M.Tech.Ph.D.,
Assoc. Professor

HEAD OF THE DEPARTMENT

Dr. S. N. Tirumala Rao, M.Tech., Ph.D.,
Professor & HOD

EXTERNAL EXAMINER

DECLARATION

We declare that this project work titled "**MAKING SLEEP DISORDER CLASSIFICATION USING OPTIMIZED MACHINE LEARNING MODELS**" is composed by ourselves that the work contain here is our own except where explicitly stated otherwise in the text and that this work has not been submitted for any other degree or professional qualification except as specified.

M. Dileep Kumar (21471A05A2)

D. Abhishikth (21471A0571)

Sk. Alimoon (21471A05C5)

ACKNOWLEDGEMENT

We wish to express my thanks to carious personalities who are responsible for the completion of the project. We are extremely thankful to our beloved chairman sri **M. V. Koteswara Rao**, B.Sc., who took keen interest in us in every effort throughout thiscourse. We owe out sincere gratitude to our beloved principal **Dr. S. Venkateswarlu**, Ph.D., for showing his kind attention and valuable guidance throughout the course.

We express our deep felt gratitude towards **Dr. S. N. Tirumala Rao**, M.Tech., Ph.D., HOD of CSE department and also to our guide **Y.Chandana**, M.Tech , of CSE department whose valuable guidance and unstinting encouragement enable us to accomplish our project successfully in time.

We extend our sincere thanks towards **Dr. M. Sireesha**, M.Tech.,Ph.D., Associate professor & Project coordinator of the project for extending her encouragement. Their profound knowledge and willingness have been a constant source of inspiration for us throughout this project work.

We extend our sincere thanks to all other teaching and non-teaching staff to department for their cooperation and encouragement during our B.Tech degree.

We have no words to acknowledge the warm affection, constant inspiration and encouragement that we received from our parents.

We affectionately acknowledge the encouragement received from our friends and those who involved in giving valuable suggestions had clarifying out doubts which had really helped us in successfully completing our project.

By

M. Dileep Kumar (21471A05A2)

D.Abhishikth (21471A0571)

Sk. Alimoon (21471A05C5)



INSTITUTE VISION AND MISSION

INSTITUTION VISION

To emerge as a Centre of excellence in technical education with a blend of effective student centric teaching learning practices as well as research for the transformation of lives and community,

INSTITUTION MISSION

M1: Provide the best class infra-structure to explore the field of engineering and research

M2: Build a passionate and a determined team of faculty with student centric teaching, imbining experiential, innovative skills

M3: Imbibe lifelong learning skills, entrepreneurial skills and ethical values in students for addressing societal problems



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VISION OF THE DEPARTMENT

To become a center of excellence in nurturing the quality Computer Science & Engineering professionals embedded with software knowledge, aptitude for research and ethical values to cater to the needs of industry and society.

MISSION OF THE DEPARTMENT

The department of Computer Science and Engineering is committed to

M1: Mould the students to become Software Professionals, Researchers and Entrepreneurs by providing advanced laboratories.

M2: Impart high quality professional training to get expertise in modern software tools and technologies to cater to the real time requirements of the industry.

M3: Inculcate team work and lifelong learning among students with a sense of societal and ethical

Program Specific Outcomes (PSO's)

PSO1: Apply mathematical and scientific skills in numerous areas of Computer Science and Engineering to design and develop software-based systems.

PSO2: Acquaint module knowledge on emerging trends of the modern era in Computer Science and Engineering.

PSO3: Promote novel applications that meet the needs of entrepreneur, environmental and social issues.

Program Educational Objectives (PEO's)

The graduates of the programme are able to:

PEO1: Apply the knowledge of Mathematics, Science and Engineering fundamentals to identify and solve Computer Science and Engineering problems.

PEO2: Use various software tools and technologies to solve problems related to academia, industry and society.

PEO3: Work with ethical and moral values in the multi-disciplinary teams and can communicate effectively among team members with continuous learning.

PEO4: Pursue higher studies and develop their career in software industry.

Program Outcomes

1. **Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem analysis:** Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. **Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and sustainability:** Understand the impact of the professional engineering

solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. **Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
12. **Life-long learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Project Course Outcomes (CO'S):

CO421.1: Analyze the System of Examinations and identify the problem.

CO421.2: Identify and classify the requirements.

CO421.3: Review the Related Literature

CO421.4: Design and Modularize the project

CO421.5: Construct, Integrate, Test and Implement the Project.

CO421.6: Prepare the project Documentation and present the Report using appropriate method.

Course Outcomes – Program Outcomes mapping

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
C421.1		✓											✓		
C421.2	✓		✓		✓								✓		
C421.3				✓		✓	✓	✓					✓		
C421.4			✓			✓	✓	✓					✓	✓	
C421.5					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
C421.6									✓	✓	✓		✓	✓	

Course Outcomes – Program Outcome correlation

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
C421.1	2	3											2		
C421.2			2		3								2		
C421.3				2		2	3	3					2		
C421.4			2			1	1	2					3	2	
C421.5					3	3	3	2	3	2	2	1	3	2	1
C421.6									3	2	1		2	3	

Note: The values in the above table represent the level of correlation between CO's and PO's:

- Low level
- Medium level
- High level

Project mapping with various courses of Curriculum with Attained PO's:

Name of the course from which principles are applied in this project	Description of the device	Attained PO
C2204.2, C22L3.2	Gathering the requirements and defining the problem, plan to develop a model for predicting Sleep Disorder using GA and GNB	PO1, PO3
CC421.1, C2204.3, C22L3.2	Each and every requirement is critically analyzed, the process model is identified	PO2, PO3
CC421.2, C2204.2, C22L3.3	Logical design is done by using the unified modelling language which involves individual team work	PO3, PO5, PO9
CC421.3, C2204.3, C22L3.2	Each and every module is tested, integrated, and evaluated in our project	PO1, PO5
CC421.4, C2204.4, C22L3.2	Documentation is done by all our four members in the form of a group	PO10
CC421.5, C2204.2, C22L3.3	Each and every phase of the work in group is presented periodically	PO10, PO11
C2202.2, C2203.3, C1206.3, C3204.3, C4110.2	Implementation is done, and the project will be utilized for automated sleep disorder classification	PO4, PO7
C32SC4.3	The physical design includes a website to classify sleep disorders using machine learning	PO5, PO6

ABSTRACT

Sleep disorders have a significant impact on human health and early diagnosis is crucial for the improvement of quality of life. In this work, machine learning models that classify sleep disorders have been developed from the data set of 374 participants that comprises features relating to sleep health and lifestyle, such as sleep duration, physical activity, stress levels, and BMI. It has used many machine learning algorithms, like GNB, KNN, SVM, RF, Logistic Regression, and ANN/MLP Classifier. Dataset was highly preprocessed to deal with missing values and normalize features, thereby offering the best performance to the models. Feature engineering and optimization through a Genetic Algorithm improved the predictive ability of these models. GA was highly effective especially in the extraction of the most relevant features, improving the classification accuracy, and dealing with the problem of small dataset size. The best accuracy by GNB was achieved in combination with GA, namely 94%, compared to all other models: SVM 93%, Logistic Regression 92%, and ANN 93%. Importantly, issues like dealing with missing data and feature extraction, where meaningful features such as systolic and diastolic blood pressure levels had to be extracted, were addressed in order to enhance the outcome. Our results show that the proposed model effectively detects patterns in the database, hence providing consequent results that can be used in real-time by health care professionals to diagnose sleep-related conditions. This consequently decreases manual effort towards diagnosis and improves its timing. It also allows for scalability, which may fit easily into wearable health devices and clinical decision systems that are intended to prevent long-term health risks due to sleep disorders left untreated. The proposed model facilitates automated classification of sleep disorders. This will be a valuable tool for advancing healthcare outcomes and fostering future innovations in sleep medicine.

INDEX

S.NO.	CONTENT	PAGE NO
1.	Introduction	01
2.	Literature Survey	05
3.	Existing System	07
4.	Proposed System	09
5.	System Requirements	13
	5.1 Hardware Requirements	13
	5.2 Software Requirements	13
6.	System Analysis	14
	6.1 Scope of the Project	14
	6.2 Analysis	15
	6.3 Data Preprocessing	16
	6.4 Feature Selection and Optimization	17
	6.5 Model Building	18
	6.6 Comparative Analysis	21
7.	Design	25
8.	Implementation	28
9.	Result Analysis	47
10.	Test Cases	51
11.	User Interface	53
12.	Conclusion	54
13.	Future Scope	55
14.	References	56

LIST OF FIGURES

S.NO.	LIST OF FIGURES	PAGE NO
1.	Fig 1.1 Sleep Disorder Statistics	02
2.	Fig 1.2 Sleep Disorder by region	03
3.	Fig 3.1 Existing Model process	08
4.	Fig 4.1 Flow chart of proposed system	10
5.	Fig 6.1 Dataset Description	16
6.	Fig 6.2 Corelation Matrix	17
7.	Fig 7.1 Sleep Disorder Classification Flowchart	26
8.	Fig 9.1 Confusion Matrix	48
9.	Fig 9.2 ROC Curve of GNB	49
10.	Fig 9.3 Model Performance Metrics	50
11.	Fig 10.1 Identified “Sleep Apnea Disease”	51
12.	Fig 10.2 Identified “Insomnia Disease”	51
13.	Fig 10.3 Identified “None Disorder”	52
14.	Fig 10.4 Form Validation	52
15.	Fig 11.1 Home Screen	53
16.	Fig 11.2 Disease Prediction Form	53

LIST OF TABLES

S.NO.	LIST OF TABLES	PAGE NO
1.	Table.9.1 Evaluation Metrics of Models with Genetic Algorithm	48

1. INTRODUCTION

Recently, sleeping disorders have emerged as a serious public health issue due to findings where sleep quality and duration directly impact the total well-being of an individual. Once a person's normal sleep pattern becomes disturbed, they begin to suffer from very critical health problems like heart diseases and metabolic syndrome or even cognitive impairment. The most common sleeping disorder is the most common form of sleep apnea. In general, it is described as the chronic disruption of sleep ventilation mainly due to upperairway obstruction. Untreated sleep apnea may lead to possibly life-threatening complications that come from health conditions such as heart disease, stroke, or hypertension. Overall treatment for sleep apnea remains Continuous Positive Airway Pressure, whereby one keeps the airway open throughout the night. More severe conditions may require surgical interference with the treatment of obstruction within the airway on a permanent basis. CPAP has proven to be one of the strongest treatments, and its symptoms have reduced the complications and problem of sleep apnea in many cases [1].

Insomnia is yet another very common sleep disorder. It makes falling asleep difficult, staying asleep, and waking up too early. There are two types: acute and chronic insomnia. Former usually does not last long. But the latter requires long time treatment; though pharmacological along with behavioural therapy is commonly used for treatment of chronic insomnia. Among the latter there is one type of cognitive behavioural therapy. This therapy, combined with the enhancement of sleep hygiene practice has been proven to result in considerable success in the treatment of the condition [2].

In the past couple of years, technology has improved much-detection and management of sleep disorders. For instance, the design for portable diagnostic systems, usually based on CNNs, has transformed big classification techniques of sleep disorders. Such designs provide much more accurate and accessible tools for assessing sleep disorders outside the traditional clinical settings. This means that patients can now seek timely interventions to enhance their long-term outcomes [3]. Ensemble machine learning algorithms have also played a major role in the advancement of developing the diagnostic accuracy for sleep disorders. Algorithm that can classify a complex pattern

of sleep can become highly useful for predictions and diagnosis related to sleep disorders. This has held special value in the differentiation of sleep disorders that are similar, but will have different treatment modalities [4].

These technological developments in practice would bring much improvement in diagnosis and treatments. Such developments do heighten the quality of lives of afflicted people suffering from sleep disorders, but they also open pathways for future discoveries in sleep medicine. The advancement of diagnostic tools and further research on treatment methodologies can help a healthcare provider deal better with the complexities of sleep-related disorders [5]. Hence, this approach can lead to better care and improved long-term health outcomes.

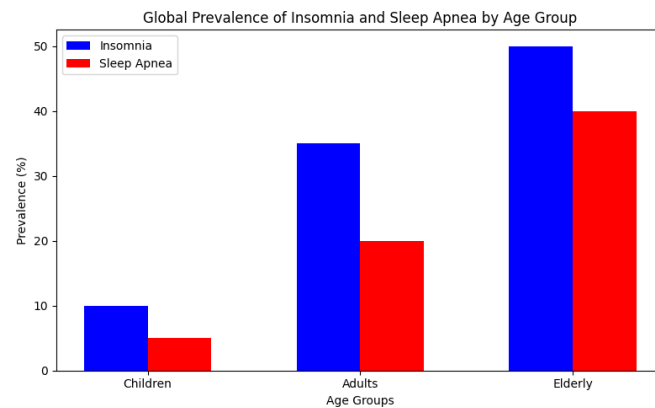


Fig 1.1 Sleep Disorders Statistics

Fig 1.1 illustrating the global prevalence of insomnia and sleep apnea across different age groups: Children, Adults, and Elderly. The prevalence of insomnia is represented in blue, while sleep apnea is shown in red. The chart indicates that both conditions increase with age, with the elderly having the highest prevalence of insomnia (50%) and sleep apnea (40%). Adults show intermediate values, while children have the lowest prevalence.

In children, they may arise due to anxiety, irregular sleep schedules, or medical conditions. Adults often experience sleep disorders due to stress, lifestyle choices, or underlying health issues, while in the elderly, age-related changes, chronic illnesses, and medication side effects further exacerbate sleep disturbances. Addressing these issues requires lifestyle modifications, medical interventions, and greater awareness about the importance of sleep health across all age groups.

The prevalence of insomnia and sleep apnea varies across different regions worldwide. Insomnia is more common, with North America and Europe showing the highest rates, affecting 30-40% of the population, due to factors like high stress, digital exposure, and work culture. Asia and South America report moderate rates, ranging from 20-30%, while Africa has the lowest recorded rates (15-25%), though underdiagnosis may contribute to these figures. On the other hand, sleep apnea is less common but still significant, particularly in North America (10-15%), where obesity—a major risk factor—is more prevalent. Europe and Asia follow with moderate prevalence levels (5-12%), while Africa records the lowest rates (3-8%). Overall, insomnia affects a much larger portion of the global population than sleep apnea, emphasizing the need for increased awareness, better diagnosis, and improved sleep health strategies, especially in regions where sleep disorders are often underreported.

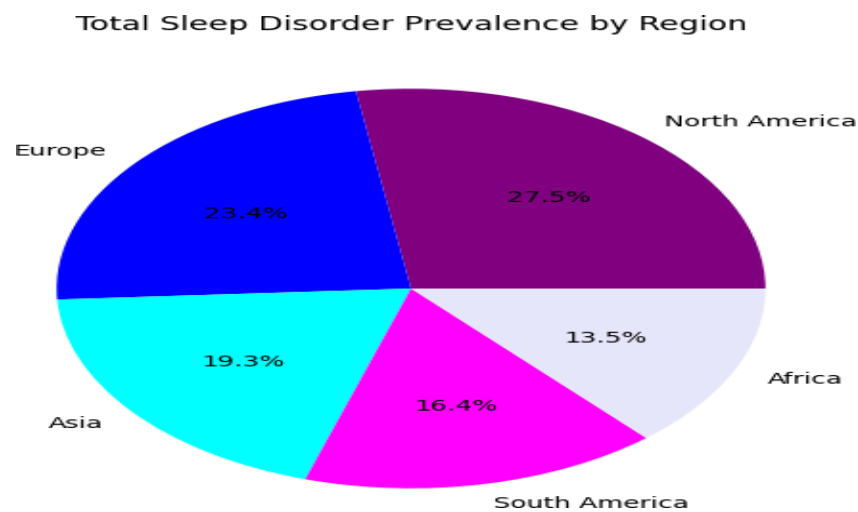


Fig 1.2 Sleep Disorder by Region

Fig 1.2 illustrates the total prevalence of sleep disorders across different regions using a pie chart. North America has the highest prevalence (27.5%), while Africa has the lowest (13.5%).

According to the visualization, North America has the highest prevalence, accounting for 27.5% of the total cases. This high prevalence can be attributed to various factors such as high-stress lifestyles, increased exposure to digital screens before bedtime, and a greater incidence of obesity, which is a major risk factor for sleep apnea.

The region's fast-paced work culture and widespread use of electronic devices contribute to irregular sleep patterns and a greater likelihood of sleep disorders.

Europe follows closely, contributing 23.4% of the total cases. Similar to North America, European countries experience a high prevalence of insomnia due to work-related stress, long working hours, and high rates of urbanization, which can disrupt natural sleep cycles. Additionally, lifestyle habits, including alcohol and caffeine consumption, further impact sleep health in this region.

