Detection of Obstructive Sleep Apnoea using Features Extracted from Segmented Time-Series ECG Signals with a One Dimensional Convolutional Neural Network

By

Steven Robert Thompson

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy

January 2025

Table of Contents

List of Figures 5

List of Tables 6

Acknowledgments 7

Abstract 9

Acronyms 10

List of Publications Resulting from this Study 13

1. Introduction 14

1.1 Aims and Objectives 18

1.2 Unique Contribution 20

1.3 Motivation 21

1.4 Thesis Structure 22

2. Sleep Disorders 24

2.1 Classifications of Sleep Disorders 24

2.2 Types of Sleep Disorders 25

2.2.1 Sleep-related breathing disorders 25

2.2.2 Insomnia 25

2.2.3 Central disorders of hypersomnolence 25

2.2.4 Circadian rhythm sleep-wake disorders 26

2.2.5 Parasomnias 26

2.2.6 Sleep-related movement disorders 26

2.3 The prevalence of Sleep Disorders 27

2.4 Children to Adolescents 28

2.5 Serious Health Issues of Sleep Disorders 28

2.5.1 Mental health and Sleep Disorders 29

2.5.2 Physical health and Sleep Disorders 29

2.6 Economic costs 29

2.7 Safety issues 31

2.8 Diagnosis of Sleep Disorders 33

2.9 Treatment for Sleep Disorders 34

3. Obstructed Sleep Apnoea 37

3.1 During Sleep and OSA 37

3.2 Anatomical structure of the Upper Respiratory Tract 38

3.2.1 The pharynx 38
3.2.2 Aetiological risk factors and symptoms of OSA 39
3.3 Diagnosis Of OSA and Limitations 41
3.4 OSA Treatment and Limitations 43
3.5 Electrocardiogram/Elektrokardiogramm 46
3.5.1 Single-Lead ECG Diagnosis of OSA (Polysomnography - PSG) 47
3.5.2 Correlations Between Single-Lead ECG Findings and OSA 47
3.5.3 OSA and ECG Tracing Deflection Waves 48
3.5.4 ECG Signals and Machine Learning 50
4. Machine Learning 51
4.1 Artificial Intelligence 51
4.2 Introduction to Machine Learning 52
4.2.1 Key Machine Learning Subcategories 54
4.2.2 Machine Learning Algorithm Process 56
4.2.3 Key Machine Learning Subsets 57
4.3 Machine Learning Classification Models 58
4.3.1 Neural Networks 58
4.3.2 Deep Learning Neural Networks 61
4.3.3 Random Forest Classifier 66
4.3.4 Support Vector Machine 68
4.4 Related Work for the Diagnosis of OSA 69
5. Methodology 80
5.1 Data Acquisition and Materials 80
5.1.1 Data 80
5.1.2 Apnoea-ECG dataset description 81
5.1.3 Dataset Files 82
5.1.4 Data viewing and analysing 83
5.1.5 Data extraction, function code and file format 83
5.1.6 Apnoea-ECG Database Structure 85
5.1.7 Annotations 87
5.1.8 Apnoea-ECG Database table of additional information 87
5.1.9 Materials 89
5.2 Data Pre-processing and Windowing Strategy 89
5.2.1 Data Profiling 90

5.2.2 Data Cleaning 91
5.2.3 Data Reduction 91
5.2.4 Data Transformation 96
5.3 Classification Model Comparison Study 100
A. One Dimensional Convolutional Neural Network Model 101
B. Random Forest Classifier Architecture and Evaluation 104
C. Support Vector Machine Architecture and Evaluation 104
5.3.1 Performance Metrics 105
5.3.2 Hyperparameter Experiments for Fine Tuning the 1DCNN Model 109
6. Training and Test Results 113
6.1 Training and Validation of the Models 113
6.2 Testing the Models 122
6.3 Results for RFC and SVM using Feature Engineering Methods 124
6.4 Comparison of results against other OSA detection methods 125
7. Discussion 128
7.1 Scope 132
8. Conclusions and Future Work 133
8.1 Conclusions 133
8.2 Future Work 134
9. References 136

List of Figures

Figure 1: Graphical examples of both an open and narrowed airway [128] 38
Figure 2: Anatomical structure of the Upper Respiratory Tract [131] 39
Figure 3: Heartbeat Cycle represented by as deflection waves, segments, and intervals [201] 48
Figure 4: Umbrella of select data science techniques [204] 51
Figure 5: Different types of ML algorithms and their use in real life applications 56
Figure 6: Algorithm Process from Raw Data to Prediction 57
Figure 7: A biological and an artificial neuron [243] 59
Figure 8: A basic model of a single node: x_i = input, w_i = weight, *f* = transfer function, *y* = output [244] 59
Figure 9: A neural network with inputs, a single layer of hidden neurons, and output neurons [238] 61
Figure 10: DL Neural Network showing input layer, 3 hidden layers and an output layer [241] 62
Figure 11: A Convolutional Neural Network layers through the network towards classification [258] 64
Figure 12: A typical Random Forest [271] 67
Figure 13: An SVM showing hyperplane, vectors and datapoints [271] 69
Figure 14: A selection of dataset files recordings for subject a01 82
Figure 15: Data Preprocessing into Datasets 90
Figure 16: Annotation file a01 - showing both Non-Apnoea and Apnoea events 93
Figure 17: Annotation file a01 - showing non-Apnoea events only. 93
Figure 18: Signal file a01.txt - **showing the first 5 rows** 94
Figure 19: Shows the first Non-Apnoea cluster, row 1 – 72002 and 0 – 12:00:000 minutes 94
Figure 20: Shows the 3 broken periods of Apnoea 95
Figure 21: Shows the extracted Apnoea event, taken from the signal file 96
Figure 22: Typical Architecture of a One-Dimensional Convolutional Neural Network. 101
Figure 23. Graphical output results from the 1DCNN-500 model using dataset W=500. Showing Training and Validation Accuracy 115
Figure 24. Graphical output results from the 1DCNN-500 model using dataset W=500. Showing Training and Validation Loss 115
Figure 25. Graphical output results from the 1DCNN-500 model using dataset W=500. Showing ROCAUC plot. 115
Figure 26. Graphical output results from the RFC-500 model using dataset W=500. Showing AUC plot. 117
Figure 27. Graphical output results from the SVM-500 model using dataset W=500. Showing AUC plot. 118
Figure 28: The average scores (Sensitivity and Specificity) of the 5 models after testing 123
Figure 29: Comparable ROCAUC plot results for the best performing model from each group, using dataset size W=500. 130

List of Tables

Table 1: Treatment options for a selection of sleep disorders....36
Table 2: OSA Symptoms and Complications....40
Table 3: Apnoea-Hypopnoea Index (AHI)....43
Table 4: ECG Wave Types and their Descriptions....50
Table 5: The Four Machine Learning Types....55
Table 6: Predominant ML subsets or types of algorithms [240]....57
Table 7: Neurons 3 key components....60
Table 8: Deep Learning Methods....63
Table 9: Types of Convolutional Neural Networks....66
Table 10: Apnoea-ECG Database – 3 Group types....82
Table 11: Software Tools for extracting and viewing....83
Table 12: Apnoea-ECG Database – 3 Group types....85
Table 13: Apnoea-ECG Database Structure....85
Table 14: Apnoea-ECG Database Information....87
Table 15: Apnoea-ECG Database Subject Information....89
Table 16: Apnoea-ECG Database Files....91
Table 17: Shows the extracted Apnoea and non-Apnoea events....97
Table 18: Shows the python scripts and their functionality....98
Table 19 – First stage of merging....99
Table 20 – Second stage of merging....99
Table 21 – Third stage of merging....99
Table 22 – 1DCNN-500 Experiments No.1 – 10....110
Table 23 – 1DCNN-500 Experiments No.35 – 50....111
Table 24 – 1DCNN-500 Experiments No.124 – 139....111
Table 25 – 1DCNN-500 Experiments No.162 – 173 and 236 – 243....112
Table 26: 1DCNN-500 Configuration and results....114
Table 27: RFC-500 Configuration & Results and Confusion Matrix....116
Table 28: SVM-500 Configuration & Results and Confusion Matrix....117
Table 29: 1DCNN Training and Validation Experiments - Configuration and Results....121
Table 30: RFC Training and Validation Experiments - Configuration and Results....121
Table 31: SVM Training and Validation Experiments - Configuration and Results....121
Table 32: Test results for the five 1DCNN models averaged over 20 runs....123
Table 33: Comparison of proposed model V's other OSA ML and DL models....124
Table 34: Comparison of proposed model V's other OSA ML and DL models....126

ACKNOWLEDGMENTS

I would like to begin by thanking my parents for all the support, guidance, and education they provided, making me into the person I am today. I will be eternally grateful to you both. I would also like to show my immense and sincere gratitude to my two directors of study. Firstly, to Professor Paul Fergus for trusting in me and giving me this opportunity, and secondly to Dr Denis Reilly for taking over the reins at such short notice and without any hesitation. For almost a decade, they have both guided me with their wide-ranging and unparalleled knowledge and support, demonstrating great patience and believing in me at every step of this very challenging, and often seemingly impossible, journey to complete. Together, they have equipped me with the required skillset to become an accomplished researcher. It has been a great honour to have been their student, and I am forever indebted to them for assisting me in successfully achieving my goal. I would further like to extend my thanks to my Co-supervisor Dr Andy Shaw, along with Professor Carl Chalmers for their invaluable input and guidance. I would also like to convey my thanks to the PhD teams (Faculty Research team and universities Doctoral Academy team), who have provided me with important and vital information, sessions, dates, timelines, milestones, and objectives, including the occasional “kick-up-the backside”. I really do appreciate all your support and guidance throughout this journey, thank you all so much. Another team I would like to thank and express my appreciation to is my work team (CSM Technicians). Without their continual support, I would have found it impossible to complete my day-to-day work and accomplish this PhD. I truly appreciate everything, thank you. Moving onto my personal family and friends, I would firstly like to thank my dear wife for her continual support, patience and understanding throughout the completion of this PhD. She, more than anyone, has had to sacrifice much of her own life to support me through these nine years. She has also been an invaluable asset in helping me refine my writing, particularly with spelling and grammar. Thank you for your understanding, encouragement and helping me realise my goals. I will be forever grateful. I would also like to convey my thanks to my very good friend and proofreader, Debbie Parker, who has

always shown a great interest in my PhD, consistently offering me support and comforting words. Thank you very much for everything. Finally, I would like to thank my family and friends for all their unwavering support and encouragement. I am also deeply grateful to Liverpool John Moore's University for giving me the opportunity, resources, and platform necessary to successfully complete my studies.

ABSTRACT

The study in this paper presents a one-dimensional convolutional neural network (1DCNN) model, designed for the automated detection of obstructive Sleep Apnoea (OSA) captured from single-channel electrocardiogram (ECG) signals. This system presents a novel idea that could provide support mechanisms in clinical practice to promptly diagnose patients suffering from OSA. Using the state-of-the-art in 1DCNNs, a model is constructed using 4 major parts, a convolutional feature layer, max pooling layer, activation layer and a fully connected Multilayer Perceptron (MLP), comprising of a hidden layers and a SoftMax output for classification. The model repeatedly learns how to extract prominent features from the one-dimensional data before mapping it to the MLP to better learn complex feature relationships and increase classification. To improve training the novel idea to produce specific window sizes of time-series ECG data was introduced. For training and validation of the model, 35 ECG signal recordings were selected from an annotated database containing 70 night-time ECG recordings. (Group A = a01 to a20 (Apnoea breathing), Group B = b01 to b05 (moderate), and Group C = c01 to c10 (normal). Performance of the model was evaluated by using various metrics. For the testing stage 35 recording are withheld. Test results showed the 1DCNN-500 model produced excellent classification results with averaged scores of sensitivity 0.9742, specificity 0.9715, accuracy 0.9728. To further evaluate the ability of the IDCNN, the same datasets and training were applied to two alternative Machine Learning (ML) classification algorithms, namely the Random Forest Classifier (RFC) and the Support Vector Machine (SVM). The RFC produced scores of (Sensitivity/Recall (0) 0.90 / (1) 0.94, Precision (0) 0.94 / (1) 0.90, Accuracy 0.91) and the SVM produced scores of (Sensitivity (0) 0.94 / (1) 0.50, Precision (0) 0.65 / (1) 0.90, Accuracy 0.72). Analysing the results from all the experiments confirmed the IDCNN model can identify the presence of Apnoea with a higher degree of accuracy and rapidity than other traditional ML models.

ACRONYMS

AHI	Apnoea-Hypopnoea Index
AI	Artificial Intelligence
ANN	Artificial Neural Networks
API	Application Programming Interface
BMI	Body Mass Index
SDB	Sleep Disordered Breathing
CPAP	Continuous Positive Airway Pressure
CNN	Convolutional Neural Network
DL	Deep Learning
DLN	Deep Learning Networks
ECG	Electrocardiogram
EEG	Electroencephalogram
EMG	Electromyography
EOG	Electrooculography
FN	False Negatives
FP	False Positives
HRV	Heart Rate Variability
HSAT	Home Sleep Apnoea Testing

KNN	K-Nearest Neighbor
LSTM	Long Short Term Memory Networks
ML	Machine Learning
MAD	Mandibular Advancement Device
MAS	Mandibular Advancement Splints
MMA	Maxillo-Mandibular Advancement
MLP	Multilayer Perceptron
NN	Neural Networks
OSA	Obstructed Sleep Apnoea
1DCNN	One-Dimensional Convolutional Neural Network
OSAS	Obstructed Sleep Apnoea Syndrome
PSG	Polysomnography
RFC	Random Forest Classifiers
ROC AUC	Receiver Operator Characteristic, Area Under the Curve
RELU	Rectified Linear Unit
REI	Respiratory Event Index
SSL	Semi- Supervised Learning
SL	Supervised Learning
SVM	Support Vector Machines

TSD	Time-Series Data
TN	True Negatives
TP	True Positives
UL	Unsupervised Learning
UPPP	Uvulopalatopharyngoplasty
WHO	World Health Organisation
WOF	World Obesity Federation

LIST OF PUBLICATIONS RESULTING FROM THIS STUDY

1. **Thompson, S**.; Fergus, P.; Chalmers, C.; Reilly, D. *Detection of Obstructive Sleep Apnoea Using Features Extracted from Segmented Time-Series ECG Signals Using a One Dimensional Convolutional Neural Network*. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8. https://ieeexplore.ieee.org/document/9207470

2. **Thompson, S**.; Reilly, D.; Fergus, P.; Chalmers, C. *Detection of Obstructive Sleep Apnoea Using Features Extracted from Segmented Time-Series ECG Signals with a One Dimensional Convolutional Neural Network*. IEEE Access Journal, December 2023; pp. 1–16. https://ieeexplore.ieee.org/abstract/document/10373019

1. INTRODUCTION

Obstructive Sleep Apnoea (OSA) is a sleep disorder that interrupts the natural rhythm of a person's breathing whilst they are sleeping. In the International Classification of Sleep Disorders Third Edition (ICSD-3) report, published in April 2014, OSA is classified as the most common subtype of sleep disordered breathing (SDB) and is characterised by episodes of complete or partial upper airway obstruction during sleep. The symptoms of Sleep Apnoea include chronic snoring, insomnia, gasping, and breath holding, unrefreshing sleep, and daytime sleepiness [1]. OSA can affect anyone regardless of age or gender, however, most studies show the condition to be more prevalent amongst 30- to 60-year-olds [2][3]. The Apnoea–Hypopnoea Index (AHI) is used to indicate the severity of OSA with an AHI value <5 classed as normal. Estimates have shown that OSA affects 20% of the general population, where AHI is ≥5 [4], [5]. with approximately 20% of men and 10% of women developing moderate to severe OSA over a 5-year period [6]. However, despite this high prevalence, the vast majority of OSA sufferers go undiagnosed [7][8].

Current diagnostic techniques for OSA can be expensive, cumbersome, complex, and lengthy, meaning sufferers do not receive the required treatment and therapy needed. It has been suggested that over 80% of patients remain incorrectly diagnosed [9][10]. Consequently, OSA represents a major public health concern and left untreated can lead to numerous negative health-related consequences and mortality [11][12]. OSD results in a lack of sleep and/or poor sleep quality, which can affect an individual's function and decision-making capabilities. This can often lead to accidents at home, in the vehicle and in the workplace [13]. Globally, the direct and indirect costs of OSA, such as health care costs, accidents, decreased productivity, and sickness, reaches into the billions annually [14].

Diagnosing OSA is determined through consultation with a physician or sleep specialist. A physical examination is performed to consider the blood pressure, body mass index (BMI) and neck measurements of the patient. This is often followed up by a detailed discussion to gather sleep

information, typically achieved using a sleep log or sleep diary to record sleep times, nightly bedtimes, time to fall asleep, morning arising-times, wake-up-time duration and number, additional nap times and any episodes of tiredness/sleepiness throughout the day [15]. Other mechanisms include self-assessment questionnaires, such as the Epworth Sleepiness Scale (ESS) [16], Berlin [17] and STOP-Bang Questionnaires [18].

A more precise diagnosis and information gathering process can be performed through non-intrusive sleep studies, also known as Polysomnography (PSG). PSGs are the best approach when identifying incidences of OSA and are the preferred method for clinicians [15]. It involves the recording and analysis of multiple physiological variables during sleep using body worn sensors to record electroencephalogram (EEG), electrooculography (EOG), electromyography (EMG), electrocardiogram/ electrocardiography (ECG or EKG), nasal cannulas, pulse oximeters and respiratory belts. Patients will generally sleep overnight in a PSG sleep centre attached to as many as 16 separate physiological sensor channels and multiple devices to monitor stages of sleep, measure oxygen levels, body movements, heart rate and breathing patterns, to provide a comprehensive analysis [19]. However, there are several major problems with this type of diagnostic testing. These include a lack of available PSG sleep centres and equipment, high costs and the employment of sleep technicians to monitor a person's sleep [20]. Furthermore, it is often an inconvenience for patients to attend and sleep in such sleep laboratories, particularly when testing on children [21].

To combat these issues alternative OSA diagnostic methods have been proposed. One example is the Home Sleep Apnoea Testing (HSAT) kit, known in Europe as polygraphy kits. HSATs are lightweight, portable, and wearable devices that use far less physiological sensors than standard PSGs [22]. However, their use as stand-alone diagnostics in routine clinical practice is yet to yield any convincing results [23]. This is primarily because HSATs find it difficult to compute the Respiratory Event Index (REI), since it is calculated against recoding-time instead of sleep-time and

this ultimately misrepresents AHI assessments, and they lack the ability to detect hypopneas which can lead to misdiagnosis of OSA severity [24].

