EXPLAINABLE AUTOMATED SLEEP SCORING SYSTEM

MINI PROJECT

By

Bhavi Dave	**60004180013**
Nirali Parekh	**60004180065**
Raj Shah	**60004180076**

Guide:

Kriti Srivastava

Asst. Professor

Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)

University of Mumbai

2020-2021

CERTIFICATE

This is to certify that the mini project entitled **“Explainable Automated Sleep Scoring System”** is a bonafide work of **“Bhavi Dave (60004180013) , Nirali Parekh (60004180065) , Raj Shah (60004180076) ”** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of B.E. in Computer Engineering

(Name and sign)

Prof. Kriti Srivastava

Guide

(Name and sign)

Head of Department

(Name and sign)

Principal

Mini Project Report Approval.

This mini project report entitled **Explainable Automated Sleep Scoring System** by **Bhavi Dave (60004180013) , Nirali Parekh (60004180065) , Raj Shah (60004180076)** is approved for the partial fulfillment of the degree of ***B.E. in Computer Engineering.***

Examiners

1.____________________________

2.____________________________

Date:

Place:

Declaration

I/We declare that this written submission represents my/our ideas in my/our own words and where others' ideas or words have been included, I/We have adequately cited and referenced the original sources. I/We also declare that I/We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my/our submission. I/We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature)

(Name of students and Roll Nos.)

Bhavi Dave (60004180013)

Nirali Parekh (60004180065)

Raj Shah (60004180076)

Date:

Abstract

Although Deep Learning approaches generate State Of The Art performance across several domains, the approach often lacks Interpretability, hindering adoption beyond academia into high stakes situations like Health Care and medicine. The potential for human scrutiny, interpretation and verification can make the model robust and reliable, increasing confidence. The study of sleep-Polysomnography- uses Electroencephalogram (EEG) readings, among other parameters, to gain a clearer picture of a patient's sleep patterns since different brain activities correspond to different stages of sleep. Monitoring and interpreting EEG signals and the body's reactions to the changes in these cycles can help identify disruptions in sleep patterns, which can in turn help with the medical prognosis of several pervasive diseases like Sleep Apnea and the predilection for seizures. For simultaneously addressing the pitfalls associated with the traditional manual review of EEG signals for disease prognosis and concerns like ethics, reliability and interpretability that arise when automation is introduced in the realm of medicine, we propose the creation of a novel XAI architecture. Using data from raw, single channel EEG signals, the project classifies the five sleep stages (Wake, N1, N2, N3, N4 and REM) and demonstrates how the the classification aids in detecting prominent sleep related diseases by detecting Sleep Apnea and Seizures with the proposed architecture. The purpose of this study is to propose a novel and Explainable architecture for both sleep stage classification and malady detection that emulates the accuracy of traditional DL models while incorporating interpretability. As a first stage towards achieving the goal, we have implemented a CNN to extract time-invariant features and learn stage transition rules among sleep stages from EEG epochs. The model successfully classifies the five stages with a 96% accuracy as well as precision, recall and f1 score of 0.96.

Contents

List Of Figures

List of Abbreviations

Sr. No.	Abbreviation	Expanded form
i	EEG	Electroencephalogram - a test that detects electrical activity in your brain
ii	EOG	Electrooculography
iii	LSTM	Long Short-Term Memory
iv	BMA	Bayesian Model Averaging
v	DNN	Deep Neural Networks
vi	UML	Use Case Diagram
vii	EDF	European Data Format
viii	CNN	Convolution Neural Networks
ix	RNN	Recurrent Neural Networks

Chapter 1 : Introduction

1.1 Background

Artificial intelligence (AI) is already delivering on making aspects of health care more efficient, with the potential to support clinical and other applications that result in more insightful and effective care and operations. AI-based solutions can effectively streamline diagnostic and treatment processes by using large amounts of structured and unstructured medical data across institutions. This can aid physicians at hospital and health systems in clinical decision-making by providing them with real-time, data-driven insights that they can alter and implement based on their personal expertise. Deep Learning approaches have generated State Of The Art performance across several domains revolutionising the AI technological landscape. As deep learning continues to be adopted, the prominence of assistance and automated decisions made by neural networks in high stake situations is an undeniable fact that stakeholders and academicians have to grapple with. Centered on the Universal Approximation Theorem, the domain often lacks Interpretability.

Even within the domain of HealthTech, the specific niche of analysis that relates the physiological effects of sleep on human health and well being is a contentious and pervasive research domain. In the inaugural issue of the *Journal of Clinical Sleep Medicine* (2005), a feature article titled *History of the development of sleep medicine in the United States(* Shepard JW, Buysse DJ, Chesson AL, et al.,*J Clin Sleep Med.* 2005) [14] traced early milestones in the developing field of sleep medicine, which slowly emerged from the older field of sleep research during the 1970s and 1980s. Sleep medicine, the article noted, was closely linked with and made possible by the discovery of electrical activity in the brain. As connections between sleep disruption and both disease and mortality have become more firmly established in *The Extraordinary Importance of Sleep: The Detrimental Effects of Inadequate Sleep on Health and Public Safety Drive an Explosion of Sleep Research* (Susan L. Worley, 2018 Dec)[15], accurate and efficient diagnosis and management of sleep disorders have become increasingly critical.

In general, much information can be extrapolated from the polysomnography; some can be directly related to the sleep, such as the sleep onset latency (SOL), the REM-sleep onset latency, the number of awakenings during the sleep-period, the total sleep duration, percentages and durations of every sleep stage, and the number of arousals. A sleep study registers a body's shifts between the stages of sleep, which are rapid eye movement (REM) sleep, and non-rapid eye movement (non-REM) sleep. Non-REM sleep is divided into "light sleep" and "deep sleep" phases.During REM sleep, brain activity is high, but only eyes and breathing muscles are active. This is the stage in which one dreams. Non-REM sleep involves slower brain activity.A person without a sleep disorder will switch between non-REM and REM sleep, experiencing multiple sleep cycles per night. Observing sleep cycles, along with the body's reactions to the changes in these cycles, can help identify disruptions in sleep patterns.

One of the most widely used tests for evaluation of seizures and Sleep Apnea is the EEG. Neurologists who are trained in reading EEGs know the normal patterns of brain waves that occur during sleep and wakefulness and Variations in these patterns can show a region that is damaged , and is consequently potentially as a source of seizures. Sleep during an EEG allows a more complete evaluation of brain activity while also increasing the chances that an abnormality will be seen if present. Sleep apnea brings about EEG arousals, and sleep for patients with sleep apnea syndrome (SAS) is thus frequently interrupted. The number of respiratory-related arousals during the whole night on PSG recordings is directly related to the quality of sleep. Patients suspected of having the aforementioned maladies often have an EEG that includes sleep, particularly, if a waking recording was normal. Therefore an automated, accurate, interpretable and robust model can significantly help in disease detection.

1.2 Motivation

One of the major challenges that clinicians face during the initial assessment of people with sleep disorders is the process of identifying and sorting out comorbidities. Untangling the causes and effects in bidirectional comorbidities can be particularly difficult. For example, insomnia—by far the most common sleep disorder—often is complicated by the presence of another sleep disorder, such as sleep apnea or restless legs syndrome. Comorbidities can complicate treatment and often require sleep specialists to collaborate with not only primary care physicians but also specialists in other therapeutic areas. These risks underscore the need to improve methods for identifying and properly diagnosing sleep disorders. Understanding the sleep stages, how one should cycle through them and the necessity of achieving healthy sleep hours, is important information for medical professionals to understand as we strive to become more knowledgeable about our own sleep health.

Additionally, sleep stage scoring using PSG is typically performed in a hospital setting that requires the subject to remain on a waiting list for an unspecified period of time. In addition to the cost and complexity of measuring sleep patterns, patients must stay overnight in a specially equipped sleep laboratory with adhesive electrodes and wires attached to the head. Because sleeping in an unfamiliar environment is uncomfortable, these procedures can also affect the patient's sleep efficiency. In *Real Time Sleep Detection System Using New Statistical Features of the Single EEG Channel* (Khald Ali I. et Al)[16] apart from the several logistical issues highlighted, technical Scoring Limitations pertaining to the field can also significantly hinder manual EEG Processing - [17] like Wake - Sleep unreliable (when the clarity of the EEG makes distinguishing the transition), REM/non-REM unreliable (when identification of Stage REM is unreliable (usually due to poor or missing EMG or when both EOGs are absent), Arousals unreliable (when the technical quality of the study does not allow one to distinguish discrete increases in EEG frequency from background changes in EEG. EEG still may be of sufficient quality to score stages), and Arousals in REM unreliable (when EMG is artifactual or absent during all or REM portion of the study) Data Loss is a particularly pernicious issue that occurs when the Recording ended before participant awoke - the last epoch of the study is any stage of sleep; or when an arousal is seen in the last few epochs of the study and there is a question if the participant actually awoke or would have returned to sleep (i.e., lack of sustained activity indicating "out of bed."). This can make a conclusive medical prognosis especially difficult.

As investments in the AI- HealthTech domain increase and AI-powered solutions become more widespread in health care settings, the industry would have to address the new set of challenges both from the data used—including cyber threats—and the potential for bias in the AI algorithms. The strategy should comply with regulations—including to assure patient privacy and other HIPAA requirements. AI algorithms present risks such as variability of output in patient diagnosis and treatment, data bias, and traditional IT risks such as change management. Trained models are usually black boxes enveloped in complexity that is deployed but not deciphered. Consequently, even accurate models sometimes fail to capture the correct reasoning in the decision process. It is this vulnerability that is exploited when engineered noises, indetectable to the model's human counterparts, can cause it to predict the wrong class with a fatally high confidence. Even when the causal relationship is correctly modeled , an architecture that is interpretable will aid data understanding while making the decision making process more transparent. The potential for human scrutiny and verification can make the model robust, increasing confidence. As the scientific community comes closer to emulating human intelligence , severe ethical conundrums are also seen at the forefront. A through understanding of model bias, accountability and security will be critical to the adoption of AI at scale. The aptitude of an AI system has to move beyond accuracy into the realm of explanation.

Chapter 2 : Review of Literature

2.1 Literature Review

In the field of neurostimulation and EEG data, considerable research has been done on the problem of seizure detection. *An explainable model for EEG seizure detection based on connectivity features* (Mohammad Mansour et al)[1] classifies whether a particular data window belongs to a seizure or not. In this work, they design an optimal deep learning model using attention, CNN, BiLSTM, and fully connected layers. The authors of this paper also computed the relevance using the weights of the learned model based on the activation values of the receptive fields at a particular layer, and were able to explain the relevance of each feature across all patients. The authors tried to achieve their explainability in the post-modeling stage using the input based explanation drivers methods where they base their feature study on the output predictions.

The potential value of XAI to the field of neurostimulation is discussed in *Explainable Artificial Intelligence for Neuroscience: Behavioral Neurostimulation* (Jean-Marc Fellous ,Front. Neurosci., 13 December 2019) [2]. In the article, the authors have suggested various XAI methods like LIME (Locally Interpretable Model-Agnostic Explanations) and Contextual Explanation Networks. In the literature survey titled *Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)*(Amina Adadi and Mohammed Berrada , IEEE Xplore ,September 17, 2018) [3] , an organised overview of the existing XAI approaches. The authors have reviewed and analysed XAI Methods like LRP, Feature Importance, LIME, Decision Trees based on Taxonomy.

Various review works have been focussed on the use of EEG waves in the problem of sleep scoring techniques. In *Deep learning-based electroencephalography analysis: a systematic review* (Yannick Roy et al.,14 August 2019 ,Journal Of Neural Engineering) [4] have reviewed 154 papers that apply DL to EEG, published between January 2010 and July 2018, and spanning different application domains such as epilepsy, sleep, brain–computer interfacing, and cognitive and affective monitoring. The review also provides detailed methodological information on the various components of a DL-EEG pipeline to inform their own implementation. In *A review of automated sleep stage scoring based on physiological signals for the new millennia*(Faust et al., April 2019)[9], have provided a comprehensive review of automated sleep stage scoring systems, since the year 2000. They analyse the system that were developed for Electrocardiogram (ECG), Electroencephalogram (EEG), Electrooculogram (EOG), and a combination of signals.

In *SLEEPNET: Automated Sleep Staging System via Deep Learning* **(**Siddharth Biswa et Al.) [6] have proposed a deployed annotation tool for sleep staging. SleepNet uses a deep recurrent neural network trained on the largest sleep physiology database assembled to date, consisting of PSGs from over 10,000 patients from the Massachusetts General Hospital (MGH) Sleep Laboratory.The best performing instance of SleepNet uses expert-defined features to represent each 30-sec interval and learns to annotate EEG using a recurrent neural network (RNN).Their model analyses and compares various algorithms like Logistic Regression, Tree Boosting, Multilayer Perceptron, CNN, RNN and RCNN.

