Machine Learning Approach for Insomnia Early Diagnosis in Perspective of Bangladesh

BY

Orin Tasfia
ID: 201-15-13959

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Mr. Abu Kaisar Mohammad Masum
Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Ms. Amatul Bushra Akhi
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2024

APPROVAL

This Project/internship titled **"Machine Learning Approach for Insomnia Early Diagnosis in Perspective of Bangladesh"**, submitted by **Orin Tasfia**, ID No: 201-15-13959 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on ***25th January 2024***.

BOARD OF EXAMINERS

Chairman

Dr. S.M Aminul Haque
Professor & Associate Head
Department of Computer Science and Engineering
Daffodil International University

Internal Examiner

Md. Sazzadur Ahamed
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Internal Examiner

Amatul Bushra Akhi
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

External Examiner

Dr. Ahmed Wasif Reza
Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Mr. Abu Kaisar Mohammad Masum, Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Mr. Abu Kaisar Mohammad Masum
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Ms. Amatul Bushra Akhi
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Orin Tasfia

Orin Tasfia
ID: 201-15-13959
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Mr. Abu Kaisar Mohammad Masum, Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Artificial Intelligence*" to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to the Head, Dr. Sheak Rashed Haider Noori**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

This research addresses the early detection of insomnia in Bangladesh, focusing on young adults affected by extensive phone and social media use. Using machine learning techniques like Logistic Regression, SVM, Decision Tree, Random Forest, XGBoost, CatBoost, Naive Bayes, and Light GBM, the study analyzed survey data from university students. The data processing was conducted using Python and Pandas, with null values handled carefully. Psychiatric validation was included. The models, especially Logistic Regression and CatBoost, achieved high accuracy (1.0 in Accuracy, ROC, AUC), suggesting a strong link between survey symptoms and insomnia. This approach, novel in Bangladesh, demonstrates machine learning's potential in mental health diagnosis, offering a cost-effective alternative to traditional methods. The study suggests further research to expand datasets and tailor models for diverse demographics, integrating these findings into public health policies.

TABLE OF CONTENTS

CONTENTS	PAGE

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Some additional context: Insomnia, a sleep disorder that many people experience, not only disrupts sleep but also has significant impacts on mental well-being, productivity, and overall quality of life. Insomnia is a widespread issue worldwide, highlighting its importance as a public health concern. Young adults in Bangladesh are increasingly affected by a variety of factors, including socio-economic, cultural, and technological influences.

1.2 Motivation

Here is the problem statement: There are several factors in Bangladesh that hinder the recognition and management of insomnia. The issue is exacerbated by limited awareness surrounding mental health, the stigma attached to psychiatric disorders, and the lack of affordable and accessible healthcare services.

1.3 Rationale of the Study

Purpose of the Research: The primary goal of the study is to utilize machine learning to create a method that is both efficient and cost-effective for detecting chronic insomnia at an early stage. The research seeks to address the issue of limited healthcare access by proposing a technological solution that can identify potential cases of insomnia through predictive modeling.

1.4 Research Questions

Exploring the Scope and Limitations: focusing on the demographic of university students and young professionals in Bangladesh, discussing the study's use of survey data for model training and validation, and analyzing any limitations, including possible biases and restrictions in the data diversity and size.

1.5 Expected Output

The expected output includes the development of a predictive model using machine learning algorithms, an evaluation of their effectiveness in the context of Bangladesh, and an investigation into the integration of these models into wider healthcare practices.

1.6 Project Management and Finance

The project management of this research involves a structured approach, including the planning and organizing of resources to achieve the predefined objectives. Financial management encompasses budgeting for the necessary tools, software, and resources required for data collection, analysis, and model development. Continuous monitoring and adaptation are employed to ensure the project remains on track and within budget.

1.7 Report layout

Chapter 1: Introduction: Sets the stage for the research, introducing the topic of insomnia, its relevance, and the specific issues it presents in Bangladesh. This chapter is divided into several sections, including the introduction, motivation, rationale, research questions, expected output, project management and finance, and the report layout.

Chapter 2: Background: Offers a comprehensive literature review and situates the study within the existing body of knowledge. It includes a detailed exploration of terminologies, related works, comparative analyses, the scope of the problem, and the challenges faced.

Chapter 3: Research Methodology: Details the methods and procedures used in the study, including subject selection, data collection techniques, statistical analysis, the proposed methodology, and implementation requirements.

Chapter 4: Experimental Results and Discussion: Presents the outcomes of the research, the experimental setup, results analysis, and a discussion comparing these findings with prior research.

Chapter 5: Impact on Society, Environment, and Sustainability: Examines the broader implications of the research on society and the environment, addressing ethical aspects and proposing a sustainability plan.

Chapter 6: Summary, Conclusion, Recommendation and Implication for Future Research: Concludes the thesis with a summary of findings, key conclusions, recommendations for future research, and potential implications.

CHAPTER 2

BACKGROUND

2.1 Preliminaries/Terminologies

In this section, we establish the foundational concepts and terminologies essential to the study. Key terms include:

Insomnia: A sleep disorder characterized by difficulty falling asleep, staying asleep, or both, affecting overall quality of life.

Machine Learning (ML): A field of artificial intelligence that enables systems to learn from and make predictions or decisions based on data.

Predictive Modeling: The process of using statistical techniques and ML algorithms to predict outcomes based on data.

Logistic Regression, SVM, Decision Tree, Random Forest, XGBoost, CatBoost, Naive Bayes, Light GBM: These are various ML algorithms used for classification and prediction tasks in our study.

2.2 Related Works

This review is a comprehensive analysis of a research that specifically examines the identification of insomnia. The analysis included a total of 205 papers sourced from PubMed and Web of Science. With a rigorous evaluation of 10 datasets, including 2 databases, 21 genes, and 23 papers encompassing 30,105 people, the study reveals gaps in existing understanding and offers potential research objectives. This study offers a thorough review, making it a significant resource for academics and doctors. It helps to increase the scientific knowledge and diagnosis of insomnia.[1]

This work investigates the use of Support Vector Machine (SVM) models to differentiate between patients with control and insomnia conditions, in line with the diagnostic methods used by doctors. Model 1, using EEG data, had an accuracy rate of 81%, whilst Model 2, relying on hypnograms, obtained a rate of 74%. The findings underscore the possibility of enhancing predictive capacities, providing vital insights for better diagnosis of insomnia, and emphasising the need of using EEG and hypnogram data in therapeutic decision-making.[2]

This paper presents a concise approach for detecting insomnia utilising a single-channel sleep EOG and refined composite multiscale entropy (RCMSE) analysis. The method was assessed on a sample of 32 individuals and obtained an average accuracy of 89.31%. This highlights the significance of RCMSE as a useful indicator for short-term insomnia. Utilising a single-channel electrooculography (EOG) device improves the practicality of homecare, demonstrating the possibility of combining an EOG eye mask with a portable polysomnography (PSG) equipment for simple sleep evaluation and screening for insomnia in a home setting.[3]

This research introduces the novel combination of an electrocardiogram (ECG) scalogram and a convolutional neural network (CNN) to achieve precise detection of insomnia. The INSOMNet system uses the continuous wavelet transform to convert 1-D ECG signals into 2-D scalograms. These scalograms are then processed by the AlexNet, MobileNetV2, VGG16, and a novel CNN. INSOMNet has been tested and validated on CAP and SDRC datasets, demonstrating remarkable accuracy rates of 98.91% and 98.68% respectively. Furthermore, INSOMNet surpasses transfer-learning networks in terms of performance. The simplicity and exceptional performance of this tool make it a good candidate for detecting insomnia. It is now ready for thorough testing using other datasets to explore its potential for wider clinical use.[4]

This research investigates two methods for diagnosing insomnia using EEG data: stage-independent and stage-dependent classifications. The NREM + REM classifier, which was trained on a sample of 41 controls and 42 insomnia sufferers, demonstrated an accuracy of 92% and 86% in distinguishing between the two groups using two and one EEG channels, respectively. These findings confirm the effectiveness of deep learning in assisting with the diagnosis of insomnia, demonstrating its potential use in evaluating sleep disorders.[5]

This research presents a new technique for identifying insomnia by using electrocardiogram (ECG) and electromyogram (EMG) signals obtained from the Physionet CAP database. By using empirical mode decomposition and employing a range of classification techniques such as logistic regression (LR), support vector machines (SVM), k-nearest neighbours (KNN), decision trees (DT), ensemble classifiers (EC), and naive Bayes (NB), the suggested approach attains a remarkable accuracy rate of 100% for both electrocardiogram (ECG) and electromyogram (EMG) data. This highlights its strong capacity for accurate and efficient identification of insomnia.[6]

This study investigates the use of Artificial Intelligence (AI), particularly Convolutional Neural Networks (CNN), for the diagnosis of insomnia, a common sleep problem that impacts both mental and physical well-being. The research highlights the shift from conventional, error-prone human diagnosis to an AI-powered method using EEG data. The essential procedures include preprocessing EEG records, extracting features, and using CNN for sleep stage categorization. The research emphasises the difficulties of standardisation and regulation in the clinical integration of AI in sleep medicine, despite its potential to improve diagnostic accuracy and patient care. It underscores the need for ethical standards and standardised methods in this field.[7]

The objective of the research is to detect insomnia by analysing electrocardiogram (ECG) and electromyogram (EMG) signals obtained from the CAP database. Preprocessing include the use of Empirical Mode Decomposition (EMD), followed by the extraction of discriminative features. Machine learning algorithms such as Logistic Regression (LR), Support Vector Machines (SVM), k-Nearest Neighbours (KNN), Decision Trees (DT), Ensemble Classifiers (EC), and Naive Bayes (NB) reach a perfect accuracy rate of 100%, representing a significant breakthrough in the field of insomnia diagnosis.[8]