Since the turn of the century, the introduction of Data Science has provided a data-driven methodology of OSA detection. Harnessing the power of advanced ML algorithms with clinical expertise, it is now possible to produce better diagnostic results using single-channel physiological signals. This approach dramatically reduces the amount of required equipment, time and costs and has provided a platform for novel studies and proposals which include the use of ECG [25], EEG [26], Nocturnal Oximetry recordings [27] Respiratory sensors [28][29] and Snoring audio segments [30][31][32]. Nevertheless, akin to the more traditional diagnosis testing systems, this approach also has its drawbacks, chiefly, the required domain expertise and consumed time [33]. In more recent years the revolutionary approach of deep learning (DL) has further propelled the concept of ML for the diagnoses of diseases [34]. Its ability to automatically learn complex features from raw data has reduced the reliance on domain expertise, making it much faster and far cheaper than traditional ML models [34]. However, an important factor to the success of any DL model is seen through the capture of true optimisation. The development of a methodical and delicate approach to fine-tuning the models hyperparameters is a crucial step to improving the model's performance and achieving the best possible results [35]. Results that ideally demonstrate high accuracy, low loss, and a high balanced sensitivity and specificity. Ultimately, a well-tuned and optimised deep learning model can significantly contribute to accurate and efficient diagnosis of OSA, leading to improved patient outcomes and healthcare delivery.

This study presents a novel system that builds on recent advancements in the field of ML using deep learning neural networks (DLNNs) for the automatic and early detection of OSA, that could provide mechanisms in clinical practice to help diagnose patients suffering from OSA. This study also presents the results and finding from two alternative ML classifier models, namely RFC and SVM, when compared against the 1DCNN model. The 1DCNN, RFC and SVM are well-known for their binary classification problem-solving and ability to capture complex patterns within the data,

therefore making them well-suited for the diagnosing of OSA using time-series ECG signals. The IDCNN excels at learning hierarchical features from ECG signals, making it adept at identifying subtle changes associated with OSA. The Random Forest, with its ensemble approach, is robust to noise and can handle high dimensional data, making it suitable for our ECG dataset. The SVM, is known for its strong generalisation performance and can effectively distinguish between OSA and normal ECG patterns, even in the present of overlapping features. Many studies using RFC and SVM have provided excellent results using time-series ECG data for the diagnosis of OSA making them perfect models for a comparison study.

1.1 AIMS AND OBJECTIVES

Aims

1. **Develop an automated system for OSA detection**: The primary aim is to create a system that can accurately detect OSA events using only single-channel ECG signals, eliminating the need for conventional OSA detection methods.
2. **Research obstructed sleep apnoea aetiology, symptoms, and diagnosis**: The aim is to conduct comprehensive research into the underlying mechanisms, clinical manifestations and diagnosis approaches of OSA.
3. **Acquire and preprocess a suitable ECG database**: The aim is to find an appropriate database to implement an effective windowing strategy to better train and optimise the model using segmented time-series ECG signals of fixed-length windows of 5 different sizes.
4. **Leverage the power of a 1DCNN for feature extraction:** The aim is to demonstrate the effectiveness of 1DCNNs for automatically extracting relevant features from ECG signals for better OSA detection.
5. **Evaluate the 1DCNN performance**: The aim is to systematically evaluate the performance of the 1DCNN model for OSA detection using a comprehensive set of performance metrics and compare its results to traditional machine learning classifiers to assess its relative effectiveness, guided by a list of experiments that systematically vary window size and hyperparameters.
6. **Improve OSA diagnosis efficiency**: The aim is to accelerate the OSA diagnosis process by providing a rapid and accurate detection method.
7. **Contribute to the advancement of OSA research**: The aim is to contribute to the development of new and improved methods for OSA detection.

Objectives

1. **Implement a 1DCNN model**: The objective is to design and train a 1DCNN model capable of classifying time-series ECG signals segments as either containing OSA events or not.
2. **Design a dataset of 5 different window sizes**: The objective is to systematically explore and evaluate different windowing sizes on the Apnoea-ECG database to identify the optimal window size that maximises the model's performance in detecting OSA events.
3. **Establish the 1DCNNs optimal performance**: The objective is to systematically tune the 1DCNN models hyperparameters, including learning rate, filter size, batch size, kernel size, optimiser, number of epochs and CCN layers that yields the best performance on the 5 separate specifically windowed datasets.
4. **Evaluate models performance**: The objective is to evaluate the model's performance using standard evaluation metrics such as sensitivity, specificity, accuracy, Loss, Validation (Acc/Loss), F1 score, Kappa score, Log loss and ROCAUC.
5. **Compare with other machine learning methods**: The objective is to better evaluate the 1DCNN using the more traditional classifiers, Random Forest and Support Vector Machine.
6. **Provide a potential solution for clinical practice**: The objective is to develop a practical tool that can assist clinicians in the diagnosis of OSA, leading to earlier intervention and improved patient outcomes.
7. **Develop a novel high scoring model:** The objective is to develop a 1DCCN model that can provide excellent accuracy of over 95%, extremely low loss, of less than 0.1%, and a perfectly balanced score of over 95% for sensitivity and specificity.

1.2 UNIQUE CONTRIBUTION

1. To deliver a classification model to diagnose patients suffering from OSA using a 1DCNN Deep Learning model that overcomes the requirement to manually extract features and uses a unique windowing strategy of time-series ECG data to better train the model. This is beneficial for the following reasons:
    - Allows the reducing of the signal to capture more OSA events.
    - Enables better training of more observations using smaller time series windows.
    - Addresses the dataset class imbalance using real data, thus avoiding the use of synthetic data.

2. To provide a comparison of different classifiers (1DCNN, SVM, RF) when utilising PhysioNet (Apnoea-ECG) dataset. The 1DCNN and the automated feature extraction associated with DL models proves significantly better than traditional machines learning models.

3. This work further led to the publishing of two peer-reviewed papers, which provide a unique contribution to the scientific community.

1.3 MOTIVATION

From listening to and witnessing family members and friends who suffer with OSA, this gave me the passion to better understand this awful disease and try to contribute to the scientific and medical community, in the hope to make a difference in the future.

OSA currently affects approximately 1.5 billion people around the planet and is on the rise [36].

This increase is due to several risk factors, including:

- **An ageing population** – Is a significant risk factor contributing to the projected increase of OSA, particularly in older men and postmenopausal women. Influential factors, include anatomical changes, physiological transformations, and underlying health conditions.
- **A rise in obesity** – A major risk factor for OSA is obesity. Studies have shown the close links between the rise in OSA and obesity, with suggestions this is almost 1 for 1 [37]. The World Health Organisation (WHO) have estimated over 2 billion people to be obese over the last 5 years [38] and the World Obesity Federation (WOF) predicts that 51% of population will be obese by 2035 [39]. These Figures imply that OSA will double to approximately 4 billion sufferers over the next 10 years.

As this disease escalates, it puts more pressure on health services and the economy. A major challenge of OSA is diagnosis [10]. This includes those sufferers who are waiting to be diagnosed, those that are unaware and may never get diagnosed [7] and those who may have been incorrectly or misdiagnosed [9]. Early diagnosis of OSA can prevent the development of more serious health issues and even morality, therefore it is important to diagnose sufferers better and quicker. This will not only save lives but also reduce future costs on treatment.

1.4 THESIS STRUCTURE

The remainder of this paper is structured as follows:

Chapter 2 – Sleep Disorders: This chapter provides an overview of sleep disorders, their classification type and characteristics. Further to this, it discusses the prevalence, risk factors and challenges of sleep disorders, such as serious health issues, economic costs, and safety issues, before finally presenting the mechanisms used for the diagnosis for sleep disorders and available treatments.

Chapter 3 – Obstructed Sleep Apnoea: This chapter presents a comprehensive discussion on OSA, providing an understanding to the aetiological risk factors and symptoms associated with OSA, as well as an insight to the anatomical structure of the upper respiratory tract and its involvement in OSA. The chapter presents a discussion on both the diagnosis and treatment of OSA including their limitations. The final part of the chapter looks at ECG machines, how they're used in the diagnosis of OSA, their configurations and the existing correlations between single-lead ECG findings and OSA.

Chapter 4 – Machine Learning:

This chapter introduces AI and ML algorithms. It provides an overview of the key ML subcategories, as well as the inner working and processes of ML algorithms and DL CNN models. The chapter then presents the 3 main ML classification models 1DCCN, RFC and SVM, which are to be later used and further discussed in a comparison study. The final part of this chapter provides a comprehensive literature review of existing ML applications used to automatically detected OSA using time-series ECG signals.

Chapter 5 – Methodology:

This chapter presents the proposed framework and methodology. It firstly looks at the data acquisition and data pre-processing techniques. Providing an insight to how the data is transformed into something usable. Following this is a look at how the novel windowing strategy

is applied to deliver the required datasets. The chapter then focuses on the architecture of the 3 main ML algorithms, 1DCNN, RFC and SVM, understanding the essential components to building each model. The penultimate part of this chapter provides an overview of each of the metrics used to measure the performance of each model. The final part of this chapter presents a list (Tables 22 – 25) of hyperparameter and fine-tuning experiments performed on the model 1DCNN-500 using the dataset W=500.

Chapter 6 – Training and Test Results: This chapter evaluates the effectiveness of all models by presenting their results through training and testing. The first part presents the training and validation results, including the details of experiments and each models optimal configurations. The second part of this chapter provides the testing results for the 1DCNN models using unseen data, determining the most suitable model for the framework system. Following this is a set of results for traditional ML models, when trained with feature engineering methods. Finally, presented is a results comparison of the proposed model against other OSA studies, discussed earlier in the Chapter 4 Related Work section.

Chapter 7 – Results Discussion: This chapter provides a discussion of the results that were presented in chapter 6, Training and Testing. It includes an analysis and evaluation of each model's performance, discussing all observations made, highlighting areas of concern and any challenges and limitations.

Chapter 8 – **Conclusion and Future Work**: The final chapter of this thesis, provides a summary of all the research and findings gathered from this study. In this chapter all conclusions are matched against the original aims and objectives found in chapter 1, discussing how well each of these have been achieved. Finally, future work is discussed to set the direction for any follow-on work.

2. SLEEP DISORDERS

Sleep is a fundamental and natural reoccurring process of the human body [40]. However, for many individuals the natural process of sleeping does not happen as it should, and they can often find sleeping a difficult thing to do. The reasons for their inabilities to sleep is commonly related to a medical condition which disrupts their natural sleeping pattern [40]. This type of condition is known as a sleep disorder, and it can affect any gender at any age. *Oxford English Dictionary – definition of a disorder: An illness that disrupts normal physical or mental functions*. Over the last couple of decades and since the condition has become better recognisable, there has been an upsurge in reported cases of sleep disorders with some now estimating may affect as much as 45% of the world's population, with many still undiagnosed and untreated [41]. To date there are over 80 recognised sleep disorders, with Insomnia and Obstructed Sleep Apnoea (OSA) making up the vast majority of all recorded cases. The diagnostics and treatment for many of these conditions can be slow with many sufferers left incorrectly diagnosed or untreated. This can lead to a decline of an individual's daily functionality, health, and longevity [42]. The direct and indirect costs of sleep disorders; health care costs, accidents, decreased productivity, and sickness now reaches well into the hundreds of billions of pounds annually [43]. With an ever-expanding population and longer life expectancy, these Figures are anticipated to continue to rise, something which is causing unease, with many expressing concerns that more needs to be done [44].

2.1 CLASSIFICATIONS OF SLEEP DISORDERS

In the most recent ICSD-3 (The International Classification of Sleep Disorders Third Edition) report, published April 2014, by the international classification of sleep disorders, it lists over 80 individual sleep disorders [45]. These sleep disorders are broadly categorised into 7 diagnostic sections, Insomnia; Sleep-related breathing disorders; Central disorders of hypersomnolence; Circadian rhythm sleep-wake disorders; Parasomnias; Sleep-related movement disorders and other sleep disorders or conditions that can't be linked to any of the initial 6 sections [46]

The ICSD-3 is the most advanced and authoritative primary diagnostic classification system of sleep disorders designed to aid sleep experts and clinicians, which includes extensive literature reviews for each diagnosis. The ICSD-3 follows on from two predecessors ICSD-1 & ICSD-2, all developed and shaped by the American Academy of Sleep Medicine (AASM) in association with the European Sleep Research Society, the Japanese Society of Sleep Research and the Latin American Sleep Society.

2.2 Types of Sleep Disorders

2.2.1 Sleep-related breathing disorders

According to the ICSD-3, sleep-related breathing disorders is divided into four sections: Obstructive sleep apnoea (OSA), central sleep apnoea (CSA) syndrome, sleep-related hypoventilation disorder and sleep related hypoxemia disorder [47]. OSA is the most common subtype of breathing disorders of sleep [48]. This is followed by Central sleep apnoea (CSA) which is characterised by a lack of drive to breathe during sleep resulting in repetitive periods of insufficient ventilation and compromised gas exchange [49].

2.2.2 Insomnia

Insomnia is a condition in which an individual can find it difficult to initiate sleep and/or maintain sleep. Consequently, the sufferer will have a decreased sleep-time or no sleep-time. Sufferers of insomnia could experience tiredness, fatigue and daytime sleepiness [50], [51]. The ICSD-3 categorises Insomnia into three classes, chronic insomnia, Short-term insomnia, or other insomnia types. The ICSD-3 duration criterion for chronic insomnia disorder is 3 months, and a frequency criterion of at least three times per week [47].

2.2.3 Central disorders of hypersomnolence

The ICSD-3 divided these into three main subtypes: Narcolepsy type 1, Narcolepsy type 2, and Idiopathic Hypersomnia (IH) [52]. These disorders are characterised by excessive daytime

sleepiness (hypersomnolence) that is not attributable to another sleep disorder, specifically those that result in disturbed sleep (e.g. sleep-related breathing disorders) or abnormalities of circadian rhythm. The central disorders of hypersomnolence are often caused by intrinsic CNS abnormalities in control of sleep-wake, although other medical conditions or substances may account for the hypersomnolence. Behaviourally induced insufficient sleep is also included in this group of disorders [53].

2.2.4 CIRCADIAN RHYTHM SLEEP-WAKE DISORDERS

The circadian system regulates the timing and expression of nearly all biological processes, most notably, the sleep-wake cycle and any change or disruption to the circadian system can influence this disorder [54]. In the latest publication from the ICSD-3, the nomenclature for Circadian Rhythm disorders have been changed to “sleep-wake”. They are divided into seven sections: Delayed sleep-wake phase disorder, Advanced sleep-wake phase disorder, Irregular sleep-wake rhythm disorder, non-24-h sleep-wake rhythm disorder, Shift work disorder, Jet lag disorder and Circadian sleep-wake disorder not otherwise specified.

2.2.5 PARASOMNIAS

The ICSD-3 has divided Parasomnias into three clusters: NREM-related parasomnias, REM-related parasomnias, and other parasomnias [55]. Some of the Parasomnias behavioural manifestations associated to partial arousal from sleep are Sleepwalking, Sleep terrors, Sleep-talking and Sleep paralysis. During sleep, the brain cycles regularly between wakefulness, nonrapid eye movement (NREM) sleep and rapid eye movement (REM) sleep [56].

2.2.6 SLEEP-RELATED MOVEMENT DISORDERS

Sleep-related movement disorders are characterised by simple, often stereotyped movements occurring during sleep. In the case of restless legs syndrome (RLS), this is a condition characterised by unpleasant sensations deep inside the legs, occurring at rest accompanied by an urge to move

the legs [57]. The ICSD has divided these disorders into ten sections: Restless legs syndrome, Periodic limb movement disorder, Sleep-related leg cramps, Sleep-related bruxism, Sleep-related rhythmic movement disorder, Benign sleep myoclonus of infancy, Propriospinal myoclonus at sleep onset, Sleep-related movement disorder due to a medical disorder, Sleep-related movement disorder due to a medication or substance and Sleep-related movement disorder, unspecified [47].

2.3 The prevalence of Sleep Disorders

Since the first comprehensive Somnology (scientific study of sleep), almost 50 years ago, there has been a dramatic rise in the number of reported sleep disorder cases, where we are now at the point of epidemic proportions with up to 45% of the world's population reported to be affected [58][59][60] and a higher prevalence than heart disease, cancer and AIDS combined [61]. For many countries, the rapid rise of reported sleep disorders and associated sleep deprivation has become a major concern. An example of this is seen in the US where The Centres for Disease Control and Prevention (CDC), the leading national public health institute of the United States, has declared it a "public health problem" [62]. In a 2006 study by the National Heart Lung and Blood Institute (NHLBI), it was estimated that between 50 to 70 million Americans chronically suffer from a disorder of sleep and wakefulness [62]. Other countries demonstrating similar percentages are the UK, where it is suggested that 1 in every 3 people suffer from sleep disorders and in Australia where it is said to affect between 33-45% of Australian adults [63]. In the developing countries that includes, Ghana, Tanzania, South Africa, India, Bangladesh, Vietnam, Indonesia, and Kenya, over 150million adults are estimated to suffer from sleep disturbances. It is thought by 2030 that more than 260 million people will possibly be experiencing sleep problems in these developing countries [64].

2.4 Children to Adolescents

When looking at children through to adolescents, it is estimated that up to 50% will experience sleep disorders [65]. Combined Figures show sleepwalking, sleep talking, sleep apnoea, enuresis (bed-wetting), bruxism (grinding teeth), and night terrors, accounts for much of this prevalence [66]. It is also known that children can suffer with narcolepsy and hypersomnia [67], [68]. A major concern with sleep disorders amongst children and adolescents is the associated symptoms and illnesses that come from lack of sleep. For any developing child, quality and quantity of sleep is very important and any child deprived of sleep, they are at risk of triggering problems. Symptoms of depression, lower self-esteem [69] behavioural problems, such as anxiety, alcohol abuse [70], and even attempted suicide [71]. Some of the more aggressive, long-term, and debilitating illnesses can include diabetes, cognitive and seizures [72]. However, treatment for any such disorders is often achieved without the use of medication and it is expected that such disorders naturally fade as the child grows through childhood to adulthood [73]

2.5 Serious Health Issues of Sleep Disorders

Deprived sleep and reduced sleep quality are the two principal elements of sleep disorders, which over time can have serious adverse effects on the human body [44][43]. According to the National Sleep Foundations 2015 statement, it recommends young adults (age 18-25 years) and adults (age 26-64 years) should receive 7 to 9 hours of sleep, but not less than 6 hours or more than 10 hours (for adults) or 11 hours (for young adults). Older adults (65 years and older) should receive 7 to 8 hours of sleep but not less than 5 hours or more than 9 hours. For children (6 years to 13 years) should receive 9 to 11 hours and for Teenagers (14 years to 17 years), this age group should receive 8 to 10 hours. Sleep specialists say, for the human body to function correctly, stay heathy and keep longevity, it is important that a minimum and maximum duration of sleep is achieved. Individuals not conforming to this are at a higher risk of health issues and complications. This is

regardless of their age, weight, non-smoker, and non-exercise [74]. Evidence shows strong links between the specific sleep disorder and a broad range of serious health issues and even mortality.

