Somno: The Artificial Intelligence Sleep Tracker

Aarit Taparia

September 3, 2025

Registration Number: Comp-007

Student: Aarit Taparia
The International School Bangalore

Teacher: Sreeja Thampy
Secondary IB and IGCSE teacher at The International School Bangalore

Abstract

Sleep is integral to human health, and poor or fragmented sleep can impair cognitive function, reduce physiological restoration and compromise well-being in general. Around 30 per cent of adults chronically suffer from sleep and wakefulness disorders (insomnia) - an epidemic linked to obesity, cardiovascular attacks and deteriorating mental health. Despite their high frequency, sleep disorders remain poorly identified; less than 20% of affected individuals receive a correct diagnosis and treatment. This diagnostic gap highlights the need for accessible tools that can flag early signs of sleep deprivation or irregularity. This paper demonstrates our personalised model for sleep recovery using CNNs to predict the level of sleep deprivation and semi-Markov chains for simulating optimal recovery sleep in this work. The CNN used is a MobileNetV2 due to its light-weightedness and terrific performance of 79.1% on test data and 85.16% accuracy on validation data. This Markov model automatically accounts for user-specific parameters such as age, sex and pre-existing sleep debt, hence making recommendations very personalised per individual. This utilised the Polysomnographs and Hypnograms from the Sleep-EDF database, which had a large diversity of genders and ages. In comparison to ground-truth sleep data, the model achieved a mean percentage error of [26.58%] for wake, [14.16%] for NREM, and [32.23%] for REM sleep stages. In comparison to many commercial sleep trackers, which often offer limited diagnosis analysis and rely on generalised estimations of time spent in sleep cycles, this system is tailored to model individual sleep debt and recovery while achieving similar, if not better, results than the advanced technological prowess of the state-of-the-art sleep trackers.

Keywords: deep learning, mobile application, sleep deprivation detection, CNN, neural networks

Acknowledgements:
https://pmc.ncbi.nlm.nih.gov/articles/PMC10926017/
https://pubmed.ncbi.nlm.nih.gov/18018450/
https://www.sleepfoundation.org/sleep-news/new-research-evaluates-accuracy- of-sleep-trackers

Commitments on Academic Honesty and Integrity

We hereby declare that we

1. are fully committed to the principle of honesty, integrity and fair play throughout the competition.
2. actually perform the research work ourselves and thus truly understand the content of the work.
3. observe the common standard of academic integrity adopted by most journals and degree theses.
4. have declared all the assistance and contribution we have received from any personnel, agency, institution, etc. for the research work.
5. undertake to avoid getting in touch with assessment panel members in a way that may lead to direct or indirect conflict of interest.
6. undertake to avoid any interaction with assessment panel members that would undermine the neutrality of the panel member and fairness of the assessment process.
7. observe the safety regulations of the laboratory(ies) where the we conduct the experiment(s), if applicable.
8. observe all rules and regulations of the competition.
9. agree that the decision of YHSA(Asia) is final in all matters related to the competition.

We understand and agree that failure to honour the above commitments may lead to disqualification from the competition and/or removal of reward, if applicable; that any unethical deeds, if found, will be disclosed to the school principal of team member(s) and relevant parties if deemed necessary; and that the decision of YHSA(Asia) is final and no appeal will be accepted.

(Signatures of full team below)

X

Name of team member: Aarit Taparia

X

Name of team member:

X

Name of team member:

X

Name of supervising teacher: SREEJA S THAMPI

Noted and endorsed by

(signature)

Name of school principal: NAVEEN TOM

Contents

Chapter 1

Introduction

Around 30 per cent of adults chronically suffer from sleep and wakefulness disorders; an epidemic linked to obesity, cardiovascular attacks and deteriorating mental health. Despite their high frequency, sleep disorders remain poorly identified; less than 20% of affected individuals receive a correct diagnosis and treatment. This diagnostic gap highlights the need for accessible tools to flag early signs of sleep deprivation or irregularity.

This paper presents a novel AI-powered sleep recovery system that provides insight into sleep patterns while intelligently refining recovery recommendations based on a user's real-life schedule. Unlike conventional solutions, it requires no wearables, headbands, or complex technology, making it simple, accessible, and affordable for all.

Sleep deprivation is a growing concern globally, yet practical and accessible tools for quantifying it remain limited. While wearables such as Fitbit or Apple Watch can estimate sleep debt using biosignals like heart rate, they are often expensive, require consistent use, and may not be personalised to the user's recovery needs. Furthermore, popular sleep-tracking applications rely on sound detection or movement or require an external device to connect for providing analysis.

Motivated by these limitations, this project explores whether facial cues alone can offer a viable, low-cost, and non-intrusive alternative for estimating sleep deprivation. Facial features such as swollen eyes, dark circles, and drooping mouth corners have been associated with fatigue in previous research. Leveraging these insights, we aim to build a CNN-based model that classifies drowsiness and assigns a sleep deprivation score from an image-based assessment. Using Markov models constructed from hypnograms affected by certain factors, we aim to simulate recovery sleep accurately to alleviate the effects of sleep deprivation on the user. The innovations of the project are summarised as follows:

- Sleep Deprivation scoring from a facial image
- Personalised Markov models for accurately simulating sleep
- Non-intrusive application which is accessible to all
- Combination of sleep deprivation score through facial image and semi-Markov models for simulating sleep.
- Integration of cognitive-behavioural theory to enhance user compliance

Chapter 2

Literature Review

2.1 Importance of Sleep and Recovery

2.1.1 Sleep Science

Sleep is a temporary, homeostatically regulated and neurophysiologically active state where the brain undergoes a sequential progression of distinct, fixed sleep stage patterns controlled by the circadian and homeostatic mechanisms.

Throughout each sleep cycle, the brain undergoes the sleep stages in the following order: Wake, N1, N2, N3, N2 and REM. Sleep stages from N1 to N3 are referred to as NREM stages, or non-rapid eye movement sleep stages, progressing into deeper sleep. The typical night consists of 4-5 sleep cycles, varying from 90 to 110 minutes each cycle. The wake stage has the highest frequency of beta waves when the eyes are open and the highest frequency of alpha waves when the eyes are closed. The table below highlights the key differences between the different stages of sleep. [1]

Table 2.1: Summary of Sleep Stages

Sleep Stage	Length of Stage	Percentage of Total Sleep	EEG Recording
N1	1–5 minutes	5%	Theta waves (low voltage)
N2	25+ minutes	45%	Sleep spindles and K-complexes
N3	30–60 minutes	25%	Delta waves (low frequency, high amplitude)
REM	10–60 minutes	25%	Beta-like waves (similar to wakefulness)

Notable Features of Sleep Stages

- **N1:** Often marked by slow eye movements and transition from wakefulness. Hypnic jerks may occur.
- **N2:** Presence of *sleep spindles* and *K-complexes.* K-complexes are large, sharp EEG waves that help suppress cortical arousal and aid memory

consolidation, while sleep spindles are neuronal firings into the superior temporal gyri, anterior cingulate, insular cortices, and thalamus, allowing for synaptic plasticity - the strengthening or weakening of the connections between synapses.

- **N3:** The stage in which one experiences the deepest sleep with low-frequency and high-amplitude delta wave signals. Crucial for physical recovery and immune function.

- **REM:** REM sleep is not classified as a restful stage, as irregular breathing, rapid eye movements, and vivid dreaming characterise it. The EEG patterns during REM closely resemble those of wakefulness, and muscle atonia occurs to prevent the physical enactment of dreams. The brain also tends to be very active during these stages.

Facial indicators of sleep deprivation

- **Hanging eyelids**
- **Red eyes**
- **Swollen eyes**
- **Glazed eyes**
- **Droopy corners of mouth**
- **Pale skin**

[2]

These research findings highlight specific facial cues for sleep deprivation, which consolidate a crucial foundation for the sleep deprivation scoring model. These research findings also emphasise certain facial features that are much more largely affected by sleep deprivation through the VAS scale (often used to score subjective experiences on a scale of 0-100mm), such as hanging eyelids, swollen eyes, and droopy corners of the mouth [2]. A combination of the intensity of these factors would help produce a more accurate score.

2.1.2 Sleep Debt

Sleep debt is chronic sleep loss without adequate recovery sleep, leading to the accumulation of sleep debt over time. Dinges et al.[3] studies showed that even minimal amounts of sleep debt cause impairment to brain function.
Sleep deprivation has other harmful effects on attention spans, reaction times, mood swings, along with many health complications such as heart attacks, kidney disease, strokes, and even death after long periods. [4].
Erratic sleep schedules can be dangerous for heart health. During our sleep, a phenomenon known as nocturnal dipping occurs, where blood pressure dips as a result of lower activity. Consequently, a lack of sleep would lead to eventual

hypertension and other cardiovascular risks, which could be fatal.

Below is a figure showcasing the difference between a non-sleep-deprived individual and one in a psychomotor Vigilance test. This tests the reaction time of individuals, and a PVT lapse occurs when the button is not pressed within 500 milliseconds. Although the values fluctuate, there is an upward trend as the hours progress. Nevertheless, there is always a large disparity between the number of PVT lapses in a healthy human and a sleep-deprived human.