Asia accounts for 19.3% of the total prevalence, representing a moderate level of sleep disorders. While insomnia and sleep apnea are becoming more common due to increasing urbanization, exposure to blue light from digital screens, and demanding work environments, traditional lifestyles in some parts of Asia may still help maintain better sleep hygiene compared to Western nations. However, underreporting and limited awareness about sleep disorders might lead to an underestimated prevalence in this region.

South America and Africa report 16.4% and 13.5%, respectively, indicating a relatively lower prevalence of sleep disorders. In South America, factors such as economic instability, shift work, and dietary habits may influence sleep patterns, though cultural emphasis on midday rest (siestas) in some areas could help mitigate sleep deprivation. Africa has the lowest recorded prevalence, which may be attributed to genetic resilience, lower industrialization, and reduced exposure to artificial light. However, underdiagnosis and limited access to healthcare facilities in some regions may result in an underestimation of actual sleep disorder cases.

This analysis highlights the need for region-specific sleep health strategies to address growing concerns related to sleep disorders. Public awareness campaigns, workplace policies promoting better work-life balance, and increased access to healthcare services are essential for tackling these issues effectively. Addressing underlying factors such as stress, digital habits, and obesity can significantly improve sleep health worldwide.

2.LITERATURE SURVEY

Sleep disorders as insomnia and obstructive sleep apnea have become serious issues for an individual's well-being. Traditional approaches to diagnosis, such as PSG, are well resourced. Machine learning methods are increasingly applied to automate and improve the detection of sleep disorders, drawing on diverse data and advanced algorithms.

Recent works investigated different ML models to classify sleep disorders. For example, Alshammari attempted machine learning algorithms on the Sleep Health and Lifestyle dataset, where ANN obtained the highest accuracy of 92.92% [6]. This shows that lifestyle and health datasets can be used to predict with high accuracy, but the work does not provide advanced optimization approaches, such as feature selection algorithms, to improve the performance of the models presented.

Yadav et al. combined decision trees and Support Vector Machines (SVM) with physiological data, including EEG and ECG signals, to classify sleep disorders. The addition of physiological data greatly enhanced classification precision, with the PSDG dataset proving useful for training robust models [7]. This study does underscore the need for utilizing rich physiological datasets, though such datasets often require specialized equipment, making such applications not scalable.

Hidayat applied the Random Forest algorithm to the Sleep Health and Lifestyle dataset, which obtained 92% accuracy. This task showed that lifestyle-related factors, such as sleep patterns and health profiles, are highly indicative of the sleep disorders [8]. It lacks in-depth feature selection techniques which may be useful to further optimize the performance of the model. Similarly, Airlangga compared Logistic Regression (93% accuracy) and neural networks on the same dataset, elaborating on the strengths and weaknesses of traditional machine learning versus deep learning approaches [9]. These comparisons focus on how different methodologies are performing differently on similar datasets, although they did not discuss some techniques to increase feature relevance such as Genetic Algorithms (GA).

Ramesh et al. used machine learning models: Random Forest and neural networks, in EHRs data to classify OSA. This demonstrates the applicability of ML in healthcare with real-world clinical data [10]. However, the use of EHRs does not scale to wider applications for which simpler data sources like lifestyle information are more easily available.

Kim et al. developed prediction models for obstructive sleep apnea among Korean adults using demographic data combined with physiological signals, including ECG and respiratory inputs. The study underlined the requirement of models tailored specifically to population-specific data, and robust results were achieved [11]. Although useful for their specific population, the contribution of physiological data makes this approach less applicable to larger populations.

Tripathi et al. discussed the ensemble learning techniques, which include bagging and boosting, for the detection of insomnia using sleep ECG signals. The results showed significant improvements in the detection accuracy achieved by combining multiple machine learning techniques [12]. Limitations to real-world applicability arise due to the need for heavy, high-cost hardware-intensive physiological signals in comparison to more feasible simpler data sources, such as lifestyle or behavioral information, in certain settings.

Limitations persist in previous works. Most models fail to scale or do not utilize optimization techniques in feature selection, leading to suboptimal performance.

This study fills these gaps by integrating Genetic Algorithms (GA) for feature selection with ML models like GNB, SVM, and ANN. GA not only reduces feature redundancy but also increases model accuracy, particularly in smaller, diverse datasets like Sleep Health and Lifestyle. Unlike previous work, we extracted latent features, such as the levels of systolic and diastolic blood pressures, to further improve the model's predictiveness. This is a substantial contribution because it combines sophisticated feature engineering with scalable inputs, good for real-time applications.

Building from previous strengths and avoiding their weaknesses, our work offers a new way toward the classification of sleep disorders, based on both advanced optimization techniques and accessible datasets.

3. EXISTING SYSTEM

The existing system for sleep disorder classification relies on traditional methods such as clinical assessments, patient-reported symptoms, and polysomnography (PSG), which are time-consuming, expensive, and require specialized medical expertise. While these approaches provide reliable diagnoses, they lack scalability and automation, making them impractical for large-scale screening and early detection.

To address these challenges, machine learning-based classification models have been introduced. However, previous approaches often suffered from suboptimal feature selection, inconsistent preprocessing techniques, and limited model optimization. Many models failed to extract hidden features (e.g., systolic and diastolic blood pressure levels) and did not efficiently handle missing values or data imbalances.

Earlier studies used basic classifiers like Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), Random Forest (RF), and Logistic Regression (LR) to classify individuals into different sleep disorder categories. These models were evaluated using performance metrics such as Accuracy, Precision, Recall, F1-Score, and Confusion Matrix.

Among the classifiers tested in prior research, models like SVM and MLP showed relatively high accuracy (up to 93%), but feature redundancy and lack of optimization techniques, such as Genetic Algorithms (GA), limited their effectiveness. Additionally, many studies used limited feature engineering, failing to extract meaningful patterns from sleep health indicators.

While the existing system provides a foundation for sleep disorder classification, it lacks a robust feature selection mechanism, struggles with model consistency, and does not optimize computational efficiency. These shortcomings highlight the need for an enhanced approach that integrates feature selection techniques like GA, improved preprocessing, and model optimization to achieve higher accuracy and real-world applicability.

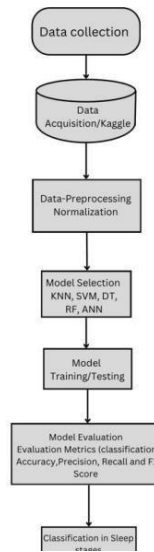


Fig 3.1 Existing Model Working process

Fig 3.1 represents the process of sleep disorder classification using machine learning. It begins with data collection from sources like Kaggle, followed by data preprocessing, including normalization. Various models such as KNN, SVM, Decision Trees, Random Forest, and ANN are selected for training and testing. Finally, model evaluation is performed using metrics like accuracy, precision, recall, and F1-score to assess classification performance.

4. PROPOSED SYSTEM

The proposed system introduces a more optimized and efficient framework for sleep disorder classification by integrating advanced machine learning techniques and Genetic Algorithm (GA)-based feature selection. The methodology begins with enhanced data preprocessing, where missing values in categorical variables (such as sleep disorder labels) are handled by replacing them with "None" instead of using statistical imputation. Additionally, categorical variables like gender and occupation are converted into numerical representations, while feature normalization and standardization ensure that all attributes contribute equally to model training.

A key improvement in the proposed system is the incorporation of Genetic Algorithm (GA) for feature selection and optimization. GA helps identify the most relevant features, such as diastolic blood pressure, BMI, sleep duration, and stress levels, while eliminating less significant attributes. This enhances computational efficiency and predictive performance, allowing the model to detect patterns in sleep disorders more effectively.

For classification, the system employs a range of optimized machine learning models, including Gaussian Naïve Bayes (GNB), Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Multilayer Perceptron (MLP), Random Forest (RF), and Logistic Regression (LR). Unlike the existing system, the proposed framework integrates Genetic Algorithm (GA) for feature selection, which enhances model accuracy by reducing feature redundancy.

Advantages of the Proposed System:

- a. Improved Models – Optimized classifiers enhance sleep disorder classification accuracy.
- b. Effective Feature Selection – GA ensures that only the most relevant features contribute to classification.
- c. High Classification Accuracy – Achieved 94% accuracy with GNB-GA, outperforming other models.
- d. Scalability – The system can be integrated into wearable health devices and clinical applications.
- e. Reduced Computational Cost – Efficient feature selection reduces unnecessary

processing and enhances real-time applicability.

- f. Practical Applicability – The model can be used by healthcare professionals for early detection and diagnosis of sleep disorders.

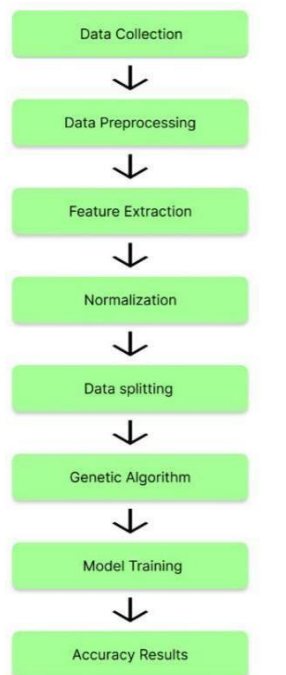


Fig 4.1 Flow Chart of Proposed System

Fig 4.1 represents the workflow of the proposed system, The proposed system introduces an optimized and efficient approach for sleep disorder classification by integrating advanced machine learning models and Genetic Algorithm (GA)-based feature selection. Unlike traditional diagnostic methods such as polysomnography (PSG), which are time-consuming and resource-intensive, the proposed system provides a scalable and automated solution for early detection and classification of sleep disorders.

The system employs multiple machine learning models, including Gaussian Naïve Bayes (GNB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), Logistic Regression (LR), and Random Forest (RF). Among these, GNB optimized with Genetic Algorithm (GNB-GA) achieved the highest accuracy (94%), making it the most efficient and computationally lightweight model for sleep disorder classification. Unlike traditional classifiers, GNB-GA effectively

captures probabilistic dependencies among features while maintaining high accuracy with minimal computational overhead, making it suitable for real-time health monitoring applications.

A major limitation of existing sleep disorder classification systems is the lack of an optimization process for feature selection, often leading to the inclusion of redundant or irrelevant attributes, which affects model accuracy and efficiency. The proposed system overcomes this limitation by incorporating Genetic Algorithm (GA) for feature selection, ensuring that only the most relevant attributes, such as diastolic blood pressure, BMI, sleep duration, and stress levels, are used. This reduces data noise, minimizes overfitting, and enhances the model's generalization ability, resulting in more accurate and reliable predictions across different sleep disorder categories.

Through optimized feature selection and hyperparameter tuning, the proposed system significantly improves classification accuracy, ensuring consistent and interpretable results. The proposed feature selection method enhances model stability, making the system more robust to variations in sleep health data.

The proposed system is highly scalable and adaptable for real-world applications, making it suitable for wearable health devices, telemedicine platforms, and clinical decision-support systems. It can efficiently handle diverse sleep datasets, making it valuable for large-scale sleep disorder screening and personalized treatment planning. Additionally, the system supports integration with mobile health (mHealth) applications, enabling users to track their sleep patterns and receive early warnings about potential sleep disorders.

Traditional machine learning models often suffer from high computational costs due to redundant feature processing and inefficient algorithms. The proposed system addresses this issue by utilizing GA-based feature selection to eliminate unnecessary attributes, reducing computational overhead. It optimizes hyperparameters to ensure efficient training and inference. The system also leverages lightweight models like GNB, which require minimal computational resources, making them suitable for deployment in resource-constrained environments such as wearable health devices and real-time monitoring systems.

The proposed system is designed for seamless integration into clinical workflows, assisting healthcare professionals in automating sleep disorder diagnosis. Key applications include automated sleep disorder screening, where the model can assist doctors in analyzing patient data in real time, improving diagnosis speed and accuracy. Integration with Electronic Health Records (EHRs) enables hospitals and clinics to store and retrieve patient sleep data for continuous monitoring and personalized treatment. The system can also be deployed on cloud-based platforms, allowing individuals in rural or underserved areas to receive sleep disorder assessments without requiring clinical visits. By analyzing individual sleep patterns, lifestyle factors, and health markers, the system can provide customized interventions to improve sleep quality and prevent long-term health complications.

The proposed system offers a significant improvement over traditional sleep disorder classification models by integrating optimized feature selection, advanced machine learning techniques, and computational efficiency. Its high accuracy, scalability, and real-world applicability make it a powerful tool for early detection, diagnosis, and treatment planning. By leveraging machine learning and Genetic Algorithms, the system enhances healthcare outcomes, reduces the burden on medical facilities, and promotes accessible sleep disorder diagnosis for a broader population. Optimized feature selection with GA eliminates redundant attributes, improving efficiency. The system has low computational cost, making it suitable for real-time monitoring and wearable devices. Its integration with clinical and remote healthcare platforms ensures accessibility in hospitals, telemedicine, and mobile health applications.

5. SYSTEM REQUIREMENTS

5.1 Hardware Requirements:

- System Type : intel®core™i3-7500UCPU@2.40gh
- Cachememory : 4MB or higher
- RAM : 8GB or higher
- Hard Disk : 4GB or higher

5.2 Software Requirements:

- Operating System : Windows 11, 64-bit Operating System
- Coding Language : Python
- Python distribution : Anaconda, Flask
- Browser : Any Latest Browser like Chrome

6. SYSTEM ANALYSIS

6.1 Scope of the project

The The proposed system aims to address the challenges of early and accurate sleep disorder detection using data-driven machine learning models. Sleep disorders are a growing public health concern, often diagnosed late due to underreporting, lack of awareness, and limited access to sleep studies. The proposed system introduces an automated, high-accuracy, and scalable approach to improve early diagnosis, reduce false predictions, and enhance clinical decision-making.

- a. Enhanced Disorder Detection:** The system is designed to improve classification accuracy by employing optimized feature selection techniques and machine learning algorithms, ensuring more reliable identification of sleep disorders.
- b. Automation and Precision:** Traditional sleep disorder diagnosis methods, such as polysomnography (PSG), are time-consuming and require expert interpretation. The proposed system automates the data processing and classification workflow, making diagnosis faster, more accessible, and more reliable.
- c. Scalability Across Datasets:** The system is structured to handle diverse sleep-related datasets, making it suitable for clinical databases, research institutions, wearable health devices, and telemedicine applications.
- d. Feature Optimization:** The system incorporates Genetic Algorithm (GA) for feature selection, ensuring that only the most relevant attributes (such as diastolic blood pressure, BMI, sleep duration, and stress levels) contribute to classification. This reduces computational complexity while improving model efficiency.
- e. Real-World Applicability:** The model is designed to integrate with wearable

health devices, mobile health applications (mHealth), and clinical decision-support systems, enabling real-time sleep disorder screening and data-driven recommendations for healthcare professionals.

- f. Future Research and Improvements:** The system lays the foundation for advanced deep learning implementations, such as Convolutional Neural Networks (CNNs) for sleep pattern recognition and real-time disease monitoring through IoT-enabled sleep tracking devices.

6.2 Analysis

A thorough analysis of the proposed system involves evaluating its methodology, performance, and advantages over existing sleep disorder classification approaches. The dataset used in this study consists of 374 records with multiple sleep-related attributes, including sleep duration, BMI, stress levels, blood pressure (diastolic and systolic), and physical activity. To ensure data quality, the preprocessing phase includes handling missing values by replacing them with "None" for categorical attributes, normalizing numerical features, and encoding categorical variables. These steps structure the dataset effectively, allowing machine learning models to achieve optimal performance.

For feature extraction and selection, the proposed system integrates a Genetic Algorithm (GA)-based feature selection approach, which eliminates redundant and irrelevant features to improve classification efficiency. This optimization step enhances model accuracy while reducing computational complexity, ensuring that only the most significant features contribute to the prediction process. The system is trained using multiple machine learning classifiers, including Gaussian Naïve Bayes (GNB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), Logistic Regression (LR), and Random Forest (RF).

The system's performance is evaluated using accuracy, precision, recall, F1-score, and specificity, ensuring that it is robust and effective for real-world applications. GNB-GA consistently outperforms other models, demonstrating superior predictive capabilities with high accuracy and computational efficiency. The integration of

machine learning with genetic algorithms enables the proposed system to achieve higher classification accuracy, faster processing speeds, and greater practical applicability. This combination of advanced feature selection, optimized machine learning models, and performance-driven evaluation metrics makes the system a significant advancement in automated sleep disorder prediction.