2.5.1 MENTAL HEALTH AND SLEEP DISORDERS

The link between sleep disorders and mental health has been known for over 30years [75]. Mental health issues range from mild mood swings and a decrease in cognitive functionality to anxiety, severe depression and suicidal thoughts [75]. Generally, people who suffer from sleep disorders portray greater levels of depression [76], anxiety and suicidal thoughts [75]. One study suggests that as many as 90% of individuals with depression will also have sleep quality complaints [77]. Other conditions that sleep disorders can trigger include, ADHD (Attention Deficit Hyperactivity Disorder) [78], OSD (Obsessive-Compulsive Disorder) [79] and mania [80]. with some studies suggesting that insufficient sleep can lead to dementia later in life [81].

2.5.2 PHYSICAL HEALTH AND SLEEP DISORDERS

Sleep disorders can have as much of a devastating effect on the human body as they do on the mind. Chronic sleep deprivation can cause serious issues, such as the weakening of a person's immune and nervous systems and increasing the risk of infections and diseases [82]. It can also affect appetite [83] decrease sex drive [84] and increase body weight [85]. Further studies show sleep disorders have strong links to obesity diabetes, high blood pressure (hypertension) and cardio vascular disease (coronary heart disease, angina, heart attack, congenital heart disease and stroke), all of which show have high mortality Figures [37] [85].

2.6 ECONOMIC COSTS

The global economic costs inflicted by sleep disorders reaches way into hundreds of billions of pounds. The costs can be direct, indirect, related, and intangible, contributing events include sickness, accidents, lost productivity, vehicle and property damage, medical care, insurance, legal action and death [86]. The expense of these events can be seen in the US where in 2015 the costs

attributable to sleep deficiency was estimated to exceed $410 billion [87]. In a recent study by AASM that looked at the costing of such events, they calculated that the annual economic burden of undiagnosed OSA among U.S. adults to be approximately $149.6 billion and of this total, $86.9 billion was attributed to lost productivity and $6.5 billion to workplace accidents [88]. Other Figures revealed through the Sleep Matters Initiative at Brigham Health for the National Safety Council (NSC), showed that an employer with 1,000 workers can lose approximately $1.4 million dollars per annum in absenteeism, diminished productivity, healthcare costs, accidents and other occupational costs associated with exhausted employees, many of whom have undiagnosed and untreated sleep disorders. It also estimated that the costs of fatigue in an average-sized Fortune 500 company consisting of approximately 52,000 employees, is about $80 million annually [89]. Other reports put the estimated comorbid costs of health care and medication for employees with undiagnosed OSA at $30 billion annually [88]. A single employee with obstructive OSA costing an employer more than $3,000 and an employee with untreated insomnia is present but not productive for more than 10 full days of work annually, and accounts for at least $2,000 in excess healthcare costs each year [88]. Data from the National Safety Council in the US, revealed in the year 2000, that 810,000 motor vehicle crashes were attributable to OSA, costing roughly $15.9 billion [90] and a national transportation company with 1,000 employees likely loses more than $600,000 annually in decreased productivity because of tired employees. Motor vehicle crashes are the leading cause of workplace deaths, OSA costs the US $26.2 billion in motor vehicle accidents. In Australia, economics examined costs associated with the three most common sleep disorders, OSA, insomnia and RLS. The total health care costs for these three conditions totalled $818 million. Of these costs, $657 million per year related to OSA: these conditions include hypertension, vascular disease, depression, and motor vehicle and workplace accidents. The analysis suggested that 10.1% of depression, 5.3% of stroke, 4.5% of workplace injuries and 4.3% of motor vehicle accidents are attributable to a sleep disorder. The indirect financial and non-financial costs associated with sleep disorders are much greater than the direct costs. The indirect

financial costs were estimated to be $4.3 billion in 2010. These included $3.1 billion in lost productivity and $650 million in informal care and other indirect costs resulting from motor vehicle and workplace accidents. Of these indirect costs, OSA accounted for 61% ($2.6 billion), primary insomnia for 36% ($1.5 billion) and restless leg syndrome for 3% ($115 million).

2.7 SAFETY ISSUES

Individuals who suffer with chronic sleep disorders are likely to have insufficient sleep time when compared to a non-sufferer. This decreased sleep time can lead to episodes of excessive daytime sleepiness, drowsiness and feeling fatigued, which in turn may interfere with an individual's day-to-day performance levels, effecting their overall awareness, alertness, and decision-making [86]. Such episodes pose real safety concerns, at home, in the car and in the workplace and the general estimation is that people that suffer with sleep disorders are twice as likely to have an accident than a non-sufferer [44][88][86]. Similar evidence of this ratio is found in one large 20-year prospective study in Sweden of nearly 50,000 individuals, this study found that individuals who suffered with disturbed sleep were nearly twice as likely to die in a work-related accident [91]. A release from the ASSM suggested that highly sleepy workers are 70 percent more likely to be involved in accidents than non-sleepy workers, something which through the years has led, in parts, to some of the world's major disasters [74].

A contributing factor in both the development of a sleep disorder and to provoking the symptoms of a sleep disorder is the influence of unusual working hours [92]. Todays 24/7 industrialised world can in some cases require round the clock schedules, rotating shift patterns and unconventional shift work. It is estimated that 15-20% of working people throughout Europe and the US are affected by such unusual working hours, however trying to adapt the human body to these economic pressures can worryingly alter an individual's circadian rhythms (body clock) thus leading to insufficient sleep time and resulting in more work-related and MVAs accidents.

Evidence of the damage caused by sleep disorders can be seen across all industries and professions, from construction, which is reported to have the highest number of on-the-job deaths each year to the medical profession where the largest volume of accidents is reported. A large survey performed in Canada, looked at the damaging impact of sleep deprived workers on construction sites. The results indicated an average increase in risk of accident was 9% when working on construction. Both circadian rhythms changes and shift work were seen to be key factors [93]

A cross-sectional study that targeted 4407 nurses working shifts in eight large general hospitals in Japan was found that 26% suffered from sleep disorders and excessive daytime sleepiness, which increased risk of work injury or fatalities [94]. A second study conducted in Boston hospitals, examined 1876 surgical procedures performed by surgeons who had slept less than 6 hours when compared to surgeons who had more than 6hrs. The results showed 2.7-fold more complications when sleeping less. In a New Zealand study, 42% of the 1,366 interns interviewed reported a fatigue-related medical error in the last 6 months and 24% reported that they had already fallen asleep at least once driving home from work [95].

When looking at Figures attained from A&E, it showed sleep deprivation was the leading cause of motor vehicle accidents (MVAs) and mortalities in the workplace[86][89]; and almost 20 percent of all serious car crash injuries in the general population are associated with driver sleepiness [44]. Further studies have shown that, compared to normal individuals, individuals with OSA have a higher risk of falling asleep while driving and are three times more likely to cause accidents [96].

Approximately 1.3 million people die every year worldwide from MVAs and an estimated 20 to 50 million people are left with a significant disability. Although it is difficult to evaluate the precise amount of these fatalities attributed to sleepiness, some modelling studies have put the Figures between 15% and 33% [97]. Surveys performed in the UK, Australia, and Brazil, showed between 16% - 20% of highway accidents and almost one-third of fatal and serious accidents are known to

have occurred because of driver fatigue, as a result of sleep disorders, especially obstructive sleep apnoea, which has a three to sevenfold increase in the rate of road traffic accidents [98][99]. In a study conducted by The AAA Foundation for Traffic Safety in the US estimated that drowsy driving may cause 328,000 MVAs and 6,400 fatal crashes on U.S. roads each year [100]. And, data used from the National Safety Council in the US, estimated that 810,000 motor vehicle crashes a year are attributable to OSAS, resulting in 1400 fatalities [101].

A particular area where a high volume of MVAs and NMAs are seen involves commercial HGVs (Heavy Goods Vehicles), with a major reason for this being the round-the-clock work schedule that HGV drivers must perform. An older study conducted by The National Transportation Safety Board (NTSB) in the US, determined that fatigue and sleepiness was the probable cause of 57 percent of crashes leading to a truck driver's death and for each truck driver fatality, another three to four people are killed [44]. A more recent road safety study performed by an international action group and Italy's ministry of transport, found truck drivers across Italy had a high level of sleep-related disorders with insomnia affecting almost a third. These insomniac truck drivers had an almost two-fold risk of driving accidents and a more than three-fold increased risk of near-miss accidents when compared to non-insomniac drivers [102].

2.8 Diagnosis of Sleep Disorders

Sleep disorder diagnosis generally follows similar patterns and procedures that starts with a gathering of information about the individual sleeping habits. Following a consultation with a doctor, information is gathered through a sleep log that includes, sleep times, nightly bedtimes, time to fall asleep, morning arising-times, wake-up-time duration and number, additional nap times and any episodes of tiredness/sleepiness throughout the day and records these to the sleep log for further investigation [103].

A second mechanism for capturing information about the individuals sleeping habits is through a specific questionnaire and comprehensive interview. The questionnaires used are self-

assessment types and typically done using the highly reliable measuring tool/system such as an Insomnia Severity Index [104] or Epworth Sleepiness Scale (ESS) which helps identify OSA and Narcolepsy. There is no measurement tool available for diagnosing RLS. Following this, an in-depth discussion will take place about the individual's health & wellbeing, medical history, sleep patterns and sleep hygiene. It will also involve an understanding of signs and symptoms, family history of sleep disorders, current medication, and any longstanding comorbid disorders. In addition to this, blood tests might be required to rule out any other causes [105].

Following this, physical examinations will be performed on the individual that would firstly consider measurements of the body, (body mass index (BMI)) and neck [106]. If required, a more precise diagnosis process will be performed through non-invasive sleep studies, such as a Polysomnography, where the recordings of multiple physiologic variables will be gathered during sleep in a controlled environment. Alternative diagnostic methods proposed are Home Sleep Apnoea Testing (HSAT) kits, which are portable and less complex [107].

Another considered non-invasive test, also said to be a gold standard, is the Multiple Sleep Latency Test (MSLT). MSLT is accurate when testing for Narcolepsy. Alternative and cheaper method of testing is an Actigraphy. This is a simple technique that has been used to diagnosis many of the more prevalent sleep disorders [108].

To further assist with diagnosis of any sleep disorders, physicians and sleep specialists will use The International Classification of Sleep Disorders 3 as a key reference point to match up an individual's symptoms and results to the severity of the sleep disorder. This will enable them to plot a better route to treating of the found condition.

2.9 Treatment for Sleep Disorders

Treatment for sleep disorders can vary depending on the type of disorder and underlying causes. This generally starts with a combination of low-level common treatments [109] that includes therapy and lifestyle changes, such as altering the patients sleep hygiene and day-time activities,

introducing exercise, healthier diets and good habits [110][111]. Further treatment may involve recommendation of therapeutic and relaxation training [112] along with alternative and herbal medicines [113] herbal medicines [113]. Following this, more serious and long-term interventions may be required, such as multidisciplinary management plan, which includes drug therapy, supplementary devices, or surgery [114]. Pharmaceutical measures can range from antihistamines [115] to antidepressants [116], stimulants [117], sleeping pills and hormone blocking tablets [118] all with varying effects to the individual. Combined with this could be the use of application and supplementary devices therapy that includes a range of oral pressure therapies [119][120], stimulation and compression. Failing this, a final treatment would be surgery, which could involve reconstructive, reductive and even brain surgical procedures [118][114].

Sleep Disorder	Treatment
Insomnia	• Cognitive behavioural therapy • Stimulus control therapy • Relaxation training • Sleep hygiene education • Paradoxical intention • Pharmacologic management strategies
Sleep-related breathing disorders	• Lifestyle changes (exercise, healthier diet, good habits) • Sleep hygiene • Pressure (CPAP) • Oral appliances (MAD, MAS) • Hypoglossal nerve stimulation (HGNS) • Myofunctional therapy (oropharyngeal exercises) • Surgery
Central disorders of hypersomnolence	• Stimulants • Sodium Oxybate

	• Antidepressants
Parasomnias	• Medications • Improved Sleep Hygiene • Treat underlying triggers • Psychological approaches
Circadian rhythm sleep-wake disorders	• Improved Sleep Hygiene • Light Therapy • Melatonin Supplementation • Medications • Cognitive Behavioural Therapy
Sleep-related movement disorders	• Behavioural Modifications • Sleep Hygiene • Exercise • Weight Management • Medication • Therapy • Surgery

Table 1: Treatment options for a selection of sleep disorders

3. OBSTRUCTED SLEEP APNOEA

OSA is a common sleep disorder that causes a person to stop breathing for short periods of time (a duration of at least 10 seconds) whilst sleeping, resulting in recurrent episodes of intermittent hypoxemia and arousal from sleep [48]. Severe apnoea sufferers can have up to 600 episodes of apnoea per night, with episodes lasting up to 40 seconds [121]. Each time this happens, their brain briefly wakes them up to restart their breathing [122]. Persistent occurrences of OSA episodes results in fragmented, nonrestorative sleep [121]. The medical term for OSA is characterised by episodes of either apnoea, which is a complete collapse of the upper airway (pharyngeal closure), or a hypopnoea, which is a partial collapse of the upper airway (pharyngeal narrowing) [123]. Many studies have been performed on this topic and evidence suggests the pathophysiology of OSA is one of complexity that involves a combination of anatomical factors and physiological changes during sleep [124].

3.1 DURING SLEEP AND OSA

During the onset of normal sleep, breathing rate will typically slow down and the depth of breath becomes shallower [125]. Additionally, all the body's muscles relax, including the muscles at the back of the throat (Pharynx) called the pharyngeal muscles [124]. However, it is these Pharynx muscles that are responsible for keeping the airways open and therefore persistently contest the onset of sleep, with enough resistance to allow sufficient airflow through and maintain Pharyngeal patency. For OSA sufferers during sleep, maintaining Pharyngeal patency isn't always possible and the anatomic and physiological factors can cause pharyngeal instabilities and failure [126]. The Pharynx lacks rigid support and has 3 soft tissue areas, the posterior pharyngeal wall, the soft plate and the base of the tongue. It is these 3 areas that combine to cause partial or complete collapse of the upper airways [127].

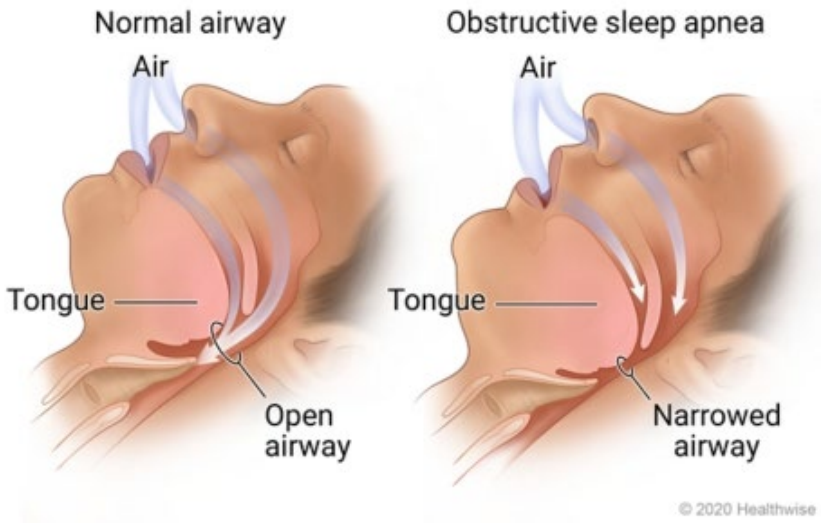

Figure 1: Graphical examples of both an open and narrowed airway [128]

3.2 Anatomical structure of the Upper Respiratory Tract

The anatomical structure of the Upper Respiratory Tract (upper airway) is one of both complexity and importance. It is part of the respiratory system that allows air to move from the nose or mouth to the lungs. It consists of the following structure: Nose (Nostrils), Mouth, Throat (Pharynx) and Voice box (Larynx), which are responsible for filtering air, warming and humidifying air, swallowing, olfaction (smell), producing sound (speech) and protecting the airway [129].

3.2.1 The pharynx

The pharynx is divided into 3 major parts, the nasopharynx, oropharynx, and laryngopharynx, which connects the nasal cavity, oral cavity, and larynx. It consists of two major muscle groups, constrictors, and elevators. The constrictor muscles (outer circular muscles), include the superior constrictor, middle constrictor, and inferior constrictor, which are responsible for swallowing. The elevator muscles (Inner longitude muscles), include the stylopharyngeus, the salpingopharyngeus and the palatopharyngeus. Of these 3 muscles, the palatopharyngeus responsible for breathing. Also in this region, other important upper airway patency muscles can be found in the Larynx (supraglottis, glottis and subglottis) and at base of the tongue (the airway dilator muscles),

geniosglossus (GG), hyoglossus, genioglossus, styloglossus, levator veli palatini, and the geniohyoid muscle that operates the hyoid bone at the base of the neck [130].

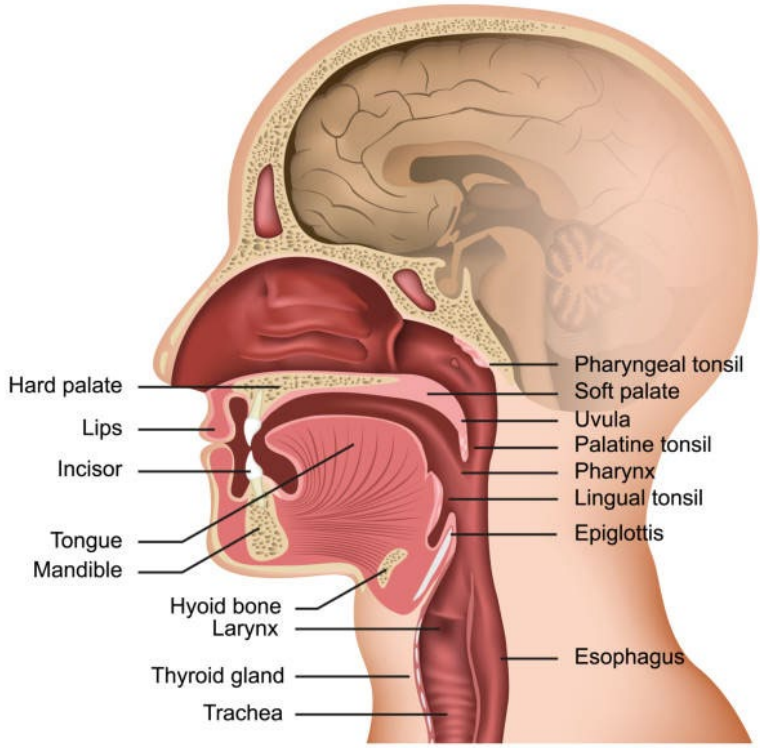

Figure 2: Anatomical structure of the Upper Respiratory Tract [131]

3.2.2 AETIOLOGICAL RISK FACTORS AND SYMPTOMS OF OSA

There are several etiological factors that can increase the risk of OSA, these include Anatomical deformities of the craniofacial structure, such as Micrognathia, Retrognathia, Mandibular hypoplasia and Facial elongation [132]. Physiological changes, such as pharyngeal muscular hypotonia [133], Adenoid hypertrophy [134], tonsillar hypertrophy [135], Inferior displacement of the hyoid [136] and the Bernoulli effect [137]. Defective respiratory control mechanisms and functions, that includes impaired chemical drive, defective inspiratory load responses, abnormal upper airway protective reflexes and functional impairment of the upper airway dilating muscles [138]. Further to this, there are environmental triggers, such as smoking, allergens, diet, exercise

times, air quality and sleeping environment [139]. However, the highest risk factor is obesity and has shown to be present in up to 90% of OSA suffers, based on their BMI status [37].