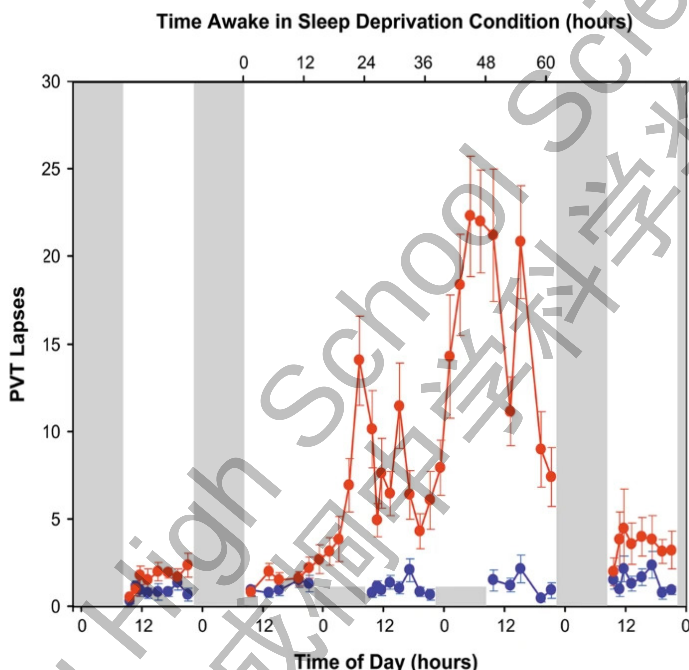


Figure 2.1: A graph showing the number of PVT lapses as the number of hours in the sleep deprivation condition progresses [5]

This figure helps to highlight the dangers to sleep-deprived individuals themselves and others. Plane crashes, nuclear meltdowns, grounding of large ships are just a few of the large-scale events occurring as a result of sleep deprivation.

Recovery of sleep debt

Although sleep debt can accumulate quickly from short sleep durations during weekdays, recovery from this debt is complex and often incomplete. The full extent of how one recovers has not been discovered yet. Studies suggest that common recovery strategies, such as catch-up sleep on weekends and short naps, provide limited benefits and do not fully reverse the cognitive and physiological impairments caused by chronic sleep restriction. [6]

Population-Level Recovery Behaviors

In a representative study of 12,637 adults [6], it was found that:

- 35.9% of the population were short sleepers (< 6 hours per day),
- 27.7% had sleep debt exceeding one hour (18.8% had *severe* debt > 90 minutes),
- Only 18.2%

A significant 75.8% of those with severe sleep debt did **not** balance it through any form of recovery during the week. Even among the general population, less than half (46.1%) engaged in weekend catch-up sleep, and only 24.7% napped.

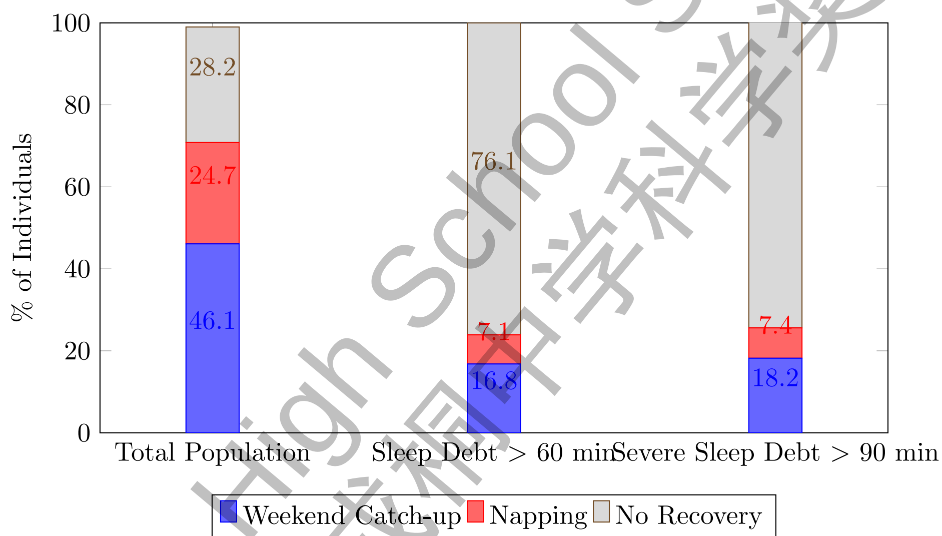


Figure 2.2: Sleep debt recovery strategies across population groups. Data from [6].

This bar chart shows that individuals suffering from severe sleep debt curiously make the least effort at compensating for the lost sleep debt. As those with lower amounts of sleep debt take more action, we can correlate this lack of initiative to taking action as a result of a lack of awareness or concern about it. It illustrates the need for free, accessible applications that plan for their recovery to alleviate the effects of sleep deprivation. Furthermore, an application accessible to all could promote the further dissemination of knowledge regarding sleep.

Effectiveness of Recovery Sleep

A study conducted by Mikael Sallinen and his colleagues proved that recovery sleep was the most effective way to improve performance. Another form of

recovery they tested was small breaks between tasks.[7] While this did prove slightly beneficial, recovery sleep drastically improved performance. However, it is important to note that recovery sleep never fully removes the effects of sleep debt, as the control group performed better on all occasions.
Furthermore, this study provides an important takeaway: recovery rates and the effects are different from person to person. This highlights the importance of the personalisation of the transition matrices.

Plateauing of Recovery

Another study [8] observed that after four consecutive nights of sleep restriction (5 hours/night), daytime sleepiness plateaued despite continued sleep loss. Upon one night of 10-hour sleep recovery, sleepiness levels returned to baseline; however, some effects persisted if recovery included only napping or partial sleep.

These findings support the idea that the body's recovery mechanisms are non-linear: sleepiness increases quickly during restriction but reaches a plateau, and recovery does not occur in full symmetry. In other words, gaining back lost sleep is neither immediate nor always proportional to the debt incurred.

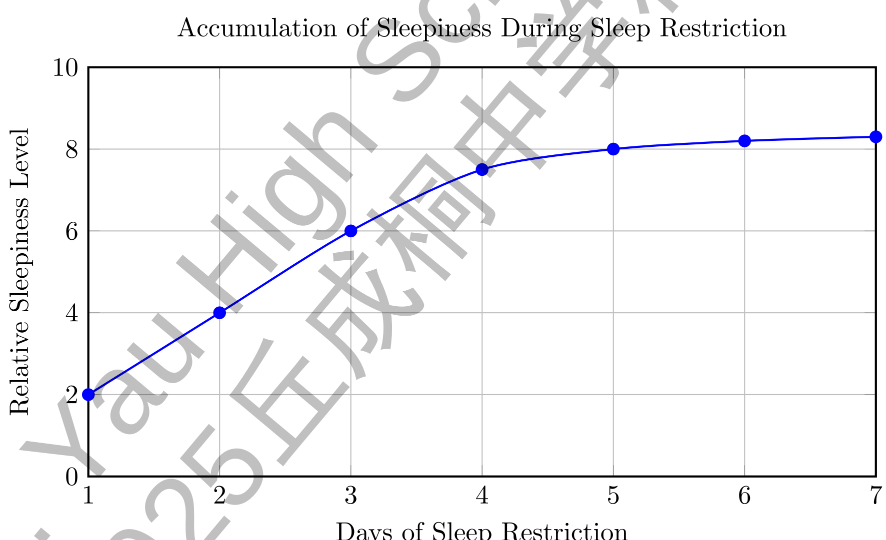


Figure 2.3: Sleepiness accumulates rapidly during the first few days of sleep restriction, but plateaus after approximately Day 4, based on Stanford Sleepiness Scale and Multiple Sleep Latency Test data [8].

2.2 Background and Current Methods in this Field

2.2.1 Non-invasive Sleep Tracking and Estimation Techniques

Ballistocardiography

This clinical technique, discovered in 1877, was one of the pioneer methods of tracking sleep, which is gaining resurgence. This method involves the detection of blood flow signals from the heart [9]. It is a non-invasive technique and can be integrated into beds or chairs, but the large number of sensors drives up the overall price and needs to be adjusted to account for many factors. [9]

Actigraphy

Actigraphy is a different yet widely supported technique for tracking sleep. Consisting of accelerometers, gyroscopes and magnetometers [9], this approach involves recording movements during sleep to determine whether the user is in a sleep state or wake state. Additionally, it can provide insight into an individual's circadian rhythms, but it cannot determine which sleep state the subject is in with much accuracy. On the other hand, it relies solely on the absence or presence of motion, provides limited data, and can also provide discomfort as it will be attached to either your wrist or arm [9]. Despite this, it is regarded as a non-invasive method.