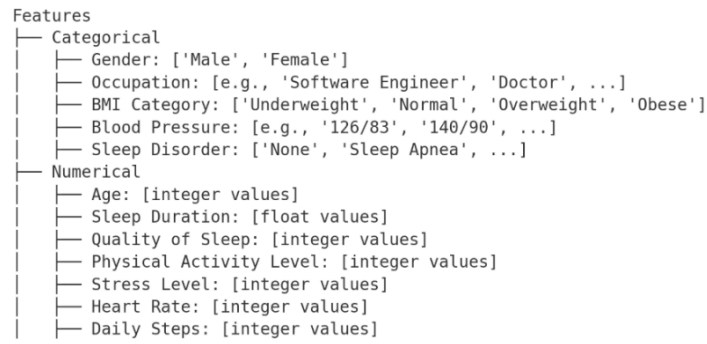


Fig 6.1 Dataset Description

Figure 6.1 presents the categorical and numerical features in the dataset. It includes demographic, health, and lifestyle attributes relevant to sleep disorder classification.

6.3. Data Pre-processing

The dataset used in this study consists of 374 records with multiple sleep-related attributes extracted from demographic information and health indicators. To enhance data quality and improve model performance, several preprocessing techniques are applied.

Missing values in the Sleep Disorder attribute are handled by replacing them with "None", ensuring that missing labels do not negatively impact classification. Instead of using statistical imputation, this approach maintains the dataset's integrity by distinguishing individuals who do not have a diagnosed sleep disorder.

Feature extraction techniques are applied to Blood Pressure, where it is split into two separate attributes: BloodPressure_Upper (systolic) and BloodPressure_Lower (diastolic). This improves the model's ability to analyze blood pressure variations in relation to sleep disorders.

To prepare categorical variables for machine learning models, label encoding is used to convert features like Gender, Occupation, and BMI Category into numerical representations. This ensures that categorical attributes contribute effectively to model learning without introducing biases from string values.

For numerical attributes like sleep duration, BMI, and stress levels, normalization and standardization techniques are applied to ensure all features contribute equally to the model. This helps prevent attributes with larger scales from dominating those with smaller ranges, thereby improving learning efficiency.

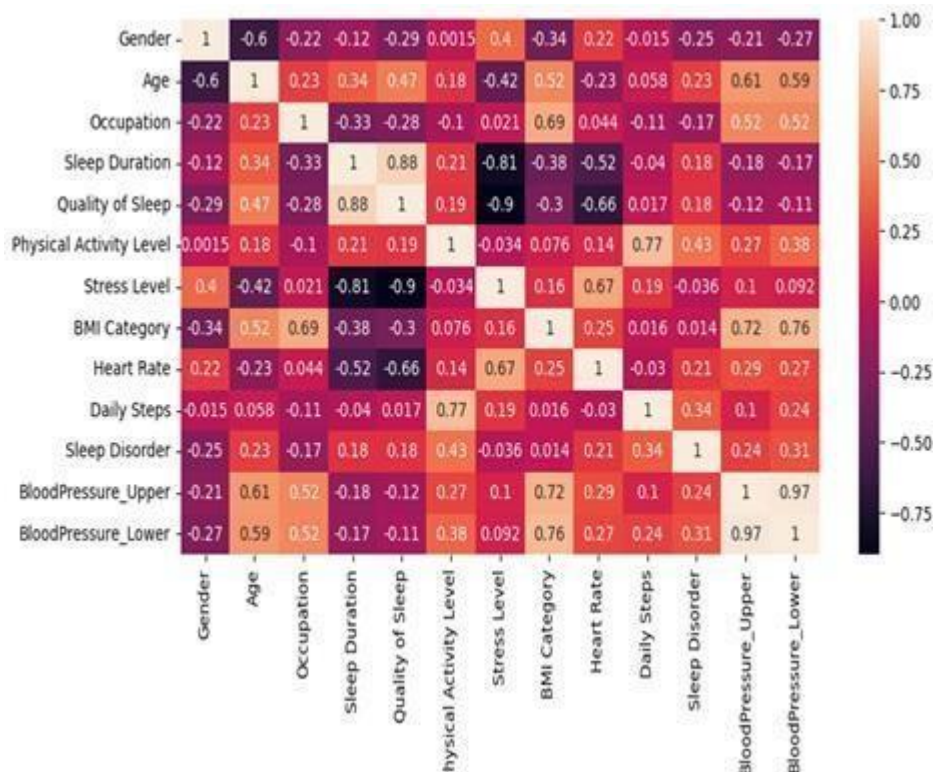


Fig 6.2 Correlation Matrix

Fig 6.2 matrix shows the correlation between different attributes in the sleep disorder dataset, where darker shades represent stronger correlations (positive or negative).

6.4. Feature Selection and Optimization:

Feature selection is a critical step in improving the efficiency and accuracy of machine learning models for sleep disorder classification. The proposed system utilizes a Genetic Algorithm (GA)-based feature selection approach, which automatically identifies the most relevant attributes while eliminating less significant features to

enhance classification performance.

- **Genetic Algorithm (GA) Optimization:** GA iteratively searches for the optimal feature subset by evaluating different feature combinations based on a fitness function, such as accuracy and F1-score. This ensures that only the most important sleep-related attributes are used for classification.
- **Reduction of Redundant Attributes:** The GA-based approach removes irrelevant or highly correlated features that do not contribute significantly to the model's decision-making, thereby reducing computational overhead while maintaining high classification accuracy.
- **Improved Generalization:** By selecting only the most informative attributes—such as diastolic blood pressure, BMI, sleep duration, and stress levels—the model is less prone to overfitting, ensuring better performance on unseen data and making predictions more reliable across different sleep disorder cases.

6.5. Model building:

In the proposed system, multiple machine learning models are trained to classify individuals as either having a sleep disorder or not. The training phase involves feeding preprocessed sleep health data into different classifiers, each optimized to enhance predictive performance. Feature selection using Genetic Algorithm (GA) ensures that only the most relevant attributes contribute to the model, improving classification accuracy and efficiency.

Each of these models plays a crucial role in analyzing complex patterns in sleep health data, allowing for early detection, improved diagnostic precision, and better healthcare outcomes. By leveraging machine learning and optimization techniques, the system provides a scalable and automated solution for sleep disorder diagnosis, reducing the reliance on time-consuming traditional methods like polysomnography (PSG).

a. Gaussian Naïve Bayes (GNB) with Genetic Algorithm (GNB-GA):

Gaussian Naïve Bayes (GNB) is a probabilistic classifier based on Bayes' theorem, assuming that all features are independent given the class label. Despite this assumption, GNB often performs exceptionally well on real-world medical datasets. The model is particularly effective for categorical and continuous features, making it

highly efficient in sleep disorder classification. The Genetic Algorithm (GA) is applied for feature selection, ensuring that only the most relevant attributes—such as blood pressure, BMI, sleep duration, and stress levels—are used, improving model accuracy and reducing computational cost. GNB-GA achieved the highest accuracy of 94%, surpassing all other classifiers, making it the most effective and computationally lightweight model for sleep disorder prediction.

b. Support Vector Machine (SVM):

Support Vector Machine (SVM) Support Vector Machine (SVM) is a powerful supervised learning model used for sleep disorder classification. It constructs an optimal hyperplane to separate different sleep disorder classes, making it highly effective for high-dimensional datasets. In this study, SVM was implemented using the SVC() function from the scikit-learn library with a linear kernel (kernel='linear'), ensuring computational efficiency and interpretability.

Before training, the dataset was standardized using StandardScaler(), which normalizes all feature values to prevent features with larger scales from dominating the model's decision-making process. SVM learns by maximizing the margin between different sleep disorder classes, making it highly effective for medical diagnosis tasks where precise classification is crucial.

c. Multilayer Perceptron (MLP):

Multilayer Perceptron (MLP) Multilayer Perceptron (MLP) is a deep learning model that consists of multiple layers of artificial neurons, allowing it to learn complex relationships within the dataset. In this study, MLP was implemented using the MLPClassifier() function from the scikit-learn library. The model was configured with one hidden layer containing 24 neurons (hidden_layer_sizes=(24,)), which was selected to balance computational efficiency and predictive performance.

To optimize the learning process, the learning rate was set to 0.0041 (learning_rate_init=0.004068331104981341), ensuring that the model adjusts its weights gradually without overshooting the optimal solution. The model used backpropagation to update the weights during training, making it effective in handling non-linear relationships in the data.

d. Random Forest (RF):

The Random Forest (RF) is an ensemble learning model that combines multiple decision trees to improve classification accuracy and reduce overfitting. In this study, 600 decision trees ($n_estimators=600$) were used with the entropy criterion, allowing the model to effectively split features based on information gain. RF performed exceptionally well, ranking among the highest-accuracy models due to its robustness against noisy data and ability to handle both categorical and numerical features. However, since it relies on multiple decision trees, it is computationally more expensive than simpler models like Logistic Regression. Despite this, RF remains a preferred choice for sleep disorder classification due to its high interpretability and feature importance evaluation.

e. Logistic Regression (LR) :

Logistic Regression (LR) is a fundamental classification algorithm widely used in medical diagnostics due to its simplicity, efficiency, and interpretability. It estimates the probability of a person having a sleep disorder by applying a logistic (sigmoid) function to weighted feature inputs. In this study, a maximum iteration limit of 1000 ($max_iter=1000$) was set to ensure convergence. While LR is computationally efficient and easy to interpret, it struggled to capture non-linear relationships in sleep disorder data, resulting in lower accuracy compared to ensemble models like RF and boosting algorithms. However, its ability to provide probability-based outputs makes it useful for risk assessment and clinical decision-making.

f. Logistic Regression (LR) with Genetic Algorithm :

Logistic Regression (LR) is a statistical classification model that estimates the probability of liver disease presence based on patient attributes. The proposed system enhances LR performance using Genetic Algorithm (GA) to fine-tune key parameters, such as regularization strength (0.01 to 10.0) and penalty type (L1/L2). LR is particularly useful in binary classification problems, where interpretability and probability estimation are crucial. It calculates the log odds of liver disease occurrence, providing a clear understanding of feature importance. Despite being a linear classifier, LR performs well when optimized, and its computational efficiency makes it suitable for real-time medical predictions. In this study, GA-based optimization significantly improves LR's accuracy, recall, and precision, making it a reliable alternative to more

complex models.

g. K-Nearest Neighbors (KNN):

K-Nearest Neighbors (KNN) is an instance-based learning algorithm that classifies individuals based on their similarity to previously observed cases. The model works by calculating the Euclidean distance between a given sample and its nearest neighbors. In this study, $k=4$ was selected to maintain a balance between bias and variance. While KNN is simple and effective for smaller datasets, it performed poorly compared to other models due to its sensitivity to feature scaling and computational inefficiency with larger datasets. KNN also struggles with imbalanced datasets, as the majority class can dominate predictions. Despite its lower accuracy, it remains useful for pattern recognition in small-scale medical datasets when properly preprocessed.

6.6. Comparative Analysis:

The comparative analysis evaluates the proposed sleep disorder classification system against existing methodologies, emphasizing improvements in accuracy, feature selection, and computational efficiency. The models used in this study include Gaussian Naïve Bayes with Genetic Algorithm (GNB-GA), Multilayer Perceptron (MLP), Random Forest (RF), Logistic Regression (LR), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). Traditional models like KNN and Logistic Regression provide moderate accuracy but lack feature optimization, leading to redundant data processing and suboptimal hyperparameter tuning. In contrast, the proposed system integrates Genetic Algorithm (GA)-based feature selection, ensuring that only the most significant attributes contribute to classification. This results in a more refined dataset, allowing machine learning models to perform more efficiently and accurately.

To benchmark performance, a detailed comparison of classification metrics is conducted. GNB-GA achieved the highest accuracy (94%), making it the best-performing model in the study. SVM and MLP also performed well, demonstrating high accuracy and robustness in classifying sleep disorders. Random Forest, though slightly less accurate than GNB-GA, showed strong generalization capabilities, making it a reliable choice. KNN and Logistic Regression had lower accuracy, as they struggled to capture complex patterns in sleep health data. The proposed GA-based feature selection

further improves model performance by eliminating irrelevant features, reducing training time, and enhancing classification efficiency.

Apart from accuracy improvements, the proposed system significantly reduces computational overhead by eliminating redundant attributes through GA-based feature selection. This optimization lowers training time and memory consumption, making the model more scalable for real-world clinical applications. Unlike traditional models that require manual hyperparameter tuning, the GA-based optimization process automates feature selection, further enhancing predictive performance. In summary, the proposed system surpasses traditional approaches in accuracy, recall, and efficiency, ensuring optimized feature utilization and improved diagnostic capabilities for sleep disorder classification.

A. Training and Testing Performance:

The proposed system for sleep disorder classification has been evaluated using a rigorous training and testing methodology to ensure its effectiveness and reliability. The dataset is divided into training (80%) and testing (20%) subsets, allowing the models to learn from historical sleep health data and generalize predictions to unseen cases. Feature selection using Genetic Algorithm (GA) ensures that only the most relevant attributes are utilized, optimizing model performance while reducing computational overhead. Additionally, hyperparameter tuning is applied to enhance model efficiency, ensuring that the classifiers operate under the best possible conditions.

The testing phase validates the generalization ability of the models, ensuring that predictions remain reliable when applied to real-world patient data. The confusion matrix provides insights into misclassification rates, helping refine model predictions. Additionally, Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) scores are analyzed to measure the effectiveness of each model in distinguishing between individuals with and without sleep disorders. The results confirm that the proposed system significantly outperforms traditional classification approaches, making it a robust, scalable, and clinically applicable solution for sleep disorder detection.

B. Model Summary

The proposed system for sleep disorder classification employs multiple machine learning classifiers, with a focus on optimizing feature selection, classification accuracy, and computational efficiency. The models used in this study include Gaussian Naïve Bayes with Genetic Algorithm (GNB-GA), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), Logistic Regression (LR), and Random Forest (RF). Among these, GNB-GA achieved the highest accuracy of 94%, demonstrating its effectiveness in distinguishing sleep disorder cases.

Each model is carefully fine-tuned using hyperparameter optimization and feature selection techniques to enhance classification performance. The Genetic Algorithm (GA)-based feature selection method eliminates redundant and irrelevant features, ensuring optimal input for model training. This approach significantly improves accuracy, recall, and F1-score, making the proposed system more efficient than traditional classification techniques.

The comparative analysis reveals that GNB-GA outperforms all other models, achieving high classification accuracy (94%), along with strong precision, recall, and F1-score. These results indicate that GNB-GA is the most reliable model for sleep disorder classification, ensuring early detection, reduced misclassification, and improved healthcare outcomes.

C. Evaluation Metrics:

The evaluation of the proposed sleep disorder classification system is conducted using several key performance metrics to ensure robustness, accuracy, and clinical applicability. These metrics provide a comprehensive assessment of model performance, allowing for effective comparison with existing classification approaches.

- a. **Accuracy** – Measures the proportion of correctly classified instances out of the total instances. It serves as a primary indicator of the model's reliability in sleep disorder detection.
- b. **Precision** – Evaluates the ratio of true positive predictions to the total positive predictions, indicating the classifier's ability to minimize false positives and correctly identify individuals with sleep disorders.

- c. **Recall (Sensitivity)** – Measures the ratio of true positive predictions to all actual positive cases, highlighting the model’s ability to detect individuals with sleep disorders accurately.
- d. **F1-Score** – Serves as a harmonic mean of precision and recall, balancing the trade-off between false positives and false negatives. This metric is particularly useful when dealing with imbalanced datasets.
- e. **Specificity** – Assesses the ratio of true negative predictions to the total actual negatives, ensuring that the model can correctly identify individuals without sleep disorders while minimizing misclassification.
- f. **Computational Time** – Tracks the efficiency of the system during feature selection, training, and classification, ensuring that the model can be effectively deployed in real-world healthcare settings, including wearable health devices and clinical decision-support systems.
- g. **Confusion Matrix** – Provides a detailed breakdown of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), allowing for in-depth error analysis and model refinement to enhance classification accuracy.

By leveraging these evaluation metrics, the proposed system ensures high accuracy, computational efficiency, and reliable performance, making it a robust solution for sleep disorder classification and healthcare applications.