OSA sufferers will often show common symptoms that can affect their day to day lives; gasping, choking, fatigue, anxiety, mood swings and memory loss, are just some of these symptoms that reduce their quality of life, which left untreated can lead to serious complications [2]. Evidence shows strong links between OSA, and a broad range of major health issues, which includes, obesity [120], hypertension [98], depression [98], diabetes [140], stroke and heart dysfunction [141].

Common Symptoms	Potential Symptoms	Serious Complications
Gasping for air during sleep	Difficulty swallowing	Hypertension
Choking or waking up with a feeling of choking	Dry mouth	Stroke
Daytime sleepiness or fatigue	Sore throat	Heart dysfunction (attack or failure)
Morning headache	Heartburn	Diabetes
Irritability	Loud and excessive snoring	Depression
Mood swings		Obesity
Difficulty with short-term memory		
Impotence		
Increased heart rate and blood pressure		

Table 2: OSA Symptoms and Complications

3.3 Diagnosis Of OSA and Limitations

Evidence suggests OSA symptoms were being observed over 2000 years ago [142].By the late 19th century, such observations were still not fully understood and symptoms of apnoeic episodes (a temporary cessation of breathing, called apnoea) were being incorrectly diagnosed as a disease that only affected obese males who under-breathed and experienced day-time sleepiness (alveolar hypoventilation) [143][144]. Back then, such symptoms were associated to specific bodily characteristics, eventually to be named "Pickwickian Syndrome (drowsiness caused by obesity)", a name, first coined by Physician Sir William Osler in 1918 and derived from Charles Dickens first novel "The Posthumous Papers of the Pickwick Club" in 1836 [145]. In this novel Dickens describes a breathless (hypoventilation) and overweight (obese) character who struggled with day-time sleepiness (hypersomnia). Unaware to dickens at the time, he was describing what later became the medical term, Obstructed Sleep Apnoea [146].

The theme of misdiagnosis continued throughout most of the 20th century. Examples of this are seen in [147], where it was agreed amongst the medical fraternity, that any patients suffering with symptoms of hypersomnia, should be linked to Hypercapnia (carbon dioxide (CO_2) poisoning of the blood). It wasn't until 1965, that the first significant characterisation of OSA would be discovered by two separate Neuroscientists research groups, *Gastant et al* and *Jung and Kuhlo et al* [148]. For the first time ever, a continuous recoding of night-time sleep was undertaken on Pickwickian patients. Combined with this, a host of diagnostic tests, functions and measurements (polygraphic study) was performed over a duration of days. These diagnosis studies included Neurological examinations, Spirometry and Arterial investigations, X-rays of the skull and diaphragm, recording of chest movements, EEG, EMG, ECG, and recurrent monitoring of the heart, providing a more accurate diagnosis. From these studies it was found that incidents of hypersomnolence were not caused by hypercapnia or obesity as previously suggested, but instead by airway blockages with episodes of apnoea [149]. From reading how these case studies were

performed, it is clear to see advanced sleep study and psychophysiology methods were being used at this time.

It would be almost two decades, with the essential additions of cardiac and respiratory monitoring [150], before the first fully functioning polysomnography would be used to diagnose obstructed sleep apnoea. This method was soon to be regarded as the gold standard, a sophisticated yet complex tool, it was used for measuring multiple physiologic parameters, greatly improving the diagnosis of apnoea and the method through which sleep could be studied [151].

Over the coming decades, the popularity of sleep medicine began to outweigh the amount of available polysomnography sleep labs, leading to delays in patient diagnosis [152]. Furthermore, there was no standardisation to their findings, which caused confusion and misdiagnosis [153]. For a short period, the move from analogue to digital polysomnography helped ease some of the pressures and brought down costs. However, by the 1990s, in the US alone, there were approximately 1.3 million patients waiting to be diagnosed [154]

To tackle the issues of complexity, inconvenience and expense, a variety of portable OSA diagnostic systems were proposed. A well-established example of this is the Home Sleep Apnoea Testing (HSAT) system, known in Europe as polygraphy kits. These systems were lightweight, portable, and in some cases wearable, which meant a reduction in physiological sensors and clinical expertise usually required when using standard PSGs [23]. This introduction helped reduced waiting lists and lowered overall costs. However, some studies suggest their use as stand-alone diagnostics in routine clinical practice is yet to yield any convincing results [24]. This is primarily since HSATs find it difficult to compute the REI [25]. Plus, HSATs sensors suffer from the inability to detect hypopneas, which are associated to cortical arousals and may cause a misdiagnosis to the severity of OSA [155]. Further limitations of HSATs include lack of leg movement, lack of sleep staging and lack of body position tracking (in most HSAT devices) [156].

The lingering issues of PSG's and home kits brought about new diagnostic ideas. The introduction of ML in 1990's, opened up an entirely brand new and alternative way to the traditional rules of medical diagnosis [157]. The large complex stores of data were now being used by ML algorithms to solve problems with great success and in a much quicker fashion [157]. The earliest algorithms to take advantage of this new data driven world included Support Vector Machines [158], K-means Clustering [159], Naïve Bayes [160] and Data Mining techniques [161]. These systems laid the foundation for Random Forests [162] and Artificial Neural Networks (ANNs) [163], that led to rapid development of medical diagnosis algorithms we see today [164]. ML applications for OSA are considered later in *Chapter 4, Machine Learning.*

3.4 OSA Treatment and Limitations

For individuals diagnosed with OSA, numerous treatment routes are available. However, this is part depending on the type of OSA and the severity of their condition. Treatment can be as simple as making lifestyle changes, from exercise, healthier diet and better sleep hygiene [165], to a more serious interventions that requires long-term multidisciplinary management, such as drug therapy, supplementary devices and even surgery [166]. The severity of an individual's OSA is determined by how often their breathing is affected, whilst sleeping, over the course of an hour and is measured using the apnoea-hypopnoea index (AHI) in Table 3 [167].

Grade	Severity	AHI
-1	Normal	<5
0	Mild	5 – 15
1	Moderate	15 – 30
2	Severe	30 – 80
3	Very Severe	>80

Table 3: Apnoea-Hypopnoea Index (AHI)

Recommended treatment usually starts with continuous positive airway pressure (CPAP) [168]. This treatment is noted as an effective treatment for moderate to severe OSA and is labelled as the gold standard for treatment [149], [168]. Evidence suggests a 95% reduction of OSA symptoms when applying CPAP measures [169]. However, a common reported problem with CPAP, is poor adherence [170]. This can affect the efficiency of the treatment between 39% to 50% [171] and it is estimated between 30 to 70% of users cannot tolerate wearing the device [172], with varying issues that include inconvenience, pressure irritation, poor mask fittings, discomfort, skin irritation, mask leaks, sore eyes, airway drying, nasal problems, complaints of noise, frequent awakening and claustrophobia. Between 5% and 50% of patients simply reject CPAP treatment before even trying it and a further 12-25% of patients commencing CPAP abandon treatment within 3 years [173]. Over the years, there has been changes to the dimensions and contour of the mask and apparatus [174]. However, many issues remain and therefore, using CPAP remains a difficultly for management of patients with OSA [171] which questions the validity of the notion that CPAP is the gold standard for OSA therapy [175].

These issues led to the development of alternative solutions, such as Oral appliances (OAs) also known as MAD (mandibular advancement device) or MAS (mandibular advancement splints) [172]. These portable devices can be custom fit, allowing for improved tolerance and adherence [176]. They have become increasingly popular and recognised as both an alternative and supplementary treatment OSA [177]. It is reported Patients tend to prefer oral appliances to CPAP and self-reported adherence rates are typically higher [178]. In the latest guidelines set out by the AASM American Academy of Sleep Medicine, it recommends that MAD devices should primarily be used in patients with mild to moderate OSA only [179], since OAs are generally less efficacious for severe OSA, although some evidence displayed from newer models that there are some benefits to severe OSA sufferers [177]. But overall OA's are more accepted and tolerated by patients, which, in turn, may lead to a comparable level of therapeutic effectiveness [179].

Alternative, but a slightly more intrusive treatment is Hypoglossal nerve stimulation (HGNS). This is an implantable device that delivers an electrical impulse to anterior branches of the hypoglossal nerve in response to respiratory variation, resulting in tongue base protrusion that alleviates upper airway obstruction in adults [180]. This treatment has a high adherence rate, with reports showing daily use at 12 months was 86% and at 18 months was 84% [181]. It also has comparable results as CPAP when treating individuals with moderate to severe OSA with one study showing a greater than 50% reduction in AHI in all patients [182]. However, limitations and drawbacks come in costs of the implants, the size of the implanted device, the exclusion of patients with a BMI of 32 kg/m2 and currently, only one device is FDA approved for HGNS treatment [171].

There are a variety of surgical procedures; however, these are usually reserved for people who have not responded to other treatments. Many of these have high success rates, but also some with drawbacks. These surgical procedures include tracheostomy, uvulopalatopharyngoplasty (UPPP), velouvulopharyngeal lift, hyoid suspension, partial glossectomy, lingual suspension, tongue base resection, genioglossus advancement and maxillomandibular advancement (MMA) [183].

Tracheostomies were performed in the late 1960's to early 1980's as the main surgical therapy for OSA [183] and are still considered as a final option for severe cases of OSA. This treatment has a success rate of 100% [184]. However, reasons such as aesthetic and inconvenience of cleaning as well as high infections rates swayed individuals to look at other alternative surgical options.

A more popular choice of surgery is MMA [145]. MMA surgeries have been extensively studied and accepted as a successful therapy for the treatment of patients with OSA [120] [158], MMA provides an alternative for OSA patients who cannot or will not use CPAP [145] and for quality of life after surgery MMA remains the most effective operation for the treatment of OSA [159].

The most common and conventional techniques of surgery performed in the treatment of severe OSA is UPPP [145] [160] [161]. For individuals with a Class I status the success rate in some studies has been around 80% or higher. However, for Classes II and III this rate declines [161] [162].

Other less common surgical procedures include: Hypopharyngeal procedures (tongue reduction, partial glossectomy, lingual tonsillectomy, mandibular advancement, Genioglossus advancement) [185]; Laryngeal procedures (epiglottoplasty, hyoid suspension) [186]; Global airway procedure (MMA bariatric surgery, Orthognathic surgery) [187] and Oropharyngeal and nasopharyngeal procedures (Uvulopalatoplasty (LAUP), Palatal advancement pharyngoplasty, tonsillectomy, excision of tori mandibularis) [188] [189].

Non-surgical method of treatments can include Myofunctional therapy (oropharyngeal exercises), which can be used on its own or in conjunction with other treatment approaches or devices [182]. There are different slants on this method with different procedures, techniques and results. Although not as effective as treatment using devices like CPAP and aimed only at individuals who suffer with mild to moderate OSA [190], studies still showed some good results. For example, a systematic review and meta-analysis that includes data across nine studies involving a total of 120 adult patients with OSA showed that oropharyngeal training reduces the AHI by approximately 50% [191].

A more recently developed approach and an alternative to surgery in the future, could be through a drug treatment and technique called DREADDs (Designer Receptors Exclusively Activated by Designer Drugs). These drugs developed by John Hopkins laboratories have been tested on mice with fantastic results and therefore offer promise for development of pharmacologic agents to treat OSA [192].

3.5 ELECTROCARDIOGRAM/ELEKTROKARDIOGRAMM

Electrocardiogram/Elektrokardiogramm (ECG/EKG) is a specialist piece of electrical equipment used as a medical apparatus to analyse a person's heart function. First developed and used 125yrs

ago by Augustus Waller, a physiologist of St Marys Hospital, London, using a Mercury Capillary Electrometer he recorded the first ever electrocardiogram [193]. The process of interpreting an electrocardiogram is called Electrocardiography, which is performed by a clinical heart specialist who expertly interprets the produced ECG tracing to diagnose any abnormalities or diseases. There are 3 types of electrocardiogram tests that can be performed to diagnose the heart: a Rest ECG (lying down comfortably), a Stress ECG (exercising) and an Ambulatory ECG (normal daily activity 24/48hrs). The most common ECG test is referred to as the standard 12 lead ECG. This test requires a total of 10 electrodes strategically positioned around the body which combine to produce a 12-lead output. These electrodes capture electrical activity from the body and feed it back to the ECG monitor for analysis [194].

3.5.1 Single-Lead ECG Diagnosis of OSA (Polysomnography - PSG)

ECG machines are an integral component in the diagnosis of OSA. They are part of a wider collection of physiological variables and devices that record and measure brain waves, body movements, eye movement, heart rate, chest and adnominal movements, oxygen levels and breathing patterns. During a PSG, a single-lead ECG is typically recorded, often from the second or third chest lead (V2 or V3). A single-lead is the simplest combination of all the ECG arrangements, generally requiring 3 cables to monitor one channel or lead [195]. It provides a baseline assessment in identifying arrhythmias and measuring heart rate and rhythm, which can be affected by sleep disturbances and underlying medical conditions [196].

3.5.2 Correlations Between Single-Lead ECG Findings and OSA

Single-lead ECG findings have been shown to provide valuable insights into the impact of OSA on the cardiovascular system [196]. An indicator is Heart Rate Variability (HRV), which is a measure of the variation in time interval between heartbeats [197] Studies have shown that HRV is significantly reduced in people who suffer with OSA [198]. This is because OSA disrupts the normal functioning of the anabolic nervous system, leading to an imbalance between the sympathetic

and parasympathetic branches of the nervous system [198]. A second indicator to the presence of OSA is the increased Arrhythmias. Several types of Arrhythmias have been associated with OSA which include Atrial Fibrillation, Ventricular Tachycardia and Bradycardia [199].

3.5.3 OSA AND ECG TRACING DEFLECTION WAVES

When performing an ECG study, the electrical signals captured from a heartbeat cycle are recorded as a tracing in the form of deflection waves, segments, and intervals [200]. With each directional beat of the heart, a deflection wave is imprinted. If electrical activity is flowing away from the lead, this causes a negative or downward deflection wave and electrical activity flowing towards a lead causes a positive or upward deflection wave [194]. The deflection waves are described as 6 major waves, segments, and intervals, which include, P-wave, QRS-wave (complex) T-wave, PR-interval, PR-segment, ST-segment, and QT-interval [200].

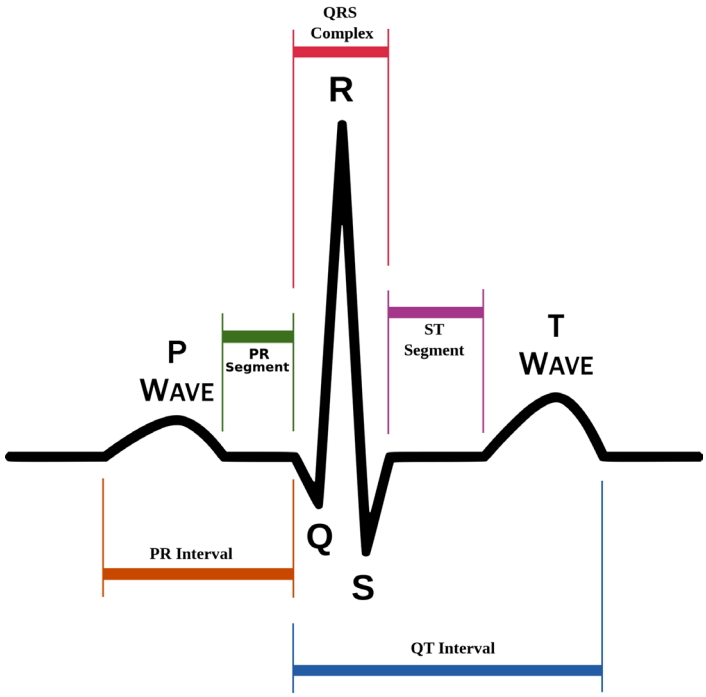

Figure 3: Heartbeat Cycle represented by as deflection waves, segments, and intervals [201]

Clinical specialists read ECG tracings to identify different kinds of conditions including OSA. Using their expertise, each wave deflection, segment, and interval is analysed and understood [202].

Deflection Type	Description
P-Wave	• **Prolonged P-wave duration**: increases in OSA due to left atrial enlargement, which occurs due to the increased workload on the left atrium caused by OSA. • **Widened P waves**: in OSA due to the increased distance between the sinoatrial node (the heart's natural pacemaker) and the left atrium. • **Increased P-wave amplitude**: increases in OSA due to the increased electrical activity of the left atrium. • **Inverted P waves**: can be seen in OSA, especially in lead V1. This is due to the presence of left atrial enlargement and altered electrical conduction in the atrium. • **Abnormal P waves:** in OSA can be a sign of atrial enlargement and increased risk of atrial fibrillation
PR Interval	• **Prolonged PR interval**: is an abnormal ECG finding that can be seen in people with obstructive sleep apnoea (OSA). PR prolongation is defined as a PR interval that is greater than 0.20 seconds.
PR Segment	• **Prolonged PR segment**: can be seen in OSA due to left atrial enlargement, which occurs due to the increased workload on the left atrium caused by OSA. • **Widened PR segment**: can be seen in OSA due to conduction abnormalities in the heart
QRS Wave Complex	• **Deep S waves** may be more common in people with OSA who also have other risk factors for right-sided heart problems, such as hypertension and pulmonary hypertension • **RS pattern with deep S waves**: is seen in OSA due to right ventricular hypertrophy and left ventricular strain. • **Shift of R-wave progression**: can be seen in OSA due to right ventricular hypertrophy.
ST Segment	• **ST-segment depression**: can be seen in OSA due to ischemia, which is a lack of blood flow to the heart muscle.

