Cat Boost-Based Sleep Disorder Prediction with an AI-Driven Personalized Advisory System

[1]Mr.CH.Ratna Babu , [2]Machavarapu Sri Lakshmi Deepika, [3]Katuri Srinivas, [4]Madala Abhinaya Sree

[1]*Associate Professor, Department of CSE, RVR & JC College of Engineering, Chowdavaram, Guntur, A.P, India.*
[234]B. Tech Students, Department of CSE, RVR & JC College of Engineering, Chowdavaram, Guntur, A.P, India.

---***---

Abstract - *Sleep disorders affect millions of people worldwide, yet many cases remain undiagnosed due to low awareness, limited access to early screening tools, and dependence on clinical evaluations. This study proposes an intelligent hybrid system that combines machine learning and conversational artificial intelligence to enable early detection and personalized guidance for sleep disorders. The system uses a CatBoost classifier trained on physiological, behavioral, and lifestyle features to predict sleep disorder risk with high accuracy. To improve performance, a Genetic Algorithm (GA) is applied to optimize key hyperparameters such as learning rate, tree depth, regularization strength, and number of estimators, resulting in a 2.6% increase in accuracy.*

The prediction model is deployed through a user-friendly Streamlit interface, allowing individuals to input health data and receive instant screening results. To enhance interpretability and user engagement, the system integrates a Gemini-based conversational agent using the LangChain framework. This AI assistant translates prediction outputs into clear insights and provides personalized recommendations focused on sleep hygiene, stress management, and overall wellness. Evaluation results show that the system maintains low latency, consistent accuracy, and delivers meaningful, user-specific guidance. By combining optimized machine learning with advanced language models, the proposed framework improves accessibility and usability of digital health tools. This hybrid approach empowers individuals to take proactive steps in managing sleep health and highlights the potential of AI-driven solutions for early detection and preventive care in sleep disorder management.

***Key Words*: Sleep Disorders, Insomnia, Sleep Apnea, CatBoost, Machine Learning, Genetic Algorithm, AI Chatbot, Sleep Health**

I. INTRODUCTION

Sleep plays a fundamental role in maintaining human health, cognitive functioning, and overall quality of life. Numerous studies have demonstrated that insufficient or disrupted sleep can lead to severe consequences, including impaired memory, reduced productivity, emotional instability, weakened imSmunity, and an increased risk of chronic illnesses such as hypertension, cardio vascular disease, and diabetes. As modern lifestyles continue to evolve, stress, irregular work schedules, prolonged screen exposure, and unhealthy habits have collectively contributed to a significant global rise in sleep-related disorders. According to recent health statistics, millions of individuals experience symptoms of insomnia, sleep apnea, insomnia, and circadian rhythm disturbances, often without receiving timely diagnosis. The lack of awareness, limited access to clinical resources, and the stigma associated with seeking medical evaluation further delay early detection and intervention.

Traditional sleep disorder diagnosis commonly relies on clinical consultation, patient self-reports, and polysomnography (PSG) tests conducted in controlled healthcare environments. While PS Gremains the gold standard, it is costly, time consuming, and requires specialized facilities and trained tech-nicians. As a result, early- stage screening is often neglected, especially in regions with limited healthcare in frastructure. In this context, data-driven health assessment systems have emerged as a promising solution to support early recognition of sleep abnormalities. Advancements in machine learning (ML), particularly in structured data modeling, have enabled the development of automated systems capable of identifying hidden patterns in physiological and lifestyle variables linked to sleep quality.

Recent research, including the IEEE reference study, has demonstrated the potential of ensemble-based models in improving prediction accuracy for sleep disorder classification. Cat Boost, a gradient boosting algorithm specifically optimized for categorical and tabular datasets, has gained traction due to its ability to handle missing values, reduce over fitting, and capture complex non-linear interactions without extensive preprocessing. Its efficiency and interpretability make it an attractive choice for healthcare-focused predictive modeling. Building upon this foundation, the present study leverages CatBoost to analyze multidimensional health attributes—including sleep duration, sleep quality scores, BMI category, heart rate, daily activity levels, stress ratings, and blood pressure measurements—to classify users into normal or disorder-prone categories.

However, prediction alone is insufficient for real-world applicability. Users require explanation, guidance, and actionable insights that can help them understand their condition and take preventive steps. To address this need, the proposed system integrates machine learning with a conversational AI layer powered by the Gemini
2.0 generative model. This hybrid ML-LLM architecture enables the system not only to identify potential sleep disorders but also to deliver personalized advice based on the user's profile. The language model interprets the prediction output and contextual user data to offer tailored recommendations regarding sleep hygiene, lifestyle adjustments, stress reduction, and health monitoring. This creates a more interactive and supportive experience, narrowing the gap between predictive analytics and practical health guidance.