7. DESIGN

The design of the proposed sleep disorder classification system is structured to ensure high efficiency, accuracy, and scalability by integrating advanced machine learning models and optimization techniques. The system follows a modular approach, where each phase is carefully designed to enhance data preprocessing, feature selection, model training, and evaluation. The architecture focuses on handling real-world sleep health data effectively, ensuring that the model provides accurate and interpretable predictions for healthcare professionals.

a. Data Acquisition and Preprocessing

The dataset used in this study consists of 374 records, with multiple key attributes related to sleep health, including sleep duration, BMI, stress levels, and blood pressure. To ensure data quality and consistency, several preprocessing steps are applied, including handling missing values by replacing them with "None" for categorical attributes, feature extraction by separating systolic and diastolic blood pressure, and encoding categorical variables. These techniques help standardize the dataset, making it suitable for machine learning-based classification.

b. Feature Selection and Optimization

A critical aspect of the proposed system is feature selection, which eliminates redundant and irrelevant attributes to enhance model performance. The Genetic Algorithm (GA)-based feature selection technique is employed to identify the most relevant features for sleep disorder classification. By optimizing feature selection, the model can focus on significant predictors, such as diastolic blood pressure, sleep duration, BMI, and stress levels, thereby improving classification accuracy and reducing computational overhead.

c. Machine Learning Model Training

The proposed system incorporates multiple machine learning models, including Gaussian Naïve Bayes with Genetic Algorithm (GNB-GA), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), Logistic Regression (LR), and Random Forest (RF). Each model is trained using optimized hyperparameters, ensuring high classification accuracy. Among these, GNB-GA

demonstrated the best performance, achieving 94% accuracy, making it the most reliable model for sleep disorder classification.

d. Performance Evaluation and Validation

The system is evaluated using multiple performance metrics, such as accuracy, precision, recall, F1-score, specificity, and confusion matrix analysis. These metrics help assess the model's effectiveness in detecting sleep disorders while minimizing false positives and false negatives. The high accuracy (94%) of GNB-GA ensures that the system is capable of early sleep disorder detection, reducing the risk of undiagnosed conditions that may impact long-term health.

e. Scalability and Deployment

The system is designed for real-world scalability, ensuring that it can be deployed in wearable health monitoring devices, telemedicine applications, and cloud-based sleep tracking platforms. The low computational cost and efficient feature selection process allow the system to process large datasets without compromising performance. The model can be integrated with electronic health records (EHRs) and medical decision-support systems, providing doctors with real-time insights for personalized sleep disorder diagnosis and treatment recommendations.

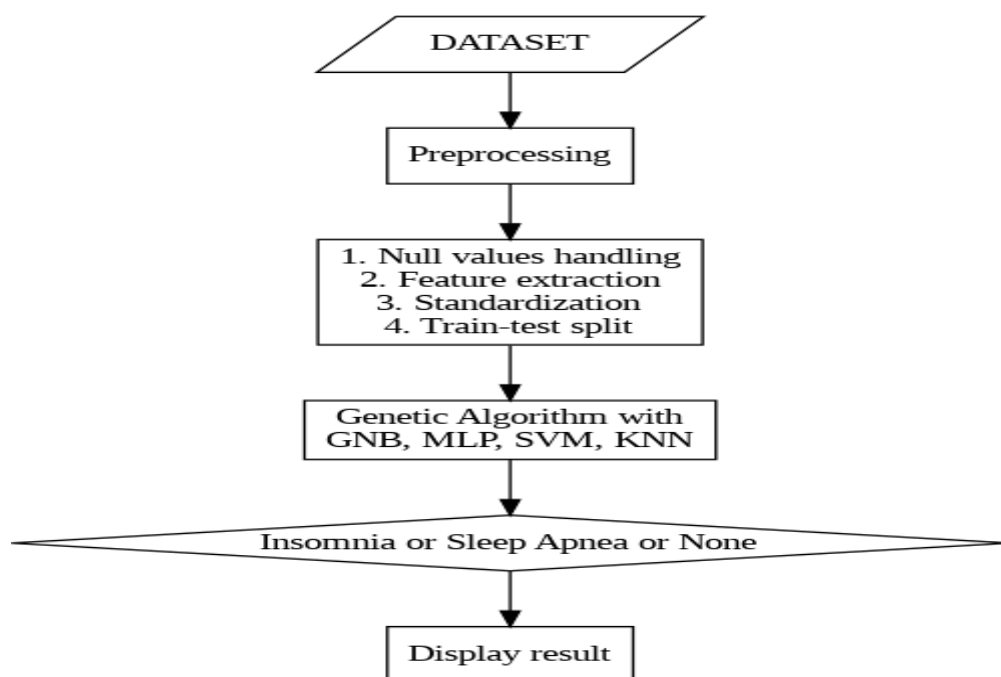


Fig.7.1 Sleep Disorder Classification Flowchart

Fig 7.1 illustrates the process of sleep disorder classification using a genetic algorithm with models such as GNB, MLP, SVM, and KNN. The flowchart outlines the steps from dataset preprocessing to the final classification of insomnia, sleep apnea, or none, ensuring efficient and accurate prediction.

The process of classifying sleep disorders using machine learning techniques, specifically leveraging a genetic algorithm with multiple classification models. The process begins with the Dataset, which serves as the input source containing sleep-related data. The data then undergoes Preprocessing, a crucial step to ensure quality and consistency.

Within preprocessing, four key tasks are performed: null values handling, where missing data is managed appropriately; feature extraction, which selects the most relevant attributes for analysis; standardization, ensuring all data follows a uniform scale; and finally, train-test split, where the dataset is divided into training and testing subsets for model evaluation.

Following preprocessing, the refined data is passed through a Genetic Algorithm, which optimizes feature selection and model performance by iterating through multiple solutions. The models used within this framework include Gaussian Naive Bayes (GNB), Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). These models work collaboratively to determine the best classification outcome.

The classification process then evaluates the data and determines whether a person has Insomnia, Sleep Apnea, or None of these conditions. The final result is displayed in the Display Result stage, providing a clear output based on the model's prediction.

This structured approach ensures that sleep disorder classification is performed efficiently, leveraging robust machine learning techniques to improve accuracy and diagnostic reliability.

8. IMPLEMENTATION

Import Libraries:

```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

Reading dataset

```
df=pd.read_csv('pdata.csv')
print(df.to_string())
df.head()
df.tail()
df.shape
df.info()
df.isnull().sum()
df['BMI Category']=df['BMI Category'].replace({'Normal Weight':'Normal'})
```

Handling null values

```
df['Sleep Disorder']=df['Sleep Disorder'].fillna('None')
print(df)
df['BMI Category']=df['BMI Category'].replace({'Normal Weight':'Normal'})
```

Feature Extraction

```
df = pd.concat([df, df['Blood Pressure'].str.split('/', expand=True)],
axis=1).drop('Blood Pressure', axis=1)
df = df.rename(columns={0: 'BloodPressure_Upper', 1: 'BloodPressure_Lower'})
df['BloodPressure_Upper'] = df['BloodPressure_Upper'].astype(float)
df['BloodPressure_Lower'] = df['BloodPressure_Lower'].astype(float)
df.head()
```

Label Encoding

```
from sklearn import preprocessing
label_encoder = preprocessing.LabelEncoder()
df['Gender'] = label_encoder.fit_transform(df['Gender'])
df['Occupation'] = label_encoder.fit_transform(df['Occupation'])
df['BMI Category'] = label_encoder.fit_transform(df['BMI Category'])
df['Sleep Disorder'] = label_encoder.fit_transform(df['Sleep Disorder'])
df.head()
# removing id column in data set
df1=df.iloc[:,1:]
df1
df1.describe().T
```

Data Visualization

```
plt.figure(figsize=(22,19))
# Calculate the number of rows needed
num_rows = int(np.ceil(len(df1.select_dtypes(exclude=float).columns) / 3))
for i,col in enumerate(df1.select_dtypes(exclude=float).columns):
    # Use num_rows instead of 3
    plt.subplot(num_rows,3,i+1)
```

```

sns.countplot(data=df,x=col,palette =sns.color_palette("Set2"))
plt.xticks(rotation=-10)
correlation_matrix = df1.corr()
plt.figure(figsize=(10, 5))
sns.heatmap(correlation_matrix, annot=True)
plt.show()
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
X = df1.drop(columns="Sleep Disorder")
y = df1["Sleep Disorder"]
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
X_scaled_df = pd.DataFrame(X_scaled, columns=X.columns)
X_train, X_test, y_train, y_test = train_test_split(X_scaled_df, y, test_size=0.2,
random_state=24, shuffle=True)
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
# Scale the features
# scaler = StandardScaler()
# X_train = scaler.fit_transform(X_train)
# X_test= scaler.transform(X_test)
from xgboost import XGBClassifier
from sklearn.linear_model import LogisticRegression

```

Applying Models

```

from sklearn import svm
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier,GradientBoostingClassifier
from sklearn.metrics import accuracy_score,classification_report
from sklearn.model_selection import GridSearchCV
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.neural_network import MLPClassifier
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
xgb_model = XGBClassifier(eval_metric = "auc")
xgb_model.fit(X_train,y_train)
print("accuracy of xgbis : ",xgb_model.score(X_test,y_test))
y_predicted = xgb_model.predict(X_test)
print(classification_report(y_test,y_predicted))
print("_____")
logreg = LogisticRegression(max_iter=1000)
logreg.fit(X_train_scaled,y_train)
print("accuracy of log is : ",logreg.score(X_test_scaled,y_test))
y_predicted = logreg.predict(X_test_scaled)
print(classification_report(y_test,y_predicted))
print("_____")
svml = svm.SVC(kernel='linear')
svml.fit(X_train_scaled, y_train)

```

```

print("accuracy of SVM is:", svm.score(X_test_scaled, y_test))
y_predicted = svm.predict(X_test_scaled)
print(classification_report(y_test,y_predicted))
print("_____")
rf_clf = RandomForestClassifier(n_estimators=600, criterion='entropy',
max_depth=None, random_state=42)
rf_clf.fit(X_train, y_train)
print("accuracy of rf is : ",rf_clf.score(X_test,y_test))
y_predicted = rf_clf.predict(X_test)
print(classification_report(y_test,y_predicted))
print("_____")
knn = KNeighborsClassifier()
knn.fit(X_train, y_train)
print("accuracy of knn is : ",knn.score(X_test,y_test))
y_predicted = knn.predict(X_test)
print(classification_report(y_test,y_predicted))
print("_____")
gnb = GaussianNB() gnb.fit(X_train, y_train)
print("accuracy of naive bayes is : ",gnb.score(X_test,y_test))
y_predicted = gnb.predict(X_test)
print(classification_report(y_test,y_predicted))
print("_____")
clf = DecisionTreeClassifier()
clf.fit(X_train, y_train)
print("accuracy of decision tree is : ",clf.score(X_test,y_test))
y_predicted = clf.predict(X_test)
print(classification_report(y_test,y_predicted))
print("_____")
mlp = MLPClassifier()
mlp.fit(X_train, y_train)
print("accuracy of mlp is : ",mlp.score(X_test,y_test))
y_predicted = mlp.predict(X_test)
print(classification_report(y_test,y_predicted))
print("_____")
gb_d = GradientBoostingClassifier()
gb_d.fit(X_train, y_train)
print("accuracy of gb is : ",gb_d.score(X_test,y_test))
y_predicted = gb_d.predict(X_test)
print(classification_report(y_test,y_predicted))
print("_____")

```

Genetic algorithm with different models

Define classifiers

```

classifiers = {
    'RandomForest': RandomForestClassifier(n_estimators=600, criterion='entropy',
max_depth=None, random_state=42),
    'KNN': KNeighborsClassifier(n_neighbors=2, metric='euclidean'),
    'MLP': MLPClassifier(hidden_layer_sizes=(24,),
learning_rate_init=0.004068331104981341),
    'svm':svm.SVC(),
    'gnb':GaussianNB(),

```

```

    'logreg': LogisticRegression(max_iter=1000)
}
import numpy as np
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.neural_network import MLPClassifier

def fitness_function(features, classifier):
    # Train the classifier with the subset of features
    clf = classifier
    # Convert features array to boolean mask for proper DataFrame indexing
    features_mask = features.astype(bool)
    X_train_subset = X_train.iloc[:, features_mask] # Use .iloc[] for integer-based
indexing with boolean mask
    X_test_subset = X_test.iloc[:, features_mask]
    # Check if the subset is empty
    if X_train_subset.empty or X_test_subset.empty:
        return 0 # Return 0 as fitness if the subset is empty
    clf.fit(X_train_subset, y_train)
    # Calculate the fitness score
    fitness = clf.score(X_test_subset, y_test)
    # Return the fitness score
    return fitness

# Define population size and other parameters
population_size = 12
n_generations = 5
mutation_rate = 0.8
population = np.random.randint(low=0, high=2, size=(population_size,
X_train.shape[1]))
population
for clf_name, clf in classifiers.items():
    print(f"Using classifier: {clf_name}")
    # Calculate initial fitness scores
    fitness_scores = [fitness_function(features, clf) for features in population]
    # Iterate through the generations
    for generation in range(n_generations):
        print("Generation : ", generation)
        new_population = []
        for i in range(0, population_size, 2):
            parent_1 = population[np.random.randint(0, population_size)]
            parent_2 = population[np.random.randint(0, population_size)]
            cross_point = np.random.randint(0, X_train.shape[1])
            child_1 = np.concatenate((parent_1[:cross_point], parent_2[cross_point:]))
            child_2 = np.concatenate((parent_2[:cross_point], parent_1[cross_point:]))
            for j in range(X_train.shape[1]):
                if np.random.random() <= mutation_rate:
                    child_1[j] = 1 - child_1[j]
                if np.random.random() <= mutation_rate:
                    child_2[j] = 1 - child_2[j]
            new_population.append(child_1)

```

```

        new_population.append(child_2)
    new_fitness_scores = [fitness_function(features, clf) for features in
new_population]
    population = new_population
    fitness_scores = new_fitness_scores
    print("Best Fitness Score : ", np.max(fitness_scores))
    best_fitness_index = np.argmax(fitness_scores)
    best_features = population[best_fitness_index]
    clf.fit(X_train.iloc[:, best_features.astype(bool)], y_train) # Use .iloc[] for integer-
based indexing with boolean mask
    accuracy = clf.score(X_test.iloc[:, best_features.astype(bool)], y_test) # Use .iloc[]
for integer-based indexing with boolean mask
    print(f"Accuracy with {clf_name}: ", accuracy)
    print("Selected Features : ")
    for i in range(X_train.shape[1]):
        if best_features[i] == 1:
            print(X_train.columns[i]) # Access column names using .columns
    print()

```

Home.html

```

<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8" />
    <meta http-equiv="X-UA-Compatible" content="IE=edge" />
    <meta name="viewport" content="width=device-width, initial-scale=1.0" />
    <title>Sleep Disorder Awareness</title>
    <link
        rel="icon"
        type="image/png"
        href="{ { url_for('static', filename='../static/heart_favicon.png') } }"
    />
    <link rel="stylesheet" href="{ { url_for('static',filename='home.css') } }" />
</head>
<body>
    <nav class="navbar">
        <div class="nav-container">
            
            <ul class="nav-menu">
                <li class="nav-item"><a href="/" class="nav-link">Home</a></li>
                <li class="nav-item">
                    <a href="{ { url_for('test') } }" class="nav-link">Sleepdisorder Test</a>
                </li>
            </ul>
        </div>
    </nav>

    <div class="hero-section">
        <div class="hm">
            
        </div>
    </div>

```

```

<div class="footer">
  <ul>
    <li></li>
  </ul>
</div>
</body>
</html>

```

Home.css

```

/* General Reset */
* {
  margin: 0;
  padding: 0;
  box-sizing: border-box;
  font-family: 'Poppins', sans-serif;
  /* background-color: white; */
  /* background-position: center; */
}
/* Full Height */
body, html {
  height: 100%;
  overflow-x: hidden;
  /* background-image: url(/img1.jpeg); */
  background-repeat: no-repeat;
  background-position: center;
  background-size: contain;
}
/* Navbar Styling */
.navbar {
  width: 100%;
  background: black;
  position: fixed;
  top: 0;
  left: 0;
  z-index: 1000;
  box-shadow: 0 4px 8px rgba(0, 0, 0, 0.3);
}
.nav-container {
  display: flex;
  justify-content: space-between;
  align-items: center;
  padding: 0 40px;
  height: 60px;
}
.nav-logo {
  color: #ff6f61;
  font-size: 1.8rem;
  text-decoration: none;
  font-weight: 600;
  font-family: 'Lucida Sans', 'Lucida Sans Regular', 'Lucida Grande', 'Lucida Sans
Unicode', Geneva, Verdana, sans-serif;

```



```

    transition: color 0.3s ease;
  }
  .nav-menu {
    display: flex;
    list-style: none;
  }
  .nav-item {
    margin-left: 25px;
  }
  .nav-link {
    color: #fff;
    text-decoration: none;
    font-size: 1rem;
    transition: color 0.3s ease;
    position: relative;
  }
  .nav-link:hover {
    color: #ff6f61;
  }
  .nav-link::after {
    content: "";
    display: block;
    width: 0;
    height: 2px;
    background: #ff6f61;
    transition: width .3s;
    position: absolute;
    bottom: -5px;
    left: 0;
  }
  .nav-link:hover::after {
    width: 100%;
  }
  /* Hero Section */
  .hero-section {
    /* background-size: cover; */
    /* background-position: center; */
    /* background-repeat: no-repeat; */
    height: 100vh;
    /* display: flex; */
    /* justify-content: center; */
    /* align-items: center; */
    /* text-align: center; */
    /* color: #fff; */
    /* padding: 20px; */
    /* position: relative; */
    margin-top: 60px;
    background-color: white;
  }
  .hero-section::before {
    content: "";

```

