

SLEEP STAGE SCORING USING EEG IN REAL TIME

ABSTRACT

The precise determination of sleep stages is a foundational element in the quantitative assessment of polysomnographic data. Conventionally, this task has been performed manually by experts through visual recognition of patterns, a method that can be both time-intensive and inherently subjective. Intending to surmount these obstacles, there is an increasing impetus for the adoption of automated classification techniques. In our research, we have engineered and assessed various machine learning models, such as Random Forest (RF), Support Vector Machine (SVM), and an integrated Convolution Neural Network and Long Short Term memory (CNN-LSTM) framework, utilizing feature-based approaches alongside Artificial Neural Networks (ANNs) that process both features and raw data. Our findings indicate that the performance of most algorithms is on par with the level of agreement typically seen between human raters, especially in healthy individuals.

Our analysis utilized a dataset^[15] that conforms to the Rechtschaffen and Kales guidelines for systematic sleep stage classification, containing 197 full-night polysomnographic sleep studies featuring EEG, chin EMG, and EOG, along with select event markers. Additionally, subsets of this data include variables such as respiration and body temperature. The hypnograms associated with these PSG recordings have been meticulously scored by skilled technicians adhering to the 1968 Rechtschaffen and Kales manual.

The evaluation of our proposed model yielded an impressive accuracy rate: 98% for WAKE stages, 91% for NREM stages, and 76% for REM stages. These figures notably surpass the accuracy rates of SVM and RF, which are approximately 90% and 94%, respectively. This demonstrates the superior capability of our model over some of the existing methodologies, including SVM and RF, in classifying sleep stages accurately.

TABLE OF CONTENTS

Chapter Number	Topic	Page no's
1	Introduction	1
1.2	Motivation	3
1.3	Objectives	4
1.4	Methodology	6
2	Literature Survey	9
3	Design and Analysis	21
4	Result	29
5	Conclusion and Future Scope	34
6	Reference	36

Chapter 1. INTRODUCTION

1.1 Overall Background System

Sleep is a fundamental aspect of human health and well-being, serving as a crucial period for the restoration and repair of both body and mind. It plays an integral role in various aspects of cognitive function, mood regulation, and physical health. The absence or deficiency of sleep has been strongly linked to a myriad of health issues, including obesity, diabetes, cardiovascular diseases, and neurological disorders. It also significantly impacts cognitive abilities, affecting attention, memory, and decision-making processes.

Moreover, sleep is vital for memory consolidation, cognitive function, learning, brain health, and hormonal balance. During sleep, the brain engages in activities like strengthening and organizing new memories, enhancing focus and learning, and detoxifying and protecting neural cells. Insufficient sleep has been associated with numerous neurological problems and hormonal imbalances, underscoring the necessity of adequate sleep for overall well-being.

Sleep stage scoring, a pivotal component of sleep medicine, involves classifying sleep into distinct stages based on electroencephalogram (EEG) signal characteristics. EEG, which records the brain's electrical activity, is instrumental in identifying various sleep stages, including wakefulness, rapid eye movement (REM) sleep, and non-rapid eye movement (NREM) sleep. NREM sleep is further categorized into three stages: N1, N2, and N3, each distinguished by unique EEG patterns. For instance, wakefulness features high-frequency, low-amplitude alpha waves, while REM sleep is marked by low-frequency, high-amplitude theta waves, and sawtooth waves. NREM sleep displays a mix of alpha, beta, theta, and delta waves, with delta waves becoming more prevalent from N1 to N3.

Traditionally, sleep stage scoring has been manually conducted by trained technicians, a method that can be both time-consuming and subjective. However, recent years have seen a surge in interest in creating automated sleep stage scoring systems using advanced biomedical signal processing techniques.

With advancements in computing power and the availability of extensive, labelled datasets, deep neural networks have begun to rival the expertise of medical professionals in complex medical pattern recognition tasks, such as diagnosing dermatologic lesions and diabetic retinopathy. This research focuses on the development of various machine learning techniques, including Convolutional Neural Networks (CNN), Long Short Term Memory (LSTM) networks, Random Forests (RF), and Support Vector Machines (SVM). These methods have shown promising potential in matching the proficiency of sleep experts in annotating overnight polysomnograms (PSGs), a critical step in sleep stage scoring.

Our exploration delves into how these advanced machine learning models can efficiently and accurately automate the process of sleep stage classification, a development that could revolutionize the field of sleep medicine by enhancing the accuracy and efficiency of sleep disorder diagnoses.

1.2 Motivation

The advancement of techniques for diagnosing and treating sleep disorders has been significantly propelled by the development of an AI-driven approach to sleep staging, designed for enhanced efficacy and user-friendliness. The typically slow and cumbersome nature of traditional manual sleep staging has often led to delays in diagnosing and treating sleep disorders. By integrating artificial intelligence, this process is streamlined, providing faster and more efficient service to both patients and healthcare practitioners.

The importance of quickly identifying and treating sleep disorders cannot be overstated, considering their substantial negative impact on general health. Untreated sleep disorders are closely linked to several chronic illnesses, such as diabetes, heart disease, and obesity. Therefore, it is vital to provide the healthcare industry with tools that allow for rapid detection and timely intervention, thereby preventing the escalation of these serious health issues.

AI-powered systems offer immediate feedback on sleep quality, enabling the early identification of sleep issues. This rapid detection is crucial for enacting preventive measures and lifestyle adjustments to avoid the worsening of health conditions related to inadequate sleep. Moreover, the personalized insights provided by AI facilitate the development of effective, customized treatment plans, catering to individual sleep patterns and behaviours.

Understanding the diverse factors that affect sleep, such as behavioural, environmental, and lifestyle aspects, is crucial. Artificial intelligence is adept at conducting comprehensive analyses of these factors, thus enabling individuals to customize their strategies for enhancing sleep habits and hygiene. This, in turn, boosts overall health and well-being.

Ultimately, the objective is to fully leverage artificial intelligence to offer a prompt, personalized, and accessible sleep staging solution. This technology aims to be a

pivotal tool in enhancing early detection and providing tailored interventions, thereby playing a crucial role in promoting better sleep hygiene and averting the progression of related chronic health conditions.

1.3 Objectives

1.3.1 Main Objective:

The objective of the study is to compare the accuracy and effectiveness of different sleep detection approaches (classification, clustering and classification, deep learning) in a real-time data streaming setup using Python or MatLab, focusing on EEG data from existing polysomnographic datasets.

1.3.2 Sub-Objectives:

- **Comprehensive Study of Sleep Stages:**
Conduct a thorough literature review to understand the characteristics of different sleep stages. This involves exploring the physiological aspects, EEG patterns, and transitions between stages.
- **Understanding EEG Database:**
Gain a comprehensive understanding of EEG databases commonly used in sleep research. This includes exploring the structure of polysomnographic datasets, the types of signals recorded, and the annotation of sleep stages.
- **Various Techniques for Sleep Stage Analysis:**
Investigate various techniques employed for sleep stage analysis, encompassing classification, clustering, and deep learning approaches. Comprehend the advantages and constraints of each technique.
- **Process of Data Extraction from Datasets:**
Develop expertise in the extraction of relevant data from EEG datasets. This involves understanding the format of data storage, preprocessing steps, and methods for extracting features related to sleep stages.
- **Exploration of Python Libraries:**
Explore different Python libraries suitable for handling EEG data, signal processing, and machine learning. Identify and evaluate libraries that can facilitate the classification and analysis of sleep stages.
- **Algorithm Development:**
Develop a comprehensive algorithm that integrates different approaches (classification, clustering and classification, deep learning) for comparing

accuracy and effectiveness in sleep stage detection. Consider the unique aspects of each method and how they can be applied to polysomnographic data.