In *Mixed Neural Network Approach for Temporal Sleep Stage Classification*(Hao Dong et Al.)[7] have proposed use of a Mixed Neural Network (MNN) to solve both the population heterogeneity and temporal pattern recognition problems. Their MNN is composed of a rectifier neural network which is suitable for detecting naturally sparse patterns, and a long short-term memory (LSTM) for detection of temporally sequential patterns.The proposed Mixed Neural Network and the corresponding training method work well for sleep stages classification problem compared with SVM, RF and MLP.

SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach **(**Sajad Mousavi et Al. ,2019 May 7)[8] have proposed an automatic sleep stage annotation method called SleepEEGNet using a single-channel EEG signal . The authors have applied sequence of deep learning models - CNN for feature extraction, BiRNNs for capturing temporal inf and an attention network to let the model learn the most relevant parts of the input sequence while training .The authors used the synthetic minority over-sampling technique (SMOTE) to tackle class imbalance problems.

A deep learning approach to detect sleep stages (Klara Stuburić et al. , Procedia Computer Science)[10] have presented the implementation of deep learning methods for sleep stage detection by using three signals: heartbeat signal, respiratory signal, and movement signal.

They employ two different neural networks for this classification problem: the convolutional neural network (CNN) and the long-short term memory network (LSTM). However their model performs poorly giving only an accuracy of 40% and F1 score of 37% in classification of the 5 stages of sleep.

SeqSleepNet: End-to-End Hierarchical Recurrent Neural Network for Sequence-to-Sequence Automatic Sleep Staging (Huy Phan et al., 2019 Jan 31) [11] has proposed a hierarchical recurrent neural network. The author treated automatic sleep staging as a sequence-to-sequence classification problem to jointly classify a sequence of multiple epochs at once. They achieve an overall accuracy and F1 score of 87.1% and 83.3% respectively.

Towards More Accurate Automatic Sleep Staging via Deep Transfer Learning (Huy Phan at al) [12] presented a deep transfer learning approach to address the problem of insufficient data in many sleep studies and to improve automatic sleep staging performance on small cohorts. They adopted the MASS dataset as the source domain and three different sleep databases are used as the target domains.

In their paper, *Intra- and Inter-epoch Temporal Context Network (IITNet) Using Sub-epoch Features for Automatic Sleep Scoring on Raw Single-channel EEG* that proposed a novel deep learning model (Hogeon Seo et al.)[13] considered the inter- and intra-epoch temporal contexts using raw single-channel EEG. They model a deep CNN based on a modified ResNet-50 extracts the sleep-related features and the RNN via two-layered BiLSTM learns the transition rules.Their results show that the proposed temporal context learning at both the intra- and inter-epoch levels is effective to classify the time-series inputs.

Table 2.1 : Literature Review Analysis

Ref	Task	Dataset used	Feature type	Deep Learning Architecture	Accuracy	XAI used?
[11]	Sleep Scoring	Montreal archive of sleep studies (MASS) open access dataset	Raw signals	SeqSleepNet. Hierarchical RNN	87.1%	No
[12]	Sleep Scoring	Sleep-EDF SC and ST	learned	Transfer Learning Using CNN	84.3%	No
[13]	Sleep Scoring	SleepEDF, Montreal Archive of Sleep Studies (MASS) and SHHS	Sub epochs	IITNet - transfer learning + bidirectional LSTM	86.2%	No
[6]	Sleep Scoring	Massachusetts General Hospital Sleep Laboratory	Raw signal	SleepNet - RNN	85.76%	No
[7]	Sleep Scoring	Montreal archive of sleep studies (MASS) open access dataset	handcrafted	SVM	79.7%	No
[5]	Sleep Scoring	Montreal Archive of Sleep Studies (MASS) and Sleep-EDF	learned	DeepsSleep Net - CNN	80.7%	No
[1]	Seizure detection	CHB-MIT EEG dataset	Connectivity measures	CNN, BiLSTM	95.54%	Yes

[22]	Seizure detection	CHB-MIT EEG dataset	Raw data	CNN	98.05%	Yes

2.2 Research Gap

After the literature survey, following gaps were identified in the previous works:

- Most of the existing research work focus only on EEG waves for the classification of sleep stages, or the diagnosis of disorders. They do not take into consideration other aspects and features like the age group of patients, oxygen level, heartbeat rate, etc. while detection. Polysomnography i.e. sleep study comprehensively provides all this data which can be utilized well to make the classifications.
- Trained models are usually black boxes enveloped in complexity that is deployed but not deciphered. Most of the previous work fails to capture the correct reasoning in the decision process. They do not implement Explainable AI for deep learning black box models in domains like sleep scoring, and seizure detection.
- Most research work uses only single-channel EEG to make classifications, and does not utilize all the available signal data (Fpz, Pz and EOG). Using all the signals can boost the accuracy as it can interpret both temporal and spatial features of the waves.
- The existing architectures are very bulky i.e. containing numerous layers and heavy preprocessing. Hence, the classification process takes excessive time to preprocess and classification in real-time especially for large amount of data.

2.3 Problem Statement

- To address both the pragmatic shortcomings associated with a manual review and disease prognosis with EEG data and the subsequent question of ethics, reliability and interpretability that arises when automation is introduced in the realm of medicine, our project proposes the development of an XAI architecture to classify Raw Single-channel EEG data into 5 corresponding sleep stages.

- To demonstrate how the classification aids in detecting prominent related diseases, we aim to detect two pervasive diseases with the proposed architecture- Sleep Apnea and Seizures. The development can potentially deliver personalized healthcare at a comparable accuracy while simultaneously automating the process of repetitive logical inferences and expensive health care operations. The XAI component is so crucial to the project, since the consequences of a wrong AI decision can be far-reaching and extremely damaging.

2.4 Objective of the Research Work

The project moves forward in three stages-

1. The first stage is associated with a review and comparative analysis of both XAI and Deep Learning approaches alongside the creation of a traditional black-box DL model on the dataset that classifies the sleep stages for further steps of the project.
2. The second stage will be the detection of the maladies - Sleep Apnea and Seizure using the EEG signals analysis done in the stage 1.
3. The third stage will be the implementation of XAI architectures for sleep stage classification and specific malady detection respectively in order to incorporate reliability and interpretability to our proposed solution.

The minor project addresses the first stage.

Chapter 3 : Design

3.1 Architectural Design

The architectural design of our proposed system would represent the software needs and design of the system. IEEE defines architectural design as “the process of defining a collection of hardware and software components and their interfaces to establish the framework for the development of a computer system.”
The software that is built for computer-based systems can exhibit one of these many architectural styles. Each style will describe a system category that consists of:

1. A set of components (for example: a database, computational modules) that will perform a function required by the system.

2. The set of connectors will help in coordination, communication, and cooperation between the components.

3. Conditions that how components can be integrated to form the system.

4. Semantic models that help the designer to understand the overall properties of the system.

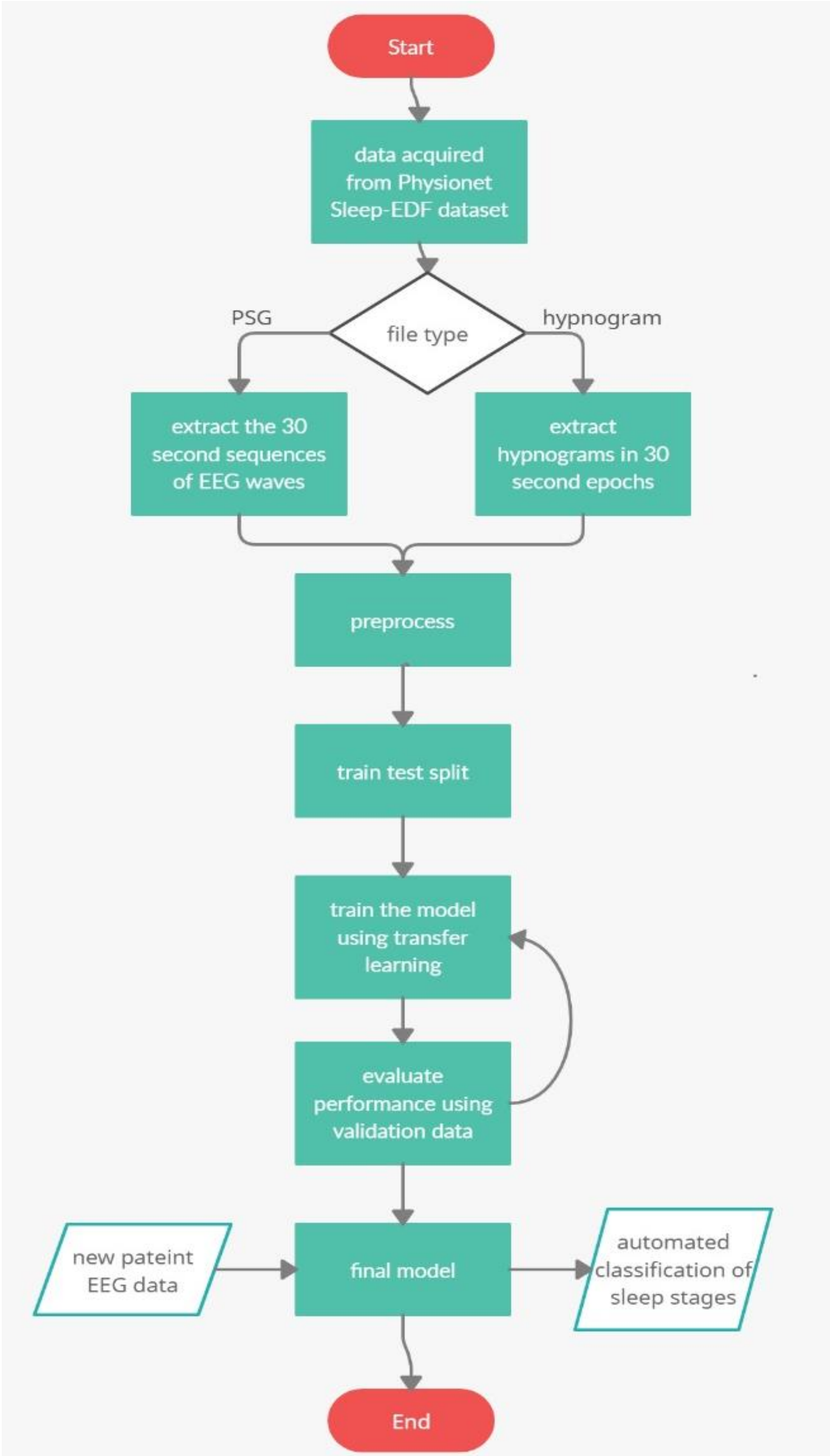


Fig 3.1 Proposed Architectural Design for sleep stage classification

3.1.1 Deep Learning Architecture

In a CNN, The convolutional layers closer to the input layer of the model learn low-level features such as lines, that layers in the middle of the layer learn complex abstract features that combine the lower level features extracted from the input, and layers closer to the output interpret the extracted features in the context of a classification task. Hence, a level of detail for feature extraction from an existing pre-trained model is chosen.

We propose a Deep Learning Convolution Neural network (CNN) based on transfer learning technique. Using the State of the Art ResNet-50 architecture. The Residual Network, or ResNet for short, is a model that makes use of the residual module involving shortcut connections. It was developed by researchers at Microsoft and described in the 2015 paper titled "Deep Residual Learning for Image Recognition." [21]

ResNet-50 is flexible, allowing the use of pre-trained models directly as feature extraction preprocessing and integrated into entirely new models.

The Figure below shows The Resnet-50 architecture.