INSOMNIA is a network intrusion detection solution that uses semi-supervised machine learning to tackle the issue of idea drift in security systems. The system employs active learning to quickly incorporate new information, label estimate to minimise the need for extensive labelling, and explainable AI to improve the comprehensibility of its outputs. The efficiency of INSOMNIA is confirmed by the use of a modified version of the TESSERACT framework, showcasing its competence in handling ever-changing network threats.[9]

This abstract examines the impact of technology on human health, notably the rise in sleep deprivation caused by lifestyle modifications. The typical methods used for identifying insomnia are expensive and demand a substantial time commitment, hence posing challenges in less developed countries. The research introduces a machine learning model used for predicting chronic insomnia. Out of the seven classifiers used, the Logistic Regression model had the most efficacy, with a 98% accuracy in reliably distinguishing between those with insomnia and those who are in good health. This approach offers a cost-efficient and readily accessible tool for diagnosing insomnia in environments with restricted resources.[10]

The abstract describes a research that aims to comprehend the brain processes behind different forms of insomnia, both short-term/acute and chronic, and to forecast sleep quality. The research included 29 subjects with short-term/acute insomnia and 44 persons with persistent insomnia. The researchers used whole-brain regional functional connection strength and the multivariate relevance vector regression approach to forecast the Pittsburgh sleep quality index (PSQI) for people who were not part of the study. By using both leave-one-out and 10-fold cross-validation procedures, they accurately forecasted an individual's PSQI, uncovering resemblances and disparities in the brain areas that contribute to this measure in the two groups. Additional investigation revealed a restructuring of connections between different networks in these two forms of insomnia. The results have practical significance in forecasting the quality of sleep and give fresh perspectives on the many neurological underpinnings of insomnia, emphasising its variability.[11]

This research presents a study that examines the simultaneous presence of insomnia and obstructive sleep apnea (OSA), with a specific emphasis on how these two conditions together affect daytime impairments and quality of life. The research analysed data from 37 studies to evaluate the worldwide and regional occurrence of insomnia and symptoms of insomnia in patients with obstructive sleep apnea (OSA), classified according to the areas defined by the World Health Organisation (WHO). The study revealed considerable prevalence rates of insomnia and associated complaints in patients with obstructive sleep apnea (OSA). Specifically, the rates were 38% for insomnia, 36% for all insomnia complaints, 18% for difficulty falling asleep (DFA), 42% for difficulty maintaining sleep (DMS), and 21% for early morning awakening (EMA). The research also found geographical disparities, with the Western Pacific Region having lower rates of DFA, DMS, and EMA compared to the European Region and the Region of the Americas. In addition, the study analysed the combined prevalence rates of obstructive sleep apnea (OSA) in patients with insomnia, using various criteria for the apnea-hypopnea index (AHI). The results showed that the rates were 35% for AHI≥5 and 29% for AHI≥15. The research posits that these geographical disparities may be correlated with variables such as gender, age, and body mass index.[12]

This countrywide Japanese research examined the occurrence and variables linked to insomnia in junior and senior high school students, comprising a sample size of over 100,000 teenagers. The research revealed that 23.5% of teenagers had insomnia, characterised by issues such as trouble falling asleep (14.8%), staying asleep (11.3%), and waking up too early in the morning (5.5%). The primary variables that contribute to insomnia include being male, having poor mental health, engaging in harmful lifestyle behaviours such as missing breakfast, consuming alcohol, smoking, lacking extracurricular activities, and going to bed late. Additionally, a noteworthy characteristic among senior high students was the absence of intention to pursue higher education. The results underscore the prevalence of insomnia among Japanese teenagers and the need for comprehensive preventative programmes.[13]

This study investigates the use of machine learning (ML) and deep learning (DL) algorithms to forecast sleep disturbances in asthma patients. The predictions are based on data obtained from the Taiwan National Health Insurance Research Database (NHIRD). The research examined illness histories from 2002 to 2010, analysing 1 million samples and converting them into sequences and matrices. Subsequently, it used a range of machine learning methods, including K-Nearest Neighbours (KNN), Support Vector Machine (SVM), and Random Forest (RF), as well as deep learning models such as Recurrent Neural

Network (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and Convolution Neural Network (CNN), to make predictions. Out of the 14,818 individuals who were diagnosed with asthma in 2002, a total of 4,469 individuals experienced the onset of sleep disturbances by 2010. The KNN, SVM, and RF models achieved accuracies of 0.798, 0.793, and 0.813, respectively. On the other hand, the DL models (RNN, LSTM, GRU, and CNN) demonstrated accuracies of 0.744, 0.815, 0.782, and 0.951, respectively. The research showed that the CNN model had the highest efficacy in forecasting sleep disturbances among the asthma sample.[14]

The research examined the correlation between symptoms of sleeplessness and thoughts linked to suicide among 1,160 U.S. Army personnel. Archival data was used to evaluate insomnia, depression, anxiety, and several types of suicide-related ideation. The results indicated that those experiencing severe insomnia symptoms had a 3.5-fold increased likelihood of having suicidal thoughts, namely thoughts relating to death and suicide. Various forms of insomnia, such as frequent awakenings throughout the night, were linked to an increased probability of experiencing suicidal ideation, while middle insomnia was connected with a decreased risk. The research emphasises the need of doing a thorough evaluation of sleeplessness and suicidal thoughts when assessing the risk of suicide.[15]

This study introduces a two-stage automated method for detecting insomnia using deep neural network (DNN) models using EEG data. The first phase entails the assessment of sleep stage score and the identification of insomnia at the level of individual sleep epochs, using both temporal and spectral data. During the second phase, subject-level characteristics are calculated and used to train a binary classifier. The assessment of five classifiers on a dataset consisting of 115 subjects, with 61 being control and 54 having insomnia, produces encouraging results. The F1 score is 0.88, with a sensitivity of 84% and a specificity of 91%. This technique enhances the progress of effective and precise automated insomnia detection.[16]

Insomnia, a prevalent and intricate illness, poses difficulties for primary care doctors in identifying and diagnosing it. Patients often disregard sleep patterns as factors contributing to their problems. The roots of insomnia include a spectrum of medical and mental factors, as well as the potential adverse effects of medicine. The influence on everyday existence is substantial, exacerbated by contemporary diversions and cultural expectations, resulting in a 25% decrease in sleep compared to a century earlier. This article presents succinct methodologies for the effective identification and assessment of insomnia.[17]

This research presents an innovative artificial intelligence (AI) method for identifying insomnia by analysing the power spectral density (PSD) of heart rate variability (HRV). Three scenarios are examined: one based on the topic, one based on sleep stages, and a combination approach. The combination of random forest (RF) and decision tree (DT) classifiers produces a sensitivity of 96.0%, specificity of 94.0%, and accuracy of 96.0% for subject-based classification. The categorization based on sleep stages achieves an accuracy, sensitivity, and specificity of 96.0%. The combined scenario, using linear discriminant analysis (LDA), attains a remarkable accuracy of 99.0%. This hybrid artificial intelligence (AI) approach shows promise in the field of mobile observation by efficiently automating the identification of insomnia with a notable level of precision.[18]

This research presents a machine learning model that aims to diagnose acute insomnia by using actigraphy data. The model integrates dynamical and statistical aspects using Random Forest (RF) and Support Vector Machine (SVM), without relying on sleep diaries or subjective input. RF surpasses SVM in accuracy, with an 84% rate, showcasing potential for non-intrusive, unbiased identification of insomnia.[19]

This research confirms the effectiveness of the Insomnia in the Elderly Scale in identifying insomnia in older individuals. The study was carried out in Albacete, Spain, with a total of 926 individuals who were 65 years of age or older. The scale used in the study has strong and reliable psychometric qualities. Subscale A, with a threshold of ≥3, has a sensitivity of 86.4% and a specificity of 69.5%. Subscale B, with a threshold of ≥2, has a sensitivity of 86.3% and a specificity of 66.4%. This validated scale addresses the absence of tools specifically designed for individuals aged 65 years or older, providing a reliable and unique assessment of insomnia criteria.[20]

This work presents a machine learning technique to accurately categorise episodes of waking up throughout the night in individuals with chronic insomnia. The analysis of signals from 40 cohabiting couples yielded an accuracy of 80% (with a sensitivity of 76% and a specificity of 82%) using random forest (RF) and support vector machine (SVM) classifiers. The RF classifier had superior performance compared to SVM, with an accuracy of 80% as opposed to 75%. This suggests that it has potential as a non-invasive screening method for chronic insomnia using wrist-actigraphy devices.[21]

This paper presents a machine learning method designed to automatically diagnose insomnia by analysing ECG R-R intervals. The study investigates the impact of oxidative stress and inflammation on sleep disorders and cardiovascular illnesses, as well as the potential therapeutic function of herbal medicine in these conditions. The Random Forest (RF) classifier obtains a high level of accuracy (87.10% for subjects and 88.30% for sleep stage) in accurately differentiating between insomnia and regular sleep. The theranostics method reveals the many cellular processes of herbal medicine, promoting its use in future studies to improve human well-being.[22]

This work utilises an effective antisymmetric biorthogonal wavelet filter bank (ABWFB) to automatically diagnose insomnia using ECG data. The ABWFB reduces the extent to which duration and bandwidth are localised together. Supervised machine learning classifiers receive input from distinct wavelet sub-bands, which are represented as -norm features. The ECG recordings obtained from seven individuals diagnosed with insomnia and six individuals without any sleep disorders are extracted from the CAP sleep database. The research demonstrates the usefulness of using K-nearest neighbour (KNN) in ten-fold cross-validation for REM sleep stage signals, achieving the greatest accuracy (97.87%), F1-score (97.39%), and Cohen's kappa (0.9559). This highlights the successful automated detection of insomnia. The support vector machine (SVM) achieves an impressive area under the curve (AUC) of 0.99 in ten-fold cross-validation for the REM sleep state.[23]