Polysomnography

Polysomnography is the most accurate and insightful tool for monitoring sleep. It will collect EEG, EOG, EMG, nasal airflow, and pulse oximetry, which is recorded through the sensors [9]. Although classified as a non-invasive technique, the subject is required to stay overnight in a specialised, controlled environment [9]. Furthermore, the subject will be required to wear a multitude of sensors, which not only raise costs but also cause discomfort during sleep. As sensor data must be processed to provide a detailed sleep analysis, it is an inefficient method.

2.3 Sleep Modelling with Markov chains

Initially, a Markov Chain model was proposed by Zung et al. [10] for representing sleep patterns in 1965. However, as a standard Markov model assumes a geometric distribution [11] and sleep stages do not progress in an orderly fashion (emphasising more on specific stages and less on others), it proved to be inadequate.

In 1973, Mark Yang et al. [11] advanced this idea through a semi-Markov

Method	Advantages	Disadvantages	Invasiveness
Ballistocardiography	Can be integrated into beds or chairs; non-invasive	Requires multiple sensors; expensive; must be adjusted for individual factors	Non-invasive
Actigraphy	Tracks sleep/wake cycles; insights into circadian rhythm; portable	Cannot accurately detect sleep stages; relies on motion only; can cause discomfort	Non-invasive
Polysomnography (PSG)	Most accurate; detailed insight into sleep structure	Expensive; uncomfortable due to many sensors; must stay overnight in lab; inefficient data processing	Non-invasive

Table 2.2: Summary of Comparison of Sleep Tracking Techniques

model to describe sleep patterns. The benefit of using a semi-Markov model is that, unlike standard Markov models, it is not memoryless; the probabilities are weighted based on the length of the period for which the model has been at a certain state (e.g. average time spent in REM stage in the sleep process) to represent complex processes.

Advancements and refinements were made on the Semi-Markov model to improve efficiency and curve fitting. Wang et al. [12] tested out different fitting functions: Exponential Density, Power Law Density and Weibull Density functions to improve their representation of sleep patterns. This was tested by calculating the sleep stage distributions of the 244 hypnograms which they had access to and smoothing through Kernel Density estimation [12]. This involves removing local estimations and thus noise, to smooth the irregularity of the plotted hypnogram by averaging nearby sleep stage transitions over time.

Given a set of observed data points x_1, x_2, $\ldots$,x_n, the KDE at point x is given by:

$$\hat{f}h(x) = \frac{1}{nh} \sum i = 1^n K\left(\frac{x - x_i}{h}\right)$$

- K: kernel function
- h: bandwidth (smoothing parameter)
- x_i: observed data points (e.g., times at which stage transitions occur)

It will assign weights to points in the near vicinity based on the bandwidth and K value to represent a smoother curve during plotting, effectively blurring

the transitions across time.
This is extremely useful for analysing and extrapolating data.

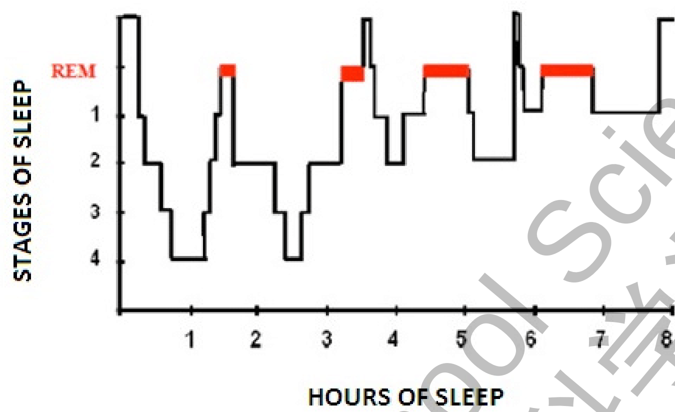


Figure 2.4: Hypnogram of a sleep cycle in a healthy young adult. Normal sleep involves cycling through stages of light sleep, deep sleep, and REM sleep approximately every 90 minutes.

2.4 Sleep deprivation through Artificial Intelligence and Machine Learning

While there are an extremely large number of drowsiness detection and fatigue detection research papers, most are inadequate for use. Firstly, these datasets only differentiate whether one feels drowsy or rested, which doesn't provide much insight. Additionally, the non-invasive datasets rely on inconsistent cues such as yawns or blinking frequency or eye closure, which are unreliable indicators of the level of sleep deprivation.

Furthermore, a study conducted by Benjamin et al. [13] in 2019 aimed to assess the effects of sleep deprivation on facial and skin features on 181 subjects, such as skin colour, eye openness, mouth curvature and periorbital darkness through objective measures.

The findings of this article showed that facial indicators alone are unable to classify sleep conditions accurately. Therefore, additional input will be required to scale the accuracy of the model.

It is important to mention that Lyu et al. [14] reached an accuracy of 90.05 % with the Long-term Multi-granularity Deep Framework approach on the NTHU

Driver Drowsiness detection dataset. However, this data set cannot be considered fully, as participants in this study actively tried to mimic sleep-deprived effects rather than actual sleep deprivation.

2.5 Conclusion

The effects of sleep debt are harmful and pose a threat to themselves and their surroundings due to the severe reaction time and health complications. Highlighting the need to spread awareness, there is an expansive array of methods to test sleep quality, which indirectly detect sleep (e.g. through motion) or cause discomfort. Solutions which have tried to solve this have reached impressive accuracy rates of up to 90.55%. Although this may be due to ethical concerns, many physiological and facial features of true sleep deprivation cannot be consciously mimicked: circles under eyes, dull skin tone, hanging eyelids, etc. This suggests that the model has been trained on features which do not correlate with sleep deprivation but are rather similar to the other participants in the study.

Miguel et al. tested a convolutional neural network in real-life scenarios and reached an accuracy of 72.23%, which can be optimised through the growing field of AutoML. The DROZY dataset used, which contains pictures primarily.

There was a gradual advancement from standard Markov models being inadequate to represent sleep patterns to semi-Markov models. After Wang et al. [12] tested and found the Weibull fitting function to perform best.

Markov chains have only been used to simulate sleep patterns, but not in adaptive recovery planning. Furthermore, Current literature lacks personalisation in Markov models. Often, a transition matrix is constructed from multiple hypnograms; however, this cannot accurately simulate recovery sleep and provide valuable insights to the users.

Chapter 3

Methodology

3.1 Proposed Approach

3.1.1 Sleep Deprivation Scoring model

Convolutional Neural Networks Overview

An image is provided as input to a convolutional neural network as a third-order tensor in the form

$$x \in \mathbb{R}^{H \times W \times D}$$

D represents the number of channels of matrices with the size $H \times W$, which in this case is 3 - Red, Green, and Blue. Furthermore, the matrix $H \times W$ contains red, green, or blue values [15].

A convolutional layer is made up of many convolutional kernels or filters. These filters are applied to the images that move across the image based on the stride. However, if $s > 1$, the convolution is executed horizontally and vertically on all s pixels [15].

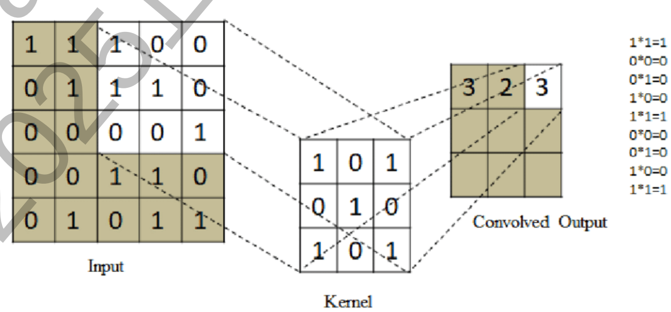


Figure 3.1: This figure shows how a kernel is applied to an image. [16].

Additionally, in mathematics, the convolution process is expressed by the following notation:

$$y_{i^{l+1},j^{l+1},d} = \sum_{i=0}^{H} \sum_{j=0}^{W} \sum_{d^l=0}^{D^l} f_{i,j,d^l,d} \times x^l_{i^{l+1}+i,\ j^{l+1}+j,\ d^l}$$

The left side: $y_{i^{l+1},j^{l+1},d}$ represents a single scalar value at the spatial location (i^{l+1}, j^{l+1}) in the d-th channel.

The right side: Looping over the height from row 0 to row $H-1$, the width of the filter from column 0 to column $W-1$ through each of the RGB channels, multiplying the receptive field $x^l_{i^{l+1}+i,\ j^{l+1}+j,\ d^l}$ which is a slice of the input volume matching the filter size. i^{l+1} and j^{l+1} correspond to the position of the filter in the input volume. i and j are added to each of them respectively to denote the position of the particular pixel in the channel d^l for a dot product multiplication with the filter $f_{i,j,d^l,d}$ (which represents the filter weight at row i, column j, input channel d) which will be added to the output feature map d.

These kernels help highlight certain image features, making it more easily recognisable (e.g. edge detection) and speeding up the learning process.

Activation functions are essential to the deep learning field for tasks such as Image recognition or Speech, etc. Activation functions prevent linearity as the linear functions are passed through a non-linear function. These non-linear functions can be much more easily modelled on complex data points which do not follow a straight line.