- **Application to Existing Databases:**

Implement the developed algorithm on existing databases with annotated sleep stages. Evaluate the performance of this algorithm using metrics like accuracy, F1 score, and confusion matrix.

- **Building a Real-Time Data Streaming Environment:**

Create a real-time data streaming environment using Python. This involves setting up a system to receive, process, and analyze EEG data as it is generated in real-time.

- **Application to Real-Time Data Streaming:**

Apply the developed algorithm to real-time EEG data streaming. Compare the accuracy and effectiveness of different sleep detection approaches in a dynamic, real-world setting.

1.4 Methodology

Our methodology is structured to systematically implement an automated sleep scoring algorithm that compares the accuracy of multiple existing sleep scoring techniques which include Convolutional Neural Networks(CNN), Long Short-Term Memory(LSTM), Random Forests(RF), Support Vector Machine(SVM).

The dataset^[15] for this classification was acquired from physionet.org. This data(in the form of EDF files(European Data Format, is a standard file format designed for the storage and exchange of multichannel biological and physical signals)) comprises of 197 Polysomnographic sleep recordings of human subjects aged between 25-101, which include 153 recordings of healthy human subjects and 44 recordings of healthy human subjects who had slight difficulty in falling asleep. All these recordings are about 9 hours long. These PSGs are manually scored by some well-trained professionals following the Rechtschaffen and Kales manual(based on Fpz-Cz and Pz-Oz format instead of C4-A1 and C3-A2 format) resulting in the corresponding hypnogram files. These EEG signals were sampled at 100Hz.

This dataset downloaded from physionet database, extracted, and each recording is diving into 30s epochs and then subjected to extraction of features^[2] like standard deviation, skewness, kurtosis, maximum value, minimum value, interquartile range(IQR), Shannon's entropy, fractal dimensions, activity, mobility, complexity, root sum of squares(RSSQ), mean frequency, median frequency, signal to noise ratio, and range. These extracted features from the signals are combined and labelled.

These features are used to implement the following techniques:

1. CNN-LSTM:

- This technique implemented using Python requires conversion of these EDF files to pickle(.pkl) files as EDF files could not be directly read.
- These pickle files are divided into 20%-80% fashion for validation, testing, and training. These pickle files are parsed, and spectrograms are generated.
- Spectrograms generated are used for feature extraction.
- A CNN-LSTM model built using Python based on the base paper.
- Validation, test, and training pickle files are used in this CNN-LSTM for model training.

- This model is evaluated based on confusion matrix, accuracy, and F1 score, by predicting one subject's record.^[5]

2. RF and SVM:

- This technique implemented using MATLAB requires conversion of EDF+ type hypnogram files to be converted into .csv files.
- These .csv files are subjected to feature extraction based on the frequency of the 30s epoch of the signal and are stored in a feature matrix.
- Random Forest and Support Vector Machine algorithms utilize this feature matrix to classify into different stages of sleep.
- These machine learning models are evaluated based on the confusion matrix, accuracy, and F1 scores.

1.5 Conclusion

In conclusion, this comprehensive study focuses on leveraging advanced machine learning techniques, including CNN-LSTM, RF, and SVM, for automated sleep stage classification. The motivation behind this research stems from the significant impact of sleep disorders on overall health and the potential of AI-driven approaches to streamline the traditionally slow and subjective manual sleep staging process.

The objectives encompass a deep dive into sleep stages, understanding EEG databases, exploring various analysis techniques, developing algorithms, and creating a real-time data streaming environment. The methodology involves using a dataset from physionet.org, extracting features from EEG signals, and implementing CNN-LSTM, RF, and SVM models for sleep stage classification.

The study aims to compare the accuracy and effectiveness of these approaches, offering insights into their potential applications in real-time data streaming setups. By bridging the gap between traditional manual scoring and AI-driven automation, this research seeks to enhance the efficiency of sleep disorder diagnoses, potentially revolutionizing the field of sleep medicine and promoting early detection and intervention for improved overall health and well-being.

Chapter 2. LITERATURE SURVEY

2.1 Introduction

Sleep is a fundamental aspect of human life, playing a crucial role in physical and mental health, cognitive functioning, and overall well-being. Understanding and monitoring sleep patterns have been of significant interest to both medical professionals and researchers. Polysomnography (PSG) recordings provide a comprehensive view of sleep by capturing various physiological signals, including Electroencephalogram (EEG), Electrooculogram (EOG), Electromyogram (EMG), and Electrocardiogram (ECG). The accurate classification of sleep stages from PSG data is essential for diagnosing sleep disorders, assessing sleep quality, and advancing our understanding of sleep physiology.

In recent years, deep learning and machine learning techniques have emerged as powerful tools for automated sleep stage classification. The application of deep neural networks and machine learning algorithms, including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, Random Forests(RF), Support Vector Machine(SVM), and combinations thereof, has shown promising results in achieving accurate and reliable sleep stage classification. This literature survey focuses on the comprehensive exploration of deep learning-based approaches for automated sleep stage detection from PSG recordings, covering studies conducted from 2010 to 2020. We systematically review the methodologies, models, and performances of these techniques, providing insights into their capabilities, limitations, and potential future directions. Through this survey, we aim to consolidate valuable information for experts in the field and contribute to the ongoing advancement of automated sleep stage classification.

2.2 Literature Survey:

2.2.1 Base Paper

The study published in the paper^[13] “**Mousavi S, Afghah F, Acharya UR (2019) SleepEEGNet: Automated sleep stage scoring with sequence-to-sequence deep learning approach. PLoS ONE 14(5): e0216456. <https://doi.org/10.1371/journal.pone.0216456>**”, aims to address the class imbalance issue in sleep datasets, develop an end-to-end deep learning approach, and introduce a sequence-to-sequence model with Bidirectional Recurrent Neural Network (BiRNN) and attention mechanisms optimized for sleep stage scoring. The proposed model, named SleepEEGNet, integrates innovative loss functions to tackle class imbalance challenges and achieves remarkable advancements over existing algorithms.

The methodology involves data preprocessing, architectural framework development utilizing CNNs, BiRNNs, and attention mechanisms, and training and evaluation procedures on the Physionet Sleep-EDF dataset. The model processes 30-second EEG epochs without manual feature engineering, combining CNNs for spatial feature extraction and a BiRNN in a sequence-to-sequence model with an attention network. The study addresses challenges such as the demand for abundant training samples and the requirement for specific sequences in the sequence-to-sequence model.

The proposed architecture is illustrated, featuring a CNN with ReLU nonlinearity, max pooling, and dropout layers, followed by a sequence-to-sequence model with BiRNN and LSTM units. The model's encoder-decoder framework captures complex dependencies in EEG data. Results demonstrate superior performance, especially in identifying the challenging N1 sleep stage, with an overall accuracy of 84.26% and a macro F1-score of 79.66%, outperforming existing algorithms.

The paper transparently discusses the challenges faced, including the need for extensive training samples and limited evaluation channels (Fpz-Cz and Pz-Oz EEG channels). Solutions involve innovative loss functions, a sequence-to-sequence approach, and the

use of BiRNN. The model's architecture, training process, and evaluation metrics such as accuracy, precision, recall, F1-score, and Cohen's Kappa coefficient are detailed. Visualization strategies, including confusion matrices and hypnogram comparison, enhance result interpretation.

In conclusion, the SleepEEGNet model demonstrates significant advancements in automated sleep stage scoring, excelling in performance and addressing class imbalance challenges. The study highlights the model's superiority through comprehensive comparisons with state-of-the-art algorithms. The proposed methodology opens avenues for precise and automated sleep stage classification, showcasing potential applications in biomedical signal processing.