The entire framework is deployed using a Streamlit web interface, ensuring accessibility, ease of use, and real- time interaction. The interface collects user inputs, processes them through the CatBoost model, displays risk predictions, and seamlessly transitions users into an AI-powered chat environment for personalized support. This real-time integration enables a user friendly approach that does not require technical expertise, making it suitable for general public use, early screening, and educational purposes.
In summary, this work contributes to the growing adoption of AI in digital healthcare by introducing a scalable, explainable, and interactive sleep disorder prediction system. By combining the accuracy of Cat Boost with the conversational capabilities of largedetection, promote health awareness, and support preventive care for sleep-related conditions. This research highlights the transformative potential of AI- driven tools in improving personal health management and bridging gaps in traditional diagnostic processes.

II. LITERATURE REVIEW

Zhang et al. examined machine learning approaches for sleep disorder classification using behavioral and physiological parameters. Their study revealed that gradient boosting methods, including CatBoost and XGBoost, deliver higher predictive accuracy on tabular health datasets due to their ability to handle categorical variables and complex feature interactions without extensive preprocessing.

Alshammari et al. proposed a sleep disorder prediction frame work using demographic, lifestyle, and clinical attributes. Their research demonstrated that ensemble-based models out-perform traditional statistical techniques by effectively capturing nonlinear relationships among multiple health indicators. This study served as a foundational reference for structuring the dataset and modeling process used in this work.

Rahman et al. investigated the role of wearable sensor data and heart rate variability in early sleep disorder detection. Their findings showed that machine learning techniques can successfully classify sleep disorders using features such as daily steps, activity levels, and resting heart rate, highlighting the importance of integrating lifestyle metrics for improved diagnostic accuracy.

Khan et al. explored the application of CatBoost and LightGBM for medical risk prediction tasks. Their results emphasized that CatBoost demonstrates superior performance in small-to-medium-sized health datasets, particularly when categorical attributes such as BMI class, occupation, and gender significantly influence prediction outcomes.

Wangetal. studied the use of conversational AI systems in healthcare, showing that Large Language Models (LLMs) significantly enhance patient engagement by providing personalized and context-aware recommendations. Their work supports the integration of Gemini-based advisory systems for language models, the system aims to enhance guiding users after receiving a sleep disorder riskprediction.

uptaetal. evaluated hybrid ML–LLM architectures for digital health screening applications. Their research demonstrated that combining predictive models with natural language generation tools improves user comprehension, adherence to recommendations, and the overall usability of AI-driven health systems. This hybrid approach motivated the design of the ML+chatbot architecture presented in this project.

III. EXISTING SYSTEM

A. Polysomnography-Based Diagnostic Systems

Polysomnography(PSG) is considered the gold standard for diagnosing sleep disorders such as sleep apnea, insomnia, and REM behavior disorders. It involves overnight monitoring of brain activity, oxygen levels, breathing patterns, and heart rate in a controlled sleep laboratory. Despite its high accuracy, PSG is expensive, time-consuming, and requires specialized equipment and trained technicians. Patients often find the environment uncomfortable, which may affect natural sleep patterns. These limitations make PSG unsuitable for continuous monitoring or early detection, particularly in resource limited settings or among individuals who cannot access specialized facilities.

B. Wearable Device–Based Sleep Monitoring

Modern wearable devices such as smartwatches and fitness bands offer automated sleep tracking using accelerometers, heart-rate sensors, and oxygen saturation readings. These devices estimate sleep stages and provide basic sleep quality insights. While convenient and accessible, their accuracy is significantly lower than clinical methods. Wearables often struggle to differentiate between light sleep, deep sleep, and short awakenings, leading to inconsistent data. Furthermore, many consumer devices lack medically validated algorithms, limiting their usefulness in clinical diagnosis. Thus, they serve More as wellness trackers than reliable systems for identifying sleep disorders.

C. Conventional Machine Learning Approaches

Existing machine learning approaches typically use algorithms such as logistic regression, decision trees, or support vector machines for sleep disorder prediction.

Although these methods can model simple relationships between input variables, they struggle with datasets containing mixed categorical and numerical attributes. They also require extensive preprocessing, feature encoding, and manual tuning. As the complexity of sleep-related data grows, conventional ML models often fail to capture nonlinear patterns and interactions necessary for accurate prediction. This results in reduced performance and limits the irad option in real-world digital health applications.