	• **ST-segment elevation**: can be seen in OSA due to pericarditis, which is an inflammation of the pericardium, the sac that surrounds the heart. • **ST-segment shift**: can be seen in OSA due to electrolyte imbalances.
T-Wave	• **Flattened or inverted T waves**: can be seen in OSA due to ischemia, which is a lack of blood flow to the heart muscle. • **Widened T waves**: can be seen in OSA due to electrolyte imbalances. • **Notched T waves**: can be seen in OSA due to ventricular hypertrophy or other heart abnormalities
QT Interval	• **Shortening of the QT interval** can be seen in people with obstructive sleep apnoea (OSA). QT shortening is defined as a QT interval that is less than 0.33 seconds.

Table 4: ECG Wave Types and their Descriptions

3.5.4 ECG SIGNALS AND MACHINE LEARNING

Already discussed, is how the electrical signals captured from an ECG study can provide clinical experts with a rich source of information when diagnosing an illness. However, in more recent years, combining these ECG signals with ML technology has greatly transformed the way medical diagnosis is approached [203]. ECG electrical signals can now be fed into an ML model to be analysed and interpreted, thus providing a meaningful insight to the diagnosis of illnesses, but without the requirement of clinical expertise of the past. Today there are numerous applications developed that use ECG electrical signals for diagnostic [203].

4. MACHINE LEARNING

This chapter introduces AI and ML algorithms. It provides an overview of the key ML subcategories, as well as the inner working and processes of ML algorithms and DL CNN models. The chapter then presents the 3 main ML classification models 1DCCN, RFC and SVM, which are to be later used in a comparison study. The final part of this chapter provides a comprehensive literature review of existing ML applications used to automatically detected OSA using time-series ECG signals.

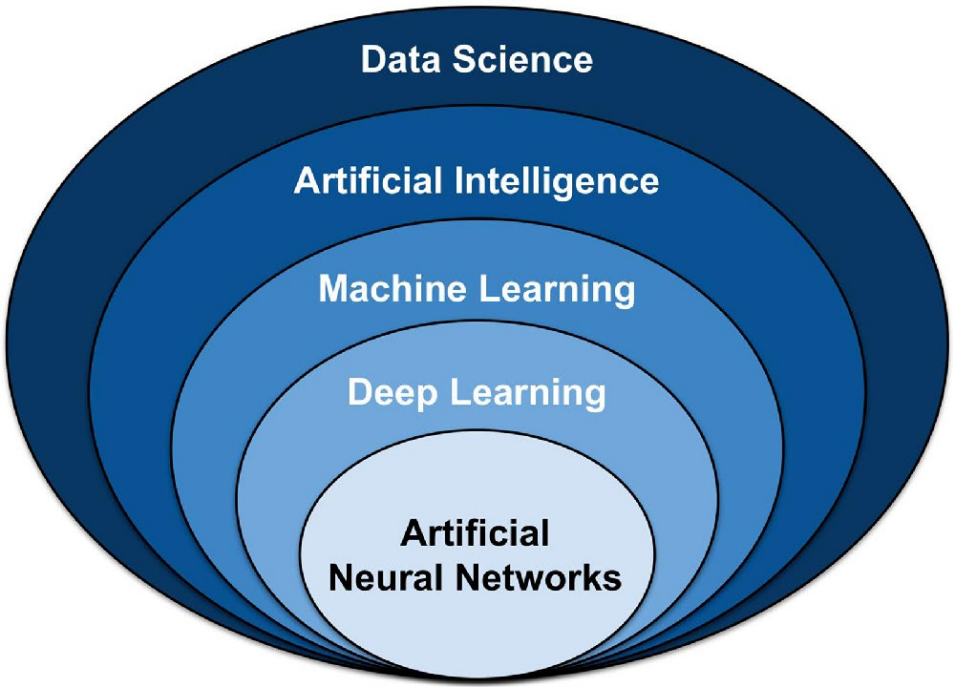

Figure 4: Umbrella of select data science techniques [204]

4.1 ARTIFICIAL INTELLIGENCE

AI is a predominant system within Data Science that encompasses ML [204]. The main goal of AI is to make machines intelligent, machines that can continuously learn and think for themselves, machines that can gather experience and adapt to their environment, machines that can self-improve and solve real world problems…basically machines that can perform human-like tasks,

but without human intervention [205]. Early envisages of such automated and mechanised thinking can be dated back to the 17th century, when Gottfried Wilhelm Leibniz, published the dissertation on "Combinatorial Art" [206] [207]. However, the first real glimpse into AI was in the 1834 when Mathematician, Charles Babbage, designed the Analytic Engine "Thinking Machine", a mechanical computer that could be programmed using punched cards [208]. Through the years, new AI concepts and research continued to be developed, such as the Turing Machine, built by Alan Turing [209] in 1937 and the Z3, built by Konrad Zuse in 1941 [210]. These advances were the birth of the modern computer that inspired a whole generation, however, there was still no actual "machine intelligence". Many of these designs focused on how machines could be programmed to perform human-like tasks and not machines that could program themselves, with this, the concepts of ML and NNs were born [211].

4.2 Introduction to Machine Learning

Historically the aim of ML can be organised around three primary research points that are the basis for problem solving and ideas: Task Oriented Studies, Cognitive Simulation and Theoretical Analysis [212][213]. The current form of ML started under the guise of "Machine Intelligence" in the early 1940's [214]. At that time, AI research focused on cognitive learning mechanisms of neurons, prompting logician, Walter Pitts and neuroscientist Warren McCulloch to collaboratively publish the world's first mathematical modelling of a neural network that mimics the human thought processes. This was a supervised learning classifier algorithm, called "Perception", that was instrumental in the development of artificial neurons [215]. In 1949 Donald Hemp published a pioneering book "The Organisation of Behaviour", which offered an even greater understanding to how the human brain related to NNs [216]. The following year Alan Turing designed a programme to establish if a human could be fooled by computer intelligence [217]. By 1952, the world's first self-learning computer programme was developed by Arthur Samuel. The programme, "computer checkers", was a ground-breaking adaptive computer game that brought

the term “Machine Learning” to the forefront [218]. Building on the works of McCulloch & Pitts and Hemp and Samuel, in 1957 Frank Rosenblatt developed a dedicated hardware version of Perceptron, called “Perceptron Mark 1”. This highly anticipated machine would now use weighted neurons and a decision tree method [219]. However, like McCullen-Pitts neuron, this development had many limitations, and therefore over the following decade the progression of ML slowly started to wane [220][221]. It wouldn’t be until 1967 when the most credible supervised ML classification method, K-nearest neighbor (KNN) was developed [222]. Initially published in 1951 as “Non-Parametric Classification” by Evelyn Fix and Joseph Lawson Hodges [223], but now with the addition of Upper Bound Error Function and Multiclass KNN Classification, this method was further advanced by Thomas Cover and Peter Hart into a reliable system [224]. The 1990s started a remarkable transformation in the field of ML, evolving from the theoretical concepts to the ubiquitous technology we see today. These years seen the advancement of backpropagation by Jurgen Schmidhuber [225] and the development of the SVM, by Vladimir Vapnik [226], to the boosting algorithms by Freund & Schapire [227], that paved the way to ensemble learning methods [228] and the vital improvements to DL technology by Hinton et al [229] that reignited the interest in NNs.

In recent times, these advancements coupled with the rise of available data and increased computational power, has brought an explosion of deep DLNNs and techniques [230]. ML Models are now applied as solutions to a wide range of problems, fundamentally changing how we live, work and interact with the world around us. Some of the most prominent areas where real-world applications are used, include: technology and computers [231], business and finance [232], science and healthcare [233], manufacturing and retail [231] and creative and entertainment [234]. Before choosing an algorithm to solve a problem, to get the best results a good understanding of both the subsets and subfields is vital. For example, a company examining solutions to help solve the problem of filtering unwanted “spam” email from genuine “non-spam” emails, the application of a supervised learning classification algorithm would be the best

approach. However, for that same company, identifying solutions to solve their customer-based marketing problem, the application of an unsupervised learning clustering algorithm would be the best approach. Not one algorithm can solve all problems, but different algorithms can solve the same problem [235]. Yet, taking different approaches can come with drawbacks and limitations, often requiring compromises, such as model speed, accuracy, interpretability, complexity and including trial and error. Before selecting any algorithm, it is important to understand the objectives and goals. Choosing the correct algorithms depends on many different factors, from the business needs to the size of the dataset, and time constraints to the outputted quality. Steps will involve the analysing of the dataset, looking at the type of data it is, if it needs to be annotated and how to clean and process it. Other important considerations are speed, time, training, and accuracy. All these factors are important to solving the degree of the problem [236].

Using a combination of both software and hardware, NN have become extremely popular over the last 30yrs and have begun to dominate across many application fields, business sectors and industry [237]. This is true for many reasons, such as the abundance of available data, more powerful and cheaper computers, better predictive accuracy, and less human intervention [230]. NNs work differently to the more traditional ML models, which are reliant on being humanly fed information to recognise patterns or perform specific tasks to get the best results. NNs have the capacity and intelligence to accept large amounts of raw and unstructured data (images, text, audio, etc) and be able to automatically recognise patterns, features and differentiate, without any human involvement. NN learn much quicker and are adaptable to their environment [238].

4.2.1 KEY MACHINE LEARNING SUBCATEGORIES

Depending on the type of ML algorithm chosen to solve a problem, will determine the type of approach used. There are four key types of ML approaches: Supervised Learning, Unsupervised Learning, Reinforcement Learning and Semi-Supervised Learning (SSL). These approaches determine how the data is managed and processed through the algorithm [239].

Type	Description
Supervised Learning	• **Model types**: Classification and Regression • **Training and labelling**: Some human supervision or annotated labelling, greatly improves the training of the model. • **Variables**: Learning a mapping between a set of input variables and an output variable. • **Evaluation and optimisation**: Continually check performance using common metrics and altering hyperparameters to improve model.
Unsupervised Learning	• **Model types**: Clustering, Association, K-Means, Cluster Analysis, K-Nearest Neighbors, Hierarchical clustering, and principal component analysis. • **Self-Learning**: Doesn't rely on the value of labelled data for training. • **Pattern discovery**: Excellent at discovering clusters, structures and feature relationships hidden in the data. • **Dimensionality Reduction**: Ability to reduce large datasets making them more manageable and allowing for better analysis and visualisation.
Reinforcement Learning	• **Model/System types**: Classification, Control systems, Robotics, Games, Finance, Natural Language. • **Agent-Environment**: Uses an agent to interact with environment by taking actions and receiving feedback in the form of rewards and punishments. • **Policy Goal**: The agent's goal is to learn a policy (set of rules) that determine the action to take in each state of the environment. Its goal is to maximise the cumulative reward over time. • **Trial and Error**: The agent learns by trial and error, which allows the agent to gradually improve its policy and achieve its goal.
Semi-supervised Learning	• **Model types**: Classification and Clustering. Defined as a mixture of both supervised learning and unsupervised learning. • **Exploitation of Unlabelled Data**: Regularises the modes training, reducing overfitting and improving generalisation. • **Label Propagation**: Aims to infer the labels of unlabelled data based on their similarity to labelled data points. This improves the model's ability to better generalise new data.

Table 5: The Four Machine Learning Types

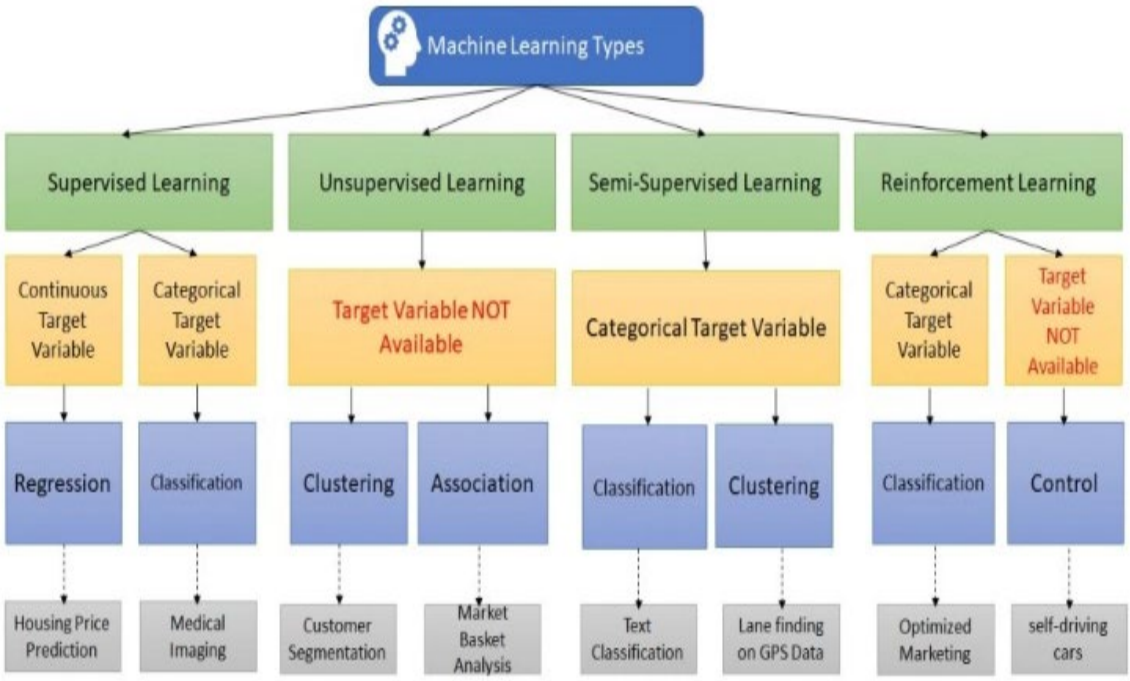

Figure 5: Different types of ML algorithms and their use in real life applications

4.2.2 MACHINE LEARNING ALGORITHM PROCESS

ML algorithms, often referred to as ML models, are a programmed sequence of instructions built from mathematical expressions. They perform a task to analyse inputted data to predict outputted values, each time learning from its operation to improve its performance. The operation of the algorithm is achieved by using the equation ***y = f(x)***. The goal is to learn the mappings between inputs (***x***) and desired results or outputs (***y***). The purpose of the algorithm is to learn the function ***f(x)***, which is based on the amount of training data presented to it. By fine tuning ***f(x)*** the model, means at the testing stage, the model will better predict (***y***) for new and unseen data.

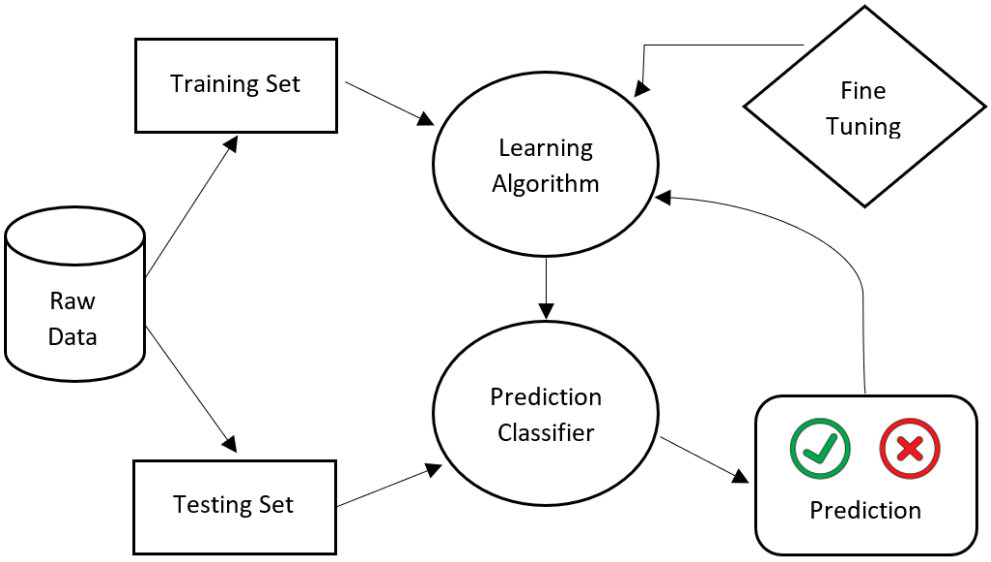

Figure 6: Algorithm Process from Raw Data to Prediction

4.2.3 KEY MACHINE LEARNING SUBSETS

Each algorithm will fall into a category of either supervised learning, unsupervised learning, reinforced learning, or semi-supervised learning and then into the subset of Classification, Regression or clustering. This section highlights these three subsets, determining how each algorithm operates, its application and technique [240].

Classification algorithms	Regression	Clustering
Neural Networks	Neural Networks	K-Means
Decision Trees	Linear Regression	Fuzzy C-Means
Random Forests	Decision Trees	Hidden Markov Model
Support Vector Machines	Support Vector Machines	Gaussian Mixture
Naïve Baynes	Random Forests	Hierarchical
Nearest Neigbor	Gaussian Process	
Ensemble Methods	Ensemble Methods	
Logistic Regression	Logistic Regression	
Discriminant Analysis	Discriminant Analysis	

Table 6: Predominant ML subsets or types of algorithms [240]

4.3 Machine Learning Classification Models

This section provides an overview of three predominant ML Models that use the classification method to solving problems. They are Neural Network (which includes, Deep Learning Convolutional Neural Networks), Random Forest Classifier and the Support Vector Machine. Each of these models fall under the category of supervised learning and have all proved to be excellent at handling non-linear data to solve classification problems across many different areas.

4.3.1 Neural Networks

NNs, also known as ANN's are the underlying technology of DL, a subfield of ML. NN's can be described as a signal processing system made up from vast collections of interconnected machine processing neurons, which combine to mimic how neurons work within the human brain to complete computational tasks [241].

- **Biological Neuron Structure**

The power of the human brain can perform highly complex, non-linear, and parallel processing to perform tasks such as accurate predictions, pattern recognition, perception, and motor control [242]. Neurons are the part of the brain that gives humans their intellect and there is an estimated 86 billion of them located within the nervous system [198]. Each neuron is made up from four major parts: Dendrites, Cell body, Axon, and Synapse. When the cell body receives enough information (inputs) through the Dendrites, this information is sent in electrical impulses (nerve signals) along the Axon. The Axon is where learning occurs; the stronger the signal pulsates through the Axon, the better the learning outcome will be when reaching the Synapse. The Synapse is the final junction where the neuron connects and communicates the information to other neurons, and the process is repeated [199].