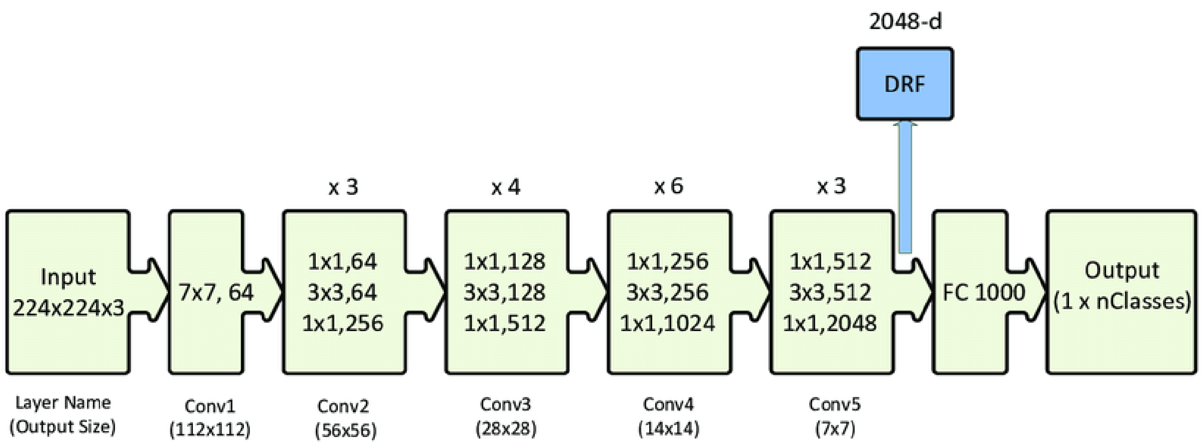


Fig 3.2 ResNet-50 Architecture

The resnet 50 architecture contain the following elements:

- A convolution with a kernel size of 7 * 7 and 64 different kernels all with a stride of size 2 giving us **1 layer**.
- Next we see max pooling with also a stride size of 2.

- In the next convolution there is a 1 * 1,64 kernel following this a 3 * 3,64 kernel and at last a 1 * 1,256 kernel , These three layers are repeated in total 3 times so giving us **9 layers** in this step.
- Next we see kernel of 1 * 1,128 after that a kernel of 3 * 3,128 and at last a kernel of 1 * 1,512 this step was repeated 4 times so giving us **12 layers** in this step.
- After that there is a kernel of 1 * 1,256 and two more kernels with 3 * 3,256 and 1 * 1,1024 and this is repeated 6 times giving us a total of **18 layers**.
- And then again a 1 * 1,512 kernel with two more of 3 * 3,512 and 1 * 1,2048 and this was repeated 3 times giving us a total of **9 layers**.
- After that we do an average pool and end it with a fully connected layer containing 1000 nodes and at the end a softmax function so this gives us **1 layer**.

We don't actually count the activation functions and the max/ average pooling layers, so totaling this it gives us a 1 + 9 + 12 + 18 + 9 + 1 = 50 layers Deep Convolutional network. Each ResNet block is either 2 layer deep (Used in small networks like ResNet 18, 34) or 3 layer deep(ResNet 50, 101, 152).

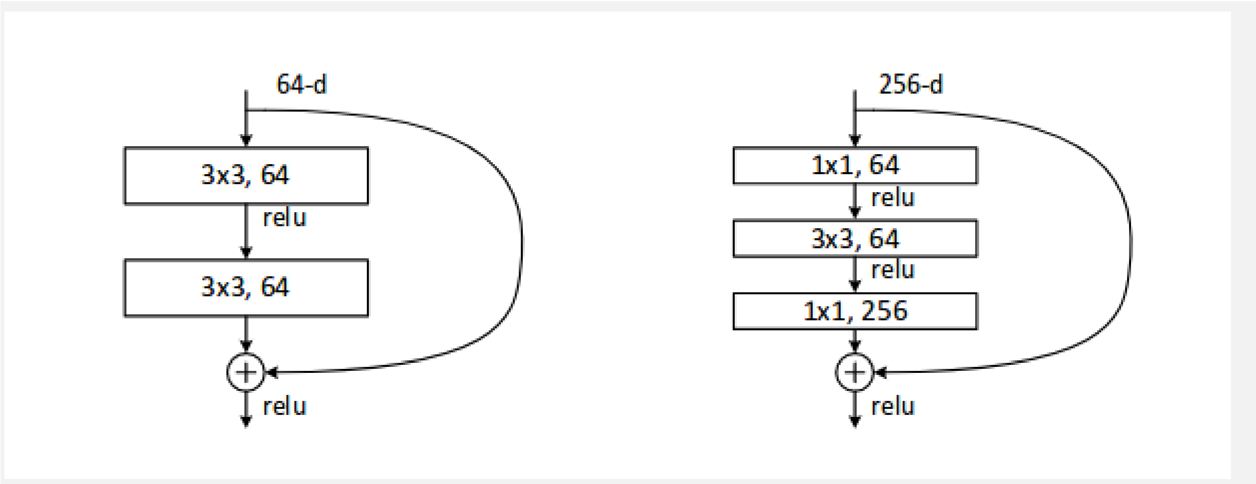


Fig 3.3 ResNet 2 layer and 3 layer Block

Advantages of ResNet:

- The skip connections in ResNet solve the problem of vanishing gradient in deep neural networks by allowing this alternate shortcut path for the gradient to flow through.

- The other way that these connections help is by allowing the model to learn the identity functions which ensures that the higher layer will perform at least as good as the lower layer, and not worse
- ResNet Network Converges faster compared to the plain counterpart of it.
- Identity vs Projection shortcuts. Very small incremental gains using projection shortcuts in all the layers. So all ResNet blocks use only Identity shortcuts with Projections shortcuts used only when the dimensions change.
- ResNet-34 achieved a top-5 validation error of 5.71% better than BN-inception and VGG.
- ResNet-152 achieves a top-5 validation error of 4.49%. An ensemble of 6 models with different depths achieves a top-5 validation error of 3.57%. Winning the 1st place in ILSVRC-2015
- The Result was pretty good on the ImageNet validation set, The ResNet 50 model achieved a top-1 error rate of 20.47 percent and achieved a top-5 error rate of 5.25 percent, This is reported for a single model that consists of 50 layers not an ensemble of it.

Architecture Used for Sleep Stage Classification

Convolution Neural Network:

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

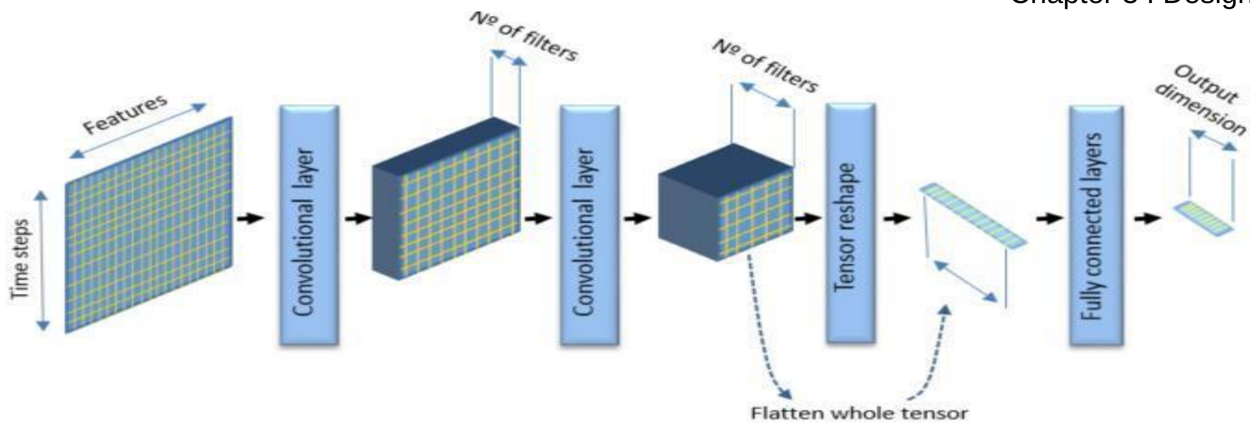


Fig 3.4. Convolution Neural Network (CNN)

Methodology - Transfer Learning

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task.

It is a popular approach in deep learning where pre-trained models are used as the starting point on computer vision and natural language processing tasks given the vast compute and time resources required to develop neural network models on these problems and from the huge jumps in skill that they provide on related problems.

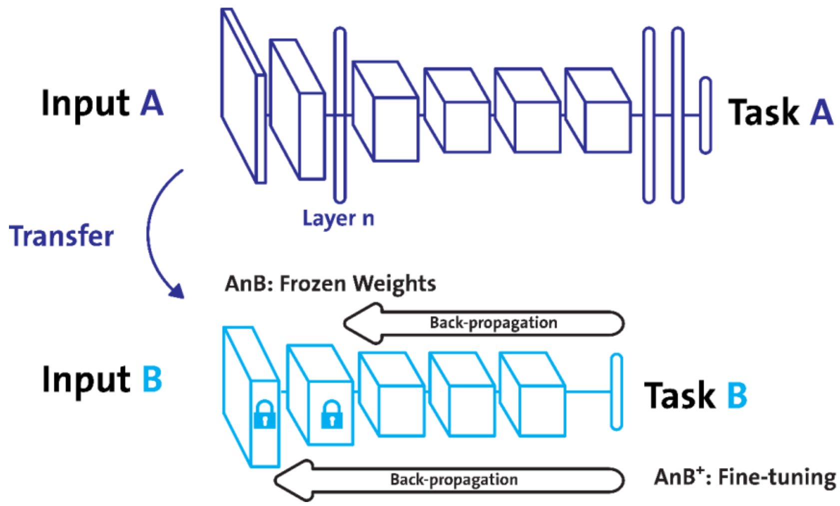


Fig 3.5. Transfer Learning Technique

Chapter 4 : System Analysis

4.1 Functional Requirements :

- The product is expected to take a edf file as an input, it should be able to process without hampering the signal quality so as to not miss out on any important values. The algorithms are expected to deliver an accuracy of over 95% for the product to be considered reliable.
- The output of the analysis algorithms should be unambiguous, complete, correct and accurate.
- The model should not lose any pixels in case of CNN that has important information. The results should be stored into the data base. The accuracy of the model should be 95% or above.

4.2 Non - Functional Requirements :

1) Performance Requirements:

The software should operate seamlessly on the devices. It should provide the desired accuracy and must be available when required. It should be able to pull data from the database as and when desired and also store it in the database so it could be retrieved in the future.

2) Safety and Security Measures:

The user should ask for login ID and password to every user and verify the same with the records in the database. This is to ensure that the software is not used for inappropriate purposes. The database should be secured against attacks of SQL injections and should be accessible only to the authorised person.

3) Software System Attributes:

RELIABLE : Should have sufficient accuracy that the users can rely on it. It should meet client satisfaction standards and be able to gain and maintain their trust.

AVAILABILITY : Whenever the need is there for the product it should be available and in working condition. It should not crash when it is required the most and function seamlessly.

SECURE : It should have security to ensure software is not tampered with and is not used for illegal purposes.

MAINTAINABILITY : The software should be easily maintainable, the users should be able to add delete information and should be able to update the software with ease when newer versions are released.

4.3 Specific Requirements :

1. For signal recording, an equipment configuration that combines a low frontal electrode for EEG signal detection with another electrode for electrooculography (EOG) is recommended.
2. A GPU of 4gb or higher capacity so that all signals can be processed without losing any data or without any errors.
3. A database to store all the information in the data set as well as the results obtained for future references.
4. A simple user interface for simplicity of operation.
5. The signal data of the subject under study should be recorded with accurate equipments and under expert supervision.
6. Files should be in the correct format as per the specifications.
7. The data uploaded should be of the .edf format and of size as mentioned
8. The data must be stored in a database which is secured and should not be tampered by any unauthorised person.

4.4 Use Case diagram:

A UML use case diagram is the primary form of system/software requirements for a new software program underdeveloped. Use cases specify the expected behavior (what), and not the exact method of making it happen (how). Use cases once specified can be denoted both textual and visual representation (i.e., use case diagram). A key concept of use case modeling is that it helps us design a system from the end user's perspective. It is an effective technique for

4.4 Use Case diagram:
communicating system behavior in the user's terms by specifying all externally visible system behavior.