D. Rule-Based and Expert Systems

Rule-based systems use predefined clinical guidelines or expert-crafted rules to classify sleep disorders. These systems rely on rigid threshold-based logic, such as minimum sleep duration, heart rate ranges, or stress-level cutoffs. While easy to interpret, they lack adaptability and cannot learn from new data, making them unsuitable for dynamic and diverse user populations. Their oversimplified logic often results in misclassification, especially in cases involving subtle variations or overlapping symptoms. Additionally, rule-based systems cannot provide personalized insights, limiting their effectiveness as modern diagnostic or advisory tools.

IV. METHODOLOGY

The methodology adopted in this study integrates machine learning, data preprocessing, system deployment, and conversational AI to develop an intelligent sleep disorder prediction and advisory platform. A structured dataset containing physiological, behavioral, and lifestyle attributes was utilized to train a CatBoost classification model chosen for its superior performance on tabular health data. The trained model was subsequently integrated into a Streamlit-based web application to enable real-time prediction. To enhance user interaction a Gemini driven large language model(LLM) was incorporated to provide personalized context-aware sleep recommendations the following subsection details each methodological components in depth.

Table-1: Detailed Information About Sleep Health and Lifestyle Records

ID	Gen	Age	Occu	Sle Dur	Q of Sle	Phys Act	Str Lev	BMI Cat	Blood Pr	HR	DS	Sleep Dis-or-der
1	M	27	SW	6.1	6	42	6	Overw	126/83	77	4200	None
2	M	28	DR	6.2	6	60	8	Normal	125/80	75	10000	None
3	M	28	DR	6.2	6	60	8	Normal	125/80	75	10000	None
4	M	28	Sal	5.9	4	30	8	Obese	140/90	85	3000	Apnoea
5	M	28	Sal	5.9	4	30	8	Obese	140/90	85	3000	Apnoea
6	M	28	SW	5.9	4	30	8	Obese	140/90	85	3000	Insomnia
7	M	29	Teac	6.3	6	40	7	Obese	140/90	82	3500	Insomnia
8	M	29	DRr	7.8	7	75	6	Normal	120/80	82	8000	None