Here is an example:
- Layer 1: $z^1 = W^1 x + b^1$
- Layer 2: $z^2 = W^2 z^1 + b^2$

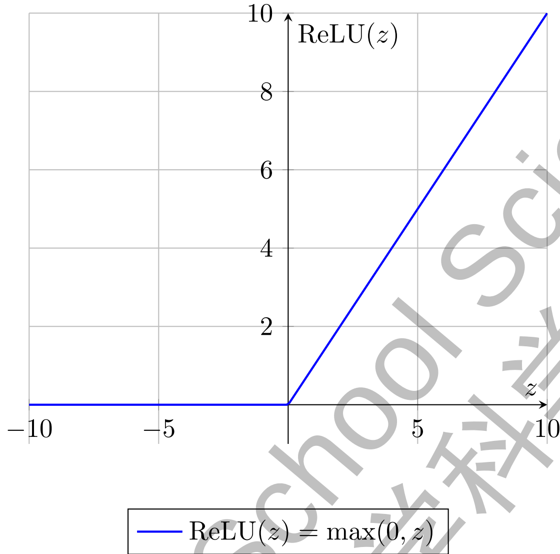
$z^2 = W^2(W^1 x + b^1) + b^2 = (W^2 W^1)x + (W^2 b^1 + b^2)$
$z^2 = W' x + b'$ (Still a linear function)

ReLU stands for rectified linear unit.

$$y = \mathrm{ReLU}(z) = \max(z, 0)$$

The reason why the ReLU activation function is so widely used is that it prevents the vanishing gradient problem. With tanh and sigmoid functions, for very large or small values, they are at either end of the range 0 to 1 for sigmoid and -1 to 1 for tanh. Thus, the output stops changing for very large and small values, making it "saturated" [17]. This creates a problem with learning and adjusting weights during backpropagation, as the derivative (providing the slope at that point) is 0.

ReLU avoids this vanishing gradient problem. As a ReLU function has a gradient of 1 when the output is > 0, the gradient does not vanish as it is not multiplied by a small number < 1.



For a reference:

Layer Type	Function / Description
Input Layer	Raw pixel values of the image as a tensor (Height*Width*Channels) $\in \mathbb{R}^{H\times W\times D}$
Convolutional Layer	Applies kernel filters to ease the extraction of local features from the input image. Sliding at the determined stride; Eventually outputs a feature map.
Activation Layer	Prevent linearity to support the modelling of complex data which does not follow a linear pattern
Pooling Layer	reduces the spatial dimensions of the feature maps, which decreases computational power, increases speed and helps prevent overfitting on the data.
Batch Normalisation	normalises the output of a previous layer to stabilise and speed up training as it prevents any of the inputs from weighting the final output more than usual
Dropout Layer	randomly drops a precise percentage of the neurons during training to reduce overfitting.
Dense Layer	Each neuron is connected to every neuron in the previous layer, which allows a combination of local features to produce high-level global features detected by the model
Output Layer	Produces final prediction probabilities. Softmax for multi-class, Sigmoid for binary classification.

Table 3.1: Summary of CNN Architecture Layers

Why CNNs Outperform Traditional Neural Networks in Image Processing

- The convolutional layers extract spatial features [18] by capturing local patterns in the input, allowing the model to recognise textures, edges, and shapes.

- The pooling layers introduce translation invariance [18], allowing the network to recognise objects regardless of their position or orientation. In contrast, traditional neural networks lack this adaptability due to their reliance on fixed input structures.
- CNNs offer spatial invariance, enabling robust object recognition even when images contain slight variations. Standard neural networks, on the other hand, require inputs to closely match training data, limiting generalisation. [18]
- Transfer learning - leveraging a model trained on large-scale datasets - greatly enhances the performance of CNNs by reducing the need for extensive task-specific data and training time.

Implementation in project

This will significantly improve the accuracy of the sleep deprivation scoring model to recognise certain facial features in comparison to traditional neural networks. CNN will recognise certain indicators of sleep deprivation on the face: hanging eyelids, redder eyes, darker circles under the eyes [2] and combine these results to form a sleep deprivation score.

Let $\mathcal{D} = \{(I_i, y_i)\}_{i=1}^{N}$ be a training set of N labeled facial images $I_i \in \mathbb{R}^{H\times W\times C}$, where each RGB color image is associated with a sleep deprivation label $y_i \in \{0, 1, 2\}$ representing low, moderate, and high levels of sleep deprivation respectively.

The objective is to train a convolutional neural network $f_\theta : \mathbb{R}^{H\times W\times C} \to \{0, 1, 2\}$, parameterized by θ, that maps an input image I to a predicted label $\hat{y} = f_\theta(I)$.

This prediction $\hat{y}$ acts as an estimated sleep deprivation score.

Data Preprocessing

The dataset used is the UTA Real-Life Drowsiness Dataset from the University of Texas at Arlington, which consists of 30 hours of RGB videos of 60 healthy participants. Each of the participants recorded a video for 3 different levels of sleep deprivation: Alert, Low vigilance and Drowsiness. All participants were adults and a vast variety of ethnicities from Caucasian to Hispanic (Non-white) to Indo-Aryan, Dravidian, Middle-Eastern and even East Asian. Furthermore, in this dataset, women represent 15% of all data (assuming equal video lengths). Moreover, this dataset also takes glasses, facial hair and different age categories from (20-59 years old) into consideration. Lastly, it was recorded in real-life environments and backgrounds, which makes it very realistic and useful for the model to identify sleep deprivation levels in the users, who will most likely not take a photo in a controlled environment. [19]

This data set is divided into 2 subsets: development groups: training, validation, and testing. This ensures that the testing process tests the model's learning capability rather than relying on memorised data and performs hyperparameter tuning as well. It was cleaned to remove any empty, unlabeled pictures, and a Label encoder was used for the x levels of sleep deprivation:x,t,y, all spatially scaled to fixed dimensions of 224x224.

As these were videos, every 20 frames, an image was extracted and labelled (based on the video title) before it was added to the dataset.

Data augmentation was applied only to the training images, using the ImageDataGenerator class from the Keras library. This data augmentation was applied only to the training images to improve the model's performance further when dealing with unseen images [20]. This augmentation involves rotations, varying levels of brightness, and a white noise filter [21].

Figure 3.2: Canny Edge Detection applied on a facial image.

Using Neural Architecture Search to optimise CNN performance

The architecture of the connected layers within a Convolutional Neural Network. If purely modified manually, determining the optimal structure for a convolutional neural network is extremely time-consuming and difficult. Neural Architecture search, a subfield of the growing AutoML field, generates high-performance models that are also memory-efficient to run on low-power devices through the use of reinforcement learning algorithms.

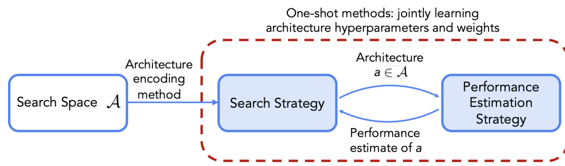


Figure 3.3: Represents the breakdown of how Neural Architecture search is conducted [22].

The search space contains all possible architectures that can be represented in a principle that can be expanded to extremely large sizes. Although these search spaces can "incorporate domain knowledge" [22] to simplify the search by setting some boundaries for what to search for, too many restrictions can prevent a NAS method from finding the highest-performance models.

Additionally, Reinforcement learning, part of the black box optimisation-based techniques, is used as the search strategy to find the highest-performing CNN architecture models.

Moreover, the performance estimation strategy is the method which is used to quickly predict how the model would perform on the dataset. It would be too computationally expensive and inefficient for the computer to train the model on the dataset and analyse its accuracy on the test data.

The concept behind NASNet, introduced by Zoph et al. [23], is that the majority of handmade model architectures are designed with repetitive microarchitectural blocks, which are reused multiple times throughout the network. This method reduces complexity and improves performance by reducing the search space much more effectively.

The policy gradient is a method used to optimise the policy in a reinforcement learning algorithm. The function of a policy $\pi(s)$ determines the optimal action to be taken in a certain state to maximise the total rewards [24]. However, the research paper requires 100s of hours of GPU computational power.

After careful evaluation, AutoKeras, which acts as a superior alternative, is a library which runs on significantly less computational power through the use of methods such as hyperparameter tuning and designing a neural network architecture through random search. This method entails defining a search space (ranges and limits of possible parameter values) and randomly picking combinations to try. Although an unorthodox method, it has proven to perform very efficiently.

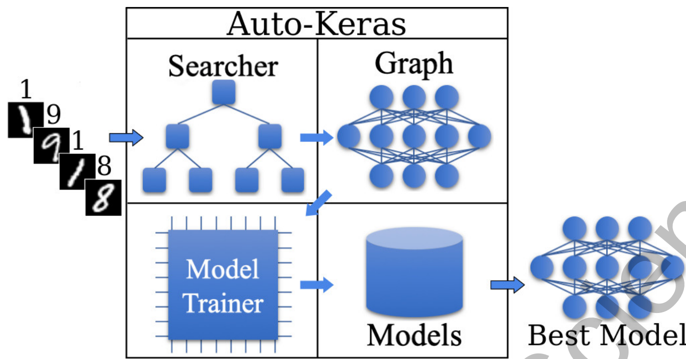


Figure 3.4: Figure displaying the flowchart of an AutoKeras class [25].