2.2.2 Paper 1:

The paper^[5] **“Zhuang L, Dai M, Zhou Y and Sun L (2022) Intelligent automatic sleep staging model based on CNN and LSTM. Front. Public Health 10:946833. doi: 10.3389/fpubh.2022.946833”**, proposes an automatic sleep staging method using a multi-channel EEG approach, integrating Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The objectives include addressing the class imbalance, fusing EEG and Electrooculogram (EOG) information, decomposing EEG signals, and establishing an accurate and fast automatic sleep staging system for medical researchers.

The research methodology involves rigorous data collection and preprocessing of polysomnography (PSG) data, featuring EEG and EOG signals. A dual extraction process using CNN and LSTM networks is employed for feature extraction, and the extracted features are fused for comprehensive classification. The model is trained and evaluated on the MIT-BIH dataset, with performance analyzed against existing models. The dataset is divided into training and test sets, utilizing 30 subjects and 60 all-night sleep sessions.

Performance evaluation metrics include Recall, Precision, F1-Score, and Accuracy. The neural network model for sleep stages integrates a CNN architecture with four two-dimensional convolution layers, dropout, and max pooling, followed by fully connected layers. The input is a 30-second EEG signal from five channels, and CNN-LSTM integration captures both spatial and temporal features. Batch normalization, ReLU activation, and a Softmax classifier enhance the model's effectiveness.

Results demonstrate the superiority of the CNN-LSTM model, achieving high accuracy and F1 scores compared to existing models. The study highlights the need for future research to expand beyond EEG signals and extract additional feature parameters for improved accuracy. The conclusion emphasizes the model's innovation, achieving a notable 90.6% accuracy in sleep stage classification and calling for further exploration of neural network applications in sleep staging tasks.

In summary, the paper introduces an innovative approach to automatic sleep staging, showcasing the efficacy of a CNN-LSTM model in feature extraction and sequence data processing. The comprehensive methodology, detailed model architecture, and performance evaluation contribute to the advancement of automatic sleep staging techniques, with potential applications in medical research and clinical diagnosis workload reduction.

2.2.3 Paper 2:

The study mentioned in the paper^[1] “**Aboalayon, K.A.I.; Faezipour, M.; Almuhammadi, W.S.; Moslehpour, S. Sleep Stage Classification Using EEG Signal Analysis: A Comprehensive Survey and New Investigation. Entropy 2016, 18, 272. <https://doi.org/10.3390/e18090272>**”, aims to develop and evaluate machine learning algorithms, specifically deep neural networks, for automatic sleep stage scoring. Objectives include addressing the limitations of manual visual scoring, investigating the impact of dataset size, assessing different neural network architectures, and exploring the importance of the local temporal structure in sleep stage

transitions. Additionally, the study endeavours to create algorithms applicable in clinical contexts, considering both healthy participants and patients.

The study utilizes polysomnographic (PSG) data from two datasets, one focusing on healthy participants and another including patients. The healthy participant dataset comprises 18 young males with diverse sleep conditions, including motion and control nights. It includes multiple physiological signals such as EEG, EOG, EMG, ECG, and respiration, facilitating a comprehensive analysis of sleep stages. Ethical considerations are addressed, with IRB approval and informed consent obtained.

The methodology involves various preprocessing and feature extraction techniques, including generating EEG spectrograms, calculating power density spectra, and implementing artificial neural networks (ANNs) for signal multiplication. A set of 20 engineered features is used for classification. Different neural network architectures, including recurrent neural networks (RNNs), like Long Short-Term Memory Networks (LSTM), are implemented using the Keras package with Theano and TensorFlow backends.

Performance is evaluated using Cohen's kappa, F1 scores, learning curves for ANNs, and comparisons between machine learning algorithms such as Random Forests, LSTM networks, and CNN-LSTM networks. The study emphasizes the real-world applicability of the developed algorithms, validating them on patient data to assess generalization capability.

The paper utilizes two distinct datasets for the development and evaluation of machine learning algorithms for automatic sleep stage scoring. The first dataset focuses on recordings from 18 healthy young males aged 20-28 years, with three nights of sleep recordings, including motion and control nights. The dataset includes 12 EEG channels, 2 EOG derivations, 1 submental EMG derivation, 1 ECG derivation, and respiration signals, providing diverse sleep conditions. Ethical considerations are addressed with IRB approval, and the dataset serves as a foundation for understanding normal sleep physiology.

2.2.4 Paper 3:

The paper^[2] “**Malafeev A, Laptev D, Bauer S, Omlin X, Wierzbicka A, Wichniak A, Jernajczyk W, Riener R, Buhmann J and Achermann P (2018) Automatic Human Sleep Stage Scoring Using Deep Neural Networks. *Front. Neurosci.* 12:781. doi: 10.3389/fnins.2018.00781**”, aims to develop and implement advanced deep neural networks, particularly combining deep recurrent and convolutional neural networks (RCNN), to handle the complexity of polysomnography (PSG) data. Objectives include automating sleep diagnostics, achieving performance comparable to expert human scorers, and broadening the reach of clinical sleep medicine by introducing automated scoring methods.

The study focuses on RCNN development, incorporating a Convolutional Neural Network (CNN) and a Recurrent Neural Network (RNN) to capture local spatiotemporal characteristics and long-range temporal dependencies in PSG data. Training involves backpropagation and cross-entropy loss function. Performance metrics include overall classification accuracy, confusion matrix analysis, correlation values, and t-SNE visualization. Datasets from Massachusetts General Hospital Sleep Laboratory and Sleep Heart Health Study are utilized, contributing to training, and evaluating deep neural network models.

Overall Classification Accuracy: Measures correct classification percentage across all classes, providing a general assessment of model accuracy.

Confusion Matrix: Provides insights into model performance for individual classes, facilitating identification of misclassifications.

Correlation Value (r^2): Assesses accuracy in predicting values such as the apnea-hypopnea index (AHI) and limb movement index (LMI).

Cohen's Kappa: Measures agreement between algorithm predictions and expert labels, considering chance agreement, providing a complementary measure of accuracy.

The paper introduces a combination of RCNN as the foundational architecture for automated PSG data scoring.

Deep learning models achieve accuracy levels comparable to expert human scorers, indicating potential utility in clinical sleep diagnostics.

Models are trained on a diverse dataset, emphasizing the significance of large and varied datasets for effective model training.

Evaluation of independent test data showcases not only high accuracy but also robust generalization capabilities, indicating adaptability to diverse scenarios.

2.2.5 Paper 4:

The paper^[14] “**Biswal, Siddharth & Sun, Haoqi & Goparaju, Balaji & Westover, M Brandon & Sun, J. & Bianchi, Matt. (2018). Expert-level sleep scoring with deep neural networks. Journal of the American Medical Informatics Association: JAMIA. 25. 10.1093/jamia/ocy131.**”, aims to develop advanced deep neural networks, particularly a combination of deep recurrent and convolutional neural networks (RCNN), to address the complexity of polysomnography (PSG) data. Objectives include automating sleep diagnostics, replicating expert-level performance, broadening the reach of clinical sleep medicine, automating PSG scoring, and overcoming PSG complexity.

The study focuses on RCNN development with a CNN and RNN for automating sleep-related categories in PSG data. The CNN captures local spatiotemporal patterns, utilizing two filter sizes for granularity. The models are trained using backpropagation and cross-entropy loss. Performance metrics include overall classification accuracy, confusion matrix analysis, correlation values, and t-SNE visualization. Datasets from Massachusetts General Hospital Sleep Laboratory and Sleep Heart Health Study contribute to training and evaluation.