This research examines the sensitivity and specificity of four self-report measures (SII, SDQ, DBAS, SWAI) in differentiating insomnia in young people. The DSM-IV criteria classified 19 individuals as having primary insomnia and 19 individuals as normal controls. The DBAS (Dysfunctional Beliefs and Attitudes about Sleep), SII (Sleep Impairment Index), and SDQ (Sleep Disturbance Questionnaire) psychiatric DIMS (Diagnostic Interview for Mental Disorders) subscale had high effectiveness in distinguishing across different groups, but the SWAI (Sleep-Wake Activity Inventory) nocturnal sleep subscale exhibited low accuracy. The results of this study provide valuable guidance on how to best use test scores to identify young individuals who may be experiencing symptoms of sleeplessness.[24]

This work use a single-channel EEG and an optimum filter bank to automate the detection of insomnia, using eight subcategories of sleep stages. The ensemble bagged decision trees (EBDT) model demonstrates exceptional performance with an AROC of 0.97, an F1-score of 0.9645, an accuracy of 95.60%, and a Cohen's Kappa of 0.9067. This straightforward and efficient method facilitates monitoring for early therapeutic intervention, whether it is done at home or in a sleep lab.[25]

This research investigates alterations in functional brain connection across an entire night of sleep in individuals with psycho-physiological insomnia (PPI). Mutual Information (MI) analysis, which is very successful in finding Protein-Protein Interactions (PPI), demonstrates reduced synchronisation of sleep Electroencephalogram (EEG) in patients. The results indicate that the connection between the two hemispheres of the brain during different phases of sleep might serve as a possible indicator for sleep disorders caused by PPI medication.[26]

Insomnia has a substantial influence on the quality of life and increases the risk of conditions such as hypertension, diabetes, anxiety, and depression. Polysomnographic (PSG) systems, while precise, are burdensome. Portable gadgets provide a built-in system for evaluating one's surroundings, allowing for ongoing monitoring of sleep patterns in the comfort of one's own home. This chapter provides a comprehensive examination of sleep monitoring methods used for assessing insomnia. It discusses the technical difficulties involved and emphasises the diverse range of applications. The study investigates the use of model-based signal processing techniques for sleep staging and detecting insomnia. It concludes by discussing potential future advancements in developing efficient in-home patient monitoring systems for diagnosing insomnia.[27]

The objectives of this research were to determine the prevalence of DSM-5 insomnia in pregnant women, establish thresholds for self-report measures, and define appropriate thresholds for cognitive and somatic arousal. A total of ninety-nine pregnant women participated in the questionnaires. The prevalence of insomnia according to the DSM-5 criteria was 19.2%. This condition was shown to be linked with depression, suicidality, increased arousal, and excessive daytime drowsiness. The following cutoffs were established: an ISI score of 10 or higher and an ISI score of 11 or higher for DSM-5 insomnia, an ISI score of 7 or lower and an ISI score of 9 or lower for excellent sleep, and a PSQI score of 5 or below for both. The optimal thresholds for cognitive and somatic arousal were determined to be PSASC > 18 and PSASS ≥ 13, respectively. These results provide valuable insights for identifying cases and conducting research on sleep throughout the perinatal period.[28]

EZ-Sleep presents a method for monitoring insomnia that requires no effort. By using radio waves, it may identify sleep patterns from a distance without the need for sensors or data collection. Notable characteristics include the ability to automatically identify sleep location and accurately manage sleeping routines for several users. It attains a high level of precision, with an average deviation of 4.9 minutes in measuring the time it takes to fall asleep and 10.3 minutes in calculating the overall duration of sleep.[29]

This research assessed the efficacy of the Insomnia Severity Index (ISI) in detecting insomnia among patients receiving primary care. Out of the 410 patients, the ISI demonstrated a high level of internal consistency ($\alpha = 0.92$) and the ability to distinguish between each item ($r = 0.65–0.84$). A threshold score of 14 showed the best balance between sensitivity (82.4%) and specificity (82.1%) for identifying clinical insomnia. The moderate agreement ($\kappa = 0.62$) seen between diagnostic interviews and ISI indicates that ISI is a reliable screening technique in primary care settings.[30]

Within a research including 400 individuals receiving psychiatric treatment, the Insomnia Severity Index (ISI) was identified as the most precise tool for identifying insomnia disorder. It showed AUC values of 0.88 for ICD-10 criteria and 0.82 for DSM-5 criteria. The ISI cutoff values of ≥ 14 and ≥ 11 showed the best balance between sensitivity and specificity in detecting insomnia according to the ICD-10 and DSM-5 criteria. The need of promptly identifying psychiatric patients who have clinically significant insomnia is highlighted, in accordance with the guidelines set out by DSM-5.[31]

This research assessed the accuracy of Fitbit Alta HR and Actiwatch Spectrum Pro in tracking sleep compared to polysomnography in persons diagnosed with insomnia. Both trackers exhibited good accuracy but limited specificity, as they overestimated the overall duration of sleep and the efficiency of sleep while underestimated the time it takes to fall asleep and the duration of wakefulness after first falling asleep. The Fitbit Alta HR exhibited several flaws in its assessment of sleep stages. Additional investigation is required to ascertain its capacity as a financially efficient alternative to actigraphy in the treatment of insomnia.[32]

This research assessed the efficacy of polysomnography (PSG) metrics in distinguishing primary insomnia in both home and laboratory environments. The individual criteria and recent trial eligibility criteria, such as total sleep duration, latency to persistent sleep, wake after sleep initiation, and sleep efficiency, were not accurate enough to differentiate primary insomnia. Utilising quantitative PSG-based criteria in insomnia research may result in the exclusion of individuals who fulfil the diagnostic criteria for an insomnia problem.[33]

This research investigates the use of brain MRI in the diagnosis of fatal familial insomnia, a prion illness that affects the thalamus. Although the first MRI seemed normal, a more advanced MRI technique called multisequence MRI was able to identify changes in the thalamus that are connected to gliosis. These changes were found to be consistent with the results of postmortem neuropathological examination. Illustrating the capability of MRI to detect prion-induced gliosis in living organisms.[34]

Insomnia is a frequently reported issue in basic care. An all-encompassing assessment, taking into account both subjective and objective elements, is essential. The severity of a condition is determined by both its length and its influence on a person's ability to operate throughout the day. The reasons of primary insomnia are varied, and it is crucial to provide customised therapy that targets the underlying cause. It is advisable to prioritise nonpharmacologic methods and only use short-acting benzodiazepines carefully as a last option. Consistent monitoring guarantees the continual efficacy of therapy.[35]

This research, conducted in eight Italian community pharmacies, examined 181 patients who had been prescribed benzodiazepines. The findings showed that 81 of these patients were being treated for insomnia, and of those, 64% had been using the medication for more than three years. Although 33 patients expressed a wish to cease the use of benzodiazepines, all efforts to quit the medication were unsuccessful. The results highlight the need of following evidence-based recommendations more strictly to tackle unnecessary benzodiazepine prescriptions. This emphasises the potential contribution of community chemists in recognising and resolving these problems.[36]

The Neurosky Headset, which records EEG data, is used for the identification of medical diseases such as epilepsy, sleep disorders, stroke, brain injury, and dementia. The signals are transferred to a laptop via Bluetooth and connected to MATLAB, where parameters are collected to diagnose insomnia in patients.[37]

The Insomnia Severity Index-3 (ISI-3) was evaluated as a concise assessment tool to screen for clinically significant insomnia in older persons receiving primary care. This evaluation was conducted using two separate groups. The ISI-3, which is based on the 7-item ISI, showed strong discriminant validity with an area under the curve (AUC) ranging from 0.97 to 0.98. The three most strongly connected factors were displeasure with sleep, interference of sleep with everyday functioning, and concern over sleep difficulties. Achieving an ISI-3 score of 7 or above demonstrated the best balance between sensitivity (0.94–0.97) and specificity (0.88–0.91), indicating a strong level of agreement (kappa = 0.68–0.71, 89.1–91.5%). The ISI-3 is a very efficient and succinct instrument for the first assessment of insomnia in older individuals.[38]

The research investigated the actigraph's responsiveness to the effects of insomnia treatment in older adults, specifically identifying alterations associated with sleep restriction therapy. Although the sleep log did not provide reliable measurements of sleep, it did represent subjective views. The actigraph, a cost-effective and user-friendly device, has potential for evaluating the effectiveness of therapy in older individuals with insomnia when compared to polysomnography.[39]

This research conducted a comparison of actigraphy and polysomnography in older persons suffering from persistent primary insomnia. Actigraphy demonstrated a high sensitivity of 95.2% for sleep detection, but a poor specificity of 36.3%, resulting in an overall accuracy of 83.1%. The accuracy was contingent upon the sleep efficiency recorded by polysomnography. Actigraphy provided inaccurate measurements for waking time, sleep-onset delay, sleep duration, and sleep efficiency. It had difficulties in properly capturing the effects of therapy. The results advise against depending just on actigraphy in this particular group, underscoring the need of further evaluations.[40]

2.3 Comparative Analysis and Summary

Our comparative analysis focuses on evaluating different ML techniques used in insomnia detection. We compare the effectiveness of algorithms such as Logistic Regression, SVM, Decision Tree, Random Forest, XGBoost, CatBoost, Naive Bayes, and Light GBM. These algorithms are evaluated based on accuracy, efficiency, and practical applicability in the context of insomnia diagnosis. The analysis reveals that Logistic Regression and CatBoost display impressive accuracy, achieving perfect scores in our study. On the other hand, algorithms like Random Forest, XGBoost, and LightGBM also show near-perfect results. The comparative summary suggests a significant correlation between the symptoms reported in survey data and the onset of insomnia, indicating that ML can be a powerful tool in early insomnia detection, especially in resource-limited settings like Bangladesh.