This is mainly because there is a small number of parameters which greatly affect performance. Grid search involves testing combinations with all parameters; it is a much more exhaustive effort, which does not yield as great results. Additionally, using a random combination almost ensures that the entire search space is covered, which cannot be guaranteed through grid search.

3.1.2 Recovery of sleep through a Semi-Markov chain

A Markov chain is a stochastic process that satisfies the Markov property: the future state of the process only depends on the current state and not upon past states. [26]

$$X_{t+1} \text{ depends only on } X_t$$

Key terms

- The state of a Markov chain at time t is the value of X_t,
- The state space of a Markov chain, S, is the range of values X_t can be assigned to. The number of elements in the set S is in $\in \mathbb{N} \cup \{\infty\}$
- The transition matrix is a matrix of the probability of moving from one state to another in a single step.
- A trajectory in a Markov chain is the path that the Markov chain takes. Eg: The trajectory up to time $t = 2$ is 1,2,3 with $X_0 = 1$, $X_1 = 2$ and $X_2 = 3$ as two steps occurred.

The Markov property can be written in mathematical notation as follows:

$$P(X_{t+1} = s \mid X_t = s_t, X_{t-1} = s_{t-1}, \ldots, X_0 = s_0) = P(X_{t+1} = s \mid X_t = s_t)$$

This equation in literal form equates:
(1) The probability of the variable X at time $t+1$ from state s_t to s given the entire history of the trajectory from time 0 to t

(2) The probability of the variable X at time $t+1$ from state s_t to s given only the state at time t.

It can be inferred that the current state only depends on the previous state and not the entire trajectory. Only the value of the previous step x_t with the value s_t is considered, as the only difference in the two equations is the state values of the variables X_0 to X_{t-1}, and since their inclusion does not affect the outcome, they are effectively irrelevant to the transition.

3.1.3 Transition Matrix

A transition matrix is an array containing the probabilities of moving from 1 state to another. The rows correspond to the current state i, and each column corresponds to the next state j

	PLAY	PAUSE	JUMP	FFW	RWD	STOP
PLAY	0,9	0,08	0	0	0	0,02
PAUSE	0,21	0,54	0,07	0,08	0,05	0,05
JUMP	0	1	0	0	0	0
FFW	0	0,88	0	0,12	0	0
RWD	0	0,85	0	0	0,15	0
STOP	0	0	0	0	0	1

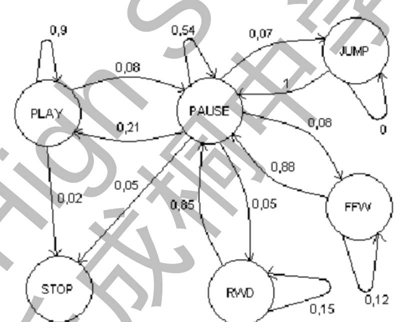


Figure 3.5: This showcases the transition matrix produced from the Markov chain diagram [27].

The sum of the probabilities in each row of a transition matrix must be equal to 1 [26]. This is because, assuming $P(i,j)$ and the current state is i, a transition must occur, whether it is from i to i or i to j in the next step.

The formula can simply represent the transition probability:

$$p_{ij} = \mathbb{P}(X_{t+1} = j \mid X_t = i), \quad \text{for } i, j \in S,\ t = 0, 1, 2, \ldots$$

This formula is the probability of going from state i to state j in one step [26]. The transition matrix is filled with these data points.

In addition to the transition probabilities, a Markov chain also requires an initial state distribution, which tells us the probability of starting in each state

at time t = 0.
This distribution is usually denoted by a vector $\boldsymbol{\pi}^{(0)}$, where: $\pi_i^{(0)} = \mathbb{P}(X_0 = i)$
For each state $i \in S$, the sum of all entries in this distribution must equal 1: $\sum_{i \in S} \pi_i^{(0)} = 1$.
To end, using matrix multiplication in this formula $\boldsymbol{\pi}^{(t+1)} = \boldsymbol{\pi}^{(t)} P$ will showcase the full behaviour of the chain over time. [26]

Semi-Markov Decision Processes

By analysing hypnograms affected by a variety of factors, Markov models can be constructed according to the distribution of the sleep stages from a hypnogram plot. Since this is a semi-Markov model, which does not follow geometric distributions with fixed periods for each epoch, the amount of time spent in each state before transitioning to another is referred to as the sojourn time.

Implementation in project

In this project, Markov models are used to simulate transitions between different sleep stages during recovery after sleep deprivation. Adapting the transition matrix to a user's sleep debt and personal information, the system creates a personalised sleep pattern. By running these simulations, the model can predict the optimal recovery sleep schedule for an individual. This probabilistic approach allows the system to generate adaptive and data-driven recommendations for regaining lost sleep.

Semi-Markov model for Sleep Recovery

The literature review provided valuable information about the feasibility of this sleep modelling between a semi-Markov chain and a Markov chain. For constructing the transition matrix used to model sleep, the base transition matrix was mapped to the "Alert" state, which will be used from the work of Wang et al. [12], who built a Semi-Markov model through analysis of 244 hypnograms (122 male, 122 female) ranging from 20 to 85 years of age.

Table 3.2: Alert-state Sleep Stage Transition Matrix

From \ To	Wake	NREM	REM
Wake	0	0.9632	0.0368
NREM	0.8093	0.0000	0.1907
REM	0.6655	0.3345	0

Additionally, a transition matrix is needed for mapping to the "Drowsiness" state. Through the support of Väinö Jääskinen et al. [28] using a dataset first mentioned by Sallinen et al. [7], a transition matrix was produced solely on hypnograms of recovery sleep of drowsy individuals.

For this transition matrix, assuming there were n state transitions between state p and state s of an epoch interval $[1, x]$, the transition probability $= n/x$, which is equivalent to its relative frequency.

As the previous transition matrix had condensed the stages S1, S2 and SWS, and we could not revert the previous Transition matrix from NREM to S1, S2, and SWS, the transition matrix needed to be collapsed from a 5-state to a 3-state matrix.

Table 3.3: Drowsy-state Sleep Stage Transition Matrix

From \ To	**Wake**	**NREM**	**REM**
Wake	0.4645	0.4905	0.0450
NREM	0.0172	0.9703	0.0124
REM	0.0170	0.0440	0.9390

Addressing the transition matrix for the "Low vigilance" state, it is calculated using convex mixing. An input x represents the sleep deprivation score as a decimal, which will be entered. This sleep deprivation score is obtained through the output of the sleep-deprivation level image classification software.

$$P(x) = (1 - x)(P_a) + x(P_d)$$

Where $P(x)$ represents the transition probability for "Low vigilance", P_a represents the transition probability for "Alert", and P_d represents the transition probability for "Drowsiness".

The user will self-report a variety of personal details. This can be used for the personalisation of the transition matrices.
Furthermore, justified by research papers, a variety of factors apart from sleep debt affect sleep quality and the transition states. Bump factors will be assigned to each state change, and then the data will be normalised for the probabilities in a single row of a Markov model to add up to 1. These bump factors are all relative to a "normal" night of sleep.

Firstly, with age, as an individual's age approaches 60, they have a reduced probability of falling asleep quickly (Wake to NREM x0.9), increased probability of remaining awake throughout the night (Wake to Wake x1.1), increased likelihood of NREM to Wake (x1.2) and REM to Wake (x1.1) [**?**] as a result of individuals above the age of 60 as more receptive to interruptions. [**?**].

In addition, stress has proven to be a major deterrent to sleep health, leading to frequent awakenings from NREM to Wake (x1.2), delayed sleep onset (x0.9) and shorter REM cycles, which increases REM to NREM (x1.1) or REM to Wake (x1.2) [29].

Moreover, the circadian rhythm alignment is a key factor in sleep quality. Thus, sleeping at adverse circadian phases leads to harder initiations of sleep at these phases [30] (lowering Wake to NREM x0.9) and increasing REM to Wake state transitions (x1.1) [30].

Equally important is the individual's gender. Studies have shown women tend to fall asleep quicker [31] (Wake to NREM x1.05) and have fewer awakenings as well (NREM to Wake x0.9). As women tend to cycle into REM earlier, it tends to be more often as well (NREM to REM x1.1) [31].

Lastly, any form of medication perturbs sleep transitions. Sedatives lead to fewer awakenings (NREM to Wake x0.8) and likewise, make it easier for sleep onset (Wake to NREM x1.1). SSRI Antidepressant-treated sleepers have fewer REM episodes (NREM to REM x0.7) [32].
As mentioned previously in the Literature review, Weibull fitting functions will be utilised in this semi-Markov chain for realistic bouts of sleep.