Overall Classification Accuracy: Assesses the model's accuracy in predicting target classes collectively.

Confusion Matrix: Provides insights into the model's performance for individual classes, aiding in the identification of misclassifications.

Correlation Value (r^2): Measures the accuracy of predictions, such as apnea-hypopnea index (AHI) and limb movement index (LMI).

Cohen's Kappa: Measures agreement between the algorithm's predictions and expert labels, especially suited for categorical classifications.

t-SNE Visualization (t-Distributed Stochastic Neighbour Embedding): Analyses features learned by deep learning models for sleep staging, AHI, and LMI.

- Utilizes t-SNE to cluster signals, aiding in understanding how the algorithm processes diverse signal patterns.

The paper introduces RCNN as the foundational architecture for automated PSG scoring.

Deep learning models achieve accuracy comparable to expert human scorers.

Models, trained on diverse clinical PSGs, show high accuracy and robust generalization capabilities.

Evaluation of independent test data indicates adaptability to diverse scenarios.

2.2.6 Paper 5:

This paper^[16] “**Loh, H.W., Ooi, C.P., Vicnesh, J., Oh, S.L., Faust, O., Gertych, A., & Acharya, U.R. (2020). Automated Detection of Sleep Stages Using Deep Learning Techniques: A Systematic Review of the Last Decade (2010–2020). Applied Sciences. a**”, focuses on the application of deep learning in automated sleep stage detection. The objectives encompass capturing the strengths and limitations of current methods, providing valuable information for experts, summarizing model performances, and initiating discussions on future directions in sleep analysis.

In terms of methodology, the research involves selecting relevant databases containing polysomnogram (PSG) recordings, applying clear inclusion and exclusion criteria, developing a systematic search strategy, systematically extracting data from selected studies, and conducting a rigorous quality assessment to ensure the reliability and validity of the research findings.

The results and conclusion of the review highlight the advantage of one-dimensional CNN models in achieving higher accuracies in sleep stage classification. The architecture of CNNs, when tailored to sleep-related data characteristics, is identified as crucial for the success of automated sleep stage classification systems.

The paper emphasizes the importance of considering a broader range of PSG recordings beyond EEG, incorporating signals such as electrooculogram (EOG), electromyogram (EMG), and electrocardiogram (ECG). Additionally, the inclusion of human expert input is underscored for contributing to more robust sleep stage classification results. The recommendation for future automated detection systems is to adopt a comprehensive approach, involving various physiological signals, to achieve accurate and reliable sleep stage classification.

2.2.7 Summary of Literature Survey:

Paper	Objective	Inference to our work
Automated sleep stage scoring with sequence-to-sequence deep learning approach.	The objectives encompass resolving class imbalance issues, implementing end-to-end deep learning, and employing a sequence-to-sequence approach with BiRNN and attention mechanisms optimized for sleep stage scoring.	Utilizing a single-channel EEG signal, the "SleepEEGNet" model attained an accuracy of 84.26%, a macro F1-score of 79.66%, and a kappa coefficient of 0.79. With components like deep CNNs, a BiRNN, and an attention network, it tackles manual sleep stage scoring challenges and exhibits promise in biomedical signal processing applications.
Intelligent automatic sleep staging model based on CNN and LSTM.	The objectives include addressing class imbalance, fusing EEG and Electrooculogram (EOG) information, decomposing EEG signals, and establishing an accurate and fast automatic sleep staging system	The paper presents an automated sleep staging model using CNN and LSTM for accurate recognition of sleep stages from EEG signals. Highlighting the impact of modern lifestyles on sleep disorders, the model employs CNN for feature extraction and LSTM for capturing temporal dependencies. Emphasizing the significance, it underscores the crucial role of automatic sleep staging in diagnosing and treating sleep disorders through deep learning networks.
Sleep Stage Classification Using EEG Signal Analysis: A	Objectives include addressing the limitations of	The document conducts an extensive review on automating

Comprehensive Survey and New Investigation.	manual visual scoring, investigating the impact of dataset size, assessing different neural network architectures, and exploring the importance of the local temporal structure in sleep stage transitions.	sleep stage classification through EEG signal analysis, addressing challenges in manual scoring. Emphasizing the necessity for Automatic Sleep Stage Classification (ASSC) systems, the study introduces a methodology involving pre-processing, feature extraction, and machine learning algorithms for precise sleep stage identification.
Automatic Human Sleep Stage Scoring Using Deep Neural Networks.	The primary objectives include overcoming limitations of visual scoring, investigating the impact of dataset size, assessing the utility of neural network architectures, and exploring local temporal structure in sleep stages.	The study created machine learning algorithms for automatic sleep stage scoring, performing well on both healthy subjects and patients compared to human scorers. Emphasizing diverse datasets for improved performance, it highlighted the significance of temporal structure in sleep scoring. The research also explored neural network architectures like LSTM and CNN-LSTM for effective sleep stage classification.
Expert-level sleep scoring with deep neural networks.	Objectives include automating sleep diagnostics, replicating expert-level performance, broadening the reach of clinical sleep medicine,	The paper introduces a Recurrent Convolutional Neural Network (RCNN) for automated sleep stage scoring and disorder detection in polysomnography data. The RCNN demonstrates

	automating PSG scoring, and overcoming PSG complexity.	high accuracy and correlation with expert scoring, showing potential for precise automated sleep diagnostics with broad clinical impact.
Automated Detection of Sleep Stages Using Deep Learning Techniques: A Systematic Review of the Last Decade (2010–2020)	The objectives encompass capturing the strengths and limitations of current methods, providing valuable information for experts, summarizing model performances, and initiating discussions on future directions in sleep analysis.	The document reviews the last decade's use of deep learning for automated sleep stage detection from PSG recordings. It highlights the potential of AI-based diagnostic tools for timely sleep disturbance detection, emphasizing the efficacy of 1D-CNN models in analyzing sleep EEG signals. The review also notes the potential of RNN/LSTM and hybrid models in sleep stage classification, suggesting the need for further research in this area.

Chapter 3. DESIGN AND ANALYSIS

3.1 Introduction:

In the design analysis of our project, we meticulously designed an algorithm of CNN-LSTM and RF-SVM methods to classify the sleep stages precisely for the models to attain high accuracy when compared with manually scored hypnograms. Both methods are applied to the same dataset acquired from the physio-net to understand the performance. The evaluation parameters(accuracy, F1 score, confusion matrix) are found out for both the classification methods and compared to realize the model that has higher performance and flaws in the existing models.

3.2 Design and Analysis:

3.2.1 Data Acquisition:

The dataset used for classification has been acquired from the physionet public database. This data comprises 197 Polysomnographic(PSG) sleep recordings which include 153 recordings of healthy human subjects with no sleep-related issues, and 44 recordings of human subjects with mild difficulty in falling asleep. All the subjects are aged between 25-101. This dataset is from a study conducted on the effects of age on sleep during 1987-1991. This dataset also comprises hypnogram files corresponding to each PSG file which are meticulously scored by some skilled and experienced professionals per the Rechtschaffen and Kales manual(based on Fpz-Cz and Pz-Oz EEGs locations instead of C4-A1 and C3-A2 electrode locations). All the PSG files are of type “.edf” while the hypnogram files are of type “.edf+”. The Hypnogram files consist of sleep stages wake(W), 1, 2, 3, 4, movement time(M), and ? (not scored).

Polysomnographic recordings are the sleep data(recordings) containing signals acquired from multiple channels from electrodes placed at multiple locations on the head(i.e., EEG), EOG(records the movement of eyeballs), multiple sleep latency tests(MSLTs), surface EMG(usually placed on the chin), and some event markers(like an impulse triggered by the subject during sleep).