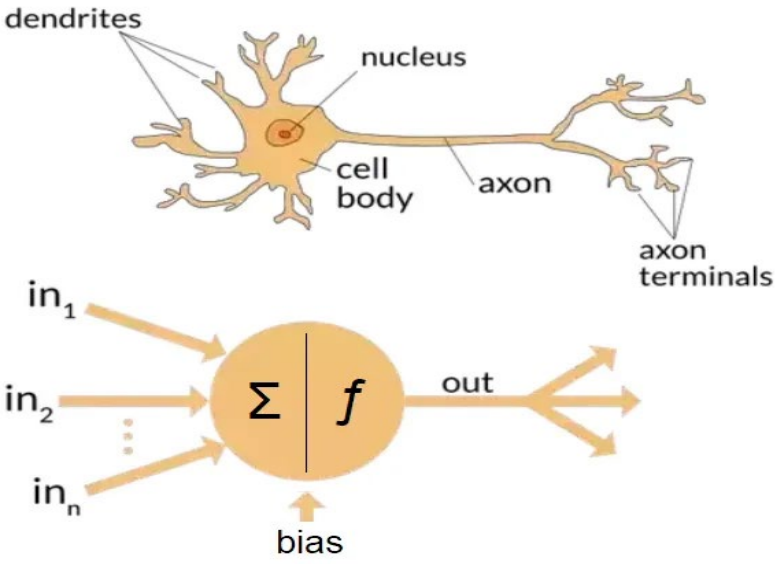

Figure 7: A biological and an artificial neuron [243]

- **NEURON (NODE OR PERCEPTRON)**

In NNs, a neuron is a mathematical processing unit that takes a set of inputs ***X*** and produces a single output y [241]. This basic setup can be seen in Figure 8. As previously mentioned, its design is inspired by the biological neuron cell, which exists in the nervous system of the human brain to transmit signals (electrical impulses) to other neurons.

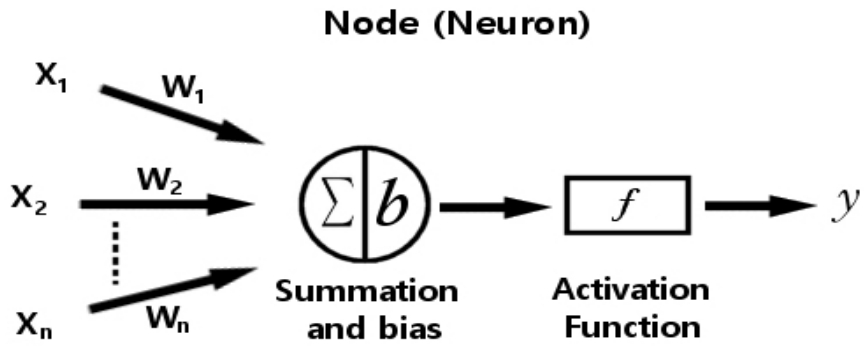

Figure 8: A basic model of a single node: x_i = input, w_i = weight, f = transfer function, y = output [244]

A neuron is made up of 3 key components (Table 7) that are used to improve the performance of the neural network: weights, bias, and activation function. Each neuron receives multiple inputs

from other neuron connections that have associated weights. If the weighted sum of inputs exceeds the threshold value of that neuron, this activates, sending the updated value to the next connected neuron. [244]

Component	Description
Weights	• Are the most important parameter in the neural network. • They have values that control the strength of the connection between two neurons. • They determine how much effect the input has on the output. • Can be either positive or negative, which are constantly being altered until the • Allows the network to learn complex relationships between inputted and outputted data
Bias	• Is a constant value. • At the summation point, is added to the input of a neuron before passing through to the activation function. • Is used to alter the activation function towards either positive or negative
Activation Function	• Is used to introduce non-linearity into the neutral network to solve complex tasks. • Takes the inputted weighted sum from summation point and produces an output value to pass onto the neighbouring neuron

Table 7: Neurons 3 key components

- **NEURAL NETWORK STRUCTURE**

A neural network structure as see in Figure 9, refers to the organisation and arrangement of its neuron processing units into layers: an active input layer, a hidden layer, and an output layer. A signal (e.g., a pixel value or numeric feature) is received by the input layer, before the hidden layer, performs a mathematical operation, typically be a Sigmoid (hyperbolic tangent) or rectified linear unit (ReLU) activation function, before giving an output [241].

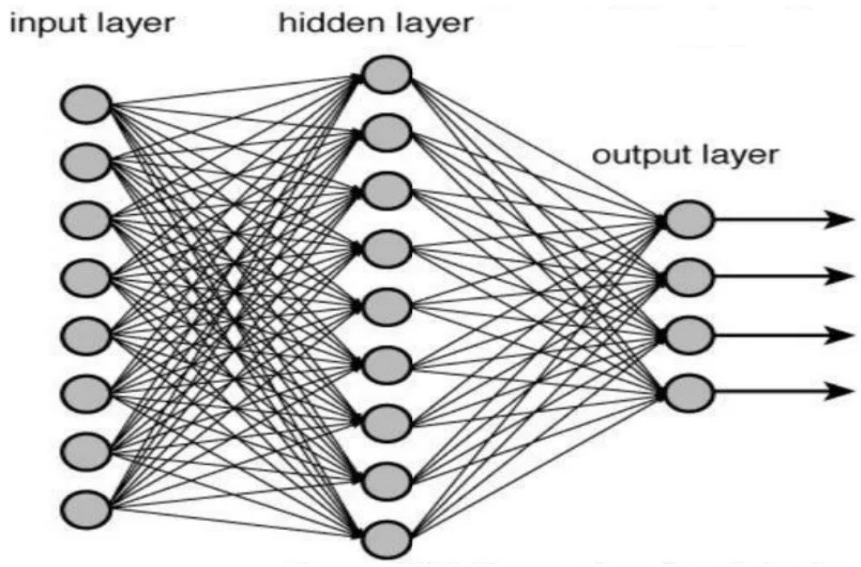

Figure 9: A neural network with inputs, a single layer of hidden neurons, and output neurons [238]

- **BACKPROPAGATION**

Backpropagation is a key component in the training of NNs. The goal of backpropagation is to minimise the error between the actual NN output and the anticipated output. Backpropagation works by first performing a forward pass through the network. This involves calculating the output of each neuron in the network, starting from the input layer to the output layer. The output of each neuron is calculated by applying a non-linear activation function to the weighted sum of its inputs. Any calculated errors are then propagated back through the network and used to update the network weights that connect the neurons together. Backpropagation repeats this process of updating weights until all errors are minimised [245].

4.3.2 DEEP LEARNING NEURAL NETWORKS

DL is comprised of NNs which combine to make DL algorithms. The main architectural difference between DL algorithms and NN algorithms is the number of layers used. A standard NN algorithm will contain three layers: Input Layer, one Hidden Layer and an Output Layer, whereas a DL algorithm, as seen in Figure 10, can contain anything from three to hundreds of hidden layers between the Input and Output [229]. DL models have proven to be very good at a wide range of complex classification and regression problems [246].

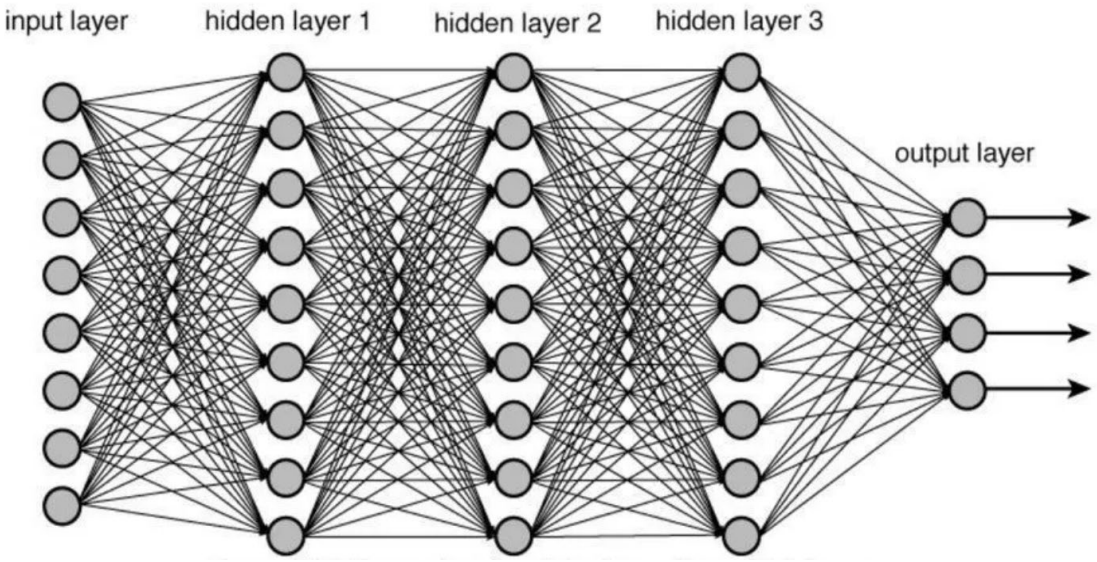

Figure 10: DL Neural Network showing input layer, 3 hidden layers and an output layer [241]

Table 8, describes some of the most popular DL methods that have been widely used, and includes CNNs and Autoencoder Networks used for solving problems with spatial and translational data, such as image recognition, Long Short Term Memory Networks (LSTMs) and Recurrent Neural Networks (RNNs) used for analysing temporal and sequential data, variance in time, such as classifying audio, speech and natural language processing, Multilayer Perceptron's (MLPs), Self Organising Maps (SOMs), General Adversarial Networks (GANs) and Radial Basis Function Networks (RBFNs).

Deep Learning Network	Method	Type	Problem
Convolutional neural network	Supervised Learning	Classification/Regression	Image, Audio, Speech
Recurrent neural network	Supervised Learning	Classification	Speech, Text
Radial Basis Function Networks	Supervised Learning	Classification/Regression	Image, Spam, Signal
Multilayer Perceptron's	Supervised Learning	Classification	Image, Speech, Machine translation
Denoising AutoEncoder	Unsupervised Learning	Classification	Image, Audio, Speech

Deep belief networks	Unsupervised Learning	Classification/Regression	Image, Video, Motion, Speech
Long Short-Term Memory	Unsupervised Learning	Classification	Speech, Text
Self Organising Maps	Unsupervised Learning	Classification/Regression	Visualisation and exploration of high dimensional data
General Adversarial Networks	Unsupervised Learning	Classification	Image and Video improvement and restoration

Table 8: Deep Learning Methods

- **CONVOLUTIONAL NEURAL NETWORKS**

Developed in the 1980's, CNNs are Feed-forwarding DL Algorithms. There is a total of three types of CNN architecture: One-Dimensional CNN (1DCNN), Two-Dimensional CNN (2DCNN) and Three-Dimensional CNN (3DCNN). The first piece of significant work was performed by Yann LeCun et al, when their 2DCNN model successfully demonstrated hand-written character recognition from the MNIST dataset [247]. However, it wasn't until the early 2000's, with the increased power of CPUs (Central Processing Units) and particularly GPUs (Graphical Processing Units), that CNNs started to become more mainstream. 2012 brought a huge surge in interest when the AlexNet model (a Two-Dimensional CNN) produced a state-of-the-Art performance, by accurately classifying complex images using the ImageNet dataset [247]. Today CNNs are one of the most established, advanced and powerful DLN available. They are mainly known for their Classification power, such as in image [248], audio [249] and speech [250]. They are designed to be Translation and Spatial Invariance, which makes them a popular choice for image processing and recognition [251]. They are extremely scalable and can process large amounts of data with very high prediction accuracy, whilst also eliminating the need for manual feature engineering. Along with their processing power, what sets CNNs apart from NNs is their key components that can been seen in Figure 11 and includes: convolutional layers, filters and kernels, pooling layers and Fully-connected (FC)

Layer [252]. These components are further discussed in Chapter 5, Classification Model Compassion Study.

In more recent times 1DCNNs have emerged as powerful tools to solving diverse medical problems, particularly where time-series data is used. Their strength lies in analysing and diagnosis classification of Bio-Signals (sequential data), making them particularly well suited to areas of Cardiology (ECGs), neurophysiological (EEGs), Respiratory, Neuromuscular and Genomic. Some instances of 1DCNNs used for diagnosis are seen here in [253] where a 1DCNN was successfully used to classify different types of cardiac arrhythmias using ECG signals. In a different study [254] that looked at monitoring brain activity, a 1DCNN was proposed for the automatic recognition of normal and abnormal EEG signals. Further examples are in presented in [255] where the automatic classification of sound signals captured through voice, cough and breath were used to diagnose Covid-19 respiratory diseases and in [256] where signals captured from neuromuscular tests were automatically classified to diagnose muscular disorders. A final instance of IDCNN classification can be seen in [257] where a 1DCNN was successful in the diagnosis of small cell lung cancer by using segmented tumour images.

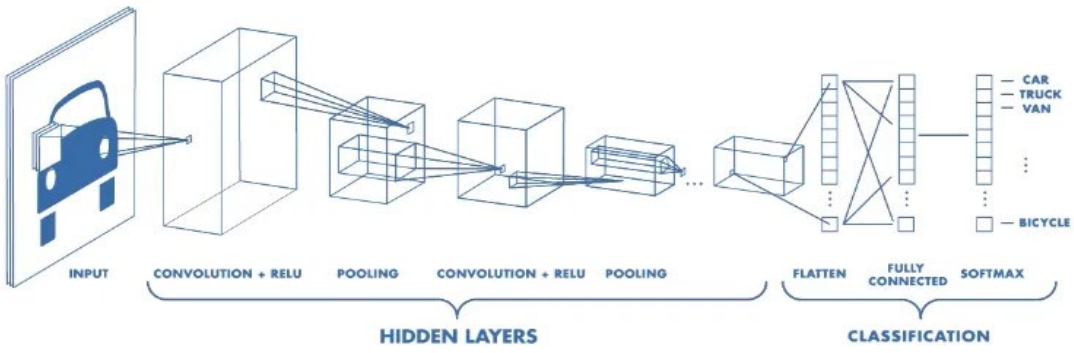

Figure 11: A Convolutional Neural Network layers through the network towards classification [258]

There are many types of CNNs, which includes: ResNet, LeNet, VGG, AlexNet, GoogleLeNet, R-CNN, Fast R-CNN, Faster R-CNN and MobileNet. Table 9, provides some brief information about CNN's.

Type of CNN	Description
ResNet	First introduced in 2015, ResNets have become of the most popular CNNs for computer vision tasks, such as Image classification, Object detection and Image segmentation. They can be trained using 1000s of convolutional layers to improve accuracy.
LeNet	LeNet-5 is a 5-layer CNN architecture developed by Yann LeCun et al in 1998. It is one the earliest and most influential CNNs for image classification, including handwritten digit recognition and face detection.
VGG	VGGs are simple but effective CNNs that typically use 16 or 19 convolutional layers with small filters. They were first introduced in 2014 for image recognition tasks and are still widely used today for objection detection, image classification and image segmentation.
AlexNet	Coming to prominence in 2012, AlexNet uses a combination of 5 convolutional layers and 3 fully connected layers. Mainly used for image recognition tasks, it is also used for objection detection, image classification and image segmentation.
GoogleLeNet	Also known as InceptionNet, it was developed in 2014 by Google AI. It is a powerful CNN and consists of 22 layers and 9 inception modules. It is widely used for image classification, objection detection, and other computer vision tasks.
MobileNet	Development by Google AI in 2017. MobileNet was designed to be lightweight for portable devices. There are 3 versions, MobileNetV1, MobileNetV2 and MobileNetV3, with later versions using improved approaches. It is applied to tasks that include: Image classification, Object detection, Face detection, Image segmentation, Video processing and Pose estimation
R-CNN Fast R-CNN Faster R-CNN	First developed in 2014, R-CNNs where one of the first approaches to use CNNs for object detection. R-CNNs are a two-stage algorithm, first a region proposal algorithm is used before object classification is performed by the CNN algorithm. Building on the back of R-CNN is Fast R-CNN and Faster R-CNN, these use slightly different approaches, each time improving on computational times. Some

	applications of the R-CNN family include; Autonomous driving, Smart surveillance systems, Robotic systems, Facial recognition and Medical image analysis

Table 9: Types of Convolutional Neural Networks

4.3.3 RANDOM FOREST CLASSIFIER

Random Forests first came into prominence approx. 20yrs ago [259]. They are classed as supervised ML algorithms that are predominantly used to solve classification problems [260]. They are highly accurate and therefore heavily used in critical areas of banking [261], finance [262], e-commerce [263] and healthcare [264] but more specifically for image recognition [265], natural language processing [266] and medical diagnosis [267]. Due to their ability to handle complex data, RFC are becoming increasingly popular in medical applications designed for early disease detection, prognosis prediction and medical diagnosis. Some instances of this can be seen here in [263], where early disease detection of retina abnormities is found in patients that suffer with diabetes, and again here [264] successfully used for the prognosis prediction of diabetes, and finally in here [265] used to better classify the diagnosis of breast cancer.

RFC is a powerful ensemble learning method, particularly effective for non-linear classification tasks. They leverage the bootstrap aggregation (bagging) technique to enhance both the accuracy and stability of predictions. Random Forests are constructed using an ensemble of decision tree classifiers, denoted as $\{h(x,\theta k),k=1,\ldots\}$. Each tree in the forest relies on a random vector $\{\theta k\}$, sampled independently and identically distributed. After each tree casts its vote for the most likely class of data point x, the final prediction is determined by a majority vote. To achieve this, the RFC create multiple training samples, each a subnet of the original dataset, for each decision tree. These trees independently make predictions, which are then aggregated to form the final prediction. This aggregation is typically achieved by taking the mean of the output values from the previous trees at each decision node before making a final decision [268].

This method makes it less susceptible to overfitting and more robust to noisy data [269]. A crucial component to Random Forests is performed by the Gini Impurity or Gini Index, which is used to measure the impurity or disorder within a group of data points. It does this by highlighting the probability of misclassifying an observation; basically, assisting how to better arrange, or split the classes, essentially reducing their impurity score to make them purer. The closer the Gini score is to zero the better it classifies the data [270].

Figure 12 presents a Random Forest. Each decision tree in the random forest is given a subset of the data to work with. When training, each decision tree generates prediction results (blue and green). The RFC predicts the final decision based on the majority of outcomes (votes) when the new data point appears.

- Pick a random ***k*** data points at random from the training set.
- Build a decision tree associated to these specified ***k*** data points (subsets).
- Choose the number ***N*** of trees required.
- Repeats step 1 and 2.
- Find the predictions of each decision tree for new data points and assign the new data points to the category with the most votes.