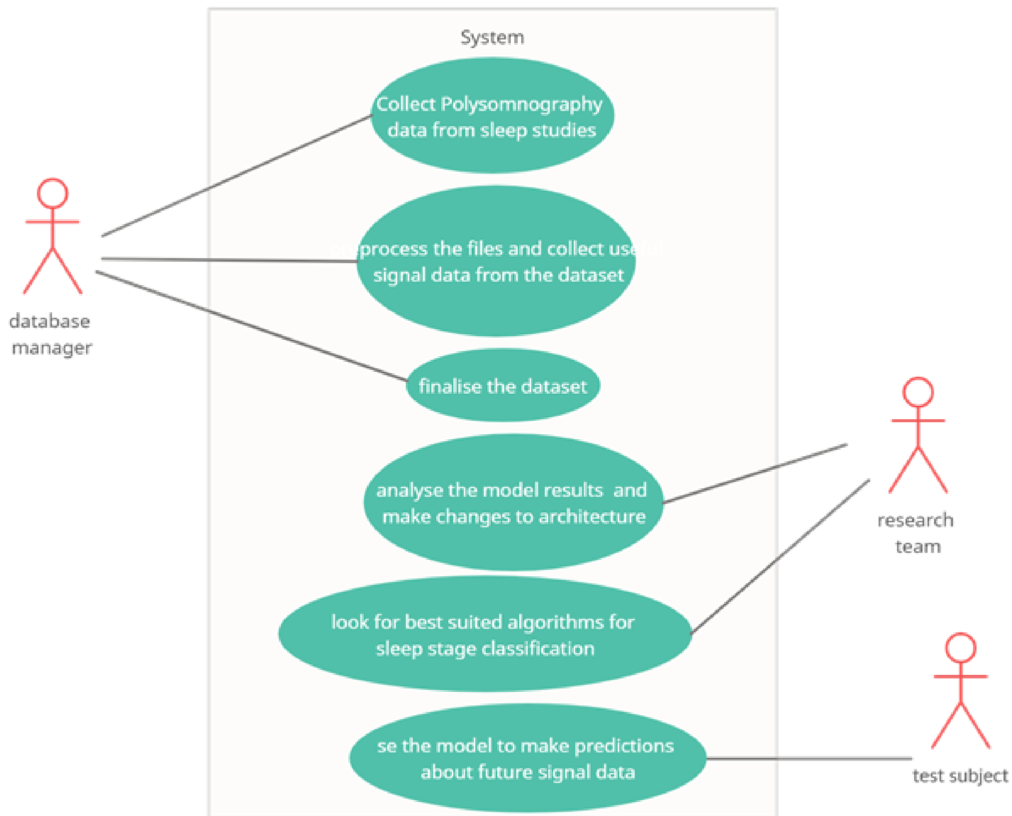


Fig 3.1 Use case diagram of the proposed system

Chapter 5 : Analysis Modeling

5.1 Data Modeling

Data modeling is the process of creating a visual representation of either a whole information system or parts of it to communicate connections between data points and structures. The goal is to illustrate the types of data used and stored within the system, the relationships among these data types, the ways the data can be grouped and organized and its formats and attributes. Data can be modeled at various levels of abstraction.

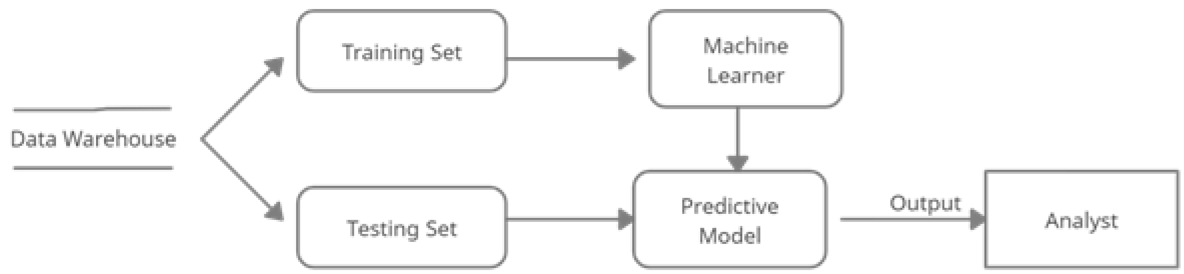


Fig 5.1 Context Level Diagram of the proposed system

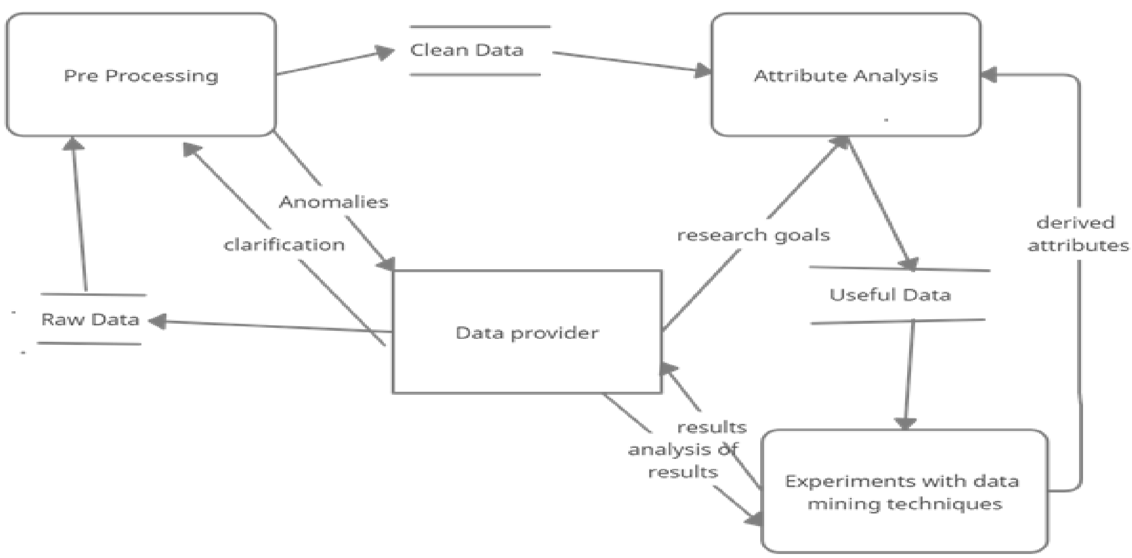


Fig 5.2 : DFD level 0 diagram for the proposed system

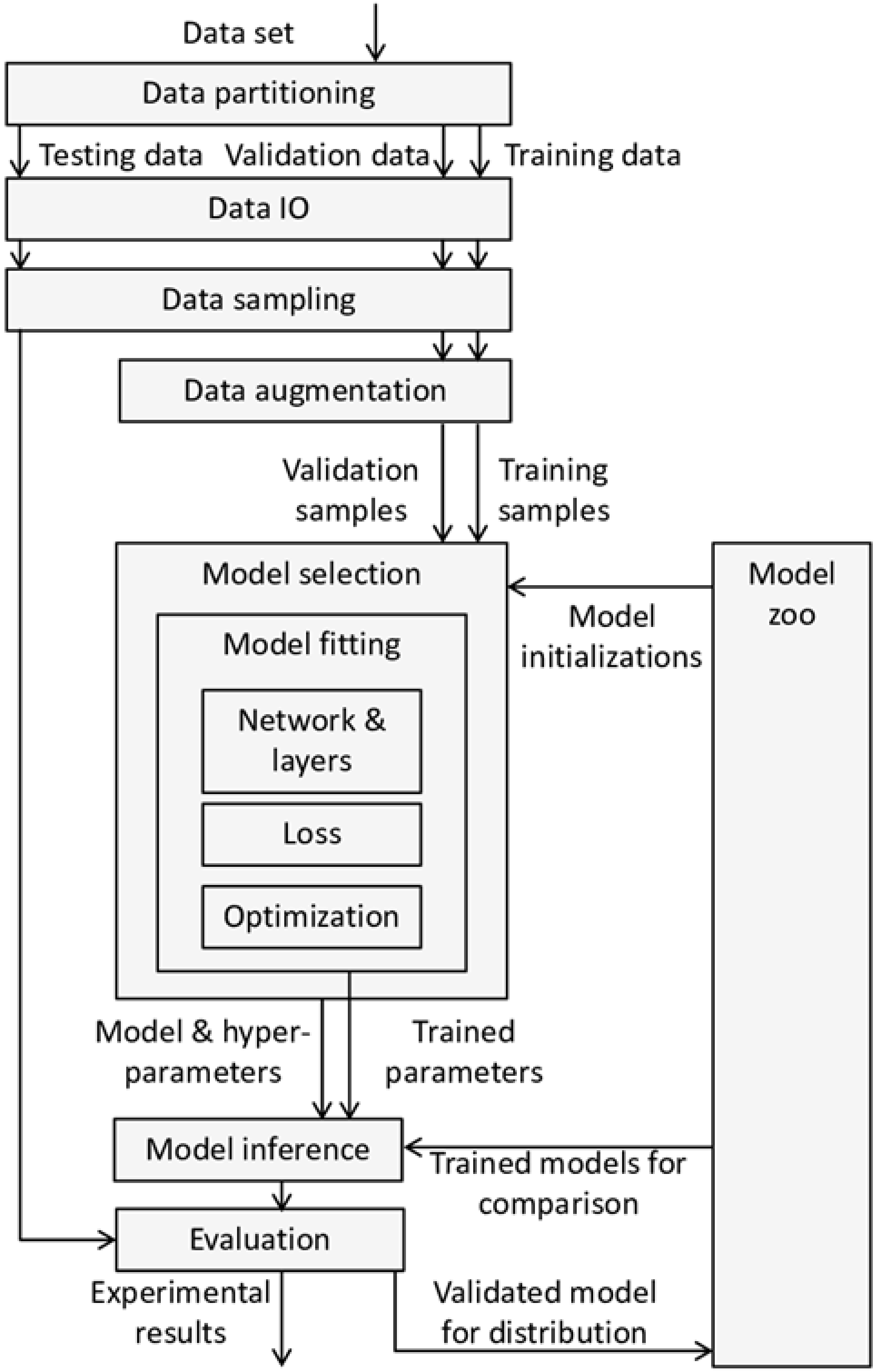


Fig 5.3: DFD level 1 of the proposed system

5.2 Class Diagram / Sequence / Collaboration / State / Activity Diagram

5.2.1 Class Diagram

Class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing, and documenting different aspects of a system but also for constructing executable code of the software application.

Class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modeling of object oriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages.

Class diagram shows a collection of classes, interfaces, associations, collaborations, and constraints. It is also known as a structural diagram.

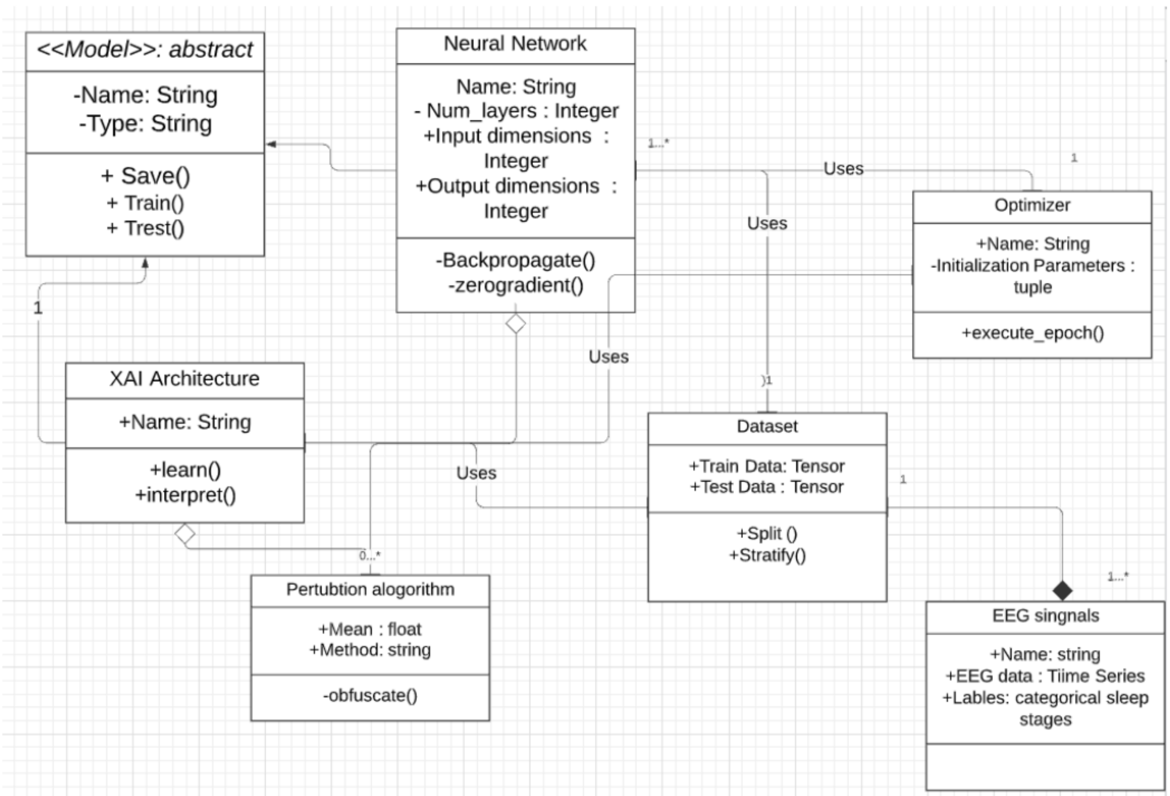


Fig 5.4 Class diagram of the proposed system

5.2.2 Sequence Diagram :

A sequence diagram simply depicts interaction between objects in a sequential order i.e. the order in which these interactions take place. We can also use the terms event diagrams or event scenarios to refer to a sequence diagram. Sequence diagrams describe how and in what order the objects in a system function. These diagrams are widely used by businessmen and software developers to document and understand requirements for new and existing systems.