```

    position: absolute;
    top: 0;
    left: 0;
    width: 100%;
    height: 100%;
    background: rgba(0, 0, 0, 0.5); /* Dark overlay */
}
.hero-content {
    position: relative;
    z-index: 1;
}
.hero-content h1 {
    font-size: 3rem;
    max-width: 80%;
    background: rgba(255, 255, 255, 0.1);
    padding: 20px 30px;
    border-radius: 10px;
    box-shadow: 0 4px 15px rgba(0, 0, 0, 0.3);
    backdrop-filter: blur(5px);
    animation: fadeInDown 1.5s ease;
}
/* Animation */
@keyframes fadeInDown {
    0% {
        opacity: 0;
        transform: translateY(-20px);
    }
    100% {
        opacity: 1;
        transform: translateY(0);
    }
}
/* Responsive */
@media (max-width: 768px) {
    .hero-content h1 {
        font-size: 2rem;
    }
    .nav-container {
        padding: 0 20px;
    }
    .nav-menu {
        flex-direction: column;
        background: rgba(0, 0, 0, 0.9);
        position: absolute;
        top: 60px;
        left: -100%;
        width: 100%;
        transition: all 0.5s ease;
    }
    .nav-menu.active {
        left: 0;
    }
}

```

```

    }
    .nav-item {
        margin: 10px 0;
    }
}
.hm{
    height: 100%;
}

```

Test.html

```

<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8" />
    <meta http-equiv="X-UA-Compatible" content="IE=edge" />
    <meta name="viewport" content="width=device-width, initial-scale=1.0" />
    <title>Stroke Test</title>
    <link rel="stylesheet" href="{{ url_for('static',filename='test.css') }}" />
    <link
        href="https://fonts.googleapis.com/css2?family=Poppins:wght@300;400;500;600
&display=swap"
        rel="stylesheet"
    />
</head>
<body>
    <div class="page">
        <h2 id="head">
            <span>Predict</span> and <span>improve</span> your
            <span>sleep.</span>
        </h2>
        <div class="container">
            <div class="stroke-test-card">
                <!-- <h2>Sleep Disorder Test</h2> -->
                <form action="{{ url_for('predict') }}" method="post">
                    <div class="form-grid">
                        <div class="form-group">
                            <label for="Gender">select</label>
                            <select id="Gender" name="Gender" required>
                                <option value="">Gender</option>
                                <option value="1">Male</option>
                                <option value="0">Female</option>
                            </select>
                        </div>
                        <div class="form-group">
                            <label for="Age">Age</label>
                            <input
                                type="number"
                                id="Age"
                                name="Age"

```

```

        placeholder="Enter age"
        min="18"
max="120"
        required
    />
</div>
<div class="form-group">
    <label for="Occupation">Occupation</label>
    <select id="Occupation" name="Occupation" required>
        <option value="0">Accountant</option>
        <option value="1">Doctor</option>
        <option value="2">Engineer</option>
        <option value="3">Lawyer</option>
        <option value="4">Manager</option>
        <option value="5">Nurse</option>
        <option value="6">Sales Representative</option>
        <option value="7">Salesperson</option>
        <option value="8">Scientist</option>
        <option value="9">Software Engineer</option>
        <option value="10">Teacher</option>
    </select>
</div>
<div class="form-group">
    <label for="Sleep Duration">Sleep Duration</label>
    <input
        type="number"
        id="Sleep Duration"
        name="Sleep Duration"
        placeholder="Enter Sleep Duration"
        min="4"
max="12"
step="0.1"
        required
    />
</div>
<div class="form-group">
    <label for="Quality of Sleep">Quality of Sleep</label>
    <input
        type="number"
        id="Quality of Sleep"
        name="Quality of Sleep"
        placeholder="Enter Quality of Sleep"
        min="1"
max="10"
        required
    />
</div>
<div class="form-group">
    <label for="Physical Activity Level">Physical Activity Level</label>
    <input
        type="number"

```

```

        id="Physical Activity Level "
        name="Physical Activity Level"
        placeholder="Enter Physical Activity Level"
        min="10"
    max="100"
        required
    />
</div>
<div class="form-group">
    <label for="Stress Level ">Stress Level</label>
    <input
        type="number"
        id="Stress Level "
        name="Stress Level"
        placeholder="Enter Stress Level"
        min="1"
    max="10"
        required
    />
</div>
<div class="form-group">
    <label for="BMI Category">BMI Category</label>
    <select id="BMI Category" name="BMI Category" required>
        <option value="">Select</option>
        <option value="0">Normal</option>
        <option value="1">Over Weight</option>
        <option value="2">Obese</option>
    </select>
</div>
<div class="form-group">
    <label for="BloodPressure_Upper">BloodPressure_Upper</label>
    <input
        type="number"
        id="BloodPressure_Upper"
        name="BloodPressure_Upper"
        placeholder="Enter BloodPressure_Upper"
        required
    />
</div>
<div class="form-group">
    <label for="BloodPressure_Lower">BloodPressure_Lower</label>
    <input
        type="number"
        id="BloodPressure_Lower"
        name="BloodPressure_Lower"
        placeholder="Enter BloodPressure_Lower"
        min="90"
    max="180"
        required
    />
</div>

```

```

<div class="form-group">
  <label for="Heart Rate">Heart Rate</label>
  <input
    type="number"
    id="Heart Rate"
    name="Heart Rate"
    placeholder="Enter Heart Rate"
    min="40"
    max="120"
    required
  />
</div>
<div class="form-group">
  <label for="Daily Steps">Daily Steps</label>
  <input
    type="number"
    id="Daily Steps"
    name="Daily Steps"
    placeholder="Enter Daily Steps"
    min="1000"
    max="30000"
    required
  />
</div>
<!-- <div class="form-group full-width">
  <label for="stroke-status">Stroke Status</label>
  <select id="stroke-status" required>
    <option value="">Select Stroke Status</option>
    <option value="0">No Stroke</option>
    <option value="1">Stroke</option>
  </select>
</div> -->
</div>
<button type="submit" class="btn">Test</button>
</form>
</div>
<div class="img">
  
</div>
</div>
</div>
</body>
</html>

```

Test.css

```

/* General Reset */
* {
  margin: 0;
  padding: 0;
}

```

```

    box-sizing: border-box;
    font-family: "Poppins", sans-serif;
    background-color: rgb(199, 196, 196);
    background-repeat: no-repeat;
  }
  .container {
    display: flex;
    align-items: center;
    /* justify-content: space-between; */
  }
  /* Body Styling */
  body {
    display: flex;
    justify-content: center;
    align-items: center;
    min-height: 100vh;
    /* background: #f5f5f5; */
    /* padding: 20px; */
    /* background-color: rgb(229, 219, 219) */
  }
  .img {
    margin: 20px;
  }
  /* Container for the Form */
  .container {
    /* background-color: hsl(0, 0%, 100%); */
    display: flex;
    justify-content: space-around;
    align-items: center;
    width: 90%;
    border-radius: 20px;
    /* box-shadow: 3px 3px 3px 3px gray; */
    max-width: 800px;
    height: 60%;
  }
  /* Stroke Test Card Styling */
  .stroke-test-card {
    width: 70%;
    /* background: #fff; */
    padding: 30px 30px;
    margin: 10px;
    border-radius: 10px;
    /* box-shadow: 2px 2px 2px 2px gray; */
    text-align: center;
    /* background-color: rgb(149, 143, 143); */
  }
  .stroke-test-card h2 {
    margin-bottom: 20px;
    font-weight: 600;
    color: black;
  }
  /* Form Grid Styling */

```

```

.form-grid {
  display: grid;
  grid-template-columns: repeat(2, 1fr);
  gap: 20px;
  margin-bottom: 20px;
}
.form-group label {
  display: block;
  margin-bottom: 5px;
  font-size: small;
  font-weight: 500;
  color: black;
}
.form-group input,
.form-group select {
  width: 100%;
  padding: 10px 15px;
  font-size: small;
  border: 1px solid #ddd;
  border-radius: 5px;
  outline: none;
  transition: border-color 0.3s ease;
}
.form-group input:focus,
.form-group select:focus {
  border-color: red;
}
/* Full-width for specific inputs */
.form-group.full-width {
  grid-column: span 2;
}
/* Test Button Styling */
.btn {
  width: 70%;
  padding: 12px 0;
  border: none;
  background: rgb(11, 178, 53);
  color: #fff;
  font-size: small;
  border-radius: 5px;
  cursor: pointer;
  transition: background 0.3s ease;
}
.btn:hover {
  background: rgb(72, 221, 8);
}
/* Responsive Design */
@media (max-width: 600px) {
  .form-grid {
    grid-template-columns: 1fr;
  }
}

```



```

}
h2 {
  text-align: center;
  padding: 8px;
  color: red;
  font-weight: 900;
}
.page {
  width: 800px;
  /* height: 80vh; */
  border-radius: 14px;
  /* box-shadow: 2px 2px 2px 2px rgb(178, 184, 181); */
}
#head {
  color: rgb(21, 1, 1);
  font-size: xx-large;
  font-weight: bolder;
}
span{
  color: rgb(43, 103, 13);
}

```

Insomnia.html

```

<!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="UTF-8" />
    <meta http-equiv="X-UA-Compatible" content="IE=edge" />
    <meta name="viewport" content="width=device-width, initial-scale=1.0" />
    <title>Stroke Prediction Result</title>
    <link rel="stylesheet" href="{ { url_for('static',filename='stroke.css') } }" />
    <link
      href="https://fonts.googleapis.com/css2?family=Poppins:wght@300;400;500;600
&display=swap"
      rel="stylesheet"
    />
  </head>
  <body>
    <div class="result-container">
      <div class="result-card">
        <h2>Sleep Disorder Prediction Result</h2>
        <!-- Display Prediction Result Here -->
        <p class="prediction">
          The result indicates:
          <span class="result-text">Insomnia</span>
        </p>
        <!-- Use JavaScript to dynamically change the text and color based on prediction
-->
        <button class="btn" onclick="window.location.href='test.html'">
          <a href="{ { url_for('test') } }" class="nav-link">Take Another Test</a>

```

```

    </button>
  </div>
</div>
</body>
</html>

```

Sleep Apnea.html

```

<!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="UTF-8" />
    <meta http-equiv="X-UA-Compatible" content="IE=edge" />
    <meta name="viewport" content="width=device-width, initial-scale=1.0" />
    <title>Stroke Prediction Result</title>
    <link rel="stylesheet" href="{{ url_for('static',filename='stroke.css') }}" />
    <link
      href="https://fonts.googleapis.com/css2?family=Poppins:wght@300;400;500;600
&display=swap"
      rel="stylesheet"
    />
  </head>
  <body>
    <div class="result-container">
      <div class="result-card">
        <h2>Sleep Disorder Prediction Result</h2>
        <!-- Display Prediction Result Here -->
        <p class="prediction">
          The result indicates:
          <span class="result-text">Sleep Apnea</span>
        </p>
        <!-- Use JavaScript to dynamically change the text and color based on prediction
-->
        <button class="btn" onclick="window.location.href='test.html'">
          <a href="{{ url_for('test') }}" class="nav-link">Take Another Test</a>
        </button>
      </div>
    </div>
  </body>
</html>

```

None.html

```

<!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="UTF-8" />
    <meta http-equiv="X-UA-Compatible" content="IE=edge" />
    <meta name="viewport" content="width=device-width, initial-scale=1.0" />
    <title>Stroke Prediction Result</title>
    <link rel="stylesheet" href="{{ url_for('static',filename='stroke.css') }}" />

```

```

<link
  href="https://fonts.googleapis.com/css2?family=Poppins:wght@300;400;500;600
&display=swap"
  rel="stylesheet"
/>
</head>
<body>
  <div class="result-container">
    <div class="result-card">
      <h2>Sleep Disorder Prediction Result</h2>
      <!-- Display Prediction Result Here -->
      <p class="prediction">
        The result indicates:
        <span class="result-text">None Disorder</span>
      </p>
      <!-- Use JavaScript to dynamically change the text and color based on prediction
-->

      <button class="btn" onclick="window.location.href='test.html'">
        <a href="{{ url_for('test') }}" class="nav-link">Take Another Test</a>
      </button>
    </div>
  </div>
</body>
</html>

```

App.py

```

# import numpy as np
from flask import Flask, request, jsonify, render_template
import pickle
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import LabelEncoder
import pandas as pd
scaler = StandardScaler()
app = Flask(__name__)
# Load the trained model
model=pickle.load(open('web/gnb_model1.pkl','rb'))
@app.route('/')
def index():
    return render_template('home.html')
@app.route('/test')
def test():
    return render_template('test.html') # Render the test.html page
@app.route('/login')
def login():
    return render_template('login.html') # Render the test.html page
@app.route('/reg')
def reg():
    return render_template('reg.html') # Render the test.html page

```

```

@app.route('/predict', methods=['POST'])
def predict():
    # Collect features from form data
    Gender=int(request.form['Gender']) Age=int(request.form['Age'])
    Occupation=int(request.form['Occupation']) SleepDuration=float(request.form['Sleep
    Duration']) QualityofSleep=int(request.form['Quality of Sleep'])
    PhysicalActivityLevel=int(request.form['Physical Activity Level'])
    stresslevel=int(request.form['Stress Level'])
    BMICategory=int(request.form['BMI Category'])
    HeartRate=int(request.form['Heart Rate'])
    DailySteps=int(request.form['Daily Steps'])

    BloodPressure_Upper=int(request.form['BloodPressure_Upper'])
    BloodPressure_Lower=int(request.form['BloodPressure_Lower'])

    feature_array=[[gender,age,hypertension,heart_disease,ever_married,work_type,Resid
    ence_type,
        # avg_glucose_level,bmi,smoking_status]]
    # feature_array = scaler.transform(feature_array)
    # features_array = np.array(feature_array).reshape(1, -1)
    # print(feature_array)

    input_data =
    pd.DataFrame([[Gender,Age,Occupation,SleepDuration,QualityofSleep,PhysicalActiv
    ityLevel,stresslevel,BMICategory,HeartRate,DailySteps,
        BloodPressure_Upper,BloodPressure_Lower]],
        columns=['Gender', 'Age', 'Occupation', 'Sleep Duration', 'Quality of
    Sleep', 'Physical Activity Level', 'Stress Level', 'BMI Category', 'Heart Rate', 'Daily
    Steps', 'BloodPressure_Upper', 'BloodPressure_Lower']) # Replace with your actual
    column names

    # Scale the input data using the previously fitted scaler
    # input_data_scaled = scaler.transform(input_data)

    # Convert the scaled data back to a DataFrame with original column names
    input_data = pd.DataFrame(input_data, columns=input_data.columns)

    # Now you can use column names for selection
    input_data = input_data[['Gender', 'Age', 'Occupation', 'Sleep Duration', 'Quality of
    Sleep', 'Physical Activity Level', 'Stress Level', 'BMI Category', 'Heart Rate', 'Daily
    Steps', 'BloodPressure_Upper', 'BloodPressure_Lower']] # Replace with your actual
    column names
    prediction = model.predict(input_data)
    print(prediction)

    if prediction == 0:
        return render_template("insomnia.html", prediction="Fake Profile...!")

```

```
elif prediction==1 :  
    return render_template("none.html", prediction="Real Profile...!")  
elif prediction==2 :  
    return render_template("sleepapnea.html", prediction="Real Profile...!")  
  