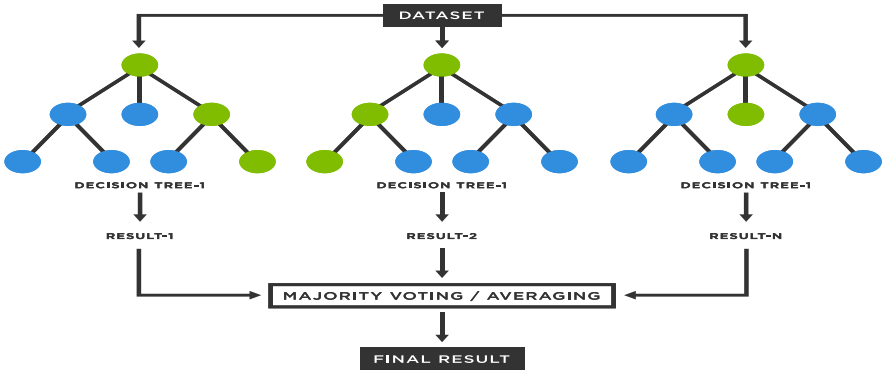

Figure 12: A typical Random Forest [271]

4.3.4 SUPPORT VECTOR MACHINE

Support Vector Machines where originally proposed in the mid-1960s, however their popularity in practical applications didn't really take off until the early 1990s [272]. SVMs are a robust and well-established supervised ML algorithm, primarily associated with classification tasks. They are widely used for tasks that include: Image recognition, object detection, text classification, bioinformatics and financial forecasting. Over time SVMs have also been a popular choice in the field of medicine, such as disease classification and diagnosis, disease prognosis prediction, monitoring of treatment and prediction of active drug compounds. Some instances of their uses are seen in [273] for multiclass molecular cancer classification and in [274] to classify skin diseases. Other uses are in [275] to predict the overall survival and disease-free survival rate and finally in [276] for drug compound classification. The power of SVMs is attributed to a technique called kernel mapping, which represents data as points in high-dimensional space. The algorithm searches for the best hyperplane to maximise the margin between the data points of different classes, thus generating critical support vectors that define the position of the hyperplane for improved classification [277].

In Figure 13, the support vectors are the 3 data points (2 blue and 1 green) laying on the dotted lines and the separating hyperplane is on the red line. The hyperplane is also represented as Linear Equation $\boldsymbol{w * x - b =}$

- w is the vector representing the direction of the hyperplane
- x is a vector representing a point in the space
- b is a scalar representing the intercept of the hyperplane with the y-axis (when dealing with 2D space)

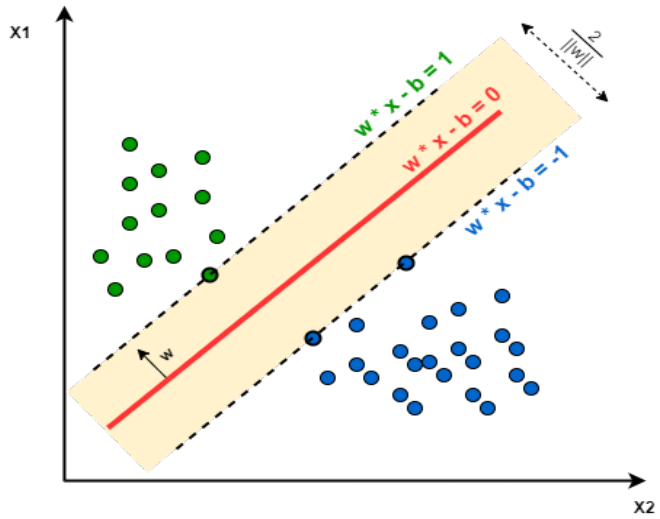

Figure 13: An SVM showing hyperplane, vectors and datapoints [271]

4.4 Related Work for the Diagnosis of OSA

In recent years, various physiological signals that include, ECG, EEG, Blood Oxygen Saturation (SpO2), snoring and airflow have been successfully used by supervised ML algorithms for the detection of OSA. A study performed in 2018, found ECG signals were more commonly used and provided the highest global classification, this was followed by SpO2, Airflow, Snoring and EEG [278]. Data preparation and machine learning methods are significant processes to how the objective of OSA detection is achieved using these physiological signals. Key limitations of traditional data preparation techniques are in the application of handcrafted methods, not only are they time consuming and costly, but also require domain knowledge and expertise [34]. This section investigates a selection of related studies, highlighting the various ML classifiers applied, the feature extraction and engineering techniques used, along with the physiological signal data employed by each study.

Traditional Machine Learning and Feature Engineering approaches

It is fair to say, many previous studies generally focus on detection methods that use feature engineering and traditional ML to diagnose OSA using time-series ECG signals. A selection of such

studies is presented here: In [279] the authors propose a novel OSA detection approach based on ECG signals by considering temporal dependence within segmented signals. A discriminative hidden Markov model (HMM) and corresponding parameters estimation algorithms are provided. A feature pool was created from various features that hold valuable information. In total 34 features were extracted from R-R-intervals and EDR (ECG Derived Respiration) signals on a minute-by-minute basis. To better understand the sequences of observations created by the HMM, a Baum-Welch algorithm was applied. Classification was performed using SVM, Logistic Regression (LR), Linear Discriminant Analysis (LDA) and KNN. Results showed the SVM+HMM produced the best results with a sensitivity 82.6, specificity 88.4 and accuracy 86.2.

A similar approach was applied in [280], again using ECG signals, the statistical method HMM with HMM kernel and SVM classifier was considered. Data preprocessing and extraction involved the removal of ECG artifacts using an ECG Decider-Block and Continuous Wavelet Transform (CWT) for the detection of R waves. The results provided an accuracy of 99.23, however the most important score of sensitivity and specificity were not produced and/or published.

Another approach that used signal analysis and feature extraction is found in [281]. This method leverages Wavelet Scattering Transformation (WST) to extract time-frequency features from the ECG data, before employing a RF for classification. The results show a very good balanced score of sensitivity 88.75 specificity 92.44 and accuracy 91.51 when testing using the hold-out validation method. Improvement to the model scores could be achieved through the combination of QRS complex detection and WST. This would have provided an enhanced feature extraction and comprehensive analysis.

A study that does use the QRS complex detection for feature extraction is seen in [282]. The method used, extracts 10 time-domain features from the ECG signal and reduces them to 5 using PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis). The best results were

found using LDA with RFC, providing a very high score for sensitivity 92.17, specificity 94.79 and accuracy 95.01, suggesting this is a well-balance model.

A second study that uses QRS complex detection is seen in [283]. The method here is to use Statistical and Time Series Feature Engineering detecting QRS complexes to determine RR intervals, before extracting statistical features. Several classifiers were applied, with three standout results, being the KNN 1.00 recall, 97.56 precision 98.75 accuracy, the SVM recall 97.50, precision 97.50 accuracy 97.50 and the RFC recall 1.00, precision 93.02 accuracy 96.25.

Employing a combination of both ECG and SpO2 signals is quite common for the detection of OSA [278]. The application of such a multimodal approach is used in [284]. Here a Finite Impulse Response (FIR) bandpass filter and Hamilton algorithm are applied to the signal data, to firstly remove any noise, secondly, to locate and correct the R-peaks, and thirdly, to extract and fuse all created features. RF Classification provides excellent results showing a sensitivity 95.9, specificity 98.4 and accuracy 97.5. A highlighted limitation of this study was the small sample size used for training and testing. Further to this it was found that preprocessing of the SpO2 features produced stronger results when compared to the ECG features.

A more complex approach that also adopted an RF Classifier is found in [285]. Feature selection and extraction are captured from ECG signals using a combination of methods that included Dynamic Time Warping, Median Frequency Features and Discrete Wavelet Transform (DWT). The RF Classifier produces respectable scores of recall 82.40%, precision 82.10%, and accuracy 82.43%. To further evaluate this study, the authors performed the same tests using a LSTM classifier, but with lesser results.

Again in [286] the approach is to apply an RF Classifier. This time using EEG data derived from polysomnography 30 overnight sleep patients. Preprocessing and denoising of the data was performed using IIR Butterworth Band Filters and feature selection employed a 10-step NCA (Neighbourhood Component Analysis) algorithm to improve the overall effectiveness of the data

and classification accuracy. A total of 13 features were identified across the whole area. Classification was performed using an RF Classifier, an SVM classifier and KNN classifier. The RF Classifier was seen to produce the best results with a respectable score of recall 89.0%, F1-score 87.0% and accuracy 86.0%.

In [287] the authors approach was to use the Random Under Sampling Boosting (RUSBoost) technique. The preprocessing of single lead ECG signals was performed by applying a novel signal processing technique called TQWT (Tunable-Q factor Wavelet Transform) to decomposed and clean the ECG segments. To assist with imbalances of the dataset, the classification algorithm, RUSBoost was seen as the best fit and provided the authors with more balanced results for sensitivity 87.5 and specificity 91.49, with an accuracy of 88.88. Dataset imbalance is seen as a common issue with medical data. Within this current PhD thesis study, the issue of imbalance was solved by applying our own unique preprocessing windowing strategy.

The proposal in [288], was to detect OSA captured from snoring audio signals. Snoring signals are the least common physiological signal employed for this type of ML task. This is since they are quite susceptibility to noise and outside interference. Preprocessing of the signals involved the use of VAD (voice activity detection) and mixed Gaussian distribution model. Feature extraction was obtained via MFCC (Mel-Frequency Cepstrum Coefficient) Meyer Cepstrum. A total of 24 classification experiments were performed using different feature vector dimensions across the 3 separate classifiers, KNN, SVM and Gaussian Naive Bayes (GNB). The KNN outperformed both the SVM and GNB, with and accuracy of 85.55 and F1 score of 85.04. The important scores of sensitivity or specificity were not captured and/or published.

Combination of Machine Learning Models

In an attempt to improve classification results, many research papers have employed the method of combining two or more ML models. A quite complex approach to this is seen in [289], where preprocessing is achieved using a Hamilton algorithm and median filter to both located and extract

R-peak features and to further remove any noise, thus preserving the features and generate R-R-intervals. The claimed novelty of this study is found within the combined approach of 3 types of ML models used; two 1DCNN's for feature extraction; Two LSTM networks for gradient vanishing problem and long-term dependency; and an SVM for Classification. The method was seen to produce very good scores of sensitivity 91.24%, specificity 90.36%, accuracy 90.92%, and F1_Score 92.76%.

A second study that combines ML models is seen in [202]. Using digitised ECG signals, this study looked to compare thirteen classic ML models and four DL models for automatic detection of OSA. Signal processing involved removing unwanted frequency noises using a digital IIR (Infinite Impulse Response) notch filter. Feature extraction codes were applied to capture nine specific features from ECG signals, reducing the data's high dimension and improving the overall performance. The results showed the 4 DL models outperformed the 13 classical ML models, with the hybrid model CNN-LSTM network producing a best performance with an recall 88.7%, precision 86.6% and accuracy 86.25%.

A further combination of ML algorithms is seen in [290], which proposed a novel convolutional neural network model based on multi-task learning for automatic OSA detection, using a combined supervised (1DCNN) and unsupervised (Auto-encoder) model and a feature fusion method. EGC signals were pre-processed using a variety of complex procedures, that included a FIR Filter for denoising, a Hamilton algorithm to extract RRIs, a median filter algorithm to remove any unexpected R-peaks, a cubic spline interpolation and HRV for better R-peak calculation and measurement. To solve class imbalance issues, the practice of oversampling using Borderline-SMOTE (Synthetic Minority Oversampling Technique) was applied. The goal of this multitask learning approach was to build-on the insufficient feature extraction ability of the single supervised learning task (1DCNN). Applying a feature fusion extraction method across the two models was seen to improve results: sensitivity 90.3%, specificity 91.6%, and accuracy 91.1%. A

comparison study showed respectable scores against other DL studies that also bettered feature engineered studies.

Deep Learning Studies

In recent years, there has been a growing interest in DL techniques, which offer the potential to surpass the limitations of earlier ANNs and traditional ML approaches. This study in [291] looks to address many of these limitation through the application of a deep learning 1DCNN model. Data was captured through the device of nasal pressure recordings captured from 179 subjects undergoing polysomnography tests. To clean the signals, preprocessing was performed using IIR Butterworth and real-time filters. The signals were then adaptively normalised and segmented using a 10-s sliding window at 1-s intervals for training and classification. All feature extractions and selections were performed by the CNN architecture which consisted of 3 convolutional layers, two max-pooling layers, and two fully connected layers, which provided the results: sensitivity 81.1%, specificity 98.5%, accuracy 96.6%. Although high scores are produced for accuracy and specificity, the imbalance of sensitivity and specificity is a concerning limitation of this model, which can cause the misidentification of apnoea events as normal, thus increasing the likelihood of misdiagnosing of patients.

A study in [292], that looked to address the limitations of feature extraction using traditional ML models. Here, a 1D squeeze-and-excitation residual group network (1D-SEResGNet) using a multi-feature (RII+RA+QA) fusion method was proposed to carefully extract the complementary information of HVR and EDR. Further to this a Hamilton algorithm and bandpass filter was applied to find R-peak from 2-minute ECG signal segments to detect OSA. The architecture of the model comprises of using 20 convolutional layers, a max-pooling layer and fully connected layers and Results of segment detection showed a sensitivity 87.6%, specificity 91.9% and accuracy 90.3%

The following 3 studies provide a similar approach to the current research found in this PhD thesis, that includes the type of deep learning model (1DCNN) applied and the dataset (Apnoea-ECG

database) employed. Reference [293] considers a 1DCNN model for the automatic detection of OSA using 70 single-lead ECG signals recordings collected from 35 subjects overnight sleep PSG tests. Pre-processing involved the complex approach of Fourth-Order Butterworth bandpass filtering to reduce baseline drift and high frequency interference. Further to this, a statistical measuring tool, z-score, was also applied for the removal of noise and normalisation of the signal data. The models architecture consisted of 10 identical convolutional layers, 5 fully connected layers and 4 identical classification layers. To help improve feature extraction, each of the 10 convolutional layers had 45 feature maps (n_filters) and a kernel length (k_size) of 32. These two parameters determined the number of output channels (feature maps) and the width of the filter (kernel). To assist in reducing network complexity and improve efficiency, each convolutional layer was equipped with a max-pooling layer. Before the classification stage, a flattened layer was applied to convert the 2D feature matrix of 45 1D feature maps into a 1D feature vector. Within the classification stage, each classification layer included a fully connected layer, which allows for learning of complex relationships between extracted features. At both the feature extraction stage and classification stage, a He Normal Initialisation technique is applied to initialise the weights. Following this, a batch normalisation layer is added to normalise the data before it enters the ReLU function, to improve the speed, performance, and stability of the neural network. Finally, a dropout layer with a rate of 0.5 was added to reduce overfitting. After the classification stage, a SoftMax activation function calculates the probabilities of the two outputs of FC-2 layer, which corresponds to the classification of both normal and apnoea events. At the training phase an Adam optimiser was applied to minimise cross entropy. The model was trained and validated for 50 epochs using ECG signals from both the released and withheld datasets. Each experiment took approx. 48 minutes to complete. When comparing to other related studies, this model was seen to show the best scores for per-minute apnoea detection, sensitivity 81.1%, specificity 92.0%, accuracy 87.9%, and AUC of 94%. A major limitation of this model is seen in the imbalance between sensitivity and specificity, which is likely to cause the misidentification of apnoea events

as normal, thus increasing the likelihood of misdiagnosing. To address this, within this current PhD thesis, issues with imbalances between sensitivity and specificity were resolved through hyperparameter changes, adjustment of neuron weights and inclusion of balanced datasets. A further limitation of this model is the duration of each experiment. These times are excessively long when considering our model completes each experiment in approx. 3 to 4 minutes. It sometimes seems there is too much focus on increasing network depth. Maybe reducing the number of convolutional layers, hidden layers and filters could speed this up.

Reference [294] also considered the more state-of-the-art approach, 1DCNN. The architecture of this model consisted of six convolutional layers with varying depths of filter and kernel sizes, a ReLU activation function, max-pooling layers, a dropout layer, a fully-connected layer and a SoftMax output. The dataset was made up from single-lead ECG signals collected from 82 subjects who took part in a standard full-night PSG test. Data processing involved the application of a bandpass filter was applied to remove undesirable noise and segment data using 10s events. The process of the model begins at the convolutional stage where batch normalisation is firstly applied before feature extraction is performed using six convolutional layers. This is followed by a ReLU activation function to improve efficiency and a Max-pooling layer to extract only key features. A drop out layer is then applied to reduce the possibility of overfitting. At the classification stage, a fully connected layer is applied to learn the complex relationships and extract higher level features before a SoftMax output shapes the predicted classes. Many of the experiments were conducted in a fully supervised manner. Optimal performance of the model was achieved through the fine-tuning of extensive and complex hyperparameters across a varied depth of convolutional layers. Optimal performance was found in the six-layer architecture which produced excellent results of: recall 96%, precision 95%, and F1_score 96%. Excellent results were also achieved using three-layer architecture all the way up to nine-layers, anything over nine layers caused both overfitting and underfitting. Further performance measurements of the 1DCNN architecture were compared to other models from previous studies, which showed the 1CDNN outperformed each of these

models: SVM, Fuzzy reasoning module, LDA, QDA AdaBoost, Bagging REPTree and Kernel density classifier. One of the limitations highlighted in this study was the small sample size of the data used.

This final study in [295], presents our previously published work. Again, using a 1DCNN model for the automated detection of OSA captured from single-lead (ECG) signals. The architecture of the model is consistence with our current model in that it used a single Convolutional Layer, Max Pooling Layer, Flattening layer, Fully Connected Multilayer Perceptron (MLP) with hidden dense layers (FC layers) and a SoftMax output. For training the network, a combination of an Adam optimiser and CrossEntropy loss function is applied. A variety of hidden layers were trialled, including various depths of Convolutional Layers, FC Layers (within the MLP) and a selection of activation functions and outputs. 70 night-time polysomnography tests using 35 subjects were recorded and split into a released set of 35 recording for training and validation and a withheld set for testing. To create the datasets, preprocessing focused on the physical segmentation of the data into two classes (Apnoea & Non-Apnoea). These classes were then merged and squared into specific window sizes for training. The model was evaluated using various metrics and showed a high classification for training and validation (sensitivity 0.9705%, specificity 0.9725%, F1_Score 0.9717%, accuracy 0.9377%, ROCAUC 0.9945%). A limitation of this model was the absence of a testing stage using the withheld set. Other limitations include the absence of a drop out layer which may have eased some of the overfitting that is evident in some of the test runs.