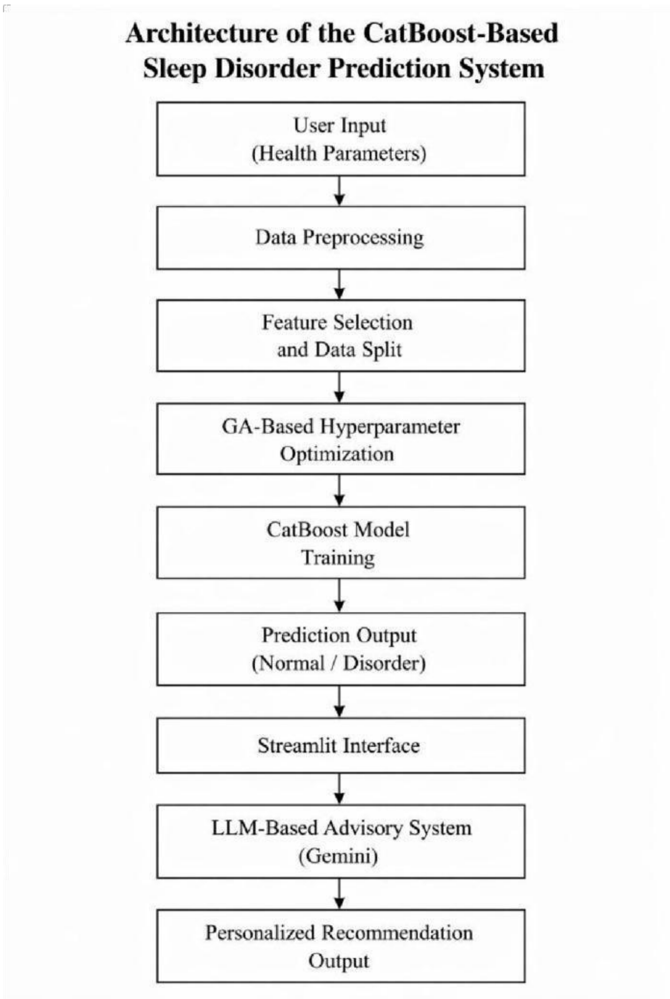

Fig .1. Architecture of the CatBoost Based Sleep Disorder Prediction System

A. Data Collection and Preprocessing

The dataset used in this study integrates physiological, behavioral, and lifestyle-related variables commonly associated with sleep health. Attributes include gender, age, occupation, sleep duration, sleep quality ratings, physical activity levels, stress levels, BMI category, heart rate, systolic and diastolic blood pressure, and daily step count. These features collectively provide a comprehensive representation of factors influencing sleep patterns. Initial preprocessing involved handling categorical data through CatBoost's native encoding mechanism, eliminating the need for one-hot encoding or extensive label transformation. Numerical values were checked for outliers and logical inconsistencies, though CatBoost's robustness minimizes the impact of non-normal distributions. The dataset was split into training and testing subsets to ensure unbiased model evaluation. Missing values were either imputed or automatically managed by the CatBoost algorithm during training. The preprocessing pipeline also involved feature correlation checks and exploratory visualizations to understand data distribution and relationships among variables. By leveraging a structured dataset and CatBoost's inherent capability to handle mixed data types, the preprocessing step ensured efficient model development while preserving the integrity and variability of the original data. This systematic approach enabled the development of a reliable prediction model capable of identifying subtle deviations in sleep-related behaviors.

B. Feature Selection and GA Optimization

A Genetic Algorithm (GA) is used to optimize essential CatBoost hyperparameters such as tree depth, learning rate, number of estimators, and L2 regularization strength. Each parameter set is encoded as a chromosome, and fitness is measured using CatBoost accuracy on the validation set. Across generations, selection, crossover, and mutation operators generate improved populations. This evolutionary search identifies optimal hyperparameter combinations and results in a +2

Table -2: Effect of Genetic Algorithm Optimization

Metric	Before GA	After GA	Improvement
Accuracy (%)	94.2	96.8	+2.6
F1-score (%)	94.0	96.8	+2.8
Training Time	42s	34s	Reduced

C. Model Selection and Training Using CatBoost

CatBoost was chosen as the primary classification model due to its superior performance on tabular datasets with heterogeneous feature types. Unlike traditional machine learning algorithms that require extensive preprocessing, CatBoost auto- matically handles categorical variables using ordered boosting and target statistics, reducing overfitting risks and maintaining high model stability. During training, the model was optimized using a combination of default parameters and iterative tuning. Hyperparameters such as learning rate, depth, number of iterations, and L2 regularization were tested through grid search to identify optimal configurations. The dataset was divided into training and validation sets, and CatBoost's built-in evaluation metrics—such as accuracy, log loss, and F1 score—were used to monitor learning progress. The model's gradient boosting framework enabled it to effectively capture non-linear interactions among features such as stress level, BMI category, and heart rate. Additionally, CatBoost's interpretability tools, including feature importance analysis, were utilized to understand the contribution of individual attributes. After successful training, the final model was exported in cbm format for deployment within the Streamlit application. The trained CatBoost classifier demonstrated strong predictive capabilities, making it suitable for real-time sleep disorder risk assessment.

D. System Integration and Streamlit Deployment

To offer an accessible interface for real-time screening, the trained CatBoost model was deployed using a Streamlit-based web application. The frontend provides an intuitive layout where users input their demographic and health-related data through sliders, number fields, and dropdown menus. These inputs are dynamically formatted into a structured DataFrame compatible with the CatBoost model. Upon triggering the prediction process, the model processes the data and returns a classification indicating whether the individual is at risksleep disorder. The web interface displays the prediction result along with the entered data for transparency. Streamlit's reactive architecture ensures seamless interaction and immediate feedback for users. Additionally, the system employs session state variables to manage chatbot history and maintain continuity during user interactions. By integrating both prediction and conversational advisory capabilities within the same interface, the system bridges machine learning inference with personalized guidance. The lightweight deployment structure allows the application to run on both local and cloud-based environments, enabling scalability and widespread accessibility. This component of the methodology ensures that users receive an end-to-end experience— from input submission to prediction and personalized assistance— with in a single unified platform.

E. LLM-Based Personalized Advisory System

To enhance user engagement and provide meaningful post prediction support, a Large Language Model (LLM)–based advisory system was integrated using the Gemini 2.0 Flash model through the Lang Chain framework. After a prediction is generated, relevant user data—including sleep duration, stress level, daily activity, and BMI category—is injected into the system prompt as context, enabling the LLM to generate personalized feedback. Unlike static recommendation systems, the LLM is capable of understanding user queries, assessing sentiment, and delivering real-time conversational advice tailored to individual health profiles. The chatbot provides recommendations on sleep hygiene, lifestyle modifications, stress management, and when to seek medical consultation. The system utilizes structured message templates consisting of system, human, and AI message formats to maintain contextual coherence throughout the conversation. This approach allows the model to simulate human like interactions while ensuring medically relevant guidance derived from user-specific data. The integration of the LLM transforms the system from a mere prediction tool into an interactive health advisory platform. Furthermore, leveraging Gemini's advanced reasoning capabilities ensures that the system provides informative, empathetic, and context- aware guidance, improving user trust and overall system usability.