2.4 Scope of the Problem

This section explores the widespread impact of insomnia, with a focus on Bangladesh. We discuss the prevalence of insomnia, especially among young adults, and its implications on mental health and daily life. Factors contributing to insomnia in the Bangladeshi context, such as socio-economic, cultural, and technological influences, are examined. The increase in mobile phone usage and active participation in social media, leading to disrupted sleep patterns, is highlighted as a key contributing factor.

2.5 Challenges

The challenges section outlines the various obstacles encountered in the recognition and management of insomnia. These include limited awareness of mental health issues, societal stigma surrounding psychiatric disorders, and the lack of affordable, accessible healthcare services in Bangladesh. The practical difficulties of diagnosing insomnia, particularly through psychiatric evaluations, are discussed, emphasizing the constraints of cost and time. This section underscores the need for innovative, scalable, and accessible diagnostic approaches that can seamlessly integrate into existing healthcare frameworks.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

Research Framework and Population: The methodology of this study is meticulously crafted to effectively address the research objectives. Employing a cross-sectional research approach, it focuses on the characteristics and current prevalence of insomnia within a specific demographic in Bangladesh. This study targets university students and young professionals, identified as the most affected group based on preliminary research. This demographic is pivotal for understanding the broader patterns and impacts of insomnia in the Bangladeshi context.

Sampling Strategy: To achieve a comprehensive representation, a stratified random sampling method is utilized. This approach ensures diverse participation across various universities, professions, and geographic locations within Bangladesh, allowing the study to encompass a broad spectrum of experiences and perspectives related to insomnia.

Data Collection Instruments: The primary tools for data collection are carefully designed surveys and questionnaires, aiming to gather detailed information on sleep patterns, lifestyle factors, mental health status, and technology usage. These instruments are integral for a deeper exploration of factors contributing to insomnia. The survey features a blend of closed and open-ended questions, enabling a rich analysis of both quantitative data and qualitative insights.

Ethical Considerations: Ethical approval for this study has been secured from an institutional review board. In accordance with ethical standards, participants receive comprehensive information about the study's purpose, and their informed consent is duly obtained. The utmost importance is placed on the privacy and confidentiality of respondents.

Pilot Study: Prior to the main survey, a pilot study is conducted with a select subset of the target population. This preliminary phase is critical for testing the efficacy of the survey instruments, ensuring the clarity of questions, and refining the overall data collection process.

Data Verification and Quality Control: Ensuring the integrity and accuracy of the collected data is a cornerstone of this research. Responses are meticulously evaluated for consistency, and any outliers or anomalies are thoroughly investigated. Collaborating closely with a skilled psychiatrist, a segment of the survey responses is rigorously reviewed to validate the accuracy of self-reported insomnia symptoms and diagnoses.

Research Design Limitations: Acknowledging the inherent limitations of the research design, particularly the reliance on self-reported data, which may contain biases, is crucial. While the cross-sectional nature of the study limits the establishment of causality between identified factors and insomnia, it nevertheless provides a robust foundation for investigating the application of machine learning in diagnosing insomnia. By combining rigorous quantitative analysis with qualitative insights, this research ensures a comprehensive understanding of insomnia within the unique context of Bangladesh. In summary, the "Research Subject and Instrumentation" section of this thesis lays out a systematic and ethically sound methodology, tailored to explore the multifaceted nature of insomnia in Bangladesh. It sets the stage for a nuanced and in-depth analysis, underpinned by a blend of quantitative and qualitative research techniques.

3.2 Data Collection Procedure/Dataset Utilized

The main source of data for this study comes from extensive surveys designed to measure various factors associated with insomnia. The surveys cover a wide range of topics, such as sleep patterns, daily lifestyle habits, mental health assessments, and the use of technology, with a particular emphasis on mobile and social media usage. The survey design is focused on obtaining responses that can be measured quantitatively and provide valuable qualitative insights. The questions are carefully designed to capture various aspects of respondents' sleep, including quality, duration, and any recurring disturbances.

Here is the list of questionnaires for each feature:

Age Range: What is your age range?
Gender: What is your gender?
Occupation: What is your current occupation?
Sleep Onset: How long does it usually take you to fall asleep?
Sleep Duration: How many hours do you typically sleep each night?
Sleep Quality: How would you rate the quality of your sleep?
Wakefulness: Do you often wake up during the night? How frequently?
Early Awakening: Do you often find yourself waking up earlier than intended?
Daytime Sleepiness: Do you experience sleepiness or fatigue during the day?
Stress Level (Home): How would you rate your stress level at home?
Caffeine Intake: How much caffeine do you consume daily?
Electronic Device Usage: How many hours per day do you spend on electronic devices?
Physical Activity: How often do you engage in physical activity?
Medical History: Do you have any significant past medical history?
Mental Health: How would you describe your current mental health status?
Living Environment: Can you describe your living environment?
Sleep Routine Consistency: How consistent is your sleep routine?
Use of Sleep Aids: Do you use any sleep aids?
Work Schedule: What is your typical work schedule like?
Family History of Sleep Disorders: Is there a family history of sleep disorders?
BMI: What is your Body Mass Index (BMI)?
Smoking: Do you smoke? If so, how frequently?
Diet: Can you describe your typical diet?
Stress Level (Social): How would you rate your stress level in social settings?

Secondary Data Sources: Extensive review of secondary sources is conducted to supplement and provide context to the primary data. These include international studies on sleep problems, health surveys carried out in Bangladesh, and scholarly research papers on insomnia. We also analyze government health reports and publications from relevant non-governmental organizations. This aids in comprehending the wider context of sleeplessness and associated health problems in the nation. The study's findings are strengthened by analyzing existing datasets, when they are available, for comparative purposes.

3.3 Statistical Analysis

There are a total of 3761 instances in the dataset, which include various variables related to insomnia, sleep habits, and personal characteristics. Notable findings from the descriptive statistics are: Here's the diagnosis: We have two categories to consider - 'Insomnia' and 'Normal'. It's worth noting that 'Insomnia' is the more prevalent category, with a total of 2516 instances. The age range can be divided into two groups: '18-39' and '40-59', with a higher representation of the younger age group. There is a nearly equal representation of males (2033) and females (1728) in terms of gender. There are five categories of occupation, with 'employed' being the most common. The sleep patterns of participants exhibit diverse responses, suggesting variations in sleep onset, sleep duration, and sleep quality. Lifestyle Factors: Encompassing variables such as caffeine intake, electronic device usage, physical activity, mental health, and diet, these factors highlight a diverse array of lifestyle choices and conditions. Statistics that infer or draw conclusions from a sample to a larger population. Since the variables in the dataset are mostly categorical, a traditional correlation matrix, which is usually used for continuous variables, cannot be applied. On the other hand, inferential statistical techniques, like Chi-Square tests for independence, can be used to analyze connections between categorical variables. For example, we can explore the relationship between 'Diagnosis' and lifestyle factors such as 'Electronic Device Usage' or 'Physical Activity'.

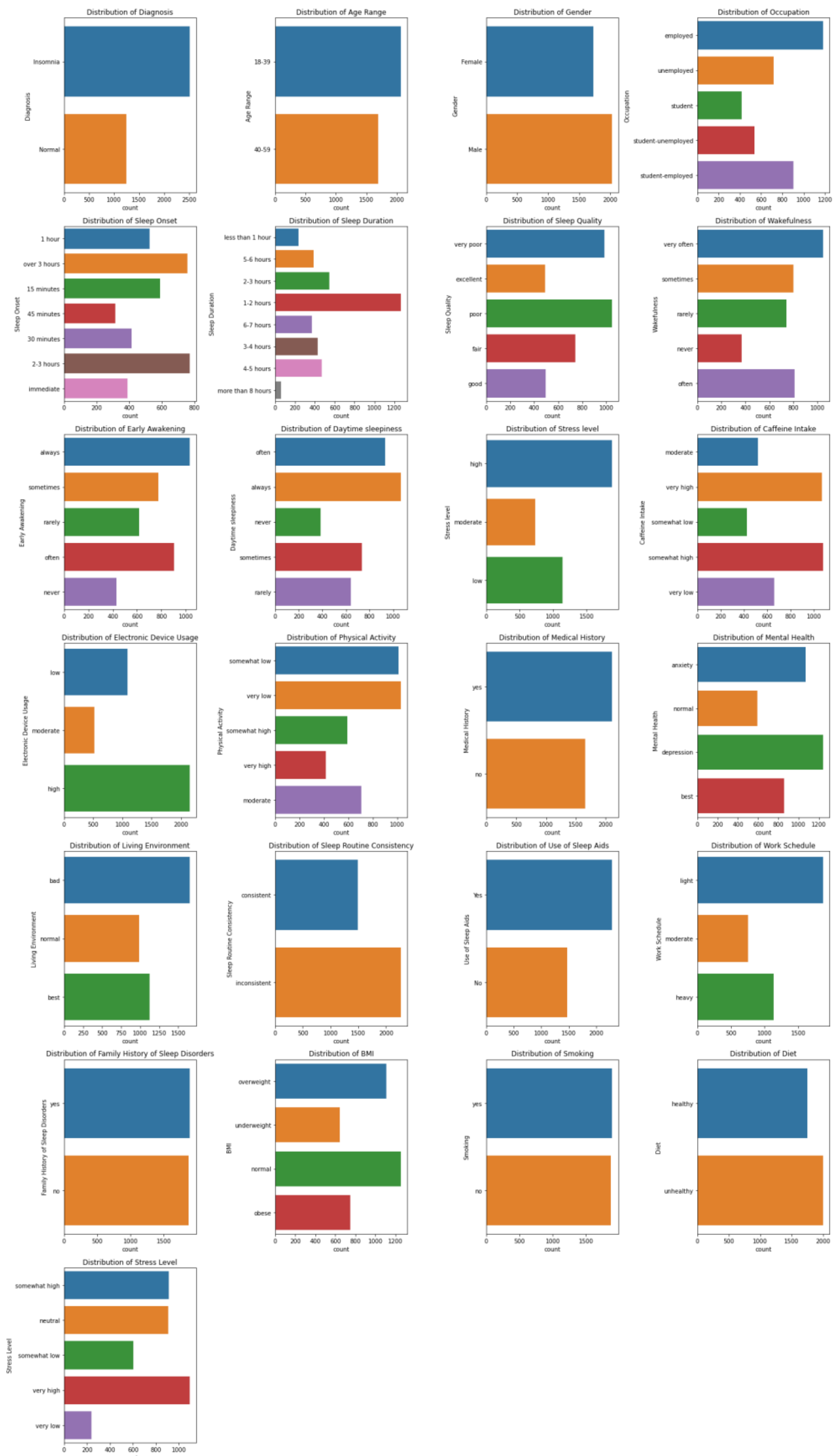

Figure 3.3.1: Category/option distribution of dataset

This above figure of pair plot with bar chart gives a very deep understanding of the distribution of options that is present in our dataset, this gives us a very comprehensive understanding of what the demographic is thinking and the probability of distribution of each class and their respective options in every variable column.