if __name__ == "__main__":  
    app.run(debug=True,port=5050)
```

9. RESULT ANALYSIS

The proposed system for sleep disorder classification was evaluated using a dataset consisting of 374 records, with individuals classified based on key sleep health indicators such as sleep duration, BMI, stress levels, and blood pressure. The performance of multiple machine learning models was analyzed to determine the most effective approach for predicting sleep disorders. The models tested include Gaussian Naïve Bayes with Genetic Algorithm (GNB-GA), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), Logistic Regression (LR), and Random Forest (RF). Among these, GNB-GA achieved the highest classification accuracy of 94%, outperforming all other classifiers.

The proposed system significantly improves classification accuracy by integrating Genetic Algorithm-based feature selection, which eliminates redundant attributes and optimizes model training. The evaluation metrics used include accuracy, precision, recall, F1-score, and specificity, ensuring a comprehensive assessment of model performance. The GNB-GA model demonstrated superior predictive ability, achieving high accuracy, recall, and F1-score, making it the most effective classifier for sleep disorder detection. The confusion matrix analysis further highlights the model's efficiency, with fewer false positives and false negatives compared to traditional machine learning classifiers.

The comparative analysis between the existing and proposed systems shows that the highest accuracy achieved using traditional models was lower, while the proposed system using GNB-GA reached 94% accuracy. This confirms that the proposed approach offers a more optimized solution for sleep disorder classification. The high recall score ensures that more actual sleep disorder cases are correctly identified, reducing the risk of misclassification. Additionally, the system's scalability and efficiency make it suitable for deployment in real-world healthcare applications, where automated sleep disorder screening can assist healthcare professionals in making more informed and data-driven decisions.

Table 9.1 Evaluation Metrics of Models with Genetic Algorithm

Models	Accuracy	F1-Score	Precision	Recall
GNB	0.94	0.94	0.94	0.94
MLP	0.93	0.93	0.93	0.93
SVM	0.93	0.93	0.93	0.93
RF	0.92	0.91	0.92	0.92
LR	0.92	0.91	0.92	0.92
KNN	0.85	0.86	0.87	0.85

Table 6.1 shows the evaluation metrics (accuracy, precision, recall, and F1 score) for different models used in sleep disorder prediction, including SVM,GNB,MLP,RF,LR,KNN with GA and resulted highest with GNB-GA.

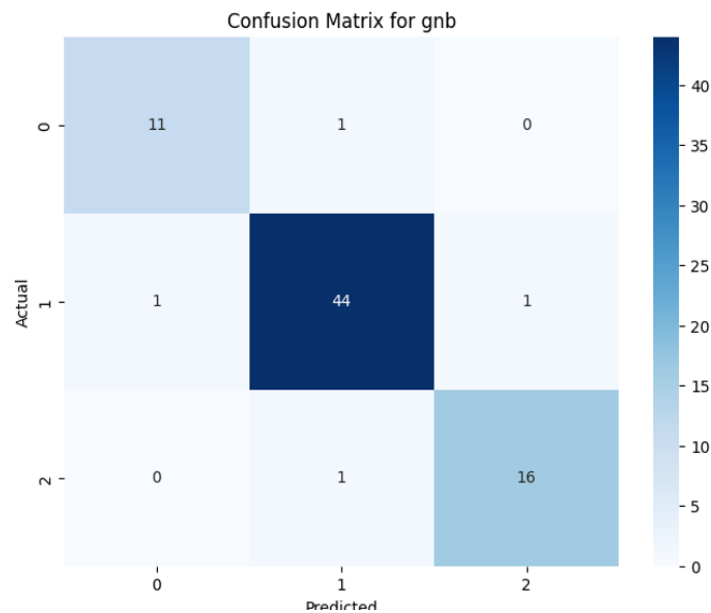


Fig 9.1 Confusion Matrix

Fig 9.1 shows a confusion matrix for a Gaussian Naïve Bayes (GNB) classifier, illustrating the model's performance in predicting three classes. The diagonal values represent correct predictions, while off-diagonal values indicate misclassifications.

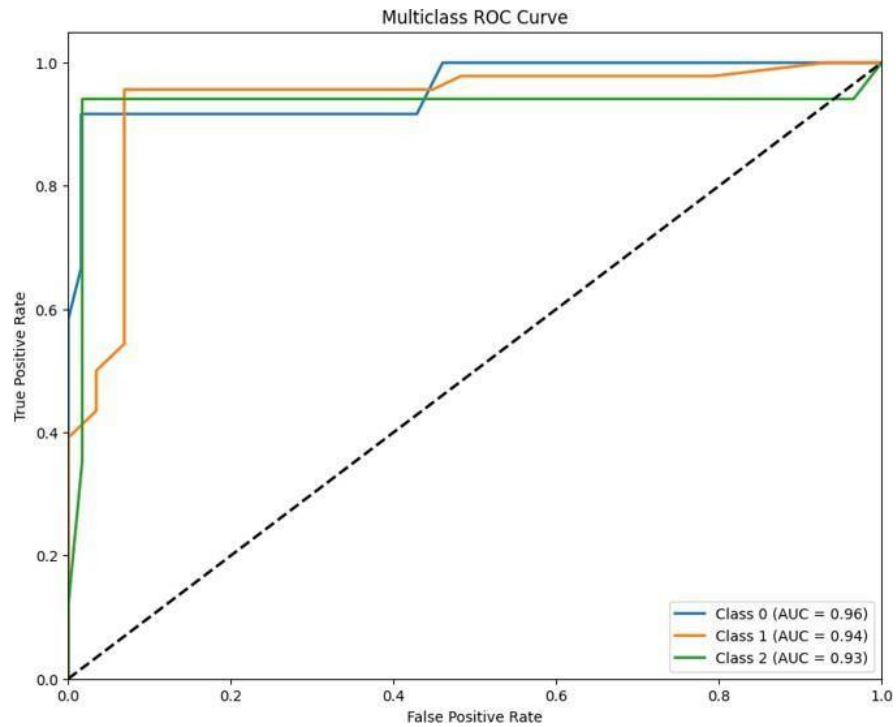


Fig 9.2 ROC Curve of GNB

Fig 9.2 ROC curve displays a multiclass ROC curve, showing the trade-off between the true positive rate and false positive rate for three classes. The AUC values indicate high classification performance, with Class 0 achieving 0.96, Class 1 at 0.94, and Class 2 at 0.93.

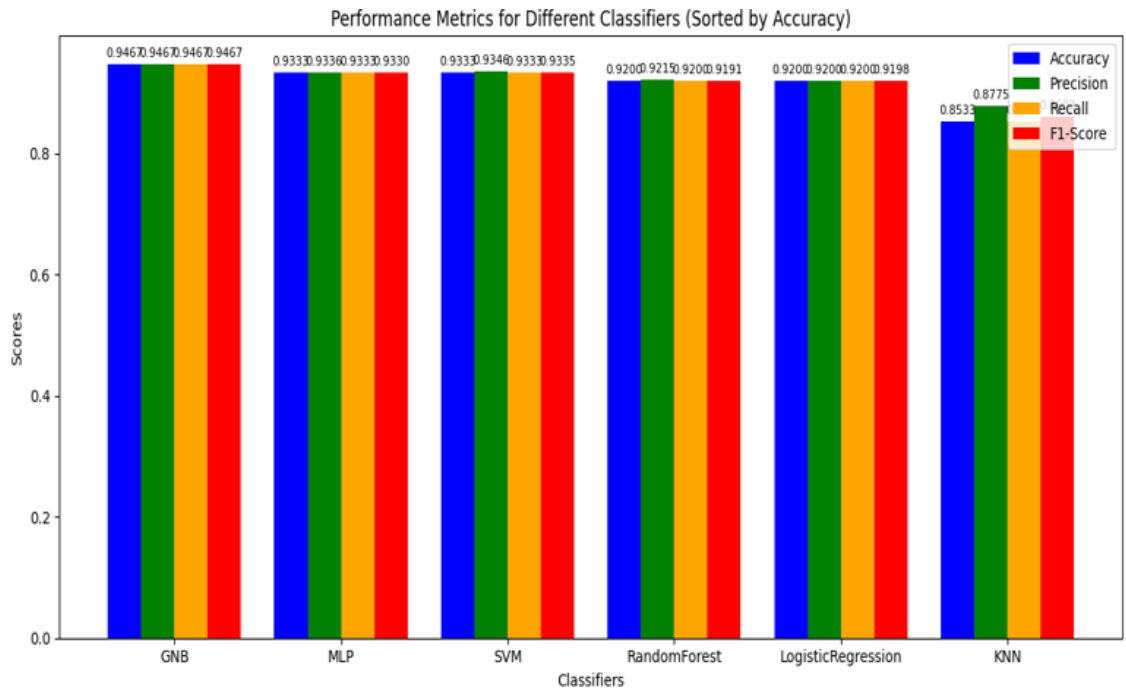


Fig 9.3 Model Performance Metrics

Figure 9.3 Model Performance Metrics – The bar chart compares performance metrics (Accuracy, Precision, Recall, and F1-Score) for different classifiers, sorted by accuracy. Gaussian Naïve Bayes (GNB) achieved the highest accuracy (0.947), followed by MLP and SVM, while KNN performed the lowest.

10. TEST CASES

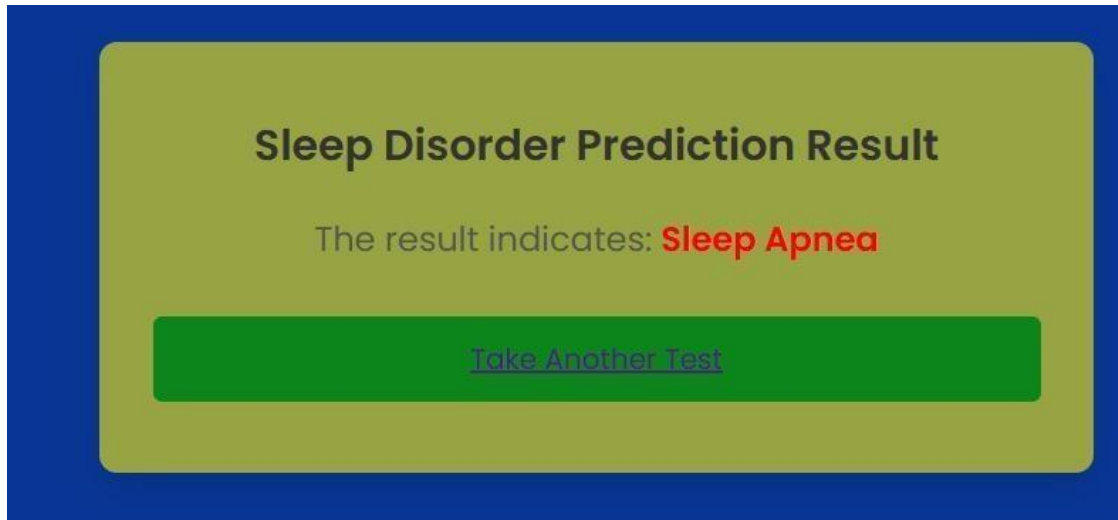


Fig 10.1 Identified “Sleep Apnea disease”

Fig 10.1 shows the result screen of the Sleep Disorder Prediction application, displaying the test outcome and providing an option to take another test.



Fig 10.2 Identified “Insomnia Sleep Disease”

Fig 10.2 shows the result screen of the Sleep Disorder Prediction application, indicating that no liver disease was detected, with an option to take another test.



Fig 10.3 Identified “None Disorder”

Fig 10.3 shows the result screen of the Sleep Disorder Prediction application, indicating that no liver disease was detected, with an option to take another test.

 The image displays a web form titled "Predict and improve your sleep." in a green and black font. The form contains several input fields and dropdown menus arranged in two columns. On the left side, the fields include: "select" (a dropdown menu showing "Male"), "Occupation" (a dropdown menu showing "Accountant"), "Quality of Sleep" (a text input field with "7"), "Stress Level" (a text input field with "4"), "BloodPressure_Upper" (a text input field with "70"), "Heart Rate" (a text input field with "76"), "Physical Activity Level" (a text input field with "90"), "BMI Category" (a dropdown menu showing "Normal"), "BloodPressure_Lower" (a text input field with "80"), and "Daily Steps" (a text input field with "2000"). A red box highlights the "Age" field, which contains the value "1". A tooltip message with an orange exclamation mark icon appears next to the "Age" field, stating "Value must be greater than or equal to 18." At the bottom of the form, there is a green "Test" button. On the right side of the form, there is a small graphic titled "Types of Sleep Disorders" showing various sleep-related icons and the hashtag #GetGoodSleep.

Fig 10.4 Form Validation

Fig 10.4 includes validation checks for user inputs to ensure data accuracy. In this case, the age field has a restriction requiring users to enter a value of 18 or higher. Similar validations may be applied to other attributes.

11. USER INTERFACE

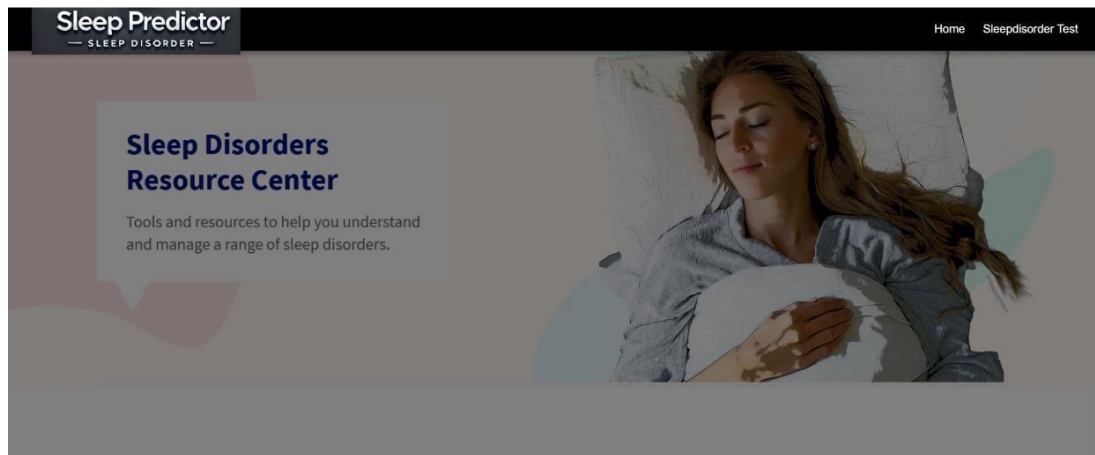


Fig 11.1 Home Screen

Fig 11.1 appears to be a webpage from "Sleep Predictor" focused on sleep disorders. It features a "Sleep Disorders Resource Center" that provides tools and resources to help understand and manage sleep disorders. The banner includes an image of a woman sleeping peacefully, with navigation options for "Home" and a "Sleep Disorder Test".

Fig 11.2 Disease Prediction Form

Fig 11.2 shows the input form for the Sleep Disorder Prediction application, where users can enter patient details and test values for analysis. The "Test" button at the bottom initiates the prediction process.

12. CONCLUSION

The best feature selection was carried out by applying a Genetic Algorithm which showed the most marked performance on various classifiers. In this, the GNB classifier reached the highest accuracy that was 94%, which was better in all other models. This reflects the feasibility of GAs in filtering out unnecessary attributes to yield efficiency on lightweight models, such as GNB, so as to attain better predictive accuracy. This coupling of GA and GNB provides an efficient and practically viable approach to enhancing performance with predictive modeling tasks sans the entrance into the realms of complex algorithms.

The implications are very high, especially in applications where wearable health devices and real-time diagnostics are concerned. GNB is light and, along with the feature reduction capability of GA. However, there are still several limitations to the study namely a relatively small size of the dataset and possible biases that may decrease the generalizability of the results; in fact, subsequent studies will be able to cross the problem with the help of larger and diversified datasets and deep learning models that can be deployed for the evaluation of performance in similar situations. Further investigation of the effect of GA on other lightweight and complex classifiers would offer broader insights into the applicability of these across domains.

The research presented here focuses on the computational effectiveness as well as real-world usability furthering the efficient integration of feature selection based on GA with lightweight classifiers in real-world applications, potentially benefitting from the same higher accuracy and speed in even more resource-constrained settings.

13. FUTURE SCOPE

The future scope of the proposed sleep disorder classification system focuses on enhancing accuracy, scalability, and real-world applicability through advanced machine learning and optimization techniques. Future research can explore higher-order optimization strategies, such as deep learning architectures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), to further improve predictive performance by analyzing sleep patterns and time-series data. Expanding the dataset with diverse patient records can enhance model generalization, ensuring better adaptability across different demographics and sleep disorder categories. Additionally, integrating the system with electronic health records (EHRs) and cloud-based healthcare platforms would enable real-time monitoring and remote diagnosis, making sleep disorder detection more accessible for individuals in rural and underserved areas.

Further refinements in Genetic Algorithm (GA)-based feature selection can optimize the selection of key sleep-related attributes, improving both accuracy and computational efficiency. Lastly, incorporating Explainable AI (XAI) techniques will enhance the interpretability of predictions, allowing medical professionals to trust and utilize the system effectively for early detection, treatment planning, and personalized healthcare interventions.

14. REFERENCES

- [1] P. Levy, V. Viot-Blanc, and J.-L. Pépin, "Sleep disorders and their classification—an overview," *Sleep Apnea*, vol. 35, pp. 1-12, 2006.
- [2] C. R. Soldatos, J. D. Kales, T. L. Tan, and A. Kales, "Classification of sleep disorders," *Psychiatric Annals*, vol. 17, no. 7, pp. 454-458, 1987.
- [3] A. Atianashie Miracle, E. Armah, and N. Mohammed, "A portable GUI-based sleep disorder system classification based on convolution neural networks (CNN) in Raspberry Pi," *J Eng Appl Sci Human*, vol. 6, pp. 13-23, 2021.
- [4] M. Sharma, J. Tiwari, V. Patel, and U. R. Acharya, "Automated identification of sleep disorder types using triplet half-band filter and ensemble machine learning techniques with EEG signals," *Electronics*, vol. 10, no. 13, p. 1531, 2021.
- [5] T. Wongsirichot and A. Hanskunatai, "A classification of sleep disorders with optimal features using machine learning techniques," *J Health Res*, vol. 31, no. 3, 2017.
- [6] T. Alshammari, "Applying machine learning algorithms for the classification of sleep disorders," *IEEE Access*, 2024.
- [7] P. K. Yadav, U. K. Singh, J. J. A. Kovilpiaali, and R. Tamilarasi, "Sleep disorder detection using machine learning method," in *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, IEEE, pp. 1530- 1532, 2023.
- [8] I. A. Hidayat, "Classification of sleep disorders using random forest on sleep health and lifestyle dataset," *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, vol. 3, no. 2, pp. 71-76, 2023.
- [9] G. Airlangga, "Evaluating machine learning models for predicting sleep disorders in a lifestyle and health data context," *JIKO (Jurnal Informatika dan Komputer)*, vol. 7, no. 1, pp. 51-57, 2024.
- [10] J. Ramesh, N. Keeran, A. Sagahyroon, and F. Aloul, "Towards validating the effectiveness of obstructive sleep apnea classification from electronic health records using machine learning," *Healthcare*, vol. 9, no. 11, p. 1450, 2021.
- [11] Y. J. Kim, J. S. Jeon, S.-E. Cho, K. G. Kim, and S.-G. Kang, "Prediction models for obstructive sleep apnea in Korean adults using machine learning techniques," *Diagnostics*, vol. 11, no. 4, p. 612, 2021.
- [12] P. Tripathi, M. A. Ansari, T. K. Gandhi, R. Mehrotra, M. B. B. Heyat, F. Akhtar,C.

- C. Ukwuoma, et al., "Ensemble computational intelligence for insomnia sleep stage detection via the sleep ECG signal," *IEEE Access*, vol. 10, pp. 108710- 108721, 2022.
- [13] Sleep Health and Lifestyle Dataset. Available at: <http://www.kaggle.com/datasets/uom190346a/sleep-health-andlifestyledataset>, 2023.
- [14] M.Sireesha, S. N. TirumalaRao, Srikanth Vemuru, Optimized Feature Extraction and Hybrid Classification Model for Heart Disease and Breast Cancer Prediction International Journal of Recent Technology and Engineering Vol - 7, No 6, Mar - 2019 ISSN - 2277-3878, Pages – 1754 – 1772
- [15] Sunayna, S. Siva, SN Thirumala Rao, and M. Sireesha. "Performance evaluation of machine learning algorithms to predict breast cancer." In Computational Intelligence in Data Mining: Proceedings of ICCIDM 2021, pp. 323-335. Singapore: Springer Nature Singapore, 2022.
- [16] Moturi, Sireesha, Jhansi Vazram Bolla, M. Anusha, M. Mounika Naga Bhavani, Srikanth Vemuru, SN Tirumala Rao, and Sneha Ananya Mallipeddi. "Prediction of Liver Disease Using Machine Learning Algorithms." In International Conference on Data Science and Applications, pp. 243-254. Singapore: Springer Nature Singapore, 2023.

CERTIFICATE



3rd Congress on Smart Computing Technologies
(CSCT 2024)
Organized by
National Institute of Technology, Sikkim, India



Certificate of Presentation

This certificate is proudly awarded to

Dileep Kumar Modepalli

for presenting the paper titled

**Making Sleep Disorder Classification using Optimized Machine
Learning Models**

authored by

**Chandana Yamani, Dileep Kumar Modepalli, Abhishikth Dola,
Yalavarthi Alimoon Shaik, S Siva Nageswara Rao, Sireesha Moturi**
in the 3rd Congress on Smart Computing Technologies (CSCT 2024)

held during

December 14-15, 2024.

Prof. Mukesh Saraswat
General Chair

Dr. Abhishek Rajan
General Chair



<https://www.scrs.in/conference/csct2024>

SCRS/CSCT2024/PC/468

Making Sleep Disorder Classification Using Optimized Machine Learning Models

Chandana Yamani^{1*}, Dileep Kumar Modepalli², Abhishikth Dola³, Yalavarthi Alimoon Shaik⁴, Dr. S.Siva Nageswarara Rao⁵, Dr. Sireesha Moturi⁶

¹Asst.Professor,Department of CSE, Narasaraopeta Engineering College,
Narasaraopet-522601, Palnadu, Andhra Pradesh, India.

^{2,3,4}Department of CSE, Narasaraopeta Engineering College, Narasaraopet-522601,
Palnadu, Andhra Pradesh, India.

⁵Professor,Department of CSE, Narasaraopeta Engineering College,
Narasaraopet-522601, Palnadu, Andhra Pradesh, India.

⁶Assoc.Professor,Department of CSE, Narasaraopeta Engineering College,
Narasaraopet-522601, Palnadu, Andhra Pradesh, India.
chandana.nrt nec@gmail.com