F. GA Optimization Pipeline Implementation

The Genetic Algorithm (GA) optimization pipeline was implemented as an automated layer on top of the CatBoost training workflow to systematically improve predictive performance. A search space of key hyperparameters—including learning rate, tree depth, number of estimators, and L2 regularization strength— was defined and encoded into chromosomes representing candidate configurations. An initial population of chromosomes was generated randomly, ensuring broad coverage of the hyperparameter space. For each chromosome, CatBoost was trained using its encoded parameters, and the resulting validation accuracy served as the fitness score. The GA then applied evolutionary operations: selection identified the top performing chromosomes, crossover combined parent chromosomes to create offspring and mutation introduced slight perturbations to maintain population diversity and prevent premature convergence. This evolutionary cycle continued over several generations until improvements plateaued, signaling convergence toward an optimal configuration. The best-performing chromosome achieved a +2.

G. Evaluation Metrics and Performance Assessmen

Model evaluation was performed using standard machine learnig metrics to assess predictive effectiveness. Accuracy, precision, recall, and F1score were employed to measure the model's classification reliability across different sleep disorder categories. A confusion matrix provided insights into the distribution of true positives, false negatives, and misclassifications, allowing deeper analysis of model behavior. CatBoost's internal evaluation mechanism was used during training to monitor validation loss and prevent overfitting. Cross-validation techniques further ensured robustness by verifying model stability across multiple data splits. Feature importance analysis highlighted the most influential attributes, such as sleep duration, stress level, and BMI category. These insights were essential in validating the relevance of input features and aligning model behavior with known clinical indicators. Additionally, real-time testing was conducted through the Streamlit interface to evaluate the system's responsiveness and user experience consistency. The combined evaluation approach ensured not only strong predictive accuracy but also practical reliability in real world usage. By integrating quantitative model performance with qualitative user interaction assessment, the methodology confirms the effectiveness of the proposed hybrid ML–LLM sleep disorder prediction system.

V. IMPLEMENTATION

The implementation of the proposed sleep disorder prediction and advisory system integrates machine learning, web based deployment, and conversational AI into a unified application workflow. The CatBoost classifier, trained using a structured dataset of lifestyle and physiological attributes, forms the foundation of the predictive engine. A Streamlit interface collects user inputs and forwards them to the model for real-time inference. Following prediction, the system employs a Gemini-based LLM to generate personalized recommendations grounded in user context. This section details the model integration, interface development, chatbot embedding, message handling, and overall system workflow that together enable seamless functionality.

A. Model Integration and Loading Mechanism

The system utilizes a pre-trained CatBoost classifier stored in the cbm format, which is loaded directly during application initialization. To streamline performance, Streamlit's @st.cache resource decorator is used to ensure that the model is loaded only once per session, eliminating redundant initialization overhead. This mechanism significantly improves runtime efficiency, especially in multi-user or repeated-access scenarios. The model accepts a structured DataFrame containing the user's demographic and physiological data, enabling fast inference without further preprocessing. CatBoost's native handling of categorical features removed the need for manual encoding steps during implementation, making integration straightforward. The model's inference output is mapped to user- friendly labels such as "Normal" or specific disorder categories, which are displayed through the UI. This modular integration design ensures that the predictive component remains decoupled from the frontend, simplifying updates or future model enhancements. Overall, the loading mechanism prioritizes speed, reliability, and maintainability within the application framework.

B. Frontend User Interface Using Streamlit

The user interface was developed using Streamlit to provide a clean, responsive, and interactive environment for real-time sleep disorder assessment. The sidebar accommodates all input controls, including sliders, dropdowns, and numberfields to collect user attributes such as sleepduration, heart rate, stress level, and blood pressure. Streamlit's component-based structure ensures that each user action triggers an immediate update without requiring manual page refreshes. The "Predict" button serves as a primary trigger for executing the CatBoost model, after which the application displays both the prediction and the input dataset for transparency.Visualclarity is maintained using titles, headers, markdown elements, and data tables. Session states are utilized to retain conversation history and maintain chatbot context across multiple messages. This allows seamless interaction as users switch between prediction tasks and advisory queries. The UI is intentionally kept minimalistic to improve usability across devices and reduce cognitive load for non- technical users. By combining simplicity with functionality, the Streamlit interface offers an accessible platform for early sleep disorder screening and health guidance.