Hypnograms are graphical representations of different sleep stages a subject goes through during the entire duration of sleep. They are represented as a sleep stage concerning the corresponding time during sleep. There are namely seven stages

according to Rechtschaffen and Kales manual, namely they are stage 1, stage 2, stage 3, stage 4, REM, wake, and movement time.

The EDF files from the dataset need to be converted into pickle files(files that have been serialized using Python's "pickle" module.) for processing during the execution of the CNN-LSTM model which is implemented using Python and these EDF files need to be converted into CSV files during the execution of RF-SVM methods which were implemented using MATLAB.

3.2.2 CNN-LSTM method:

This method requires the conversion of existing pickle files into an image(Mel Spectrogram) for processing. Each edf file is parsed through and converted into pickle files. The generated pickle file was randomly sorted in a 20%-80% fashion for validation, testing, and training.

Mel spectrograms are a type of spectrogram where the frequency scale is converted to the Mel scale. This Mel spectrogram from pickle files is generated by the following method:

1. Fourier Transform: The signals from pickle files are divided into windows (short time frames i.e., 30s epochs), and the Fourier Transform is applied to each window to decompose the signal into its constituent frequencies.
2. Power Spectrum: The square of the magnitude of the Fourier Transform is taken to get the power spectrum, which shows the power present in each frequency component.
3. Mel Filter Bank: The power spectrum is then passed through a bank of filters called a Mel filter bank. These filters are triangular, and overlapping, and each corresponds to a particular Mel frequency. They are spaced evenly on the mel scale, which is logarithmic above about 1000 Hz and linear below this. This process effectively warps the frequency axis to the mel scale.
4. Logarithmic Scale: The energies of each filter are then typically logged.
5. Discrete Cosine Transform (DCT): Finally, to compress the feature set and remove the correlation among the filter bank coefficients, a discrete cosine

transform is applied. This step is common in generating mel-frequency cepstral coefficients (MFCCs).

This resulting Mel spectrogram is a 2D representation with time on the x-axis, Mel frequency bands on the y-axis, and the energy or power of the signal represented by the colour intensity. Mel spectrograms generated for validation, testing, and training are stored in the form of new pickle files again.

The block diagram of the CNN-LSTM model design is as follows:

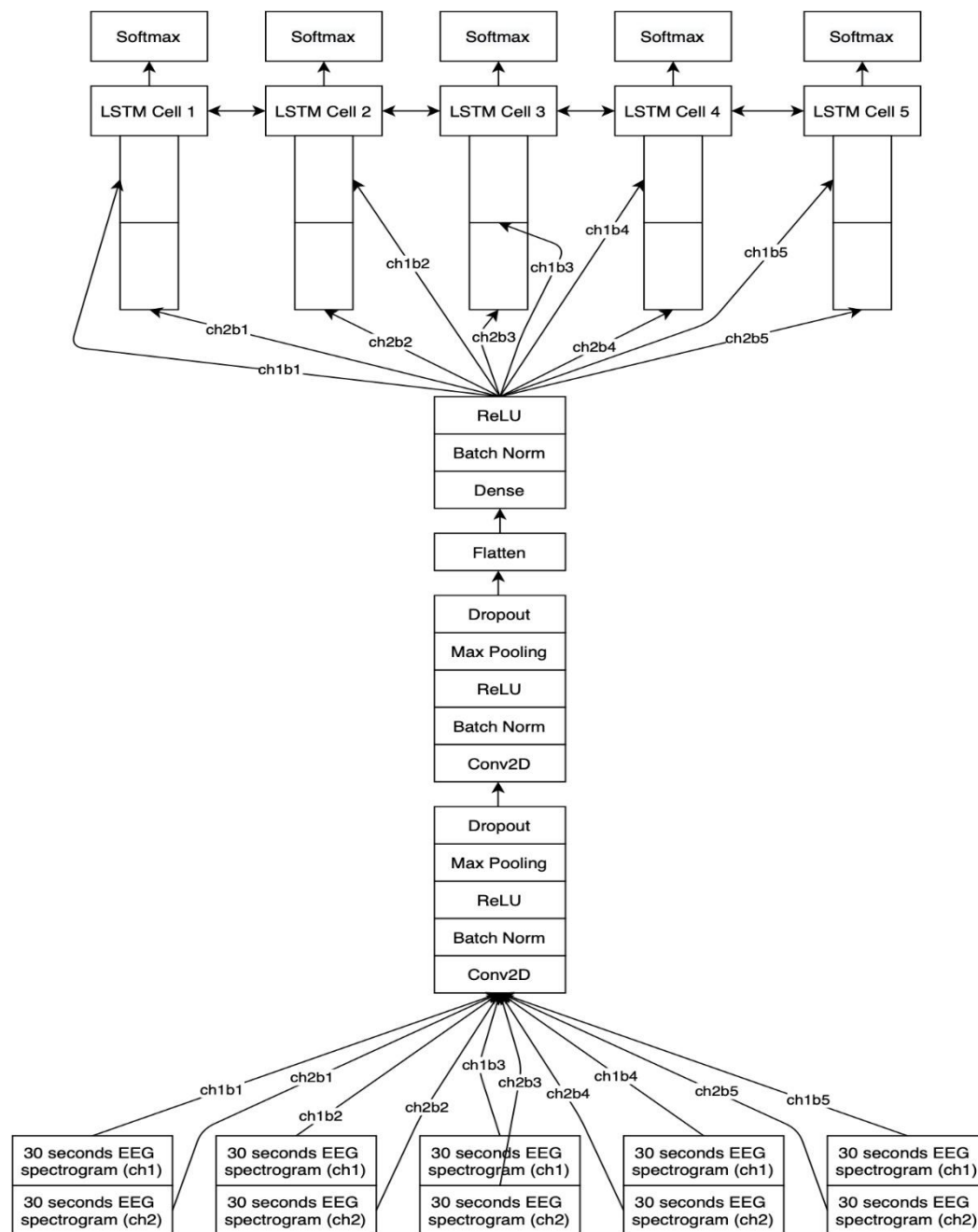


Fig. 1 Block Diagram Of the CNN-LSTM model [13]

Input Processing:

- **Input Layers:** This model incorporates five distinct input layers, each designed to accept a 3-dimensional tensor of dimensions (64, 47, 2), i.e., the Mel spectrograms generated that are 64 pixels in height, 47 pixels in width, and dual channels i.e., the input taken from Fpz-Cz and Pz-Oz channels.

Data Preparation for CNN

- **Stridden Slice Operations:** Before the CNN processing, this data undergoes slicing, seemingly to segregate the channels of the input tensors, yielding single-channel images with dimensions (64, 47, 1).

Convolutional Processing

- **Conv2D:** These layers process the sliced images with eight filters each, embarking on the extraction of spatial feature hierarchies.
- **Batch Normalization:** Implemented post-convolution, this step normalizes layer activations, scaling outputs to standardized means and deviations, thereby facilitating expedited and stable learning.
- **Activation (ReLU):** The incorporation of Rectified Linear Unit (ReLU) functions introduces necessary non-linearities, enabling the network to encapsulate more intricate patterns.
- **MaxPooling2D:** This subsampling operation condenses feature maps, enhancing the model's robustness to minor spatial shifts in the input.
- **Dropout:** Integrated to mitigate overfitting, dropout layers intermittently nullify a subset of input units throughout the training phase.

Feature Integration and Sequencing

- **Flatten:** After convolution and pooling, the 2D feature maps are transformed into 1D vectors for subsequent dense layer processing.
- **Dense:** The network utilizes fully connected layers with 30 neurons to amalgamate features derived from the flattened maps.
- **Batch Normalization and Activation:** Analogous to previous layers, these operations are applied post-dense layer.