Abstract. Sleep disorders have a significant impact on human health and early diagnosis is crucial for the improvement of quality of life. In this work, machine learning models that classify sleep disorders have been developed from the data set of 374 participants that comprises features relating to sleep health and lifestyle, such as sleep duration, physical activity, stress levels, and BMI. It has used many machine learning algorithms, like GNB, KNN, SVM, RF, Logistic Regression, and ANN/MLP Classifier. Dataset was highly preprocessed to deal with missing values and normalize features, thereby offering the best performance to the models. Feature engineering and optimization through a Genetic Algorithm improved the predictive ability of these models. GA was highly effective especially in the extraction of the most relevant features, improving the classification accuracy, and dealing with the problem of small dataset size. The best accuracy by GNB was achieved in combination with GA, namely 94%, compared to all other models: SVM 93%, Logistic Regression 92%, and ANN 93%. Importantly, issues like dealing with missing data and feature extraction, where meaningful features such as systolic and diastolic blood pressure levels had to be extracted, were addressed in order to enhance the outcome. Our results show that the proposed model effectively detects patterns in the database, hence providing consequent results that can be used in real-time by health care professionals to diagnose sleep-related conditions. This consequently decreases manual effort towards diagnosis and improves its timing. It also allows for scalability, which may fit easily into wearable health devices and clinical decision systems that are intended to prevent long-term health risks due to sleep disorders left untreated. The proposed model facilitates automated classification of sleep disorders. This will be a valuable tool for advancing healthcare outcomes and fostering future innovations in sleep medicine.

Keywords: Feature Engineering · BMI (Body Mass Index) · Optimization · Real-Time Diagnostics · Automated Sleep Disorder Detection.

1 Introduction

Recently, sleeping disorders have emerged as a serious public health issue due to findings where sleep quality and duration directly impact the total well-being of an individual. Once a person's normal sleep pattern becomes disturbed, they begin to suffer from very critical health problems like heart diseases and metabolic syndrome or even cognitive impairment. The most common sleeping disorder is the most common form of sleep apnea. In general, it is described as the chronic disruption of sleep ventilation mainly due to upperairway obstruction. Untreated sleep apnea may lead to possibly life-threatening complications that come from health conditions such as heart disease, stroke, or hypertension. Overall treatment for sleep apnea remains Continuous Positive Airway Pressure, whereby one keeps the airway open throughout the night. More severe conditions may require surgical interference with the treatment of obstruction within the airway on a permanent basis. CPAP has proven to be one of the strongest treatments, and its symptoms have reduced the complications and problem of sleep apnea in many cases [1]. Insomnia is yet another very common sleep disorder. It makes falling asleep difficult, staying asleep, and waking up too early. There are two types: acute and chronic insomnia. Former usually does not last long. But the latter requires long time treatment; though pharmacological along with behavioural therapy is commonly used for treatment of chronic insomnia. Among the latter there is one type of cognitive behavioural therapy. This therapy, combined with the enhancement of sleep hygiene practice has been proven to result in considerable success in the treatment of the condition [2]. In the past couple of years, technology has improved much-detection and management of sleep disorders. For instance, the design for portable diagnostic systems, usually based on CNNs, has transformed big classification techniques of sleep disorders. Such designs provide much more accurate and accessible tools for assessing sleep disorders outside the traditional clinical settings. This means that patients can now seek timely interventions to enhance their long-term outcomes [3]. Ensemble machine learning algorithms have also played a major role in the advancement of developing the diagnostic accuracy for sleep disorders. Algorithm that can classify a complex pattern of sleep can become highly useful for predictions and diagnosis related to sleep disorders. This has held special value in the differentiation of sleep disorders that are similar, but will have different treatment modalities [4]. These technological developments in practice would bring much improvement in diagnosis and treatments. Such developments do heighten the quality of lives of afflicted people suffering from sleep disorders, but they also open pathways for future discoveries in sleep medicine. The advancement of diagnostic tools and further research on treatment methodologies can help a healthcare provider deal better with the complexities of sleep-related disorders [5]. Hence, this approach can lead to better care and improved long-term health outcomes.

2 Literature Review

Sleep disorders as insomnia and obstructive sleep apnea have become serious issues for an individual's well-being. Traditional approaches to diagnosis, such as PSG, are well resourced. Machine learning methods are increasingly applied to automate and improve the detection of sleep disorders, drawing on diverse data and advanced algorithms.

Recent works investigated different ML models to classify sleep disorders. For example, Alshammari attempted machine learning algorithms on the Sleep Health and Lifestyle dataset, where ANN obtained the highest accuracy of 92.92% [6]. This shows that lifestyle and health datasets can be used to predict with high accuracy, but the work does not provide advanced optimization approaches, such as feature selection algorithms, to improve the performance of the models presented.

Yadav et al. combined decision trees and Support Vector Machines (SVM) with physiological data, including EEG and ECG signals, to classify sleep disorders. The addition of physiological data greatly enhanced classification precision, with the PSDG dataset proving useful for training robust models [7]. This study does underscore the need for utilizing rich physiological datasets, though such datasets often require specialized equipment, making such applications not scalable.

Hidayat applied the Random Forest algorithm to the Sleep Health and Lifestyle dataset, which obtained 92% accuracy. This task showed that lifestyle-related factors, such as sleep patterns and health profiles, are highly indicative of the sleep disorders [8]. It lacks in-depth feature selection techniques which may be useful to further optimize the performance of the model. Similarly, Airlangga compared Logistic Regression (93% accuracy) and neural networks on the same dataset, elaborating on the strengths and weaknesses of traditional machine learning versus deep learning approaches [9]. These comparisons focus on how different methodologies are performing differently on similar datasets, although they did not discuss some techniques to increase feature relevance such as Genetic Algorithms (GA).

Ramesh et al. used machine learning models: Random Forest and neural networks, in EHRs data to classify OSA. This demonstrates the applicability of ML in healthcare with real-world clinical data [10]. However, the use of EHRs does not scale to wider applications for which simpler data sources like lifestyle information are more easily available.

Kim et al. developed prediction models for obstructive sleep apnea among Korean adults using demographic data combined with physiological signals, including ECG and respiratory inputs. The study underlined the requirement of models tailored specifically to population-specific data, and robust results were achieved [11]. Although useful for their specific population, the contribution of physiological data makes this approach less applicable to larger populations.

Tripathi et al. discussed the ensemble learning techniques, which include bagging and boosting, for the detection of insomnia using sleep ECG signals. The results showed significant improvements in the detection accuracy achieved

by combining multiple machine learning techniques [12]. Limitations to real-world applicability arise due to the need for heavy, high-cost hardware-intensive physiological signals in comparison to more feasible simpler data sources, such as lifestyle or behavioral information, in certain settings.

Limitations persist in previous works. Most models fail to scale or do not utilize optimization techniques in feature selection, leading to suboptimal performance. This study fills these gaps by integrating Genetic Algorithms (GA) for feature selection with ML models like GNB, SVM, and ANN. GA not only reduces feature redundancy but also increases model accuracy, particularly in smaller, diverse datasets like Sleep Health and Lifestyle. Unlike previous work, we extracted latent features, such as the levels of systolic and diastolic blood pressures, to further improve the model’s predictiveness. This is a substantial contribution because it combines sophisticated feature engineering with scalable inputs, good for real-time applications.

Building from previous strengths and avoiding their weaknesses, our work offers a new way toward the classification of sleep disorders, based on both advanced optimization techniques and accessible datasets.

3 Methodology

The methodology in the fig.1 presents a typical workflow for machine learning. It begins with using the collected data and preprocessing followed by feature extraction/normalization to prepare it to feed into the system. Then,

the dataset needs to be divided to be split for training and testing purposes. For identifying which features were most appropriate for the training of the model, a genetic algorithm was employed with the objective of feature selection while optimizing the input variables of the model being trained. Lastly, the model was trained, and the performance of that model was estimated by evaluating accuracy results.

3.1 Dataset

The data is collected from the Kaggle named as “Sleep Health and Lifestyle” This dataset covers 374 participants with 12 features [13]. It dives into their habits, work life, physical activities, health markers like heart and blood pressure. Researchers generally use it to see how these factors might affect sleep quality or even lead to issues like sleep apnea and Insomnia. Although the dataset was useful, the number of participants in it which was just 374 necessarily limited it from the perspective of generalization. Thus, to minimize these problems, careful preprocessing and feature selection steps were followed to reduce overfitting. However, the study recognizes that including more diverse participant groups into the dataset in future research will strengthen the finding. In addition, linear patterns of the dataset and demographic representation also influenced the choice of models, though the results need validation through future studies on larger more varied datasets.

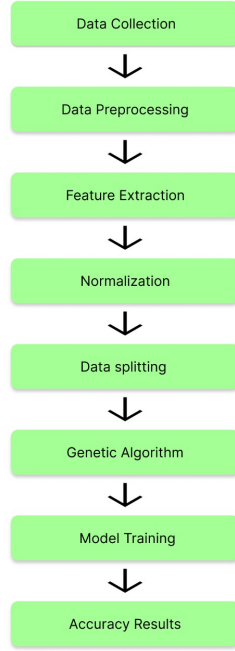


Fig. 1. Workflow of Models

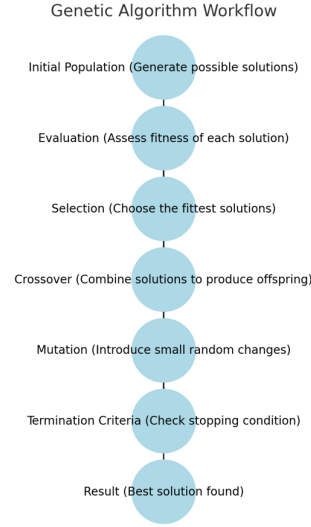


Fig. 2. steps of Genetic Algorithm

3.2 Data Preprocessing

i) Handling Null Values : Null values can cause the data inconsistency and may the model cannot be trained well at the same time there will be the downfall in achieving accuracy. In our dataset the target column contains the null values we replace it with the 'none' this might help to train the model without null values. For the target attribute we cannot use the statistic method such as mean and mode due to dataset contains the people with not having any sleep disorder.

ii) Feature Extraction : This method is mainly used to extract the new feature that is hidden and which we cannot see. we can make the meaningful variables by extracting them[14].so we can improve our model performance rapidly. As coming to our dataset there is a hidden feature that can be extracted from the blood pressure attribute can be called as systolic and diastolic which might be helpful for increasing the model performance.

iii) Label Encoding : label encoding technique is mainly used to transform the categorical columns into numerical columns. It helps in applying mathematical operations and improving performance. It assigns each category to a unique identification number. Our dataset contains the categorical attributes

like gender, occupation, bmi and sleep disorder these will be transformed into a numerical columns.

iv) Standardization : Scaling of features to have similar ranges of values is very important in machine learning. Standardization - or normalization - is the process of rescaling one or more attributes to a common range, usually 0 to 1. This method will be of much help in the data that contains various of features with various units.

3.3 Analysis of Features

Relation with Target Feature : The scatter plots are suitable for visualizing and simplifying the relationship between the two numerical variables. They can show the patterns, trends, and interconnection between the variables. Outliers can be easily identified. In this case, a scatter plot can help us to identify on how the age might be related to sleep disorders. If you have a categorical target variable, you might consider other visualizations like bar plots. In the below we can see the scatter plot that are associated with Age and Sleep Disorder and bar plot graph of Gender related with Sleep Disorder.

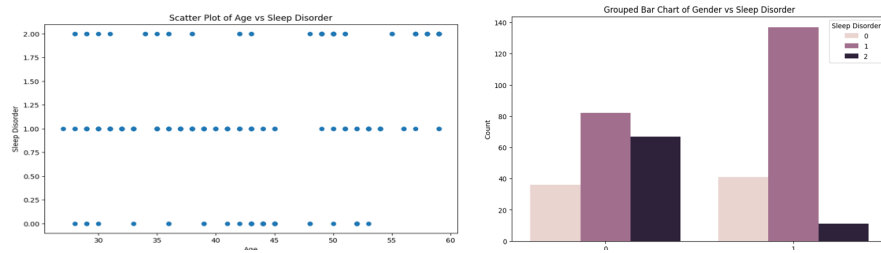


Fig. 3. Relationship between features

3.4 Models

1. Support Vector Machine (SVM) : Support Vector Machine is one among those used supervised learning algorithms that have been widely applied for the classification problems. It can adapt to the two types of data distributions, linear and non-linear, by using different kernel functions, such as a linear kernel, polynomial kernel, and Radial Basis Function kernel. In this experiment, the features of the dataset were aligned linearly, so a linear kernel(kernel='linear') was used [15].

2. Random Forest (RF) : For dealing with complex data distributions and possible overfitting, the ensemble learning technique of Random Forest has been applied. Hyperparameters have been tuned to optimize the model. These include values such as `n-estimators = 600` that creates many decision trees, `max-depth = None`; that is, letting them get as deep as possible so that splits on a node are optimal. As compared to criterion which uses entropy to maximize information gain on a split, the parameters `min-samples-split` and `min-samples-leaf` prevent growth trees from getting unbalanced [16].

3. K-Nearest Neighbour (KNN) : KNN is a non-parametric algorithm that is quite simple and yet powerful, since it bases its predictions about the class of a given data point on its nearest neighbors. For this small dataset, `n-neighbors=2` has been used as hyperparameter to optimize performance, combining the use of Euclidean for the distance metric in order to preserve very high accuracy in classification.