Once the percentages of time spent in each sleep stage are determined, these values are compared to a standard "healthy" distribution, which is generally considered to reflect an optimally recovered night. The percentages of time will be multiplied by the epochs or conversely by the total time to understand how much time was spent in each sleep stage.

The participant's observed time per stage is then subtracted from these baseline values to quantify the residual sleep debt in each stage. These differences (wholly representing the sleep debt) are standardised, ensuring that deviations in NREM, REM, or Wake are weighted appropriately by their relative contribution to physiological recovery.

This process is repeated across consecutive nights, allowing stage-specific deficits to be accumulated over time and used to update a sleep recovery score. This score decreases as the individual approaches full recovery. Notably, sleep debt often plateaus after several nights of recovery, indicating that certain deficits cannot be immediately compensated and that the recovery process is inherently non-linear, as referenced in section 2.1.2 on the plateauing of recovery sleep.

By converting stage-wise sleep periods into a quantitative recovery metric, this approach enables the model to simulate realistic recovery trajectories. Consequently, it provides a basis for generating personalised recommendations based on individual recovery progress and allows the user to be classified as fully recovered from sleep debt.

The Google Calendar API can be used for scheduling these bouts to recover from the sleep debt through naps or longer sleep durations based on the scheduling of the calendar and its precision.

Chapter 4

Experiment

4.1 Sleep Deprivation Model

In this experiment, we will test the accuracy of multiple models on the customised UTA-RLDD dataset, which was used and maximise the accuracy to reach the heights of the state-of-the-art models currently used.

4.1.1 Hypothesis

I believe that I can train and create a light-weight model which reaches a similar performance level to the currently best running models.

4.1.2 Training Environment

The experiments were conducted using Google Colaboratory's cloud-based environment, accessed from a local machine running macOS Sequoia 15.1 with 16 GB of RAM storage, accelerated by an NVIDIA T4 GPU. This model was implemented with TensorFlow, Keras, Numpy, Sci-kit learn, and Matplotlib frameworks.
The network was trained using the Adam Optimiser over 100 epochs with a batch size of 32. This prevents the risk of generalisation and facilitates the process of escaping local minima. An added The dataset is split into a 50/30/20 split (Train, Test and Validation). The key differentiator between Validation data and Test data is that Validation is used during the training of the model for hyperparameter tuning, while Test data is applied to the model after it is frozen.
ReLU activation functions were introduced after each layer in the neural network, with a softmax function at the end, since all the outputs are positive and the total sum is 1.

Furthermore, the loss function used is the Categorical Cross-entropy. This is often used for multiple mutually exclusive states and outputs a probability distribution. [33]

$$L_i = -\sum_{k=1}^{K} y_{i,k} \log\left(\hat{y}_{i,k}\right)$$

This formula displays the workings of the categorical cross-entropy function for:

- L_i: Loss for the i-th sample
- K: total number of states
- $y_{i,k}$: Truth value
- $\hat{y}_{i,k}$: Predicted probability for class k from the model for sample i

If the model predicts 100% accuracy for a class (which is a probability of 1), and as $y_{i,k}$ will only produce a value of 1 if the correct class is chosen (thus rendering the rest as 0), for a sample i, the loss can just be represented as $-\log(\hat{y}_{i,k})$. Thus, this would just return $-\log(1) = 0$ to show 0 loss. Additionally, as the model predicts a small probability for the correct class, the loss value becomes a larger positive value because of the negative sign. This allows for model improvement with validation data during training.

4.1.3 Model Architectures

As established in the Literature Review, it is difficult to capture sleep deprivation from photographs alone. This was largely due to the methods followed and the quality of the datasets, where often the reactions were mimicked but did not focus on true sleep deprivation.

After collecting 10,000+ images with extensive diversity in terms of race, ethnicity, accessories and facial hair, different models were tested to optimise the accuracy which can be achieved. This included: A MobileNetV2 model and a variety of models tested using the AutoKeras library.

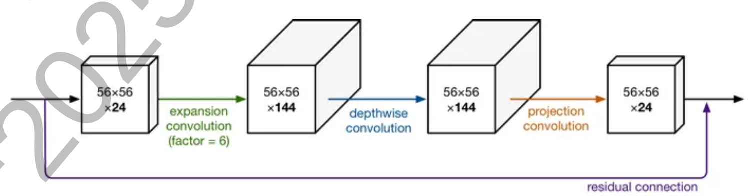


Figure 4.1: MobileNet V2 Architecture

The maxtrials parameter in the AutoKeras library was set to 3 as a result of the computational power required to run. Furthermore, the random seed was set to 42, which allows for consistency throughout the processing, as for all other parameters, such as random search for hyperparameters in tuners or weight initialisation. This is crucial as it allows for reproducibility and observation of the factors affecting the model without any biases.

Furthermore, several steps were added after the initial tries due to excessive overfitting, which occurred. For this dataset, the images are grouped based on the individuals to prevent data leakage, as the model may learn to recognise the faces itself rather than the sleep deprivation levels using the GroupShuffleSplit() function along with others. Early Stopping was also introduced, which prevents further training of the model after determining whether overfitting is occurring.

4.2 Semi-Markov model for recovery sleep

In this experiment, we will test the similarity of a sleep simulation from the Semi-Markov model and a Hypnogram produced from the analysis of a Polysomnograph by a medical professional.

4.2.1 Hypothesis

I believe that I can produce an accurate personalised semi-Markov chain to reflect sleep patterns with individuals.

4.2.2 Dataset

The Polysomnographs and Hypnograms are available at the Sleep-EDF Database. This data set represents the ground truth against which the semi-Markov model will be compared. These were obtained in a 1987-1991 study of the effects of age on sleep in healthy Caucasian adults 25 to 101 years, exempt from any sleep medication [34].

Polysomnographs of about 20 hours each were recorded during two subsequent day-night periods at the subjects' homes. We will be using the 2nd day, as users' sleep may have been affected by the cassette tape attached to them on the 1st night more than on the 2nd night.

Data Preprocessing

Furthermore, the hypnogram records the stages in the R&K format: W, 1, 2, 3, 4, R, M, ? while our transition matrix uses Wake, NREM, REM. Thus, for comparison, a sleep staging map was constructed for converting the hypnogram into a sleep sequence. The extraction of data from the .edf file was done through the MNE-Python library, a library specialised for EEG signal analysis and visualisation and saved to a .csv file.

It was necessary to identify the sections where sleep began and remove the wake stages before that so that the sleep could begin at the correct position and index it to 960 epochs (as each epoch is 30 seconds, 960 epochs is 8 hours). To accomplish this, the EDFBrowser must be used to upload hypnograms and polysomnograms for analysis.

The factors which we can utilise for personalisation are:

- **Age**
- **Gender**
- **Sleep medication (No effect as none was taken)**

4.2.3 Semi-Markov chain simulation

The transition matrix was constructed through analysis by Wang et al. [12], as there is no mention of sleep deprivation within the subjects. Furthermore, since this occurs within their home, sleep deprivation caused by discomfort is less likely. The bump factors were previously defined in section 3.1.2; however, only the age and the gender factors can be utilised. Using the applybumpsandnormalize() function, which we defined, each row will add up to 1. For Weibull fitting, which closely approximates the shape of sleep duration distributions, the Sci-Py library was used along with the parameters which were passed into the function.

To optimise these parameters over a variety of individuals, the Weibull function was fitted to each of them. However, they were adapted to individuals with those specific conditions. It was used for fine-tuning, however. Additionally, floor probabilities were implemented to prevent stages from occurring altogether.

Lastly, to determine if the results are "good", we will be comparing through Absolute Percentage error as our metric.

Chapter 5

Results

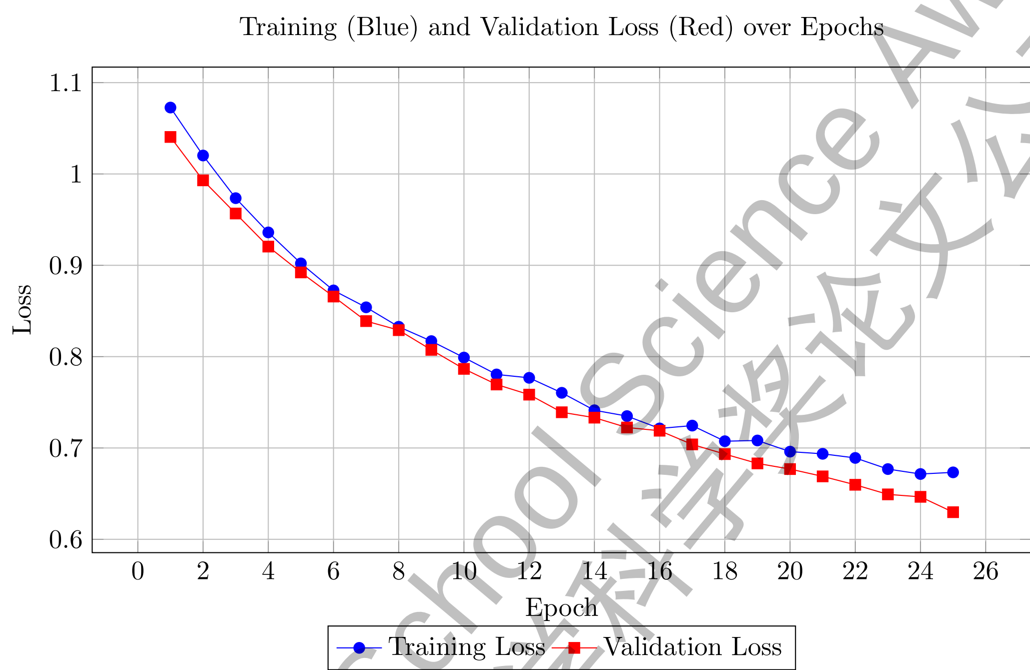
5.1 Sleep Deprivation

5.1.1 MobileNetV2 Performance

The MobileNetV2-based CNN was trained for 10 epochs using the preprocessed dataset. Training began with an accuracy of **40.49%** and validation accuracy of **56.92%** in the first epoch. Throughout the training, both of the metrics improved steadily, with the final epoch achieving **74.73%** training accuracy and **85.16%** validation accuracy.