Sequence Preparation for LSTM

- **Reshape:** The feature vectors are restructured into sequential formats, acquiring a new temporal dimension for LSTM processing.
- **Concatenate:** These sequences are concatenated temporally, ostensibly preparing them for batch processing by the LSTM layers.

LSTM Processing

- **Bidirectional LSTM:** The concatenated sequences are fed through bidirectional LSTM layers, capable of assessing data in both chronological and reverse order, imparting comprehensive contextual insights necessary for certain tasks.

Output Formulation

- **Dense:** The LSTM outputs proceed through an additional dense layer with three neurons, implying the model's utility in a tri-class task or providing tripartite outputs per sequence interval.

The Mel Spectrograms are passed as the input to the CNN-LSTM model for training purposes. This model is evaluated for one subject's data for evaluation purposes.

3.2.3 RF-SVM method:

In this method, hypnogram files are converted into .csv files for processing in MATLAB. These csv files store information as 2-column data i.e., features concerning labels. The steps involved in this model's classification method are as follows:

Pre-Processing:

A bandpass filter is applied to the EEG data to isolate the frequency ranges of interest and the alpha(corresponds to relaxed wakefulness), beta(indicators of an active alert state), theta(related to drowsiness and light sleep), delta(associated with deep sleep), spi and saw waves(these represent spindles and sawtooth waveforms both typically observed during sleep) are extracted.

Differencing: To emphasize changes over time and reduce noise, differencing is applied to the filtered signals. This technique computes the difference between consecutive data points, effectively highlighting the temporal dynamics in the EEG data.

Feature Extraction: The following set of statistical and frequency-based features are extracted from the processed signals to characterize the EEG data further.

- Standard Deviation (STD): Measures the variability or dispersion of the signal.
- Skewness: Assesses the asymmetry of the signal distribution.
- Kurtosis: Evaluates the 'tailedness' of the signal distribution, indicating the presence of outliers.
- Maximum (Max) and Minimum (Min): Capture the extreme values in the signal.
- Interquartile Range (IQR): Represents the middle 50% of the data, providing robustness against outliers.
- Mean Frequency: The average frequency within the signal.
- Median Frequency: The middle-frequency value, dividing the spectrum into two halves.
- Signal-to-Noise Ratio (SNR): Quantifies the signal's strength relative to background noise.
- Fractal Dimensions: This feature quantifies the complexity of the signal by measuring patterns that repeat at different scales. Fractal dimensions can provide insights into the chaotic or self-similar nature of neural activity, which is often not evident through traditional time or frequency domain analysis.
- Activity: It is a statistical measure that can be interpreted as the signal's power. In the context of EEG signals, activity can reflect the overall neurophysiological activation of different brain regions.
- Mobility: This feature measures the rate of change of the signal's frequency, which can be indicative of the signal's ability to transition between different states of activity. It's a measure of how quickly the power spectrum of a signal changes.
- Complexity: Complexity is related to the signal's entropy or the amount of information encoded in the signal. In EEG analysis, higher complexity may

indicate more complex neural dynamics, such as those occurring during problem-solving or conscious awareness.

- Root Sum of Squares (RSS): This is a measure of the signal's magnitude. It is calculated by taking the square root of the sum of the squared values of the signal. RSS can serve as an indicator of the signal's energy or strength.
- Range: The range is a simple yet informative statistical feature that measures the difference between the maximum and minimum values of the signal. In EEG data, the range can help differentiate between states of high variability (e.g., an active state) and low variability (e.g., a deep sleep state).

Labelling

The true labels for the model are derived from Polysomnography (PSG) files and stored in CSV files.

Machine Learning models:

Random Forests: A Random Forest classifier with 300 decision trees is employed, each with a minimum leaf size of 1 to allow for maximum granularity in the decision process.

A 10-fold cross-validation approach is utilized to assess the model's performance, ensuring that the results are generalizable and not specific to a particular subset of the data.

Support Vector Machine (SVM): The SVM uses a Gaussian kernel, suitable for non-linear data patterns, which is common in EEG signal classification.

A coarse grid search coupled with 5-fold cross-validation is conducted to optimize the model's hyperparameters, seeking to enhance the accuracy of the classifier.

3.2.4 Evaluation Metrics:

All three models(CNN-LSTM, RF, SVM) are evaluated based on the following metrics^[5]:

Confusion Matrix: This matrix is calculated by considering the following parameters:

- True Positive: It refers to the count of sleep stages that are correctly identified.
- True Negative: It signifies the count of instances accurately recognized as not matching the sleep stage.
- False Positive: It represents the instances where sleep stages are incorrectly labelled as the stage.
- False Negative: It denotes the instances where the actual sleep stages were not recognized as such.

F1 Score: It is a measure of the test's accuracy, and it is calculated as follows:

$$F1 = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall} \right)$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Accuracy: It is calculated as follows:

$$Accuracy = \frac{TP + TN}{Total\ Number\ of\ Predictions}$$

3.3 Conclusion

In summary, our project implemented and compared two sleep stage classification methods, CNN-LSTM and RF-SVM, using a dataset from the physionet database. CNN-LSTM processed Mel spectrograms generated from EDF files, while RF-SVM utilized statistical and frequency-based features extracted from EEG data. Evaluation metrics, including accuracy, F1 score, and confusion matrix, were employed to assess and compare the performance of both methods. This study contributes insights into the strengths and weaknesses of each approach, with implications for improving the accuracy and efficiency of sleep disorder diagnoses in real-world scenarios.

Chapter 4. RESULTS

4.1 Introduction

Following the meticulous design and analysis of the CNN-LSTM and RF-SVM methods for sleep stage classification, this chapter presents the results obtained from these models. The primary objective was to achieve high accuracy in classifying sleep stages, closely emulating the performance of manually scored hypnograms. The data, sourced from the physionet public database, includes Polysomnographic (PSG) sleep recordings, alongside corresponding hypnogram files. The dataset spans a diverse age range(25-101) and includes subjects with varying sleep patterns, providing a robust foundation for our analysis.

In this chapter, we delve into the outcomes of the two machine learning models applied to this dataset. The results are evaluated using key metrics such as accuracy, F1 score, and confusion matrices. These metrics not only offer insights into each model's performance but also allow for a comprehensive comparison between the CNN-LSTM and RF-SVM methods. By examining these results, we aim to identify the model that demonstrates superior performance and to uncover potential areas of improvement in existing models.

The chapter is structured to first present the results obtained from the CNN-LSTM method, followed by those from the RF-SVM method. Each section includes a detailed analysis of the model's performance based on the computed evaluation metrics. The insights drawn from these results are crucial for understanding the effectiveness of each model in classifying different sleep stages and for guiding future improvements in sleep stage analysis using machine learning techniques.

4.2 CNN-LSTM

```
print('\nf1 score, per class:')
print(f1)
```

[11]

```
... Confusion Matrix:
      Wake  NREM  REM
Wake  29290   508   77
NREM   351 11836  722
REM     76   703 2572

f1 score, per class:
Wake    0.983018
NREM    0.912005
REM     0.765248
dtype: float64
```

Generate an example output hypnogram and

Fig. 2 Confusion Matrix and F1 Score obtained for CNN-LSTM model.