4. Gaussian Naive Bayes (GNB) : Though the assumptions are rather simplistic, GNB is computationally efficient and proves to be good when dealing with small-scale datasets. The hyperparameter `var-smoothing=1e-9` had been used in the order to achieve numerical stability. Classifiers proved efficient and trained.

5. ANN (MLP Classifier) : Artificial Neural Network (ANN) or Multi-Layer Perceptron (MLP) was applied because it mimics the structure of the human brain consisting of interconnected neurons. That network was set using `hidden-layer-sizes=(24,)`, properly optimized for the size and complexity of this particular dataset. ANN finds best to capture anything but linear patterns as well as relationships between features.

6. Logistic Regression (LR) : Logistic regression was a very efficient classification algorithm, which was interpretable and suitable for a smaller dataset. `max-iter=1000` was used as a hyperparameter during the whole training process so that the model must converge to stable and reliable results.

7. Genetic Algorithm (GA) : GA Feature selection and model improvement used an optimization technique. Critical hyperparameters of the GA were used, including `population-size=12` evaluated 12 feature subsets per generation and `n-generations=9`, assuring sufficient evolution to guarantee optimal solutions; a relatively high `mutation-rate=0.8` was applied in order to maintain diversity and avoid local optima. GA assisted considerably in the optimization procedure related to feature selection, especially for a small dataset, to reduce dimensionality and improve model performance.

4 Results and Discussion

4.1 Steps in Genetic Algorithm

In the fig. 2 we can see the steps of Genetic Algorithm

- Phase 1: First Generation Generation
- Randomly creates an initial population of individuals (chromosomes), as binary strings, who carry a potential solution.
- Phase 2: Evaluation
- Evaluate individual fitness by using a problem-specific function according to which the effectiveness of an attribute is judged for those individuals.
- Phase 3: Selection
- Chooses parents for the subsequent generation based on performance ranking, where the better ones have better fitness.
- Phase 4: Crossover
- Combines two chromosomes from parent pairs using single-point, multi-point, or uniform crossover technique to produce offspring.
- Phase 5: Termination
- Its primary check is whether convergence criteria are met or not. They check for maximum fitness or minimal improvement after a certain number of generations. If not met, then reevaluate.

4.2 Confusion Matrix and ROC Curve

In fig.6 a confusion matrix for a GNB classifier that gives a visual summary of the model performance

The confusion matrix for the GNB classifier shows that this model correctly classified three classes of sleep disorder. The classification was correct in 11 instances for class 0, 44 for class 1, and 16 for class 2. There were misclassifications for 1 instance of class 0 miscoded as class 1, 2 instances of class 1 miscoded, (1 as class 0 and 1 as class 2), and 1 instance of class 2 miscoded as class 1. There is overlap in such classes that were found from the results. Further exploration of feature distributions and model tuning could lead to this reduction in errors and an improvement in the distinction among similar sleep disorders, thus giving better real-world performance in clinical applications.

In Fig.5 we can see the ROC curve analysis showed that the model performs best for Class 0 ($AUC = 0.96$), followed by Class 1 ($AUC = 0.94$) and Class 2 ($AUC = 0.93$), showing a slight drop in performance for Class 2. Curves are illustrated for misclassifications; false positives show the tendency to be higher in classes with lower AUCs, indicating greater challenges in the discrimination of particular classes. These misclassifications can have real-world consequences, particularly when Class 1 and Class 2 represent similar sleep disorders, leading to the very real consequences of mistaken diagnoses. Misclassifications could well improve the model's general effectiveness when put to use in practice.

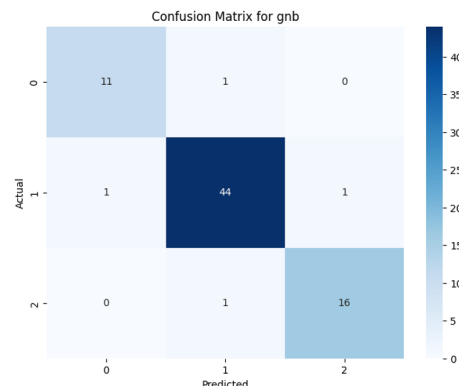


Fig. 4. Confusion Matrix of GNB

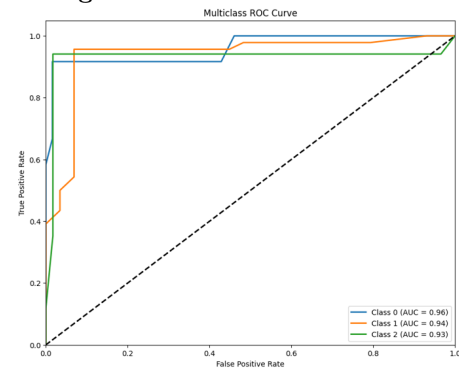


Fig. 5. ROC Curve of GNB

4.3 Performance Metrics

In the Table.1, We can see the different performance metrics like Accuracy,F1-Score,Precision and Recall related to the different classifiers with various results are mentioned.

Table 1. Performance Metrics Of Classifiers Using Genetic Algorithm.

Models	Accuracy	F1-Score	Precision	Recall
GNB	0.94	0.94	0.94	0.94
MLP	0.93	0.93	0.93	0.93
SVM	0.93	0.93	0.93	0.93
Random Forest	0.92	0.91	0.92	0.92
Logistic Regression	0.92	0.91	0.92	0.92
KNN	0.85	0.86	0.87	0.85

4.4 Anova Test

To assess the statistical significance of the performance differences between the classifiers, we conducted an ANOVA test. The results showed a significant difference in performance, with an F-statistic of 14.16 and a p-value of 3.49×10^{-12} . This indicates that at least one classifier outperforms the others in terms of accuracy or fitness score.

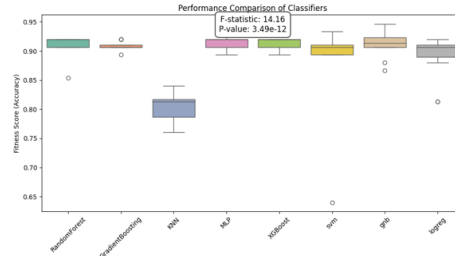


Fig. 6. Anova test for Classifiers

4.5 Computational Efficiency and Trade-offs:

GNB has greater computational efficiency than MLP. This model would train in just 0.0020 seconds and predict in 0.0012 seconds, making the same perfect for working in real-time or low-resource embedded systems. MLP takes 0.1741 seconds for training and 0.0015 seconds for predictions, thus is slow.

This means that GNB has a much smaller model size, compared to MLP which is around 1.48 KB, and MLP at 13.48 KB, which is ideal in case the devices have constraints regarding storage. GNB performance competes with MLP while giving it a higher fitness score: 0.9467 compared to 0.9333 while being significantly cheaper in terms of calculation. Therefore, in applications where speed, efficiency, or resource limitations become critical, GNB seems to be the better choice, balancing accuracy and efficiency.

4.6 Comparative Analysis

In this study, we implemented these multi-class algorithms: GNB, MLP, SVM, Random Forest, Logistic Regression, and KNN using a genetic algorithm that also performs the feature selection process and enhances it. It is so important to establish that the genetic algorithm powered the enhancement process in feature subsets for enhancing model performance. GNB achieved an accuracy of 94%, with F1-score, precision, and recall all in strong agreement. MLP and SVM followed closely with nearly balanced results at 93%, while the Random Forest and

Logistic Regression showed robust accuracy at 92%, though the F1-score was a little lesser, at 91%. KNN had very poor performance as its accuracy rested at 85%, which seems less effective even after the optimization process. These results have been found, which are underlining the effectiveness of using genetic algorithms in a direct feature selection process and more so in enhancing GNB, MLP, and SVM but may require further adjustments on KNN in order to boost accuracy.

5 Conclusion

The best feature selection was carried out by applying a Genetic Algorithm which showed the most marked performance on various classifiers. In this, the GNB classifier reached the highest accuracy that was 94%, which was better in all other models. This reflects the feasibility of GAs in filtering out unnecessary attributes to yield efficiency on lightweight models, such as GNB, so as to attain better predictive accuracy. This coupling of GA and GNB provides an efficient and practically viable approach to enhancing performance with predictive modeling tasks sans the entrance into the realms of complex algorithms.

The implications are very high, especially in applications where wearable health devices and real-time diagnostics are concerned. GNB is light and, along with the feature reduction capability of GA. However, there are still several limitations to the study namely a relatively small size of the dataset and possible biases that may decrease the generalizability of the results; in fact, subsequent studies will be able to cross the problem with the help of larger and diversified datasets and deep learning models that can be deployed for the evaluation of performance in similar situations. Further investigation of the effect of GA on other lightweight and complex classifiers would offer broader insights into the applicability of these across domains.

The research presented here focuses on the computational effectiveness as well as real-world usability furthering the efficient integration of feature selection based on GA with lightweight classifiers in real-world applications, potentially benefitting from the same higher accuracy and speed in even more resource-constrained settings.

References

1. P. Levy, V. Viot-Blanc, and J.-L. Pépin, "Sleep disorders and their classification—an overview," *Sleep Apnea*, vol. 35, pp. 1-12, 2006.
2. C. R. Soldatos, J. D. Kales, T. L. Tan, and A. Kales, "Classification of sleep disorders," *Psychiatric Annals*, vol. 17, no. 7, pp. 454-458, 1987.
3. A. Atianashie Miracle, E. Armah, and N. Mohammed, "A portable GUI-based sleep disorder system classification based on convolution neural networks (CNN) in Raspberry Pi," *J Eng Appl Sci Human*, vol. 6, pp. 13-23, 2021.

4. M. Sharma, J. Tiwari, V. Patel, and U. R. Acharya, "Automated identification of sleep disorder types using triplet half-band filter and ensemble machine learning techniques with EEG signals," *Electronics*, vol. 10, no. 13, p. 1531, 2021.
5. T. Wongsirichot and A. Hanskunatai, "A classification of sleep disorders with optimal features using machine learning techniques," *J Health Res*, vol. 31, no. 3, 2017.
6. T. Alshammari, "Applying machine learning algorithms for the classification of sleep disorders," *IEEE Access*, 2024.
7. P. K. Yadav, U. K. Singh, J. J. A. Kovilpiaali, and R. Tamilarasi, "Sleep disorder detection using machine learning method," in *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, IEEE, pp. 1530-1532, 2023.
8. I. A. Hidayat, "Classification of sleep disorders using random forest on sleep health and lifestyle dataset," *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, vol. 3, no. 2, pp. 71-76, 2023.
9. G. Airlangga, "Evaluating machine learning models for predicting sleep disorders in a lifestyle and health data context," *JIKO (Jurnal Informatika dan Komputer)*, vol. 7, no. 1, pp. 51-57, 2024.
10. J. Ramesh, N. Keeran, A. Sagahyroon, and F. Aloul, "Towards validating the effectiveness of obstructive sleep apnea classification from electronic health records using machine learning," *Healthcare*, vol. 9, no. 11, p. 1450, 2021.
11. Y. J. Kim, J. S. Jeon, S.-E. Cho, K. G. Kim, and S.-G. Kang, "Prediction models for obstructive sleep apnea in Korean adults using machine learning techniques," *Diagnostics*, vol. 11, no. 4, p. 612, 2021.
12. P. Tripathi, M. A. Ansari, T. K. Gandhi, R. Mehrotra, M. B. B. Heyat, F. Akhtar, C. C. Ukwuoma, et al., "Ensemble computational intelligence for insomnia sleep stage detection via the sleep ECG signal," *IEEE Access*, vol. 10, pp. 108710-108721, 2022.
13. Sleep Health and Lifestyle Dataset. Available at: <http://www.kaggle.com/datasets/uom190346a/sleep-health-andlifestyledataset>, 2023.
14. M. Sireesha, S. N. Tirumala Rao, Srikanth Vemuru, Optimized Feature Extraction and Hybrid Classification Model for Heart Disease and Breast Cancer Prediction International Journal of Recent Technology and Engineering Vol - 7, No 6, Mar - 2019 ISSN - 2277-3878, Pages - 1754 - 1772
15. Sunayna, S. Siva, SN Thirumala Rao, and M. Sireesha. "Performance evaluation of machine learning algorithms to predict breast cancer." In Computational Intelligence in Data Mining: Proceedings of ICCIDM 2021, pp. 323-335. Singapore: Springer Nature Singapore, 2022.
16. Moturi, Sireesha, Jhansi Vazram Bolla, M. Anusha, M. Mounika Naga Bhavani, Srikanth Vemuru, SN Tirumala Rao, and Sneha Ananya Mallipeddi. "Prediction of Liver Disease Using Machine Learning Algorithms." In International Conference on Data Science and Applications, pp. 243-254. Singapore: Springer Nature Singapore, 2023.

sleep

ORIGINALITY REPORT

7%

SIMILARITY INDEX

5%

INTERNET SOURCES

4%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1

publications.polymtl.ca

Internet Source

1%

2

scholarworks.aub.edu.lb

Internet Source

1%

3

www.hindawi.com

Internet Source

1%

4

Dinesh Kumar Maddina, Suresh Kumar Raju S, Dharmaiah Gurram, Muneerah Al Nuwairan.

"Dynamics of non-Newtonian Casson fluid and Cattaneo-Christov heat flux impacts on a rotating non-uniform surface due to Coriolis force: A comparison study of ANFIS-PSO and ANN", Results in Engineering, 2024

Publication

<1%

5

scrs.in

Internet Source

<1%

6

Submitted to Harrisburg University of Science and Technology

Student Paper

<1%

7	Submitted to Whitecliffe College of Art & Design Student Paper	<1 %
8	"International Conference on Innovative Computing and Communications", Springer Science and Business Media LLC, 2023 Publication	<1 %
9	"Proceedings of the 2nd International Conference on Big Data, IoT and Machine Learning", Springer Science and Business Media LLC, 2024 Publication	<1 %
10	www.internationaljournalsrg.org Internet Source	<1 %
11	accentsjournals.org Internet Source	<1 %
12	ijirt.org Internet Source	<1 %
13	"Data Science and Applications", Springer Science and Business Media LLC, 2024 Publication	<1 %
14	journal.ittelkom-pwt.ac.id Internet Source	<1 %
15	Ali Şenol, Tarık Talan, Cemal Aktürk. "A new hybrid feature reduction method by using MCMSTClustering algorithm with various	<1 %

feature projection methods: a case study on sleep disorder diagnosis", Signal, Image and Video Processing, 2024

Publication

-
- | | | |
|-------|---|------|
| 16 | Manish Sharma, Divyansh Anand, Sarv Verma, U. Rajendra Acharya. "Automated insomnia detection using wavelet scattering network technique with single-channel EEG signals", Engineering Applications of Artificial Intelligence, 2023 | <1 % |
| <hr/> | | |
| 17 | Recep Sinan Arslan, Hasan Ulutas, Ahmet Sertol Köksal, Mehmet Bakir, Bülent Çiftçi. "End-to end decision support system for sleep apnea detection and Apnea-Hypopnea Index calculation using hybrid feature vector and Machine learning", Biocybernetics and Biomedical Engineering, 2023 | <1 % |
| <hr/> | | |
| 18 | mdpi-res.com
Internet Source | <1 % |
| <hr/> | | |
| 19 | pdfcoffee.com
Internet Source | <1 % |
| <hr/> | | |
| 20 | www.e-informatyka.pl
Internet Source | <1 % |
| <hr/> | | |
| 21 | "Emerging Research in Data Engineering Systems and Computer Communications", | <1 % |

Springer Science and Business Media LLC, 2020

Publication

Exclude quotes On

Exclude matches Off

Exclude bibliography On