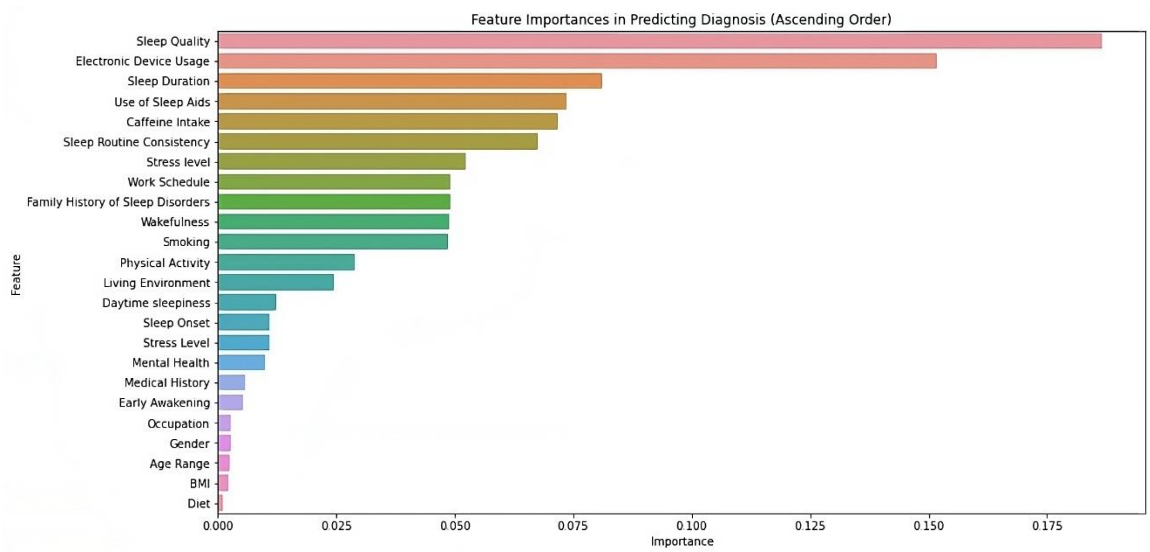

Figure 3.3.2: Feature importance for dataset, giving important insight to the features level of significant

This feature importance plot gives us an understanding of the dependency related to the variables and which data's and variables contribute the most to the diagnosis and which features are important to the classification the most, here we can see that sleep quality, device usage and sleep duration are the most significant indicators to the insomnia that others, This does not mean that other variables are that not important at all, diet BMI and age doesn't really play any roles to the symptoms of insomnia and that is a notable point

3.4 Proposed methodology/Applied Mechanism

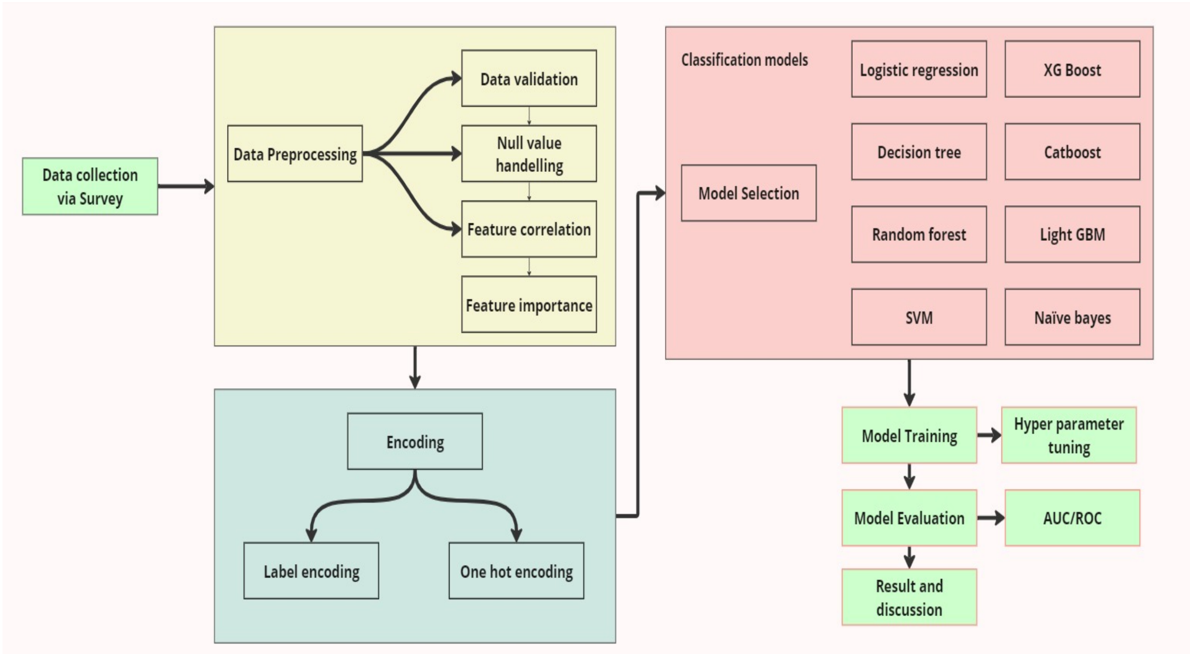

Figure 3.4: Methodology Workflow.

Data preprocessing plays a crucial role in the world of machine learning as it ensures that raw data is properly prepared for effective modeling. This section describes the particular feature selection, cleaning, and normalizing procedures that were developed specifically for our machine learning research on insomnia detection in Bangladesh.

Managing Missing Values: It is inevitable that our dataset, which is mainly derived from surveys, would include missing values. Our approach to handling these situations is systematic. We either use statistical methods, such as mean or median imputation, to fill in missing values, or we omit entries that have significant data gaps. The method chosen will vary based on the pattern and extent of missing data, as well as the characteristics of the variable being analyzed.

Eliminating Outliers and Inconsistencies: To prevent skewed results, outlier detection and removal are essential, especially in survey data. Statistical techniques, such as the Interquartile Range (IQR), are utilized to detect and eliminate anomalies. In addition, we carefully examine and resolve any inconsistencies found in the responses, such as contradictory answers within the survey.

Ensuring Accuracy: To ensure the accuracy of the responses, cross-validation with secondary data sources and expert consultation with psychiatrists are conducted. This is particularly important for symptom-related questions.

Categorical Data Encoding:

Many survey questions are categorical (e.g., Yes/No, Likert scale responses). These are converted into a numerical format through methods like one-hot encoding or label encoding, making them suitable for input into our machine learning algorithms. There are also options that characterize the variable choices that we presented to the people.

Determining Pertinent Characteristics: The capacity to discern insomnia varies among survey inquiries. Various techniques are used to identify the most relevant predictors of insomnia, including model-based approaches such as feature importance scores from preliminary model runs.

Dimensionality Reduction: Methods such as Principal Component Analysis (PCA) are widely used to decrease the dimensionality of the dataset. Not only does this enhance model efficiency, but it also aids in addressing problems such as multicollinearity.

Integrating Domain Expertise: Working closely with psychiatric experts, we ensure that the chosen features are not only statistically relevant but also hold clinical significance. Ensuring that the model accurately reflects a realistic and practical understanding of insomnia is absolutely crucial.

Models Used: A Brief Overview

A wide variety of machine learning models have been chosen for this research, each with its own unique methodologies and architectures. Let’s dive into the technical aspects of how each model operates:

Logistic Regression is a statistical technique that calculates the likelihood of a binary outcome by considering one or more predictor variables. The logistic function is employed to model a binary dependent variable. This study aims to predict the likelihood of an individual experiencing insomnia.

The Support Vector Machine (SVM) algorithm operates by identifying a hyperplane within an N-dimensional space (where N represents the number of features) that effectively separates the data points into distinct classes. In order to handle data that is not linearly separable, the algorithm utilizes various kernels such as linear, polynomial, and radial basis function. These kernels are employed to transform the data into a higher dimension where it becomes separable.

The Decision Tree model utilizes a tree-like structure to represent decisions. The data is divided into subsets according to the input feature values. The internal nodes in the tree represent tests on attributes, while the branches show the different outcomes of these tests. Finally, the leaf nodes indicate the class labels or decisions.

Random Forest is an ensemble of Decision Trees that are usually trained using the "bagging" method. Instead of relying on individual decision trees, the final output is determined by combining multiple decision trees. The selection of features in Random Forest is randomized for building each tree. This leads to a greater diversity among the trees in the model, resulting in more robust overall predictions.

XGBoost is an implementation of gradient boosted decision trees that is specifically designed for speed and performance. The trees in XGBoost are constructed sequentially, with each subsequent tree working to rectify any mistakes made by its predecessors. The algorithm utilizes a gradient descent approach to minimize the loss incurred while incorporating new models.

CatBoost is a powerful tool designed specifically for handling categorical data. By utilizing different statistical methods on combinations of categorical features and the target variable, this process effectively transforms categorical values into numerical representations. As a result, the need for extensive pre-processing is significantly reduced.