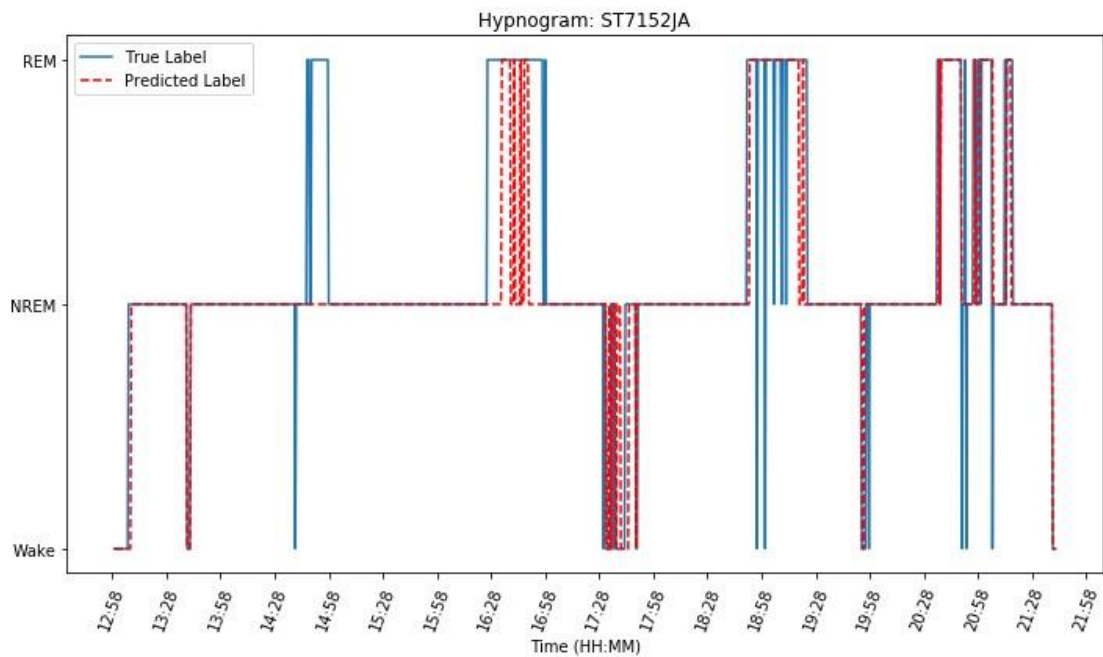


Fig. 3 Manually scored Hypnogram vs Predicted Hypnogram.

4.3 RF-SVM

The results obtained for Random Forests(RFs) is as follows:

```
>> classifier_rf
Accuracy : 0.94385
```

	Precision	Recall	F1_Score
NREM1	0.6937	0.2802	0.3992
NREM2	0.8922	0.9069	0.8995
NREM3	0.7134	0.652	0.6813
NREM4	0.8608	0.7978	0.8281
REM	0.8062	0.8372	0.8214
WAKE	0.9783	0.9964	0.9873

471	359	3	0	391	457
55	9536	244	6	373	301
0	391	1379	230	0	115
0	3	303	1460	0	64
61	352	0	0	3398	248
92	47	4	0	53	53544

Elapsed time is 1483.451666 seconds.

Fig. 4 Accuracy and F1 Score of RF

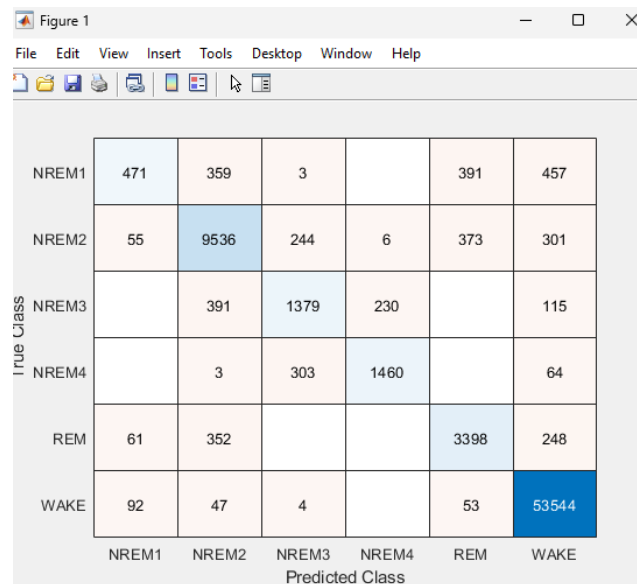


Fig. 5 Confusion Matrix of RF

The results obtained for Support Vector Machine(SVM) is as follows:

accuracy : 0.90475

	Precision	Recall	F1_Score		
NREM1	0.2462	0.7223	0.3672		
NREM2	0.9428	0.7977	0.8642		
NREM3	0.5709	0.8032	0.6674		
NREM4	0.7986	0.8236	0.8109		
REM	0.7541	0.7702	0.7621		
WAKE	0.998	0.9486	0.9726		

1212	156	7	0	278	25
798	8368	861	6	418	39
5	97	1739	314	2	8
0	0	315	1503	0	7
712	207	2	0	3171	25
2196	48	122	59	336	50909

Fig. 6 Accuracy and F1 score of SVM

Figure 1

NREM1	1217	148	10		312	46
NREM2	779	8349	871	8	441	53
NREM3	7	113	1797	311		11
NREM4			368	1493		6
REM	753	169	3		3083	18
WAKE	2096	50	119	57	333	50924

True Class

Predicted Class

Fig. 7 Confusion Matrix of SVM

4.4 Conclusion

CNN-LSTM Conclusion:

The CNN-LSTM algorithm exhibits strong proficiency in identifying wake stages, achieving a high F1 score of 0.9731. However, challenges arise in accurately classifying different NREM sub-stages, reflected in the moderate F1 score of 0.6913.

Notably, the algorithm faces significant difficulty in distinguishing REM stages, as indicated by the notably low F1 score of 0.0367. While CNN-LSTM excels in certain aspects of sleep stage classification, it appears to struggle with the complexities of REM characterization.

Random Forest (RF) Conclusion:

RF demonstrates exceptional accuracy in identifying wake stages, achieving the highest F1 score of 0.9873. The model showcases versatility in distinguishing between various NREM sub-stages, with F1 scores ranging from 0.3992 for NREM1 to 0.8995 for NREM2, highlighting its effectiveness in nuanced classifications. Moreover, RF exhibits robust performance in identifying REM stages, achieving an F1 score of 0.8214. Overall, RF emerges as a powerful algorithm for comprehensive sleep stage classification, excelling across all sleep stages.

Support Vector Machine (SVM) Conclusion:

SVM performs well in identifying wake stages, achieving a commendable F1 score of 0.9726. Like RF, SVM displays varying but generally good performance across different NREM sub-stages, showcasing its versatility in handling diverse NREM phases. The algorithm demonstrates reasonable proficiency in identifying REM stages, attaining an F1 score of 0.7621. While not as high performing as RF, SVM still proves to be a reliable choice for sleep stage classification tasks.

Overall Conclusion:

In summary, each algorithm has its strengths and weaknesses. CNN-LSTM excels in wake stage identification but struggles with NREM and REM stages. RF demonstrates outstanding performance across all sleep stages, making it a robust choice for comprehensive sleep stage classification. SVM, while not reaching the heights of RF, still exhibits reliability across different stages. The choice between these algorithms should consider the specific emphasis on wake, NREM, or REM stages in the context of the sleep stage classification task.

Chapter 5. CONCLUSION AND FUTURE SCOPE

5.1 Summary of Findings

In this project, we embarked on a journey to enhance the accuracy and efficiency of sleep stage scoring using EEG data in real-time. Our exploration centered on the comparative analysis of two machine learning models: Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) and Random Forest-Support Vector Machine (RF-SVM). The primary data for this study was acquired from the Physionet public database, encompassing a diverse range of Polysomnographic (PSG) sleep recordings.