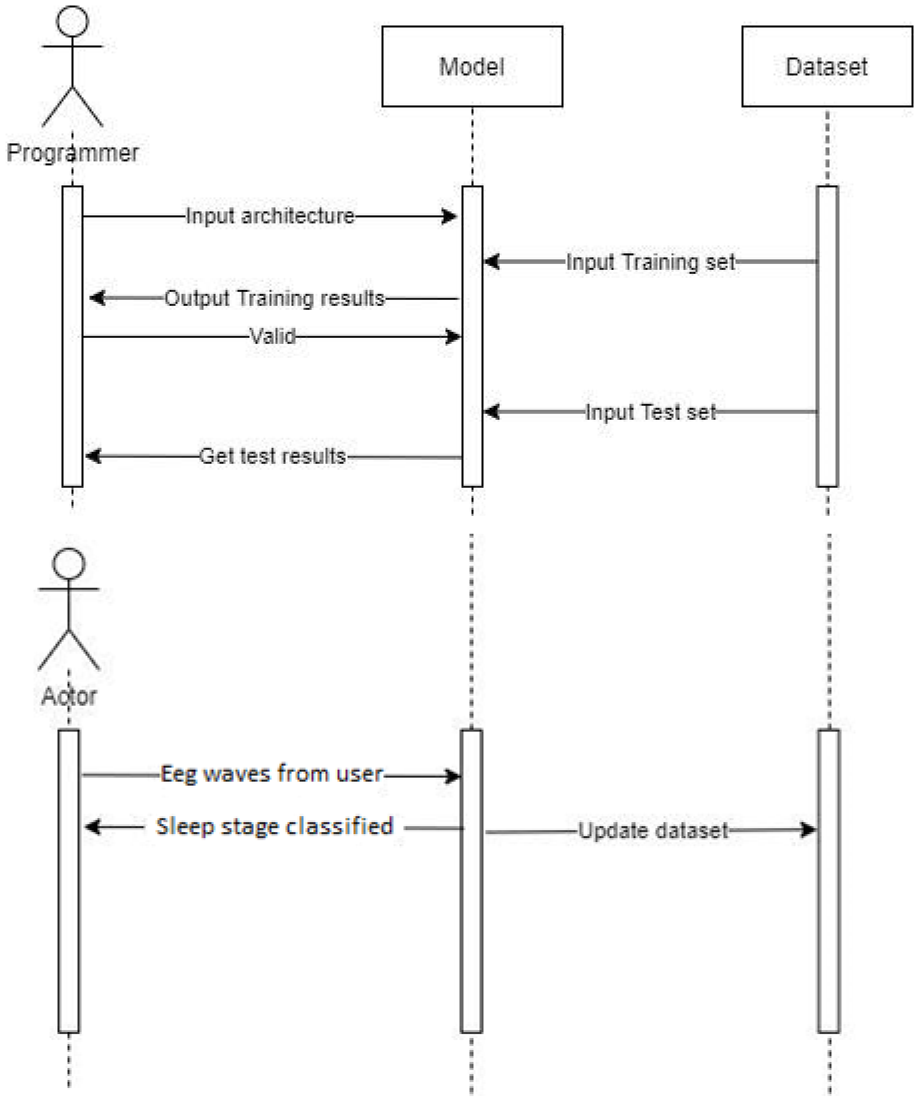


Fig 5.5 Sequence Diagram for the proposed system

5.2.3 State Diagram:

A state diagram is used to represent the condition of the system or part of the system at finite instances of time. It's a behavioral diagram and it represents the behavior using finite state transitions. State diagrams are also referred to as State machines and State-chart Diagrams. These terms are often used interchangeably. So simply, a state diagram is used to model the dynamic behavior of a class in response to time and changing external stimuli. We can say that each and every class has a state but we don't model every class using State diagrams. We prefer to model the states with three or more states.

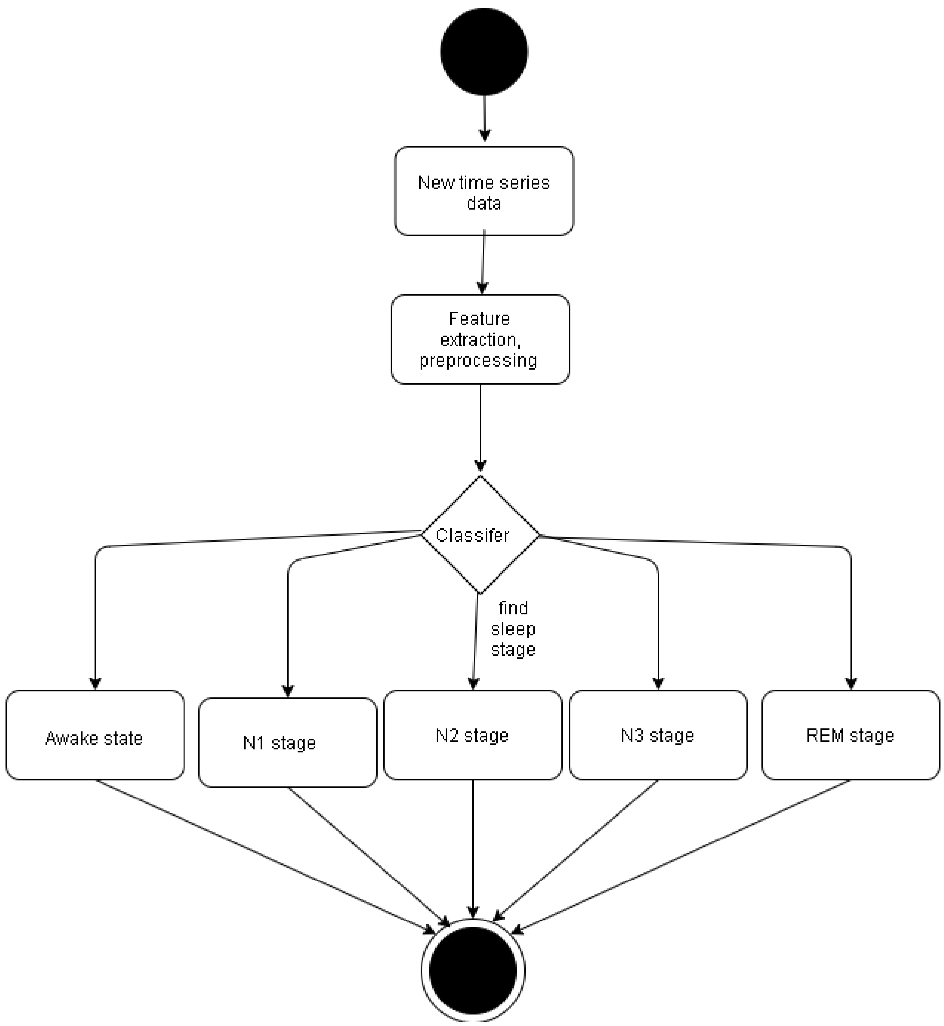


Fig 5.6 State diagram of the proposed system

5.2.4 Activity Diagram

Activity diagram is another important diagram in UML to describe the dynamic aspects of the system. Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system. The control flow is drawn from one operation to another. This flow can be sequential, branched, or concurrent. Activity diagrams deal with all type of flow control by using different elements such as fork, join, etc

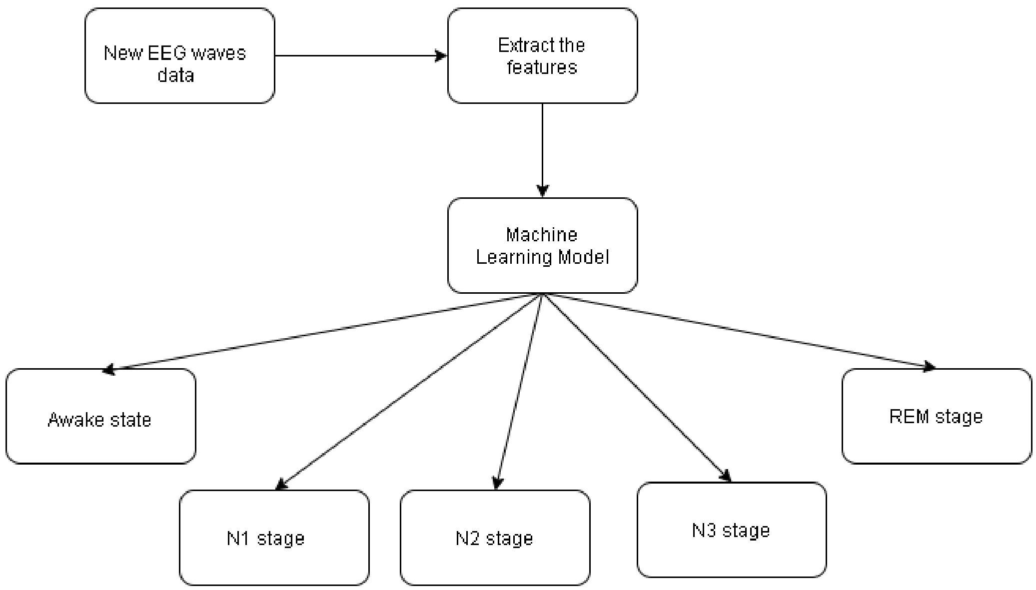


Fig 5.7 Activity Diagram of the proposed system

5.2.5 Collaboration Diagram:

The collaboration diagram is used to show the relationship between the objects in a system. Both the sequence and the collaboration diagrams represent the same information but differently. Instead of showing the flow of messages, it depicts the architecture of the object residing in the system as it is based on object-oriented programming. An object consists of several features. Multiple objects present in the system are connected to each other. The collaboration diagram, which is also known as a communication diagram, is used to portray the object's architecture in the system.

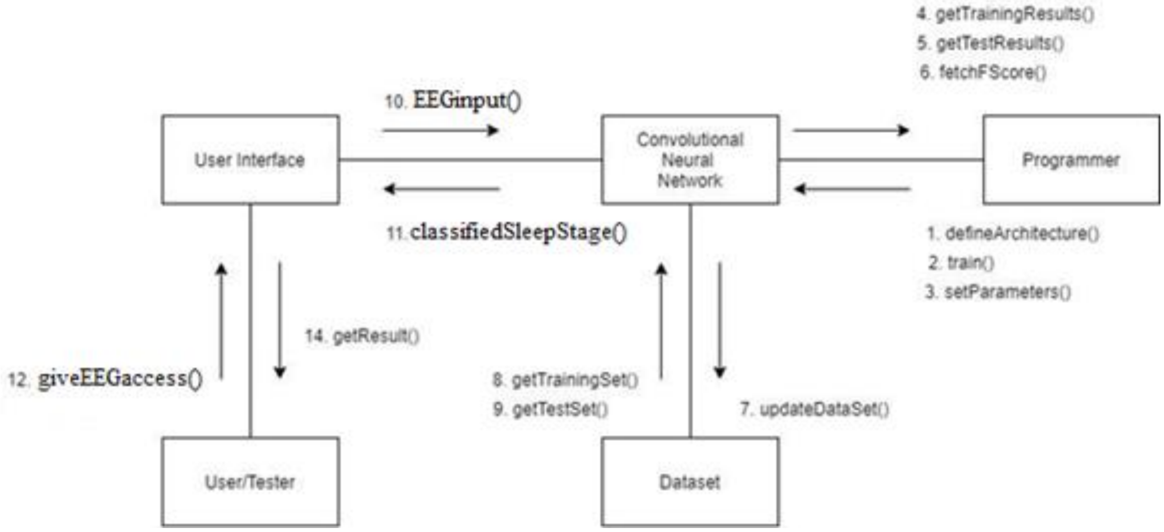


Fig 5.8 Collaboration Diagram for the proposed system

Chapter 6 : Data Description

6.1 Dataset Used

In this study, we used the Physionet Sleep-EDF extended dataset contributed in 2018 containing 197 whole-night PolySomnoGraphic sleep recordings to evaluate the performance of the proposed method for the sleep stage scoring task. Sleep-EDF dataset contains two different studies including SC (= Sleep Cassette) and ST (= Sleep Telemetry).

The dataset includes whole-night polysomnograms (PSGs) sleep recordings at the sampling rate of 100 Hz. Each record contains EEG (from Fpz-Cz and Pz-Oz electrode locations), EOG, chin electromyography (EMG), and event markers. The hypnograms (sleep stages; 30-s epochs) were manually labeled by well-trained technicians according to the Rechtschaffen and Kales standard [19].