The loss also showed a consistent downward trend, decreasing from **1.0727** (training) and **1.0405** (validation) in epoch 1 to **0.6733** and **0.6297**, respectively, by epoch 25. This indicates that the model was learning effectively without severe overfitting, as the validation loss had very similar values to the training loss, while following the same pattern through each epoch as well.

When evaluated on the test model, it still performed extremely well, with an accuracy of **0.791%**, which is key and shows that the MobileNetV2 generalised very well on the data provided.



5.1.2 AutoKeras Performance

We implemented the AutoKeras library to experiment with a variety of different models. Initially, the trials would reach 98-99% accuracy within the 1st epoch, showing clear signs of overfitting, often caused by a large number of parameters increasing the width of each layer. To prevent this, the GroupShuffleSplit function was utilised to prevent data leakage. Moreover, the EarlyStopping-Function() restores the optimal weights after quitting training, which we defined through automatic monitoring of validation data loss values. Lastly, an L2 regularizer and Dropout were added to the model layers; however, this would only come into effect after the model was trained.

5.2 Sleep Modelling through Markov chains

5.2.1 Subject 33

Subject 33 was male gender and 60 years old. This subject was chosen as it acts as a subject that best reflects the study population, as the median participant age is 57, the average age is 59, and this study consists mainly of the male gender.

Table 5.1: Comparison of simulated hypnogram to EDF hypnogram. Percentage error is calculated as (Simulated − EDF)/EDF × 100.

Sleep Stage	**EDF (%)**	**Simulated (%)**	**Percentage Error (%)**
Wake	21.98	15.83	27.9
NREM	67.29	76.98	-14.4
REM	10.73	7.19	33.0

5.2.2 Subject 9

Subject 9 was female gender and 25 years old. This subject was chosen due to their young age to provide diversity to the study to capture the reliability of the results.

Table 5.2: Comparison of simulated hypnogram to EDF hypnogram. Percentage error is calculated as (Simulated − EDF)/EDF × 100.

Sleep Stage	**EDF (%)**	**Simulated (%)**	**Percentage Error (%)**
Wake (W)	3.23	3.65	13.0
NREM (N)	69.17	80.94	17.0
REM (R)	27.60	15.42	-44.1

5.2.3 Subject 4

Subject 4 was female gender and 34 years old. This subject was chosen due to their gender to differentiate between the accuracy of simulations between males and females and further increase the diversity and reliability of the results.

Table 5.3: Comparison of simulated hypnogram to EDF hypnogram. Percentage error is calculated as (Simulated − EDF)/EDF × 100.

Sleep Stage	**EDF (%)**	**Simulated (%)**	**Percentage Error (%)**
Wake (W)	6.67	3.96	-40.7
NREM (N)	70.83	75.52	6.6
REM (R)	22.50	20.52	-8.8

5.2.4 Subject 66

Subject 6 was female gender and 101 years old. This subject represents an anomaly in age and was used to test the upper bounds of the dataset for verification of the semi-Markov simulations in extreme cases.

Table 5.4: Comparison of simulated hypnogram to EDF hypnogram. Percentage error is calculated as (Simulated − EDF)/EDF × 100.

Sleep Stage	EDF (%)	Simulated (%)	Percentage Error (%)
Wake (W)	11.04	13.85	25.44
NREM (N)	66.67	54.27	-18.64
REM (R)	22.29	31.88	43.05

Chapter 6

Discussion

6.1 Sleep Deprivation model results

For detecting sleep deprivation, due to problems caused by the AutoKeras model despite the precautions listed in 5.1.2, the AutoKeras had very noisy loss shifts and a large disparity between the validation and training accuracy, as shown in the Figure below.

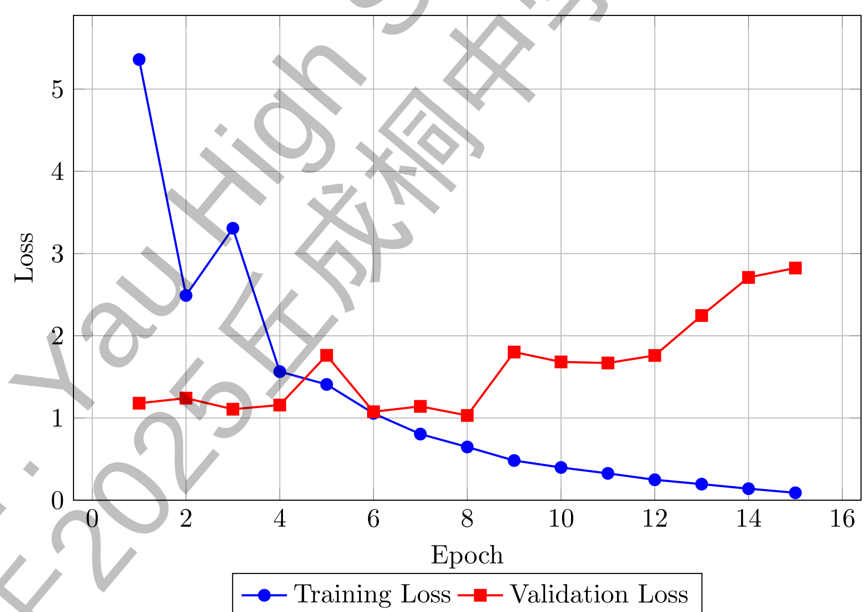


Figure 6.1: Training and validation loss over epochs for a trial using AutoKeras

Thus, as MobileNetV2 has impressive scores of 85%+ validation accuracy and approximately 80% test accuracy, similar to Lyu et al. [14] who reached

90% accuracy while using a dataset which mimicked sleep-deprived individuals driving. As mentioned in section 2.4, this may not have been detecting the correct features, as some of the features cannot be mimicked and would thus not show up in the dataset, which represents a stark contrast to the dataset we had constructed.
Additionally, the hypothesis was achieved and the goal was met, which labels this as a success, and as an advantage, using MobileNetV2 makes it much more feasible to run on phones, which would be required of this application.

6.2 Sleep Modelling through Markov chains results

Taking the average percentage errors across all the subjects of vast diversity with the given factors, is: 24.3%. This is a phenomenal performance by the personalised semi-Markov chains, considering that the top-band performers such as the Oura Ring, Fitbit Sense, Apple Watch, etc. have 71-76% accuracy [35, 36]in classifying sleep stages. This means that our probabilistic model, requiring no use of any wearables or sensors, has reached a similar accuracy to the Oura Ring, the current top contender of all commercial products. While this is a comparatively smaller dataset, it has quite a bit of diversity and small differences, such as in the REM stage, have caused high levels of inaccuracy (of 32.2%, which is very similar to the performance of the Fitbit Sense) whilst still outputting similar values.

Our hypothesis has been achieved, and this is exemplified due to the cost and discomfort that the user does not face. It is extremely important to note that this study was able to consider solely 2 factors: Age and Gender, which prevented further personalisation to the user for higher accuracy levels.

6.3 Limitations

One of the primary limitations of this study is the vast expanse of factors which influence sleep deprivation and recovery. However, due to the lack of equipment, only certain select factors can be considered and emphasised in Markov chains. Furthermore, gaps in existing sleep science literature may have ruled out possible factors contributing to recovery sleep, which could be limiting the potential of the decision processes.

Additionally, this study was limited by the absence of comparative performance analysis. Without a benchmark comparison to the 'gold-standard' methods of polysomnography, it is difficult to quantify the accuracy of recovery sleep. Future work on this topic would need to include a comparison with other methods.

The number of subjects for the experiment was also limited due to the time-intensive nature of personalised simulation and validation. Future work will require taking a larger number of subjects for proper verification.

Moreover, the dataset used is based on the Stanford Sleepiness Scale, which is self-reported by the user. In some cases, this has also led to the sleepiness case reducing over long periods of sleep deprivation [37].