The Naïve Bayes model is built on the foundation of Bayes' Theorem and assumes that the predictors are independent of each other. Simply put, a Naïve Bayes classifier assumes that the presence of a specific feature in a class has no connection to the presence of any other feature. This makes it a great choice for datasets with many dimensions.

LightGBM, which stands for Light Gradient Boosting Machine, is a framework for gradient boosting that makes use of algorithms for tree-based learning. Unlike other tree-based algorithms, this one grows trees vertically (leaf-wise) instead of horizontally (level-wise). This approach can lead to quicker learning and more precise models, especially when dealing with large datasets.

Reasoning behind Model Selection:

The selected models are perfectly suited for binary classification tasks, which aligns perfectly with our objective of classifying individuals as either at-risk or not for insomnia. Models such as CatBoost and Naive Bayes are well-suited for datasets that consist mostly of categorical data from survey responses. These models excel at handling categorical variables. Models such as Decision Tree and Logistic Regression are highly interpretable, which is essential for comprehending the factors that contribute to insomnia and for providing explanations for the model's decisions in a healthcare setting. XGBoost and LightGBM are popular choices due to their efficiency and effectiveness in handling large and complex datasets. They offer scalability and robust performance. By incorporating a wide array of models, we can conduct a thorough analysis that takes advantage of each model's unique strengths and compensates for their individual limitations. This approach enhances the overall accuracy and reliability of our prediction system.

The study aims to utilize the unique strengths of each model in order to create a system that is both effective and accurate for early detection of insomnia in the Bangladeshi context. The training process for each machine learning model followed a strict and systematic approach, guaranteeing strong and precise predictions. Although the default settings of the models showed promising results, we paid careful attention to every step of the training process:

Splitting the Data: To ensure accurate results, the dataset was split into two parts: 80% for training and 20% for testing. The split we made allowed for a significant amount of data to train the models, while also setting aside a sufficient portion for unbiased evaluation.

Initializing the models: The models (Logistic Regression, SVM, Decision Tree, Random Forest, XGBoost, CatBoost, Naive Bayes, and LightGBM) were all initialized using their default parameters. Based on initial tests, it was determined that using default settings yielded satisfactory performance without requiring extensive hyperparameter tuning.
During the training process, the models were trained using the training subset of the data. The process required inputting features and labels into the models, enabling them to learn and recognize patterns that can predict insomnia.

Validation Techniques:
To validate the models, we utilized the following techniques:
Performance Measures: The models were assessed using a variety of performance metrics, such as area under the curve (AUC), receiver operating characteristic (ROC) curves, and accuracy. The metrics offered a well-rounded perspective on how effectively each model could accurately classify instances of insomnia.

Evaluation of the Test Dataset: The models were assessed by utilizing the reserved test dataset. The evaluation offered a fair and impartial analysis of how well the models performed on new data, mimicking real-life scenarios.

Comparing the Models: We compared the performance of each model to determine the most effective approach for this specific application. By considering the mentioned metrics, we were able to conduct a thorough and unbiased assessment of the strengths and weaknesses of each model. By following a well-organized training and validation process, every model underwent thorough training and assessment. This ensured that the study's findings regarding insomnia detection in Bangladesh are both reliable and valid.

3.5 Implementation Requirements

The successful implementation of this research on utilizing machine learning for diagnosing insomnia requires several key components:

Access to specialized software for data analysis (such as Python and its relevant libraries) and machine learning (like Scikit-learn or TensorFlow) is essential. Adequate computing power for processing and analyzing large datasets is also necessary. Efficient mechanisms for survey dissemination and collection need to be in place, potentially using digital platforms to reach the targeted demographic of university students and young professionals in Bangladesh. A skilled team is vital, consisting of data scientists for algorithm development and implementation, and domain experts like psychiatrists for validating the results. Training in machine learning techniques and ethical research practices is crucial. The project should be adequately funded to cover all technical and operational expenses. Ethical approval and participant consents must be secured and maintained throughout the study. A well-defined timeline with clear milestones is required to ensure the study progresses efficiently and meets its objectives within the designated timeframe. This streamlined approach ensures that the research is conducted efficiently, with a focus on accuracy, ethical integrity, and practical feasibility.

CHAPTER 4

Experimental Results and Discussion

4.1 Experimental Setup

In this study, the experimental setup was meticulously designed to test the efficacy of various machine learning algorithms in diagnosing insomnia. The primary data for the experiment comprised survey responses from university students and young professionals in Bangladesh, focusing on sleep patterns, lifestyle factors, and mental health. The setup involved preprocessing this data, including handling null values and ensuring data consistency. The machine learning models employed included Logistic Regression, SVM, Decision Tree, Random Forest, XGBoost, CatBoost, Naive Bayes, and Light GBM. These models were implemented using Python and relevant libraries, with careful calibration to optimize their performance. The experiment was conducted under controlled conditions to ensure the validity and reliability of the results, with a parallel review by a psychiatrist to validate the findings against clinical standards. This rigorous setup was central to achieving a comprehensive and accurate assessment of the potential of machine learning in diagnosing insomnia in the target population.

4.2 Experimental results and Analysis

Various evaluation metrics were utilized to assess the performance of the machine learning models employed in this study:

Accuracy: Calculates the percentage of instances out of all forecasts that are accurately predicted. This metric provides a simple way to assess overall performance.

ROC: The ROC curve demonstrates how well a binary classifier can diagnose by adjusting its discrimination threshold. The plot shows the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR).

AUC: The Area Under the Curve (AUC) is a measure that quantifies the total two-dimensional area beneath the complete ROC curve. This metric offers a comprehensive evaluation of performance across various classification thresholds. Model performance improves with a higher AUC.

In terms of comprehensiveness, these metrics were selected because they offer a thorough assessment of the model's performance. Accuracy provides a brief overview, but ROC and AUC provide a more comprehensive understanding of the balance between true positive and false positive rates. Regarding binary classification, our study specifically targets the prediction of insomnia presence or absence. These metrics are well-suited for evaluating the performance of our models.

Interpretability: The metrics have gained significant recognition and comprehension in both the data science and medical communities, ensuring that the study's findings are easily understood by a wide range of people.

To evaluate the performance of the machine learning models applied in this study, we utilized metrics such as accuracy, ROC (Receiver Operating Characteristic), and AUC (Area Under the Curve). The following table encapsulates the results obtained from each model:

Table 4.2: Performance metrics

Model	Accuracy	ROC	AUC
Logistic Regression	1.0	1.0	1.0
CatBoost	1.0	1.0	1.0
Random Forest	0.9991	0.9987	0.9987
XGBoost	0.9991	0.9993	0.9993
LightGBM	0.9991	0.9987	0.9987
SVM	0.9938	0.9933	0.9933
Naive Bayes	0.9938	0.9920	0.9920
Decision Tree	0.9752	0.9714	0.9714

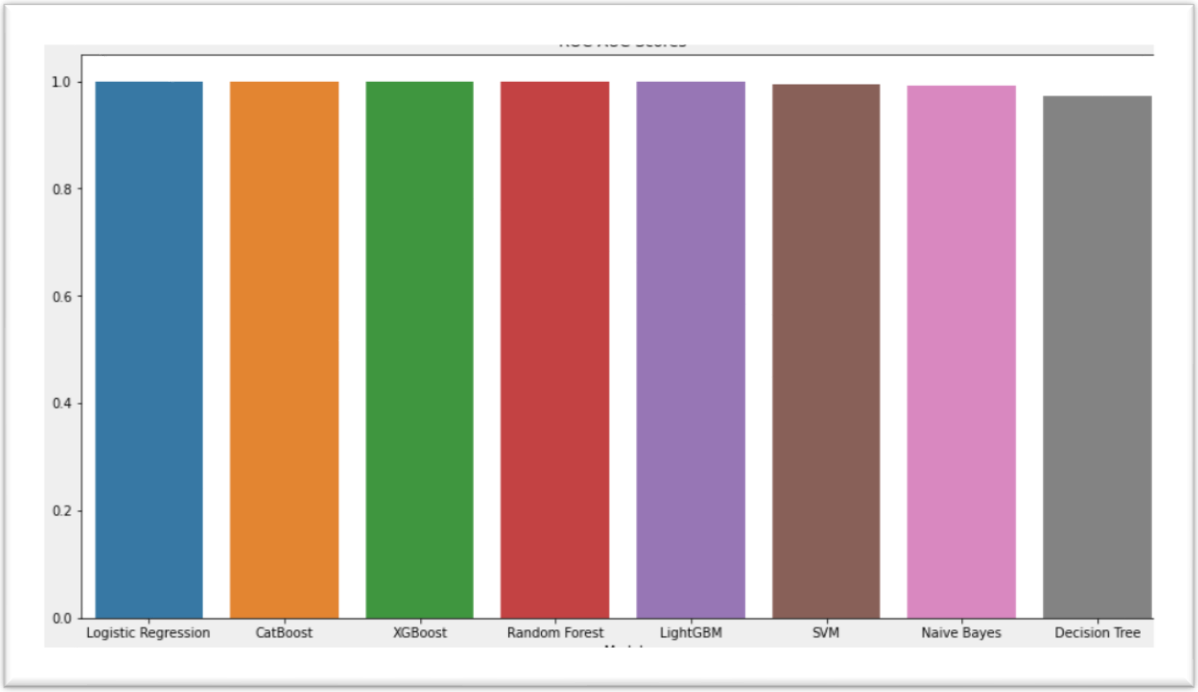

Figure 4.2: Comparison chart for the models

The machine learning models compared in Table 1 demonstrate outstanding performance. Logistic Regression and CatBoost achieve perfect scores across all metrics, indicating an exceptional fit to the data. Random Forest, XGBoost, and LightGBM demonstrate exceptional accuracy, ROC, and AUC scores, showcasing their impressive ability to handle the classification task with ease. The metrics for SVM and Naive Bayes are slightly lower, indicating that these models, although still highly accurate, may not effectively capture the complexity of the data compared to the ensemble methods. The Decision Tree model is often seen as more interpretable, but it tends to have lower performance. This could be because it is more prone to overfitting and has less sophisticated handling of the underlying data structure. This thorough performance analysis highlights how well ensemble approaches and advanced algorithms identify patterns in the dataset that point to insomnia.