6.2 Files and Data Format

For each patient, the *PSG.edf files are whole-night polysomnographic sleep recordings containing EEG, EOG (horizontal), submental chin EMG, and an event marker.
The *Hypnogram.edf files contain annotations of the sleep patterns that correspond to the PSGs. These patterns (hypnograms) consist of sleep stages W, R, 1, 2, 3, 4, M (Movement time) and ? (not scored).

6.3 AASM Sleep Scoring standards

The American Academy of Sleep Medicine (AASM) [18] modified the standard guidelines for sleep classification by Rechtschaffen and Kales and developed a new guideline for terminology, recording method, and scoring rules for sleep-related phenomena. One of the major changes is a change in terminology: in the AASM classification, the sleep stages S1 to S4 are referred to as N1, N2, and N3, with N3 including stages S3 + S4 and stage REM is referred to as stage R, and the movement time M is included as Wake stage W.

We have used the AASM standard in order to classify the data of sleep study. The comparisons are as follows:

R&K	AASM
Stage W	Stage W
N-REM stage 1	N1
N-REM stage 2	N2
N-REM stage 3 N-REM stage 4	N3
REM stage 5	Stage REM

Fig 6.1 Differences between the various sleep stages in Rechtschaffen and Kales guidelines and AASM guidelines

6.4 Exploratory Data Analysis - EDA

First, the Sleep-EDF dataset is obtained from the Physionet database. For each sleep recording, two files namely *-PSG.edf and *.Hypnogram.edf has been given. Using the Visbrain tool Sleep [20], the two files are visualized and relevant EDA is done.

Table 6.1 Hypnograms as read by VisBrain

Panels | Tools | Infos | Scoring | Detection | Annotations

	From (minutes)	To (minutes)	Stage
7	401.5	409.0	N2
8	409.0	410.0	N3
9	410.0	411.0	N2
10	411.0	412.5	N3
11	412.5	413.0	N2
12	413.0	414.5	N3
13	414.5	415.0	N2
14	415.0	416.0	N3
15	416.0	416.5	N2
16	416.5	423.5	N3
17	423.5	426.0	N2
18	426.0	426.5	N3
19	426.5	431.0	N2
20	431.0	431.5	N3
21	431.5	433.5	N2
22	433.5	434.0	N1
23	434.0	444.0	N2
24	444.0	444.5	REM
25	444.5	446.0	N2

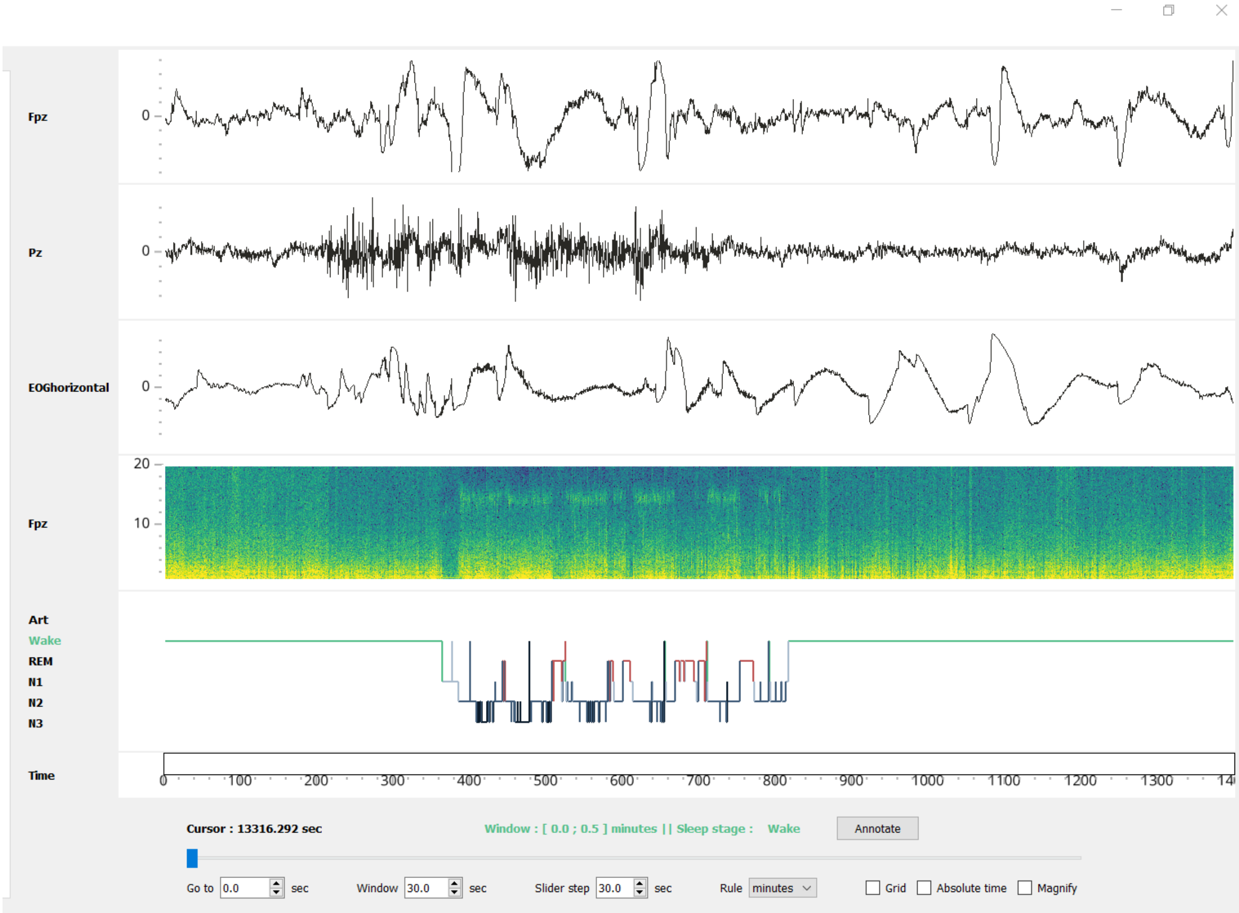
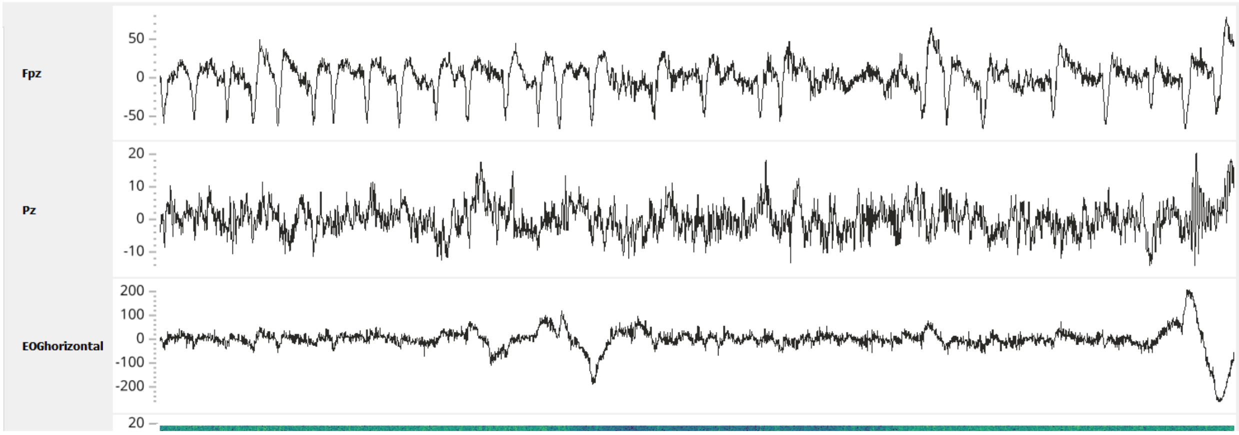


Fig 6.2. EEG and EOG Signals as read by VisBrain GUI tool

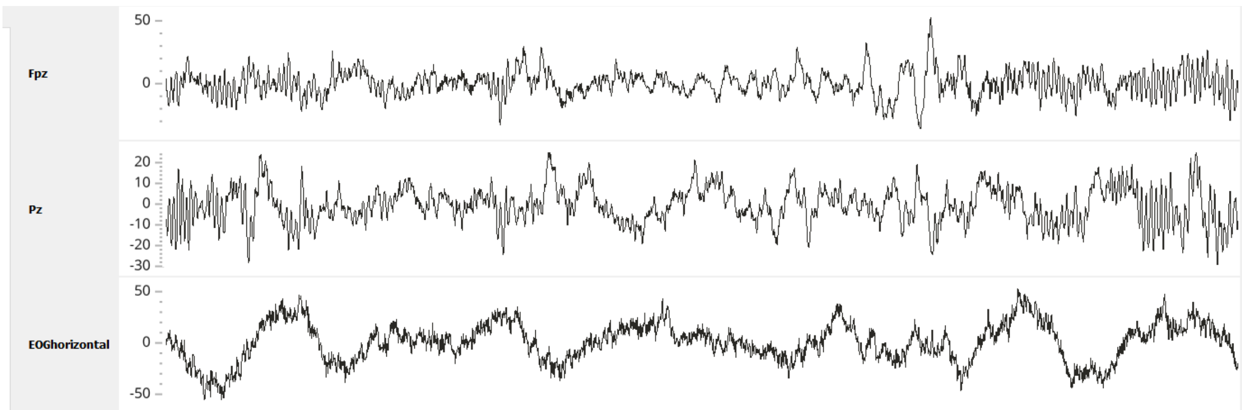
6.5 Waveforms

The waveform of EEG and EOG signals is the graph of the amplitude of the signal as a function of time. For example, waveform of different stages of sleep in an epoch of 30 second looks like:

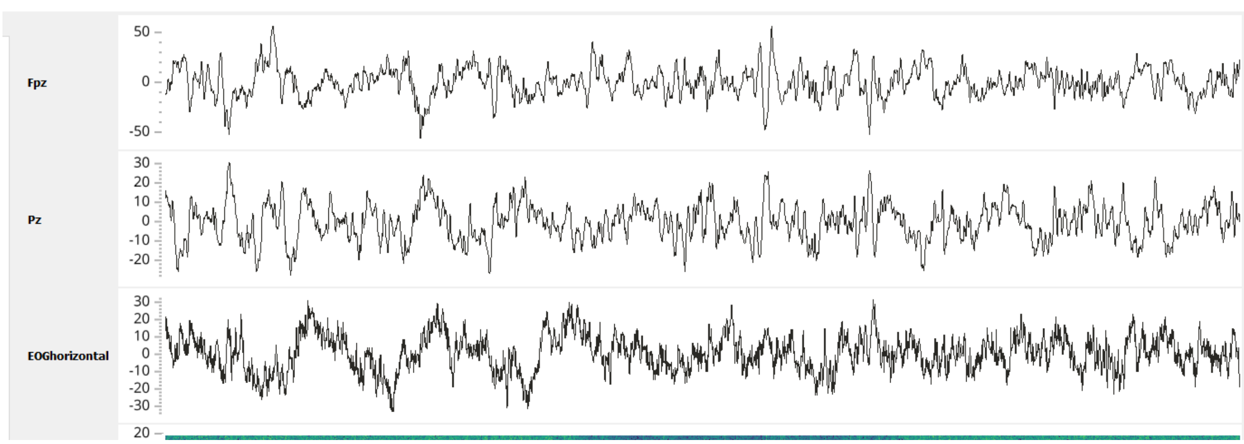
- **Wake stage**



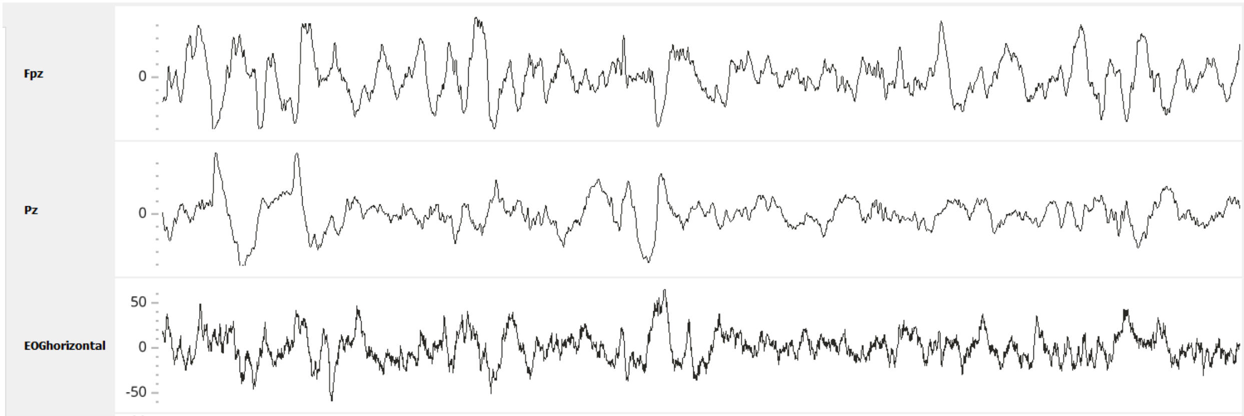
- **N1 stage**



- **N2 stage**



- **N3 stage**



- **REM stage**

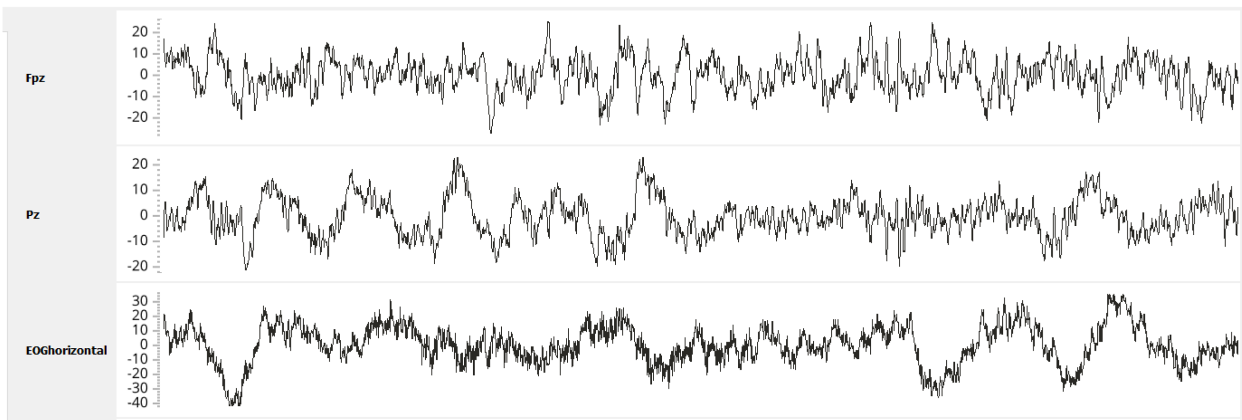


Fig 6.3 Raw EEG waveforms for each of the sleep stages : Wake, N1, N2, N3, N4 and REM

6.6 Dataset Creation

Using pyedflib toolbox, both the edf files are read. As a part of data preprocessing, the hypnograms are divided into epochs of 30 seconds each. Similarly the data from the PSG files (the EEG signals - Pz and Fpz signals) are divided into series of 30 seconds each. Then, the above two data frames are merged together to form a single one to one corresponding data file. An example of such a dataset is shown below:

	EEG_Fpz_Cz	sleep_stage
0	[11.492307692307692, 7.984615384615385, 6.2307...	0
1	[105.92307692307693, 104.07692307692308, 103.6...	0
2	[0.9692307692307693, -1.8923076923076925, 7.06...	0
3	[-6.230769230769231, -2.5384615384615388, 2.53...	0
4	[-54.323076923076925, -51.369230769230775, -51...	0

Fig 6.4. Example of final dataset after data preprocessing

And then this data is feeded into the deep learning architecture model as shown in the section Proposed Architecture. The sleep stage classification is hence obtained from raw EEG signal data.

6.7 Balancing of datasets:

It was observed that the final dataset made was very unbalanced having sleep stage 0 and extremely few data points for class 4. In this case, the dataset would have been very biased, and predicting only class 0 for most cases even leading to overfitting. Hence, sampling is done by taking 200 instances of each of the 5 classes N1, N2, N3, Wake and REM.

For CNN:

Each epoch of 30 seconds raw data of EEG signals obtained above are first converted into signal images of grayscale images. In order to obtain the data from the CNN, the Fpz signals are plotted on a graph as shown below:

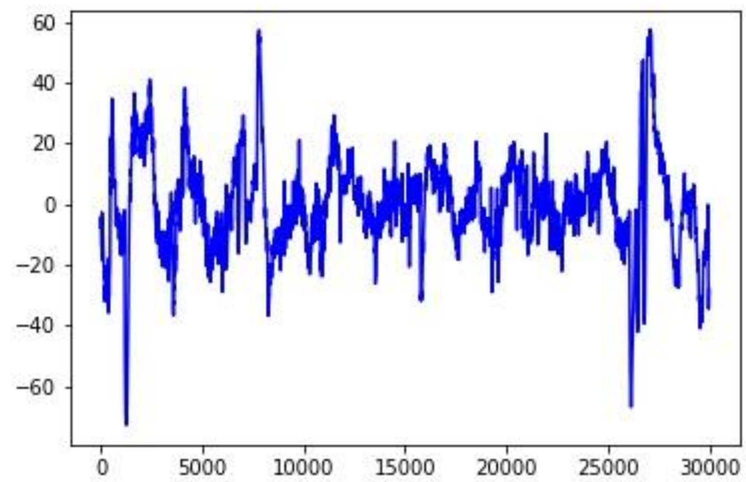


Fig 6.5. The plotted signals after raw data converted to graphs

Then, these images are feeded into the Transfer Learning model as proposed in the above section. The results and discussions are given in the next Chapter.

Results and Discussions

The aim of the project was to evaluate sleep scorings according to the new AASM standards. The model successfully classifies the stages with a **96% accuracy as well as precision, recall and f1 score of 0.96.**

```
[42] print(classification_report(y, np.argmax(preds, axis = 1)))

              precision    recall  f1-score   support

           0       1.00      0.97      0.99        36
           1       0.91      1.00      0.95        41
           2       1.00      0.93      0.96        42
           3       1.00      0.89      0.94        36
           4       0.92      1.00      0.96        45

    accuracy                           0.96       200
   macro avg       0.97      0.96      0.96       200
weighted avg       0.96      0.96      0.96       200
```

Fig 6.6. Classification Report

Table 6.2. Class Wise Accuracy

Class	0	1	2	3	4
Accuracy	0.97	1.00	0.92	0.88	1.00

1. Precision: Proportion of the positive identifications which are correct.

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

where, TP = True Positive
FP = False Positive
TN = True Negative
FN = False Negative

2. Recall: Proportion of actual positives which was identified correctly.

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

3. F1-score: It is a measure of the test's accuracy. It is calculated as the harmonic mean of precision and recall.

$$\boldsymbol{F1 - score = \frac{2 \times precision \times recall}{precision + recall}}$$

4. Accuracy: It is the fraction of predictions the model got correct.

$$\boldsymbol{Accuracy = \frac{TP + TN}{TP + TN + FP + FN}}$$

5. Support: It is the number of actual occurrences of the class in the specified dataset. Imbalanced support in the training data may indicate structural weaknesses in the reported scores of the classifier and could indicate the need for stratified sampling or rebalancing. Hence, approximately equal support in all the classes suggest a balanced dataset.

The following image presents the confusion matrix and the performances of each class achieved by the proposed method using Fpz-Cz of the Sleep-EDF data set. The main diagonals in each confusion matrix denote the true positive (TP) values which indicate the number of stages scored correctly. It can be seen from the tables (the confusion matrices' parts) that TP values are higher than other values in the same rows and columns.

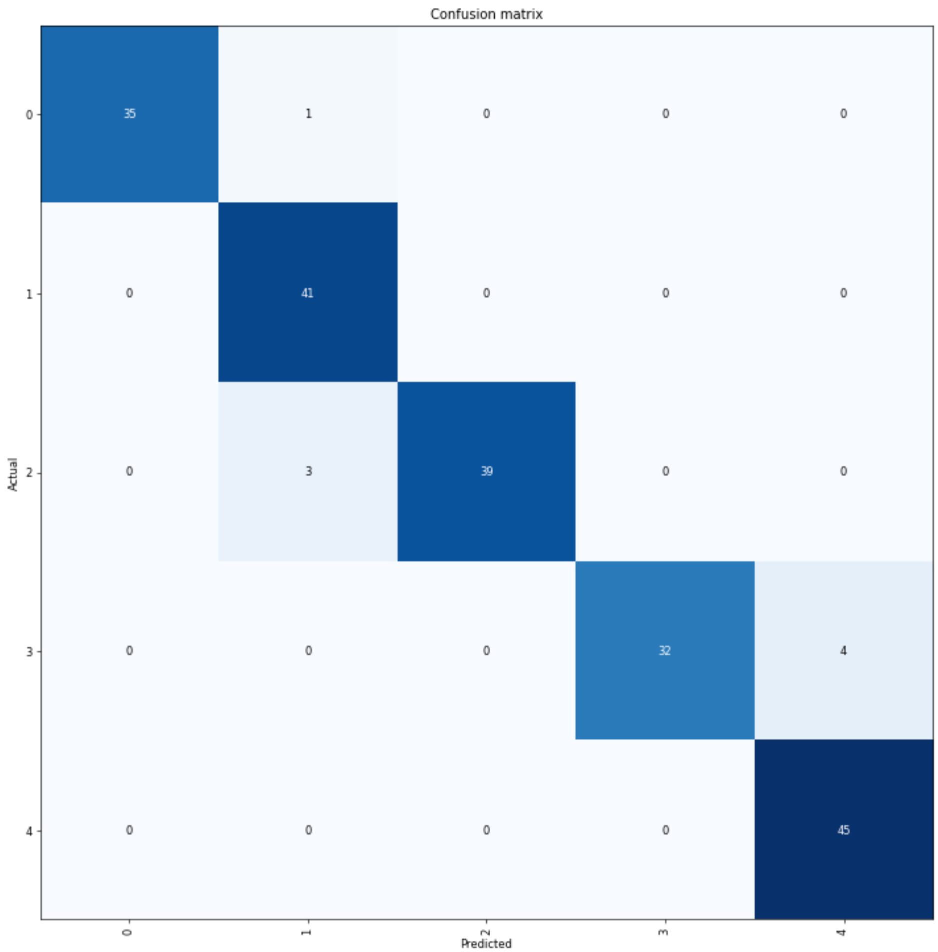
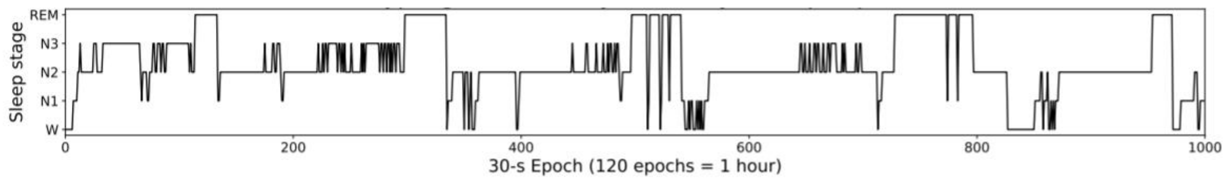


Fig 6.7.Confusion Matrix for Transfer Learning based architecture

We achieved an accuracy of 96% overall for the stage 1 ie. classification of sleep stages. It is significantly higher than most of the previous works done in the domain. The f1 score of class 3 is the least, and from the above confusion matrix, we observe that the model is confusing class 3 with class 4. i.e. stage N3 with the REM stage. With more data points of the class 3 and class 4, accuracy and f1 score may increase as our model would learn more variations in patterns of those sleep stages.