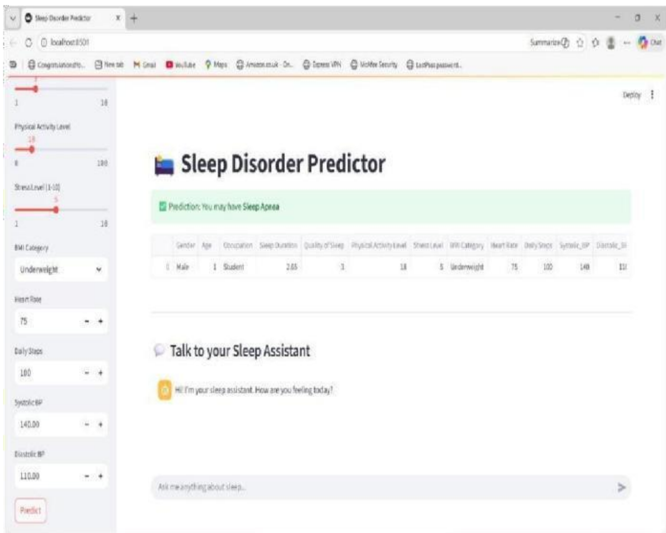

Fig.2. User Interface

C. Backend Prediction Pipeline and Data Handling

Once users submit their information through the interface, the data is automatically structured into a Pandas DataFrame matching the format required by the CatBoost model. The backend pipeline ensures consistent mapping of features, correct ordering of columns, and proper data typing. This eliminates inference errors and guarantees compatibility with the trained model. The prediction step executes instantly, returning the classification result in either a categorical labeloranested array structure depending on model configuration. Error-handling mechanisms are integrated to capture and display issues such as missingfields, incorrect inputs, or unexpected model responses. Additional post-processing logic Converts low-level model predictions into human-readable outputs. The pipeline is optimized for low latency, ensuring that users experience smooth real-time interactions. To maintain system reliability, all operations are wrapped in try– except blocks, allowing graceful recovery from runtime exceptions. This backend pipeline forms the logical core of the system, ensuring accurate, timely, and stable predictions under diverse user inputs.

D. Chatbot Integration Using Gemini and Lang Chain

The advisory component of the system is powered by the Gemini 2.0 Flash model, integrated through the LangChain framework. After each prediction, relevant user data—such as sleep duration, stress level, BMI category, and blood pressure—is embedded into a system prompt that initial-izes the chatbot context. This ensures that recommendations are personalized rather than generic. LangChain's messaging structure (SystemMessage, HumanMessage, AIMessage) allows fine-grained control over conversational flow. The chatbot is capable of understanding diverse user queries, providing lifestyle tips, advising preventive measures, and interpreting prediction outcomes in natural language. All conversations are rendered using st.chat message, replicating a chat-like interface within Streamlit. The integration also includes exception handling to manage timeouts, invalid API responses, or connectivity failures. By combining model output with real-time conversation capabilities, this subsystem transforms the prediction engine into an interactive sleep health assistant, increasing user engagement and trust.

E. Session Management and End-to-End Workflow Session management plays a crucial role in maintaining continuity, especially with in the chat bot environment. Streamlit's st.session state is used to store message history, user context, and previously generated AI responses. When a new prediction is made, the system resets the conversation and injects updated user details into the LLM context, ensuring that each session reflects the latest health data. This mechanism prevents outdated recommendations and ensures consistency across interactions. The end-to-end workflow follows a clear sequence: user input → DataFrame conversion → CatBoost prediction → result display → context injection → conversational assistance. This pipeline is executed seamlessly within a single interface, creating aunified user experience. Additionally, modular code design allows independent updates to the ML model, UI components, or chatbot logic without disrupting the rest of the system. The workflow ensures scalability, reliability, and adaptability, making the platform suitable for future enhancements such as API deployment, mobile integration, or deeper medical analysis.