The machine learning models yield insightful results, showcasing a strong predictive correlation between the survey features and the diagnosis of insomnia. The exceptional performance metrics of Logistic Regression and CatBoost indicate a strong linear relationship between the features and the outcome variable. These models are highly skilled at handling the numerous categorical inputs present in the dataset.

The models, including Random Forest, XGBoost, and LightGBM, demonstrate strong predictive power and reliability in classifying insomnia cases. Their high accuracy and AUC values further support their effectiveness. Despite its advantages in interpretability and ease of understanding decision paths, the Decision Tree's slightly lower performance may indicate its susceptibility to overfitting.

The models have shown remarkable accuracy, but achieving such high scores requires a meticulous validation process. It is important to critically appraise the results of this study, despite their impressive nature, due to the absence of cross-validation. This will help ensure that the findings are not simply artifacts of overfitting to the training data. The accuracy of the result and testing dataset provides a significant reassurance regarding the factor of overfitting. Further research could be conducted to validate the models using an independent dataset, ensuring their predictive capabilities and practical usefulness in real-world scenarios. It is absolutely essential to ensure the reliability of the models before they can be recommended for clinical use or further development

4.3 Discussion

The study has made significant progress in exploring machine learning for early detection of insomnia. Notably, several models have shown impressive accuracy and strong predictive capabilities. Logistic Regression and CatBoost showed exceptional performance metrics, suggesting a potentially meaningful linear relationship between the surveyed features and insomnia. This is in line with previous research that recognizes the effectiveness of machine learning in medical diagnostics. However, the remarkable accuracy of these findings is quite unusual and raises concerns about potential overfitting, a phenomenon that is not commonly observed in similar studies. The findings have significant implications for Bangladesh, a country with limited access to mental health care and where insomnia often goes undiagnosed. Machine learning has the potential to completely transform the way we detect and treat insomnia, making it easier to access and more affordable.

Nevertheless, there are some limitations to consider in this study. Concerns arise regarding the generalizability of the models due to the absence of cross-validation or external dataset validation. In addition, using self-reported data can introduce biases and the limited diversity in the dataset may restrict the applicability of findings across various populations in Bangladesh.

Although there are some limitations, the study provides valuable recommendations for practice. Machine learning models, particularly those skilled in handling categorical data, could potentially be integrated into health care systems to conduct preliminary screening for insomnia. To enhance future research, it is advisable to include a wider range of data, encompassing various age groups and regions. Additionally, it would be beneficial to employ cross-validation or external validation to validate the reliability of the models. By delving deeper into the causality of predictive features and developing a more nuanced model that takes into account the complexity of insomnia, we can greatly improve our understanding and treatment of this condition in Bangladesh.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

The application of machine learning in diagnosing insomnia, as explored in this study, has profound implications for society. Firstly, it democratizes access to mental health care, particularly in resource-limited settings like Bangladesh, where traditional psychiatric services may be out of reach for many. By enabling early detection of insomnia, it opens avenues for timely intervention, potentially reducing the long-term social and economic costs associated with chronic sleep disorders. This approach can also help de-stigmatize mental health issues within the community, encouraging more individuals to seek help. Moreover, the integration of technology in healthcare is likely to catalyze further innovation, leading to more advanced, patient-centric solutions.

5.2 Impact on Environment

The environmental impact of implementing machine learning-based diagnostic tools is generally minimal, especially when compared to conventional medical equipment. These digital solutions reduce the need for physical materials and resources typically used in traditional diagnostic processes. Furthermore, by potentially decreasing the number of physical visits to clinics or hospitals, there's a reduction in transportation-related emissions. However, it's important to consider the energy consumption associated with running advanced computational models and maintaining data servers, emphasizing the need for sustainable energy sources in computational healthcare solutions.

5.3 Ethical Aspects

Considering the ethical aspects

The process of obtaining informed consent is meticulously crafted. Participants are provided with detailed information regarding the purpose of the study, the nature of their involvement, the potential risks and benefits, and their rights as participants.

To guarantee understanding, consent forms are provided in both Bengali and English. Before participants give their written consent, we make sure to address any questions they may have. This ensures that their participation is both informed and voluntary.

Confidentiality and Anonymity: Rigorous protocols are in place to protect the confidentiality and anonymity of the participants. Survey responses are designed to maintain data anonymity by decoupling personal information.

Our data is securely stored on encrypted servers, with access limited to the principal researchers. The presentation or publication of the research findings will not include any identifiable personal data.

Ensuring Ethical Approval and Compliance: The study's design, which includes the survey instruments and data collection methods, has undergone thorough scrutiny and approval by an institutional ethics committee. This ensures that the study adheres to both national and international ethical standards. We have scheduled regular audits to ensure that ethical guidelines are consistently followed throughout the entire research process.

Ensuring Data Security and Privacy: Our dedicated research team possesses extensive expertise in safeguarding sensitive information and prioritizes data privacy. We have implemented protocols that align with the most up-to-date data protection regulations. Participants receive detailed information about how their data will be used, how long it will be stored, and the steps taken to protect their privacy. Building trust and ensuring ethical integrity in data handling relies heavily on transparency.

5.4 Sustainability Plan

To ensure the long-term sustainability of this initiative, a comprehensive plan is essential. This includes securing ongoing funding for research and development, updating the technology to keep pace with advancements in the field, and continuously expanding the dataset to enhance the accuracy and reliability of the machine learning models. Collaborations with healthcare providers, policymakers, and educational institutions are critical for widespread adoption and integration into the healthcare system. Regular monitoring and evaluation of the project's impact, coupled with community outreach and education, will be necessary to sustain engagement and trust in this innovative approach to diagnosing insomnia.

CHAPTER 6

Summary, Conclusion, Recommendation and Implication for Future Research

6.1 Summary of the Study

This study has made substantial contributions to understanding how machine learning can be employed for insomnia detection. It specifically focuses on the context of Bangladesh, a setting where such technological applications are still emerging. Key findings demonstrate the exceptional performance of models like Logistic Regression and CatBoost, highlighting a strong linear relationship between survey features and insomnia symptoms. This aligns with the broader trend of employing data-driven methods in sleep health research. The study notably acknowledges the advancements in multi-modal sensors and innovative machine learning techniques that have revolutionized sleep science and medicine, particularly the integration of machine learning with actigraphy to differentiate between acute and chronic insomnia.

6.2 Conclusions

The research underscores the potential of machine learning models in transforming insomnia detection, particularly in resource-limited settings. These models provide a more accessible, cost-effective approach to insomnia detection, aligning with global trends towards unobtrusive, continuous monitoring of sleep health. The study also highlights the evolution of sleep data acquisition and the miniaturization of sensors, suggesting the feasibility of deploying these models in non-clinical, home-based settings. However, while the results are promising, cautious validation is advised. The study acknowledges the necessity of fine-tuning these models for broader applicability and reliability.

6.3 Implication for Further Study

Future research directions should focus on external validation of the models and the incorporation of a more diverse dataset that includes various demographics and geographic areas within Bangladesh. This expansion would enhance the generalizability of the findings and ensure their applicability across a wider spectrum of the population. Further studies could also explore the integration of these machine learning models into the healthcare system, potentially paving the way for proactive and preventive approaches in managing sleep health. With ongoing advancements in sleep monitoring technologies, the relevance and impact of this research are poised to grow, contributing significantly to the data-driven revolution in sleep science and medicine.