An example of hypnogram comparison between the expert-scored and our model-scored is given below:

Expert scored:



Our model predictions:

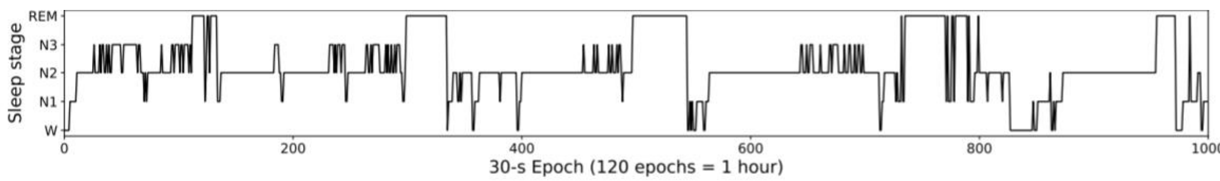


Fig 6.8. Comparison of result classifications between expert scored stages and model scored stages

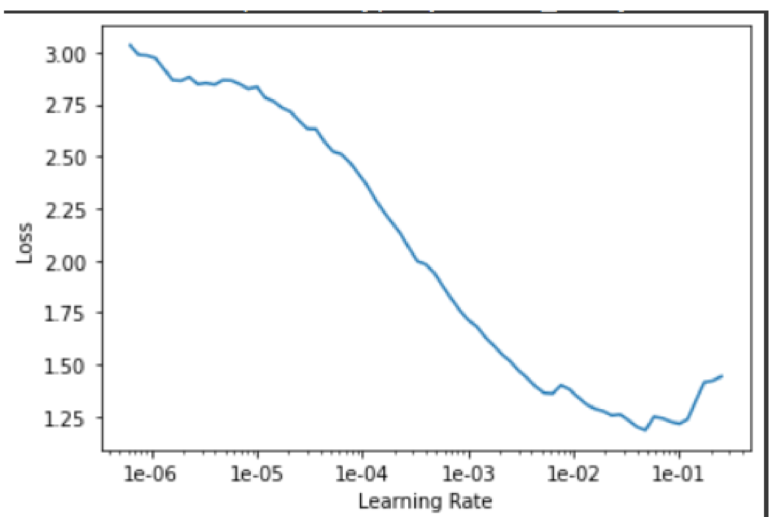


Fig 6.9. Plot of learning rate vs Validation Loss

The top losses of the model are shown below:

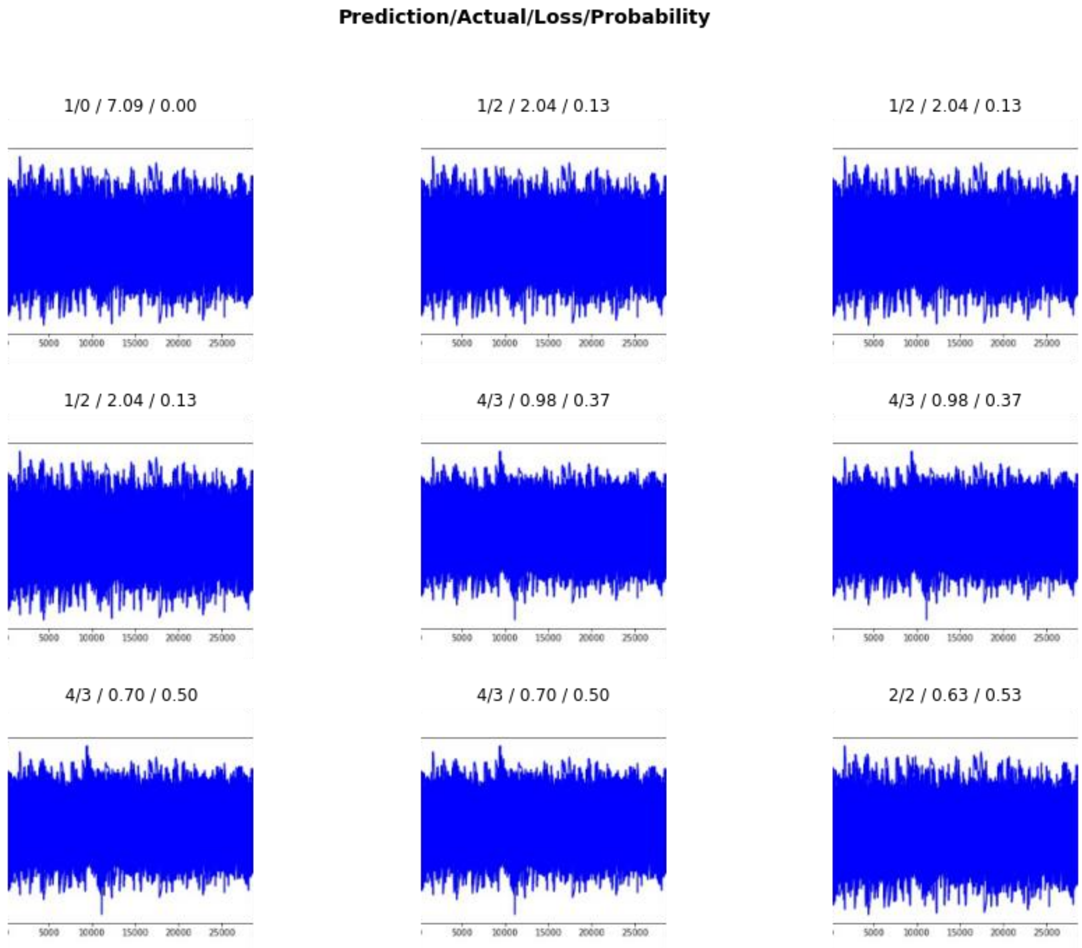


Fig 6.10. Prediction/ Actual notations and Graphs of the top losses

Chapter 7 : Conclusion & Future Scope

The project has implemented and automated the sleep staging pipeline- when given raw, single channel EEG data signals of a patient from the Fpz-Cz of the Sleep-EDF data sets, we classify the five sleep stages (Wake, N1, N2, N3, N4 and REM) using the proposed deep learning model. The preprocessing and sampling decreased helps avoid model bias, ensuring that the model evaluates sleep scorings according to the new AASM standards. We achieve an accuracy of 96% and precision, recall and f1 score of 0.96 without utilizing any hand-engineered features.Our model utilizes CNNs to extract time-invariant features and learn stage transition rules among sleep stages from EEG epochs. The results also showed that the temporal information learned from the sequence residual learning part helped improve the classification performance. These demonstrated that our model could accurately learn features for sleep stage scoring from different raw single-channel EEGs.

A deep neural network's ability to model and generalize across complex , high dimensional relationships while giving accurate results at velocity once thought to be inconceivable makes it a strong interdisciplinary contende. Although the results generated by the experiment are satisfactory for academia, one simply the deployment of such an architecture is simply untenable in a high stakes field like medicine. As a consequence of the fact that a traditional black model model was used, we simply cannot guarantee or reason results- accurate models sometimes fail to capture the correct reasoning in the decision process. This project aims to generate an architecture that can provide reasoning and insights for the inferences. The potential for human scrutiny, reasoning and verification can make the model robust, increasing confidence. The trained DL model can serve as a benchmark for the proposed XAI architecture, helping in a more accurate assessment of the tradeoff in accuracy v/s interpretability single both architectures would have trained on the same data and similar preprocessing, eliminating the potential for variability in performance due to extraneous reasons.

If successful, the architecture can not just significantly contribute to the field of polysomnography, but also create a methodology for addressing the integration of AI in Healthcare and medicine more broadly. When medical professionals and patients need to know the rationale for significant AI recommendations such as hospitalizations and diagnosis. The project can be extended to allow for Better accounting of generalization errors, to more easily identify deviations from the prediction of the algorithm. This would help practitioners determine when the model isn’t working correctly and the provided context would help determine whether the model’s suggestions are trustworthy. Further, a learned representation from the data of healthy individuals can potentially aid and provide insights for generalising and understanding a healthy norm.

The work will be extended in the major project to incorporate the aforementioned elements.

References

[1] Mohammad Mansour, Fouad Khnaisser, and Hmayag Partamian , "**An Explainable Model for EEG Seizure Detection based on Connectivity Features"**
DOI**:**https://arxiv.org/ftp/arxiv/papers/2009/2009.12566.pdf

[2]. Jean-Marc Fellous, Guillermo Sapiro, Andrew Rossi, Helen Mayberg and Michele Ferrante,
"Explainable Artificial Intelligence for Neuroscience: Behavioral Neurostimulation"
Front. Neurosci., 13 December 2019
DOI:https://doi.org/10.3389/fnins.2019.01346

[3]Amina Adadi and Mohammed Berrada
" **Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)"**
IEEE Xplore ,September 17, 2018
DOI:https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8466590

[4] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk and Jocelyn Faubert "**Deep learning-based electroencephalography analysis: a systematic review"**
14 August 2019
Journal of Neural Engineering, Volume 16, Number 5
DOI:https://iopscience.iop.org/article/10.1088/1741-2552/ab260c

[5]A. Supratak, H. Dong, C. Wu and Y. Guo, "**DeepSleepNet: A Model for Automatic Sleep Stage Scoring Based on Raw Single-Channel EEG**," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 11, pp. 1998-2008, Nov. 2017, doi: 10.1109/TNSRE.2017.2721116.
DOI:https://ieeexplore.ieee.org/abstract/document/7961240

[6] Siddharth Biswal, Joshua Kulas, Haoqi Sun, Balaji Goparaju, M Brandon Westover, Matt T Bianchi, Jimeng Sun **" SLEEPNET: Automated Sleep Staging System via Deep Learning"**
DOI:https://arxiv.org/pdf/1707.08262.pdf

[7] Hao Dong, Akara Supratak, Wei Pan, Chao Wu, Paul M. Matthews and Yike Guo **"Mixed Neural Network Approach for Temporal Sleep Stage Classification"**
DOI:https://arxiv.org/pdf/1610.06421.pdf

[8]Sajad Mousavi, Fatemeh Afghah and U. Rajendra Acharya "**SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach"**
2019 May 7.
DOI: 10.1371/journal.pone.0216456

[9] Faust, Oliver and Razaghi, Hajar and Barika, Ragab and Ciaccio, Edward and Acharya, U Rajendra "**A review of automated sleep stage scoring based on physiological signals for the new millennia"**
April 2019 Computer Methods and Programs in Biomedicine 176
DOI:10.1016/j.cmpb.2019.04.032

[10] Klara Stuburić, Maksym Gaiduk, Ralf Seepold, **"A deep learning approach to detect sleep stages"**
Procedia Computer Science,2020
DOI:**https://doi.org/10.1016/j.procs.2020.09.280**

[11]Huy Phan, Navin Cooray, Oliver Y. Chen, and Maarten De Vos, "**SeqSleepNet: End-to-End Hierarchical Recurrent Neural Network for Sequence-to-Sequence Automatic Sleep Staging"**
2019 Jan 31.
DOI: 10.1109/TNSRE.2019.2896659

[12]Huy Phan , Oliver Y. Chen, Philipp Koch, Zongqing Lu, Ian McLoughlin, Alfred Mertins, and Maarten De Vos, "**Towards More Accurate Automatic Sleep Staging via Deep Transfer Learning"**
DOI:https://arxiv.org/pdf/1907.13177v3.pdf

[13]Hogeon Seo , Seunghyeok Back , Seongju Lee , Deokhwan Park, Tae Kim, Kyoobin Lee†

“**Intra- and Inter-epoch Temporal Context Network (IITNet) Using Sub-epoch Features for Automatic Sleep Scoring on Raw Single-channel EEG**”
DOI:https://arxiv.org/pdf/1902.06562v2.pdf

[14] 1. Shepard JW, Buysse DJ, Chesson AL, et al. “**History of the development of sleep medicine in the United States.**”
J Clin Sleep Med. **2005;1(1):61–82.**
DOI:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2413168/

[15]Susan L. Worley, “**The Extraordinary Importance of Sleep: The Detrimental Effects of Inadequate Sleep on Health and Public Safety Drive an Explosion of Sleep Research**”
2018 Dec
DOI:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6281147/

[16] Khald Ali I. Aboalayon ,Miad Faezipour ,Wafaa S. Almuhammadi and Saeid Moslehpour “**Real Time Sleep Detection System Using New Statistical Features of the Single EEG Channel**”
DOI:https://www.researchgate.net/project/Real-Time-Sleep-Detection-System-Using-New-Statistical-Features-of-the-Single-EEG-Channel

[17] **Portable Monitoring in the Diagnosis and Management of Obstructive Sleep Apnea Polysomnography Scoring Manual**
https://sleepdata.org/datasets/homepap/files/m/browser/documentation/HomePAP_PSG_Scoring_Manual.pdf?inline=1

[18]Iber C, Ancoli-Israel S, Chesson A, Quan SF, editors. 1st ed. Westchester, IL: American Academy of Sleep Medicine; 2007. “**The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specification.**”
DOI:https://ci.nii.ac.jp/naid/10024500923/

[19] Rechtschaffen A. Brain information service. 1968. “**A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects**”

[20] Combrisson, Etienne and Vallat, Raphael and Eichenlaub, Jean-Baptiste and O'Reilly, Christian and Lajnef, Tarek and Guillot, Aymeric and Ruby, Perrine M. and Jerbi, Karim "**Sleep: An Open-Source Python Software for Visualization, Analysis, and Staging of Sleep Data"**
DOI**:**http://journal.frontiersin.org/article/10.3389/fninf.2017.00060

[21]Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun "**Deep Residual Learning for Image Recognition"**
DOI: https://arxiv.org/abs/1512.03385

[22] Hossain, M. S., Amin, S. U., Alsulaiman, M., & Muhammad, G. (2019). "**Applying deep learning for epilepsy seizure detection and brain mapping visualization."** ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 15(1s), 1-17
DOI: https://arxiv.org/pdf/2007.01276.pdf