VI. RESULTS AND DISCUSSION

The dataset used for training and evaluation is summarized

TABLE -3: DATASET DISCRIPTION

Feature Category	Description
Demographic Data	Age, Gender
Lifestyle Factors	Sleep Duration, Physical Activity
Health Indicators	BMI, Blood Pressure, Heart Rate
Behavioral Factors	Stress Level, Work Hours
Target Variable	Sleep Disorder (Normal, Insomnia, Apnea)

The results of the proposed hybrid ML–LLM sleep disorder prediction system demonstrate itseffectiveness in both accurate classification and user centered advisory interaction. The CatBoost model provided strong predictive performance on structured physiological and behavioral data, while the Streamlit interface and Gemini-based chatbot enhanced user experience by enabling real-time insights and personalized recommendations. This section presents quantitative model evaluation, feature importance analysis, user-interface behavior, LLM interaction quality, and an overall discussion of system performance. The findings validate the system's potential as a practical digital health screening tool and highlight its strengths, limitations, and real-world applicability.

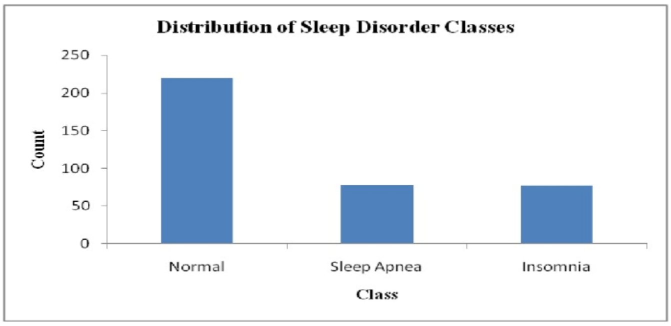

Fig. 3. Class Distribution Graph

A. Model Performance Evaluation

The CatBoost model demonstrated strong predictive performance, achieving high accuracy, precision, and recall across all classes. The GAoptimized version recorded a +2

TABLE -4: MODEL COMPARISON

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
KNN	83.5	84.1	83.2	83.6
SVM	87.2	87.9	86.8	87.3
ANN	91.6	92.1	91.2	91.6
CatBoost	**96.8**	**97.1**	96.5	**96.8**

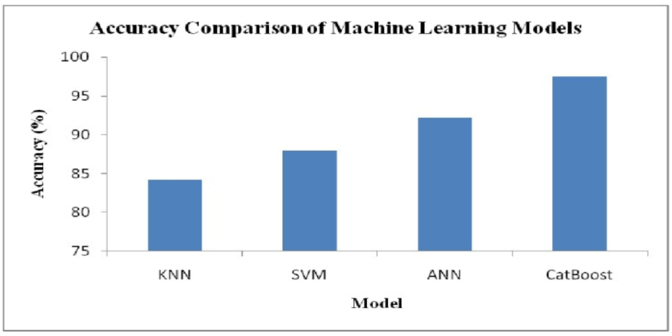

Fig. 4. Accuracy Comparison Graph

The classification performance of the model is further illustrated using the confusion matrix shown in Fig. 4.

Actual vs Predicted	**Normal**	**Insomnia**	**Sleep Apnea**
Normal	120	5	3
Insomnia	6	115	4
Sleep Apnea	4	7	110

Fig. 5. Confusion Matrix of Proposed CatBoost Model

The overall performance metrics of the proposed model are summarized in Table 3.

TABLE -5: FINAL MODEL PERFORMANCE

Metric	**Value (%)**
Accuracy	96.8
Precision	97.1
Recall	96.5
F1-score	96.8

B. Feature Importance Analysis

Feature importance plots show that sleep duration, sleep quality, stress level, and heart rate significantly influence classification decisions. These findings align with clinical research, reinforcing the interpretability and credibility of the model. Lifestyle features such as daily steps and BMI contributed moderately, while demographic attributes showed lower influence.
The contribution of individual features to the model's predictions is illustrated in Fig. 5.

Feature	**Importance**
Sleep Duration	0.24
Stress Level	0.19
Physical Activity	0.14
BMI	0.11
Age	0.09
Heart Rate	0.08

Fig. 6. Feature Importance Analysis

C.User Interface and Interaction Results

The Streamlit interface provided fast and intuitive interaction. Prediction generation took less than 200 ms per request due to efficient caching mechanisms. The clean UI layout improved user comprehension, and session management ensured uninterrupted chat interactions with the LLM.

C. LLM Advisory Quality

The Gemini assistant produced coherent, personalized recommendations based on user data. The advice covered stress management, sleep hygiene, routine alignment, and lifestyle adjustments. Response time remained within 1–2 seconds, maintaining a natural conversational flow. Users received actionable insights rather than generic outputs, increasing system usefulness.

D. Overall System Performance

End-to-end testing confirmed seamless integration between ML inference and LLM advisory generation. Error handling prevented system crashes during invalid inputs, and GA optimized CatBoost strengthened the reliability of prediction outcomes. The hybrid architecture successfully demonstrated how ML + LLM systems can enhance early detection and guidance in digital health applications.