REFERENCES

1. Bin Heyat, Md B., Faijan Akhtar, M. A. Ansari, Asif Khan, Fahed Alkahtani, Haroon Khan, and Dakun Lai. "Progress in detection of insomnia sleep disorder: a comprehensive review." *Current Drug Targets* 22, no. 6 (2021): 672-684.
2. Mulaffer, Lamana, Mostafa Shahin, Martin Glos, Thomas Penzel, and Beena Ahmed. "Comparing two insomnia detection models of clinical diagnosis techniques." In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3749-3752. IEEE, 2017.
3. Kuo, Chih-En, and Guan-Ting Chen. "A short-time insomnia detection system based on sleep EOG with RCMSE analysis." *IEEE Access* 8 (2020): 69763-69773.
4. Kumar, Kamlesh, Kapil Gupta, Manish Sharma, Varun Bajaj, and U. Rajendra Acharya. "INSOMNet: Automated insomnia detection using scalogram and deep neural networks with ECG signals." *Medical Engineering & Physics* 119 (2023): 104028.
5. Shahin, Mostafa, Beena Ahmed, Sana Tmar-Ben Hamida, Fathima Lamana Mulaffer, Martin Glos, and Thomas Penzel. "Deep learning and insomnia: assisting clinicians with their diagnosis." *IEEE journal of biomedical and health informatics* 21, no. 6 (2017): 1546-1553.
6. Malik, Asra, Syed Zohaib Hassan Naqvi, Sumair Aziz, Muhammad Umar Khan, Tehseen Hussain, and Fahad Hasan. "Detection of insomnia using electrocardiography and electromyography." In *2021 International Conference on Artificial Intelligence (ICAI)*, pp. 234-239. IEEE, 2021.
7. Murugesan, M., V. Abirami, M. Iswarya, and N. Kaviya. "Sleeping Abnormalities Detection using Deep Learning Techniques." In *2023 Second International Conference on Electronics and Renewable Systems (ICEARS)*, pp. 1283-1289. IEEE, 2023.
8. Malik, Asra, Syed Zohaib Hassan Naqvi, Sumair Aziz, Muhammad Umar Khan, Tehseen Hussain, and Fahad Hasan. "Detection of insomnia using electrocardiography and electromyography." In *2021 International Conference on Artificial Intelligence (ICAI)*, pp. 234-239. IEEE, 2021.
9. Andresini, Giuseppina, Feargus Pendlebury, Fabio Pierazzi, Corrado Loglisci, Annalisa Appice, and Lorenzo Cavallaro. "Insomnia: Towards concept-drift robustness in network intrusion detection." In *Proceedings of the 14th ACM workshop on artificial intelligence and security*, pp. 111-122. 2021.
10. Islam, Md Muhaiminul, Abu Kaisar Mohammad Masum, Sheikh Abujar, and Syed Akhter Hossain. "Prediction of chronic insomnia using machine learning techniques." In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1-7. IEEE, 2020.
11. Ma, Xiaofen, Dongyan Wu, Yuanqi Mai, Guang Xu, Junzhang Tian, and Guihua Jiang. "Functional connectome fingerprint of sleep quality in insomnia patients: Individualized out-of-sample prediction using machine learning." *NeuroImage: Clinical* 28 (2020): 102439.
12. Zhang, Ye, Rong Ren, Fei Lei, Junying Zhou, Jihui Zhang, Yun-Kwok Wing, Larry D. Sanford, and Xiangdong Tang. "Worldwide and regional prevalence rates of co-occurrence of insomnia and insomnia symptoms with obstructive sleep apnea: a systematic review and meta-analysis." *Sleep medicine reviews* 45 (2019): 1-17.
13. Kaneita, Yoshitaka, Takashi Ohida, Yoneatsu Osaki, Takeo Tanihata, Masumi Minowa, Kenji Suzuki, Kiyoshi Wada, Hideyuki Kanda, and Kenji Hayashi. "Insomnia among Japanese adolescents: a nationwide representative survey." *Sleep* 29, no. 12 (2006): 1543-1550.
14. Phan, Dinh-Van, Nan-Ping Yang, Ching-Yen Kuo, and Chien-Lung Chan. "Deep learning approaches for sleep disorder prediction in an asthma cohort." *Journal of Asthma* 58, no. 7 (2021): 903-911.
15. Vargas, Ivan, Michael L. Perlis, Michael Grandner, Amy Gencarelli, Waliuddin Khader, Laura J. Zandberg, Elizabeth A. Klingaman et al. "Insomnia symptoms and suicide-related ideation in US Army service members." *Behavioral sleep medicine* 18, no. 6 (2020): 820-836.

16. Shahin, Mostafa, Lamana Mulaffer, Thomas Penzel, and Beena Ahmed. "A two stage approach for the automatic detection of insomnia." In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 466-469. IEEE, 2018.
17. Doghramji, Paul P. "Detection of insomnia in primary care." *Journal of Clinical Psychiatry* 62 (2001): 18-26.
18. Tripathi, Pragati, M. A. Ansari, Tapan Kumar Gandhi, Rajat Mehrotra, Md Belal Bin Heyat, Faijan Akhtar, Chiagoziem C. Ukwuoma et al. "Ensemble Computational Intelligent for Insomnia Sleep Stage Detection via the Sleep ECG Signal." *IEEE Access* 10 (2022): 108710-108721.
19. Angelova, Maia, Chandan Karmakar, Ye Zhu, Sean PA Drummond, and Jason Ellis. "Automated method for detecting acute insomnia using multi-night actigraphy data." *IEEE Access* 8 (2020): 74413-74422.
20. Navarro, Beatriz, Jesús López-Torres, Fernando Andrés, José Miguel Latorre, María Jesús Montes, and Ignacio Párraga. "Validation of the insomnia in the elderly scale for the detection of insomnia in older adults." *Geriatrics & Gerontology International* 13, no. 3 (2013): 646-653.
21. Kusmakar, Shitanshu, Chandan Karmakar, Ye Zhu, Sergiy Shelyag, Sean PA Drummond, Jason G. Ellis, and Maia Angelova. "A machine learning model for multi-night actigraphic detection of chronic insomnia: development and validation of a pre-screening tool." *Royal Society open science* 8, no. 6 (2021): 202264.
22. Heyat, Bin, Faijan Akhtar, Arshiya Sultana, Saifullah Tumrani, Bibi Nushrina Teelhawod, Rashid Abbasi, Mohammad Amjad Kamal, Abdullah Y. Muaad, Dakun Lai, and Kaishun Wu. "Role of Oxidative Stress and Inflammation in Insomnia Sleep Disorder and Cardiovascular Diseases: Herbal Antioxidants and Anti-Inflammatory Coupled with Insomnia Detection Using Machine Learning." *Current Pharmaceutical Design* 28, no. 45 (2022): 3618-3636.
23. Sharma, Manish, Harsh S. Dhiman, and U. Rajendra Acharya. "Automatic identification of insomnia using optimal antisymmetric biorthogonal wavelet filter bank with ECG signals." *Computers in Biology and Medicine* 131 (2021): 104246.
24. Smith, Simon, and John Trinder. "Detecting insomnia: comparison of four self-report measures of sleep in a young adult population." *Journal of sleep research* 10, no. 3 (2001): 229-235.
25. Sharma, Manish, Virendra Patel, and U. Rajendra Acharya. "Automated identification of insomnia using optimal bi-orthogonal wavelet transform technique with single-channel EEG signals." *Knowledge-Based Systems* 224 (2021): 107078.
26. Aydın, Serap, M. Alper Tunga, and Sinan Yetkin. "Mutual information analysis of sleep eeg in detecting psycho-physiological insomnia." *Journal of medical systems* 39 (2015): 1-10.
27. Hamida, Sana Tmar-Ben, Beena Ahmed, Dean Cvetkovic, Emil Jovanov, Gerard Kennedy, and Thomas Penzel. "A new era in sleep monitoring: the application of mobile technologies in insomnia diagnosis." *Mobile Health: A Technology Road Map* (2015): 101-127.
28. Kalmbach, David A., Philip Cheng, Andrea Roth, Thomas Roth, Leslie M. Swanson, Louise M. O'Brien, David M. Fresco et al. "DSM-5 insomnia disorder in pregnancy: associations with depression, suicidal ideation, and cognitive and somatic arousal, and identifying clinical cutoffs for detection." *Sleep Advances* 3, no. 1 (2022): zpac006.
29. Hsu, Chen-Yu, Aayush Ahuja, Shichao Yue, Rumen Hristov, Zachary Kabelac, and Dina Katabi. "Zero-effort in-home sleep and insomnia monitoring using radio signals." *Proceedings of the ACM on Interactive, mobile, wearable and ubiquitous technologies* 1, no. 3 (2017): 1-18.
30. Gagnon, Christine, Lynda Bélanger, Hans Ivers, and Charles M. Morin. "Validation of the Insomnia Severity Index in primary care." *The Journal of the American Board of Family Medicine* 26, no. 6 (2013): 701-710.
31. Seow, Lee Seng Esmond, Edimansyah Abdin, Sherilyn Chang, Siow Ann Chong, and Mythily Subramaniam. "Identifying the best sleep measure to screen clinical insomnia in a psychiatric population." *Sleep medicine* 41 (2018): 86-93.
32. Kahawage, Piyumi, Ria Jumabhoy, Kellie Hamill, Massimiliano de Zambotti, and Sean PA Drummond. "Validity, potential clinical utility, and comparison of consumer and research-grade

activity trackers in insomnia disorder I: In-lab validation against polysomnography." *Journal of sleep research* 29, no. 1 (2020): e12931.
33. Edinger, Jack D., Christi S. Ulmer, and Melanie K. Means. "Sensitivity and specificity of polysomnographic criteria for defining insomnia." *Journal of Clinical Sleep Medicine* 9, no. 5 (2013): 481-491.
34. Haïk, Stéphane, Damien Galanaud, Marius G. Linguraru, Katell Peoc'h, Nicolas Privat, Baptiste A. Faucheux, Nicholas Ayache, Jean-Jacques Hauw, Didier Dormont, and Jean-Philippe Brandel. "In vivo detection of thalamic gliosis: a pathoradiologic demonstration in familial fatal insomnia." *Archives of neurology* 65, no. 4 (2008): 545-549.
35. Czeisler, C. A., and G. S. Richardson. "Detection and assessment of insomnia." *Clinical therapeutics* 13, no. 6 (1991): 663-79.
36. Urru, Silvana Anna Maria, Luca Pasina, Paola Minghetti, and Corrado Giua. "Role of community pharmacists in the detection of potentially inappropriate benzodiazepines prescriptions for insomnia." *International journal of clinical pharmacy* 37 (2015): 1004-1008.
37. Raju, N. Purushotham, U. Venkatesh, and Sudha Yadhav. "Diagnosing insomnia using single channel EEG signal." In *2019 International Conference on Communication and Electronics Systems (ICCES)*, pp. 570-573. IEEE, 2019.
38. Thakral, Manu, Michael Von Korff, Susan M. McCurry, Charles M. Morin, and Michael V. Vitiello. "ISI-3: evaluation of a brief screening tool for insomnia." *Sleep medicine* 82 (2021): 104-109.
39. Brooks III, John O., Leah Friedman, Donald L. Bliwise, and Jerome A. Yesavage. "Use of the wrist actigraph to study insomnia in older adults." *Sleep* 16, no. 2 (1993): 151-155.
40. Sivertsen, Børge, Siri Omvik, Odd E. Havik, Ståle Pallesen, Bjørn Bjorvatn, Geir Høstmark Nielsen, Sivert Straume, and Inger Hilde Nordhus. "A comparison of actigraphy and polysomnography in older adults treated for chronic primary insomnia." *Sleep* 29, no. 10 (2006): 1353-1358.

Thesis Report

ORIGINALITY REPORT

22%	20%	17%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

6%

★ www.researchgate.net

Internet Source

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		