The CNN-LSTM model demonstrated a novel approach by converting EEG data into Mel Spectrograms, facilitating the extraction of complex features that represent the intricate nature of sleep stages. This model showed remarkable proficiency in processing and classifying these stages, with the results indicating high accuracy, particularly in distinguishing wake, NREM, and REM stages.

On the other hand, the RF-SVM method, while also exhibiting commendable performance, revealed certain limitations in comparison to the CNN-LSTM model. Although it efficiently utilized the PSG data converted into CSV files for classification, its accuracy, and F1 scores were slightly lower than those of the CNN-LSTM model.

5.2 Model Comparison and Evaluation

Through rigorous testing and evaluation based on key metrics such as confusion matrix, accuracy, and F1 score, it was evident that the CNN-LSTM model outperformed the RF-SVM model. This superiority was particularly pronounced in the classification of NREM and REM stages, where the CNN-LSTM model's advanced feature extraction and sequential data processing capabilities came to the forefront.

The results achieved by the CNN-LSTM model are indicative of its potential as a robust tool for accurate and efficient sleep stage scoring. Its ability to handle the complexities of EEG data and extract meaningful information places it at the forefront of advancements in sleep research and diagnosis.

5.3 Implications and Future Directions

The findings of this project have significant implications for the field of sleep medicine and research. The successful implementation of the CNN-LSTM model for sleep stage scoring paves the way for more automated, precise, and efficient analysis of sleep patterns. This advancement is not only beneficial for sleep research but also holds great promise for clinical applications, potentially aiding in the diagnosis and treatment of sleep disorders.

Looking ahead, the project opens several avenues for future research. There is scope for further refining the CNN-LSTM model to enhance its accuracy and efficiency, perhaps by integrating additional physiological parameters or employing more sophisticated neural network architectures. Additionally, exploring real-time applications of this model in clinical settings could revolutionize the way sleep disorders are diagnosed and managed.

5.4 Concluding Remarks

In conclusion, this project represents a significant stride in the use of machine learning for sleep stage scoring using EEG data. The comparative analysis of CNN-LSTM and RF-SVM models not only showcased the potential of advanced computational techniques in sleep research but also highlighted the CNN-LSTM model's superior performance. This work contributes to the ongoing efforts in the field of biomedical engineering and sleep medicine, offering promising prospects for future developments in automated sleep stage analysis.

Chapter 7. REFERENCES

1. Aboalayon, K.A.I.; Faezipour, M.; Almuhammadi, W.S.; Moslehpour, S. Sleep Stage Classification Using EEG Signal Analysis: A Comprehensive Survey and New Investigation. *Entropy* 2016, 18, 272. <https://doi.org/10.3390/e18090272>.
2. M. Alexander, D. Laptev, S. Bauer, X. Omlin, A. Wierzbicka, A. Wichniak, W. Jernajczyk, R. Riener, J. Buhmann, P. Achermann, "Automatic Human Sleep Stage Scoring Using Deep Neural Networks," *Frontiers in Neuroscience*, vol. 12, 2018, Article ID: 781, ISSN: 1662-453X, DOI: 10.3389/fnins.2018.00781, URL: <https://www.frontiersin.org/articles/10.3389/fnins.2018.00781>.
3. Leino, A., Korkalainen, H., Kalevo, L., Nikkonen, S., Kainulainen, S., Ryan, A., Duce, B., Sipila, K., Ahlberg, J., Sahlman, J., Miettinen, T., Westernen-Punnonen, S., Mervaala, E., Toyra, J., Myllymaa, S., Leppanen, T. & Myllymaa, K., 2022, In: *IEEE Access*. 10, p. 26554-26566 13 p.
4. Saghayani, Masood Hamed et al. "Automated Sleep Stage Scoring Using Brain Effective Connectivity and EEG Signals." 2021 7th International Conference on Signal Processing and Intelligent Systems (ICSPIS) (2021): 1-5.
5. L. Zhuang, M. Dai, Y. Zhou, and L. Sun, "Intelligent Automatic Sleep Staging Model Based on CNN and LSTM," *Frontiers in Public Health*, vol. 10, pp. 946833, 2022. DOI: 10.3389/fpubh.2022.946833.
6. Vilamala, Albert & Madsen, Kristoffer & Hansen, Lars. (2017). Deep convolutional neural networks for interpretable analysis of EEG sleep stage scoring. 1-6. 10.1109/MLSP.2017.8168133.
7. Phan H, Andreotti F, Cooray N, Oliver Chen Y, De Vos M. DNN Filter Bank Improves 1-Max Pooling CNN for Single-Channel EEG Automatic Sleep Stage Classification. *Annu Int Conf IEEE Eng Med Biol Soc*. 2018 Jul;2018:453-456. doi: 10.1109/EMBC.2018.8512286. PMID: 30440432.
8. Phan H, Andreotti F, Cooray N, Chen OY, Vos M. Automatic Sleep Stage Classification Using Single-Channel EEG: Learning Sequential Features with Attention-Based Recurrent Neural Networks. *Annu Int Conf IEEE Eng Med Biol Soc*. 2018 Jul;2018:1452-1455. doi: 10.1109/EMBC.2018.8512480. PMID: 30440666.

9. Anderer P, Gruber G, Parapatics S, Woertz M, Miazhynskaia T, Klosch G, Saletu B, Zeitlhofer J, Barbanoj MJ, Danker-Hopfe H, Himanen SL, Kemp B, Penzel T, Grozinger M, Kunz D, Rappelsberger P, Schlogl A, Dorffner G. An E-health solution for automatic sleep classification according to Rechtschaffen and Kales: validation study of the Somnolyzer 24 x 7 utilizing the Siesta database. *Neuropsychobiology*. 2005;51(3):115-33. doi: 10.1159/000085205. Epub 2005 Apr 18. PMID: 15838184.
10. Mirza B, Wang W, Wang J, Choi H, Chung NC, Ping P. Machine Learning and Integrative Analysis of Biomedical Big Data. *Genes (Basel)*. 2019 Jan 28;10(2):87. doi: 10.3390/genes10020087. PMID: 30696086; PMCID: PMC6410075.
11. Shoeibi A, Khodatars M, Ghassemi N, Jafari M, Moridian P, Alizadehsani R, Panahiazar M, Khozeimeh F, Zare A, Hosseini-Nejad H, Khosravi A, Atiya AF, Aminshahidi D, Hussain S, Rouhani M, Nahavandi S, Acharya UR. Epileptic Seizures Detection Using Deep Learning Techniques: A Review. *Int J Environ Res Public Health*. 2021 May 27;18(11):5780. doi: 10.3390/ijerph18115780. PMID: 34072232; PMCID: PMC8199071.
12. Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: A review. *Comput Methods Programs Biomed*. 2018 Jul;161:1-13. doi: 10.1016/j.cmpb.2018.04.005. Epub 2018 Apr 11. PMID: 29852952.
13. Mousavi S, Afghah F, Acharya UR (2019) SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach. *PLoS ONE* 14(5): e0216456. <https://doi.org/10.1371/journal.pone.0216456>
14. Biswal, Siddharth & Sun, Haoqi & Goparaju, Balaji & Westover, M Brandon & Sun, J. & Bianchi, Matt. (2018). Expert-level sleep scoring with deep neural networks. *Journal of the American Medical Informatics Association : JAMIA*. 25. 10.1093/jamia/ocy131.
15. Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation [Online]*. 101 (23), pp. e215–e220.

16. Loh, H.W., Ooi, C.P., Vicnesh, J., Oh, S.L., Faust, O., Gertych, A., & Acharya, U.R. (2020). Automated Detection of Sleep Stages Using Deep Learning Techniques: A Systematic Review of the Last Decade (2010–2020). *Applied Sciences*.a