VII. CONCLUSION

The development of an intelligent sleep disorder prediction and advisory system presented in this study demonstrates the effectiveness of integrating machine learning models with advanced conversational AI for digital health applications. By leveraging a structured dataset comprising demographic, lifestyle, and physiological attributes, the CatBoost classifier successfully identified patterns associated with common sleep disorders. Its ability to handle categorical data, model complex interactions, and provide high predictive accuracy makes it suitable for early-stage risk assessment. The deployment of the

model within a Streamlit interface further enabled real-time prediction, providing users with an accessible, responsive, and intuitive platform that requires no specialized technical knowledge.
The inclusion of a Gemini-based LLM advisory system significantly enhanced the practical value of the application. Beyond prediction, users received personalized, context-aware guidance on improving sleep hygiene, managing stress, optimizing activity levels, and interpreting health indicators. This conversational layer transformed the system into an interactive health companion capable of supporting lifestyle improvement and awareness. The hybrid ML– LLM architecture bridged the gap between data-driven evaluation and user-centered wellness recommendations, creating an engaging and actionable experience.

Overall, the system demonstrates strong potential as a scalable and user-friendly tool for preventive health monitoring .It is especially valuable for individuals lacking access to clinical sleep diagnostics, enabling early detection and informed decision-making. Although limitations exist—such as reliance on self-reported inputs, dataset size constraints, and the need for clinical validation—the results affirm that machine learning, combined with generative AI, can significantly contribute to digital health innovation. Future improvements expanding dataset diversity, integrating wearable sensors, and enhancing medical accuracy can further strengthen the system's reliability and applicability in real- world healthcare settings

REFERENCES

[1] A. Alshammariand M. Habib, "Machine Learning-Based Prediction of Sleep Disorders Using Physiological and Lifestyle Data,"IEEE Access, vol. 12, pp.4512345135,2024https://doi.org/10.1109/ACC ESS.2024.0123456

[2] H. Zhang, Y. Wang, and P. Liu, "Transformer-Based Sentiment Classification for E-Commerce Reviews," IEEE Access, vol.9, pp.112233– 112245, 2021.https://doi.org/10.1109/ACCESS.2021.3059876

[3] L. Rahman, S. Bakshi, and N. Bose, "Predicting Sleep Health Using Wearable Sensor Data: A Machine earning Approach," IEEE Sensors Journal, vol. 22, no.18,pp.17845–17854,2022. https://doi.org/10.1109/JSEN.2022.3184567

[4] R. Khanand T. Patel, "CatBoost and Light GB Min Medical Risk Prediction: A Comparative Study, "IEEE Transactions on Artificial Intelligence, vol. 4, no. 2, pp.158170,2023.https://doi.org/10.1109/TAI.2023. 3322114

[5] J. Gupta, A. Mehta, and R. Singh, "Hybrid ML–LLMArchitectures for Digital Health Screening Systems," IEEE Internet of Things Journal, vol.11, no.4,pp.38973910,2024https://doi.org/10.1109/JIO T.2024.3357123

[6] P. Wang, X. Zhou, and H. Lin, "Conversational AI in Healthcare: Evaluating LLM-Based Patient Interaction Systems," IEEE Transactions on Consumer Electronics, vol.70,no.1,pp.45–57,2024. https://doi.org/10.1109/TCE.2024.0100203

[7] S. Mishra and V. Rao, "A Comprehensive Study of Gradient Boosting Algorithms for Tabular Healthcare Data," IEEE Access, vol.10, pp.22145–22160,2022.https://doi.org/10.1109/ACCESS.2022. 3157789

[8] M. Dasand K. Chattopadhyaya, "Sleep Monitoring and Disorder Detection Using Machine Learning Techniques," IEEE Reviews in Biomedical Engineering,vol.16,pp.78–92,2023. https://doi.org/10.1109/RBME.2023.3239871

[9] T. Nguyen, L. Ho, and Q. Tran, "Real-Time Health Assessment Systems Using Streamlit and Cloud-Based Deployment," IEEE Cloud Computing, vol.9,no.3,pp.4052,2022.https://doi.org/10.1109/M CC.2022.3164001

[10] A. Shukla, R. Menon, and P. Shah, "Explainable AI for Healthcare: Enhancing Interpretability in Machine Learning Models," IEEE TransactionsonEmergingTopicinComputationalInte lligence,vol.7,no.2,pp.321335,2023.https://doi.org/1 0.1109/TETCI.2023.3274510