6.4 Future Work

Recovery is core for athletes, and sleep optimisation can boost performance. However, most athletes do not have the resources to use high-end technology for recovery, as it is prioritised towards training. Only 14.9% of athletes who competed in the Olympic Games received direct funding from the IOC scholarship program, who are elite-level athletes. This application could prove extremely useful to an athlete as it is free and causes no discomfort.

Furthermore, truck drivers are often sleep-deprived, with the mean daily sleep duration being $5.6 + / - 1.3$ hours [38]. Such prolonged sleep restriction compounded with intensive 10-hour workdays further exemplifies sleep deprivation levels [38]. This can lead to fatal accidents as a result of sleep deprivation, which has been proven to affect human reaction times critically. Thus, this sleep recovery application can act as a benchmark for truck drivers to determine if they are ready for long-hour drives.

Similar to the truck drivers, many healthcare clinics and hospitals could drastically benefit from this. A recent Medical Defence Union survey [39] of doctors found that nearly 90% of respondents felt sleep-deprived at work. Of those, 41% experienced sleep deprivation at least weekly, and 35% said their tiredness had impacted their ability to treat patients safely [39]. This shows a critical flaw which could worsen and potentially prove fatal to the patients of these doctors. Thus, this sleep recovery application can also act as a benchmark for doctors to determine whether they can successfully perform to the necessary standards.

Lastly, in general, all workers and students can perform more efficiently if this issue is tackled through our application. This can also spread awareness to Governments, schools and companies to issue new policies to prevent further harm to more than 2/3rds of the entire global population [40] according to the International Labour Organisation from 2024.

Chapter 7

Conclusion

This research study aimed to develop a personalised sleep recovery model using CNN-based sleep deprivation scoring and semi-Markov chain simulations for optimisation of recovery sleep patterns. Unlike existing sleep trackers, which are often costly and invasive, this novel approach provides an accurate, no-cost, and fully non-contact solution through a simple application. Through integration of deep learning combined with adaptive probabilistic modelling, it offers practical applicability in the real world for populations where these expensive devices are often inaccessible. The novelty lies in the sleep recovery system, which strategically provides sleep recovery guidance available to a much wider community. Lastly, this work is scalable and can be deployed widely without specialised equipment and can produce further work and research into other solutions.

I have achieved my goal in creating a realistic prototype of an affordable, non-invasive yet effective Sleep tracker, which has the potential to improve the lives of billions cognitively, physically, emotionally and socially.

Bibliography

[1] A. K. Patel, V. Reddy, K. R. Shumway, and J. F. Araujo, "Physiology, sleep stages," 2025.

[2] T. Sundelin, M. Lekander, G. Kecklund, E. J. W. Van Someren, A. Olsson, and J. Axelsson, "Cues of fatigue: effects of sleep deprivation on facial appearance," *Sleep*, vol. 36, no. 9, p. 1355–1360, 2013.

[3] D. F. Dinges, F. Pack, K. Williams, K. A. Gillen, J. W. Powell, G. E. Ott, C. Aptowicz, and A. I. Pack, "Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night," *Sleep*, vol. 20, pp. 267–277, 04 1997.

[4] Miscellaneous.

[5] A. N. Hudson, H. P. A. Van Dongen, and K. A. Honn, "Sleep deprivation, vigilant attention, and brain function: a review," *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*, vol. 45, no. 1, p. 21–30, 2020.

[6] D. Leger, J.-B. Richard, O. Collin, F. Sauvet, and B. Faraut, "Napping and weekend catchup sleep do not fully compensate for high rates of sleep debt and short sleep at a population level (in a representative nationwide sample of 12,637 adults)," *Sleep Medicine*, vol. 74, pp. 278–288, 2020.

[7] M. Sallinen, A. Holm, J. Hiltunen, K. Hirvonen, M. Härmä, J. Koskelo, M. Letonsaari, R. Luukkonen, J. Virkkala, and K. Müller, "Recovery of cognitive performance from sleep debt: Do a short rest pause and a single recovery night help?," *Chronobiology International*, vol. 25, no. 2-3, pp. 279–296, 2008. PMID: 18533327.

[8] M. A. Carskadon and W. C. Dement, "Cumulative effects of sleep restriction on daytime sleepiness," *Psychophysiology*, vol. 18, no. 2, p. 107–113, 1981.

[9] Z. Hussain, Q. Z. Sheng, W. E. Zhang, J. Ortiz, and S. Pouriyeh, "A review of the non-invasive techniques for monitoring different aspects of sleep." 2021.

[10] W. W. Zung, T. H. Naylor, D. T. Gianturco, and W. P. Wilson, "Computer simulation of sleep eeg patterns with a markov chain model," *Recent advances in biological psychiatry*, vol. 8, p. 335–355, 1965.

[11] M. C. Yang and C. J. Hursch, "The use of a semi-markov model for describing sleep patterns," *Biometrics*, vol. 29, no. 4, p. 667–676, 1973.

[12] C. Wang, S. A. Alvarez, C. Ruiz, and M. Moonis, "Computational modeling of sleep stage dynamics using weibull semi-markov chains," in *Proceedings of the International Conference on Health Informatics*, SciTePress - Science and and Technology Publications, 2013.

[13] B. C. Holding, T. Sundelin, P. Cairns, D. I. Perrett, and J. Axelsson, "The effect of sleep deprivation on objective and subjective measures of facial appearance," *Journal of sleep research*, vol. 28, no. 6, p. e12860, 2019.

[14] J. Lyu, Z. Yuan, and D. Chen, "Long-term multi-granularity deep framework for driver drowsiness detection," 01 2018.

[15] J. Wu, "Introduction to convolutional neural networks," tech. rep., LAMDA Group, National Key Laboratory for Novel Software Technology, Nanjing University, 2017. Technical report, May 1, 2017, available at the author's website.

[16] R. Dhiman, G. Joshi, and R. Challa, "A deep learning approach for indian sign language gestures classification with different backgrounds," *Journal of Physics: Conference Series*, vol. 1950, p. 012020, 08 2021.

[17] A. T. i. Machines, "Machines of this character can behave in a very complicated manner when the number of units is large."."

[18] R. S. Chouhan, "Unleashing the superiority of cnn in image processing compared to other neural networks," May 2023.

[19]

[20]

[21] N. Deshmukh, "Low-cost device prototype for automatic medical diagnosis using deep learning methods," in *2018 9th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, p. 695–699, IEEE, 2018.

[22] C. White, M. Safari, R. Sukthanker, B. Ru, T. Elsken, A. Zela, D. Dey, and F. Hutter, "Neural architecture search: Insights from 1000 papers." 2023.

[23] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," 2018.

[24] L. Weng, "A (long) peek into reinforcement learning," Feb. 2018.

[25]

[26] U. of Auckland.

[27] S. Mongy, C. Djeraba, and D. Simovici, "On clustering users' behaviors in video sessions," pp. 99–103, 01 2007.

[28]

[29] D. A. Kalmbach, J. R. Anderson, and C. L. Drake, "The impact of stress on sleep: Pathogenic sleep reactivity as a vulnerability to insomnia and circadian disorders," *Journal of sleep research*, vol. 27, no. 6, p. e12710, 2018.

[30] S. W. Wurts and D. M. Edgar, "Circadian and homeostatic control of rapid eye movement (rem) sleep: promotion of rem tendency by the suprachiasmatic nucleus," *The Journal of neuroscience: the official journal of the Society for Neuroscience*, vol. 20, no. 11, p. 4300–4310, 2000.

[31] Y. Fatima, S. A. R. Doi, J. M. Najman, and A. A. Mamun, "Exploring gender difference in sleep quality of young adults: Findings from a large population study," *Clinical medicine research*, vol. 14, no. 3–4, p. 138–144, 2016.

[32] A. Wichniak, A. Wierzbicka, M. Walecka, and W. Jernajczyk, "Effects of antidepressants on sleep," *Current psychiatry reports*, vol. 19, no. 9, p. 63, 2017.

[33] Sept. 2024.

[34] M. S. Mourtazaev, B. Kemp, A. H. Zwinderman, and H. A. Kamphuisen, "Age and gender affect different characteristics of slow waves in the sleep eeg," *Sleep*, vol. 18, no. 7, p. 557–564, 1995.

[35]

[36]

[37] M. García-García, A. Caplier, and M. Rombaut, "Sleep deprivation detection for real-time driver monitoring using deep learning," *International Conference on Image Analysis and Recognition*, p. 435–442, 2018.

[38] R. S. N. de Pinho, F. P. da Silva-Júnior, J. P. C. Bastos, W. S. Maia, M. T. de Mello, V. M. S. de Bruin, and P. F. C. de Bruin, "Hypersomnolence and accidents in truck drivers: A cross-sectional study," *Chronobiology international*, vol. 23, no. 5, p. 963–971, 2006.

[39] Health, "Doctors more sleep deprived now than after the pandemic - the mdu."

[40] Apr. 2024.