These studies explored a diverse range of ML techniques to detect OSA using ECG time-series signals. This included both traditional ML and more state-of-the-art DL algorithms. Many of these studies employed the Apnoea-ECG database, a commonly used dataset in this field and primarily used throughout this PhD study. In addition to the more traditional ML algorithms, feature engineering, particularly Statistical and time-series feature engineering with QRS complexity, was found to be very effective in extracting relevant information from the ECG signals and producing good to excellent scores. By carefully selecting and transforming these features the ML models

were able to better capture the underlying patterns associated with OSA. Subsequently, signal processing techniques were utilised to preprocess the ECG signals, reducing noise and improving quality of the inputted data. This combination of feature engineering and signal processing enhanced the overall performance of the traditional ML models, leading to more accurate and robust OSA detection. Two noteworthy studies that demonstrated the power of this approach were seen in [283]. Which highlight the potential of ML, coupled with feature engineering and signal processing. When analysing the DL studies, these also produced good to excellent scores, with all of them employing signal processing techniques and multiple convolution layers. Some studies also applied feature engineering techniques to enhance the model's performance. However, while these studies explored a diverse range of ML and DL techniques for OSA detection, it is important to consider the computational cost and time investment associated with these approaches. Feature engineering and signal processing, while effective, can be computationally intensive and time-consuming. Additionally, the use of multiple convolutional layers in DL models further increases the model's complexity and training time. Furthermore, most of the DL studies, showed a lack of conviction or evidence towards hyperparameter testing and fine-tuning of the models. Moreover, a notable deficiency in these studies is the absence of detailed experimental timings. For instance, only one study, the top performing 1DCNN model [294], reported experimental times, which were significantly excessive. When considering the drawbacks and limitations of these studies, it is clear to see where most, if not all of them, fall-over. Therefore, the 1DCNN model I'm proposing in the following section "Methodology" offers a far simpler and more efficient approach, eliminating the need for complex feature engineering, signal processing and deep multi-layer convolutional models. Furthermore, to enhance data range and model performance the proposed 1DCCN model will be trained and tested using a unique windowing strategy to increase the amount of training data observations. This approach will reduce computational overheads, training times and improve the model's generalisability. To further evaluate the 1DCNN's effectiveness, comparative experiments will be conducted against

traditional machine learning algorithms, namely RFC and SVM. These two algorithms are well established for binary classification tasks, and particularly in the domain of OSA detection using ECG time-series signals, as already demonstrated in the above studies, producing excellent results. Notably, these models will be trained on the same Apnoea-ECG dataset utilised throughout this PhD study, providing a fair comparison.

5. METHODOLOGY

This chapter presents the proposed framework and methodology. It firstly looks at the data acquisition and data pre-processing techniques. Providing an insight to how the data is transformed into something usable. Following this is a look at how the novel windowing strategy is applied to deliver the required datasets. The chapter then focuses on the architecture of the 3 ML algorithms, chosen for the comparison study, 1DCNN, RFC and SVM, understanding the essential components to building each model. The penultimate part of this chapter provides an overview of each of the metrics used to measure the performance of each model. The final part of this chapter presents a list (Tables 22 – 25) of hyperparameter and fine-tuning experiments performed on the model 1DCNN-500 using the dataset W=500.

5.1 DATA ACQUISITION AND MATERIALS

5.1.1 DATA

The dataset selected to be analysed for this study was sourced from Physiobank, a subdivision of the publicly accessible and well renowned on-line data exchange site, Physionet, which is supported by the National Institute of General Medical Sciences (NIGMS) and the National Institute of Biomedical Imaging and Bioengineering (NIBIB) under NIH grant number 2R01GM104987-09. PhysioBank is a large archive of well-characterised digital recordings of physiologic signals and related data for use by the biomedical research community. The dataset (Apnoea-ECG) that we eventually settled on was selected from a possible eight candidates, all of which came close to the prerequisites of both the type and size of dataset required to perform our analysis. Further motivation to using this dataset was found from the widespread use in high quality journals and conferences where it has featured.

5.1.2 APNOEA-ECG DATASET DESCRIPTION

Penzel et al. [121] conducted a comprehensive study between 1993 and 1995, to investigate and record the effect of OSA on arterial blood pressure in subjects with moderate and severe Sleep Apnoea. A second study undertaken between 1998 and 1999, to create a normative set of sleep recordings, with the main research focused on multi-channel EEG recordings performed on healthy volunteers and patients suffering with Sleep Apnoea. They combined the ECG recordings and records from both of these studies to create a single database (Apnea-ECG database) and making it publicly available via Physionet. The Apnoea-ECG database contains the records of 70 patients (subjects). The dataset comprises a mixture of male and female subjects with ages ranging from 27 to 63 years (mean 45yrs). Body mass index (BMI) recordings vary between 19.2 and 45.33kg. (mean: 28.01 ± 6.49 kg.) and body weights range between 53 to 135 kg (mean: 86.3 ± 22.2 kg). Only 35 records have associated annotations - (a01 to a20 (20 ECG signals), b01 to b05 (5 ECG Signals), and c01 to c10(10 EGC signals) and used for training. The 35 non-annotated records were withheld for testing purposes. Each of the ECG annotated recordings vary in length from approx. 7hrs to 10 hours. The three groups within the recordings are defined by the AHI index. The AHI index across these groups varied between 5 and 82 events per hour. Group A (Apnoea-Set): each subject has over 100 minutes of Apnoea; Group B (Borderline-Set): this group has between 5 to 99 minutes of Apnoea; and Group C (Normal-Set): each subject in this group has between 0 to 3 minutes of Apnoea. A total of 17,125 minutes (or 285hrs 25mins) sleep-time was recorded; of this, 6,514 minutes (or 108hrs 34mins) was scored as Apnoea and 10,611 minutes (176hrs 51mins) was scored as non-Apnoea. Table 10, provides a summary for the group distributions.

Subject Recordings	EGC Files	Group Type	Apnoea (Mins)	Non-Apnoea (Mins)
A01 – A20	20	Apnoea-Set	6250	3811
B01 – B05	5	Borderline-Set	252	2060

C01 – C10	10	Normal-Set	12	4740
		Total =	6514	10611

Table 10: Apnoea-ECG Database – 3 Group types

5.1.3 DATASET FILES

Figure 14 shows the selection of file recording associated to subject a01. Each subject (a01 – c10) has the same associated file recording. The files with names of the form ***rnn.dat*** contain the digitised ECGs (16 bits per sample, least significant byte first in each pair, 100 samples per second, nominally 200 A/D units per millivolt). The ***.hea*** files are (text) header files that specify the names and formats of the associated signal files. The ***.apn*** files are (binary) annotation files, containing an annotation for each minute of each recording indicating either the presence or absence of apnoea at that time; these are available for the 35 learning set (training) recordings only. The ***qrs*** files are machine-generated (binary) annotation files, made using the single-channel QRS detector software (sqrs125).

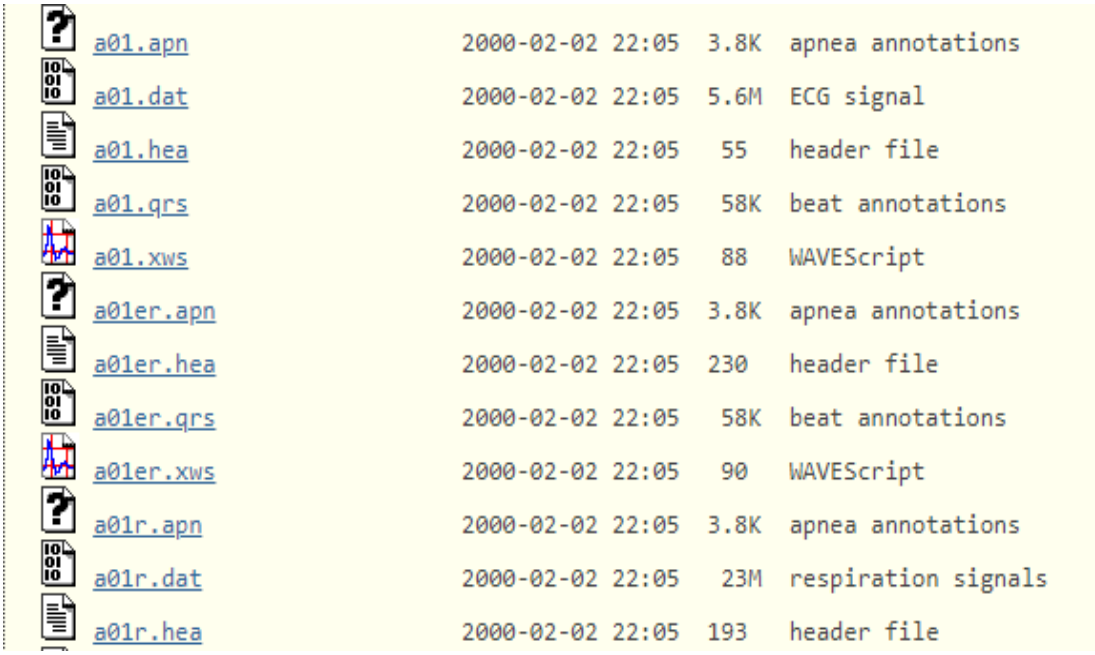

Figure 14: A selection of dataset files recordings for subject a01

5.1.4 DATA VIEWING AND ANALYSING

The data was analysed and extracted from the PhysioBank using a combination of software tools seen in Table 11.

Cygwin	Cygwin is a Unix-like platform and environment. It includes the library, cygwin1.dll, together with a vast set of GNU and other software applications. Including compilers and software development tools.
WFDB library	For an API for access to PhysioBank (for C, C++, and Fortran)
WFDB applications	For command-line tools for signal processing and automated analysis
WAVE	For software for viewing, annotation, and analysis of signal
PhysioToolkit Library	For signal processing and analysis wfdb.shtml#library

Table 11: Software Tools for extracting and viewing

5.1.5 DATA EXTRACTION, FUNCTION CODE AND FILE FORMAT

The extraction of ECG signal data and associated annotations files from PhysioNet, was achieved by passing the functions ***rdsamp*** and ***rdann*** and required switches, through the Cygwin platform. Table 12, below shows the functions and switches used for the extraction of the ECG times series signal data in csv.format and the annotation data in txt.format.

USERS+cmssthom@CMSSTHOM ~ $ rdsamp -r apnoea-ecg/a01 -c -H -f 0 -v -pd >a01.txt

USERS+cmssthom@CMSSTHOM ~ $ rdann -r apnoea-ecg/ a01 -f 0 -a apn >a01.txt

The ***rdsamp*** function exists as a simple alternative to rdrecord for the common purpose of extracting the physical signals and important descriptor fields. It reads signal files for the specified record and writes the samples as decimal numbers on the standard output. By default, each line

of output contains the sample number and samples from each signal, beginning with channel 0, separated by tabs. Below shows a list of additional options for the ***rdsamp*** function.

-r	the record name
-c	Produce output in CSV (comma-separated value) format (default: write output in tab-separated columns).
-f	(time) Begin at the specified time. By default, rdsamp starts at the beginning of the record.
-H	Read the signal files in high-resolution mode (default: standard mode). These modes are identical for ordinary records. For multifrequency records, the standard decimation of oversampled signals to the frame rate is suppressed in high-resolution mode (rather, all other signals are resampled at the highest sampling frequency).
-pd	Print time of day and date if known, as [hh:mm:ss DD/MM/YYYY]. The base time and date must appear in the header file for the record; otherwise, this format is equivalent to "e" format (below).
-v	Print column headings (signal names on the first line, units on the second). The names of some signals are too wide to fit in the columns, such names are shortened by omitting the initial characters (since names of related signals often differ only at the end, this helps to make the columns identifiable). Names of units are shortened, when necessary, by omitting the final characters, since the initial characters are usually most important for distinguishing different units
rdann	Reads the annotation file specified by record and annotator and writes a text-format translation of it on the standard output, one annotation per line. The output contains (from left to right) the time of the annotation in hours, minutes, seconds, and milliseconds; the time of the annotation in samples; a mnemonic

	for the annotation type; the annotation subtyp, chan, and num fields; and the auxiliary information string, if any (assumed to be a null-termi-r the record name
-a	annotator
-f 0(time)	Begin at the specified time. By default, rdann starts at the beginning of the record; if modification labels are present, they are not printed unless -f 0 is given explicitly.

Table 12: Apnoea-ECG Database – 3 Group types

5.1.6 APNOEA-ECG DATABASE STRUCTURE

Apnoea-ECG data contains three major parts (Table 13): digitized ECG signals, annotation scores and subject records. These parts are essential to building the required datasets.

A set of digitized ECG signals	One for each patient containing: Elapsed time (hrs, mins, secs); ECG voltage output
A set of Annotation	One for each patient containing: Elapsed time (hrs, mins, secs); Elapsed time,10 milliseconds samples; Code for the annotation type; N (Non-apnoea or A (Apnoea)
Subject Information Table	Shows a record for each patients with information containing: Patient record/number; Minutes of sleep; Non-apnoea minutes; Apnoea minutes; Apnoea index (AI); Hypopnoea Index (HI); Apnoea-hypopnea index (AHI); Age; Sex; Height (cm) ; Weight (kg)

Table 13: Apnoea-ECG Database Structure

- **Subjects Recording information.**

Each of the subject's recordings are organised into a released set (for training) of 35 records (Class a – a01 through to a20, Class b – b01 through to b05, and Class c – c01 through to c10), and a withheld set (for testing) of 35 records (x01 through x35). Class c, having at most 5 min in OSAS), borderline set (class b, having 5–99 min in OSAS), and apnoea set (class a, having at least 100 min in OSAS). The average of the ages of subjects of classes c, b, and a are 33, 46, and 50 years respectively. The subjects of class c, b, and a are 27–47, 39–53, and 29–63 years old respectively. The Apnoea-hypopnea index (AHI) of these subjects varies from 0 to 83. The ages of these subjects range from 27 to 63 years wherein the mean age was 45 years and standard deviation of age was 10 years. The body mass index (BMI) of the subjects varies between 19.2 and 45.33 kg. m2 (mean: 28.01 ± 6.49 kg. m2). The weights of the subjects range from 53 to 135 kg (86.3 ± 22.2 kg). The recordings of each set are about 7–10 h long.

Subjects	35	35 annotated recordings
Gender	25 Males	7 Females
Age	27 to 63yrs	mean 45yrs
Body Mass	19.2 to 45.33 kg	mean 28.01 ± 6.49 kg
body weights	53 to 135 kg	mean 86.3 ± 22.2 kg
AHI index	5 to 82 events p/h	
Annotated recordings	35 recording	7hrs to 10hrs
Total sleep rec	17,125 minutes	or 285hrs 25mins
Apnoea	6,514 minutes	or 108hrs 34mins
Non-Apnoea	10,611 minutes	or 176hrs 51mins

Group A	Apnoea-Set	100 mins of Apnoea
Group B	Borderline-Set	5 to 99 mins of Apnoea
Group C	Normal-Set	0 to 3 mins of Apnoea

Table 14: Apnoea-ECG Database Information

5.1.7 ANNOTATIONS

The recordings have been labelled by an expert scorer for sleep apnoea/hypopnea events. The recordings have been labelled on 60-second basis. The label of each segment indicates the presence (or absence) of OSAS events in that segment. The sampling frequency of the ECG signals is 100 Hz. The resolution of the signal is 12-bit; so, each ECG signal segment is 60 s or 6000 samples long. In each of these files, the first annotation is placed at 0 seconds and is associated with the following one-minute interval (i.e., 0-59.99 seconds elapsed time from the beginning of the record), the second annotation is placed at 60 seconds and is associated with the next one-minute interval (60-119.99 seconds), etc. Each "A" annotation indicates that apnoea was in progress at the beginning of the associated minute; each "N" annotation indicates that apnoea was not in progress at the beginning of the associated minute.

5.1.8 APNOEA-ECG DATABASE TABLE OF ADDITIONAL INFORMATION

Table 15 presents additional information about the Apnoea-ECG Database records. The first column contains the record names. For scoring, each record was divided into one-minute non-overlapping segments. The number of such segments is given in the second column of the table below. Each minute was classified as either a "non-apnoea minute" (tallied up in the third column) or an "apnoea minute" (tallied up in the fourth column). Minutes containing either apnoea or hypopnea were classified as apnoea minutes.

The one-minute segments were grouped into 60-minute non-overlapping segments for each record, and the number of these groups containing between 5 and 60 apnoea minutes is tallied

up in the fifth column, "hours with apnoea". Hours containing 1 to 4 apnoea minutes (not considered to be clinically significant) were counted as hours without apnoea for this purpose.

The apnoea index (AI) is the number of apnoea's observed per hour, and the HI is the number of hypopneas observed per hour. The apnoea-hypopnea index (AHI) is defined as the sum of AI and HI. The final 4 columns present: age, sex, height in centimetres, and weight in kilograms.

Record	Length (mins)	Non - Apnoea (mins)	Apnoea (mins)	Hours with Apnoea (hrs)	AI (PH)	HI (PH)	AHI	Age	Sex	Height (cm)	Weight (kg)
a01	490	20	470	9	12.5	57.1	69.6	51	M	175	102
a02	529	109	420	9	57.2	12.3	69.5	38	M	180	120
a03	520	274	246	9	38.4	0.7	39.1	54	M	168	80
a04	493	40	453	9	73.4	4	77.4	52	M	173	121
a05	455	179	276	8	35	6	41	58	M	176	78
a06	511	305	206	8	16.6	8.1	24.7	63	M	179	104
a07	512	190	322	9	46	17	63	44	M	177	105
a08	502	313	189	7	32	10	42	51	M	179	88
a09	496	115	381	9	23.1	8.6	31.7	52	M	178	82
a10	518	418	100	6	11	10	21	58	M	176	78
a11	467	245	222	8	11	3	14	58	M	168	103
a12	578	44	534	10	70	10.2	80.2	52	M	173	121
a13	496	252	244	9	32	10	42	51	M	179	88
a14	510	127	383	8	17.3	37.4	54.7	51	M	175	102
a15	511	143	368	9	46	6	52	60	M	176	113
a16	483	163	320	7	17	24	41	44	M	177	105
a17	486	328	158	5	21	12	33	40	M	179	96
a18	490	52	438	9	75.5	6.9	82.4	52	M	178	82
a19	503	298	205	9	34	0	34	55	M	178	90
a20	511	196	315	9	35	6	41	58	M	176	78
b01	488	469	19	2	0.12	0.12	0.24	44	F	170	63
b02	518	425	93	5	14	5	19	53	M	176	85
b03	442	369	73	4	22	2	24	53	M	176	85
b04	430	420	10	1	0.7	0	0.7	42	M	180	64
b05	434	377	57	3	2	3	5	52	M	180	135
c01	485	485	0	0	0	0	0	31	M	184	74
c02	503	502	1	0	0	0	0	37	M	180	83
c03	455	455	0	0	0	0	0	39	M	184	65
c04	483	483	0	0	0	0	0	41	F	180	65

c05	467	464	3	0	0	0	0	28	F	169	57
c06	469	468	1	0	0	0.25	0.25	28	F	171	65
c07	454	450	4	0	0	0	0	30	F	168	56
c08	535	535	0	0	0	0	0	42	M	180	64
c09	469	467	2	0	0	0	0	37	M	180	83
c10	432	431	1	0	0	0	0	27	M	184	72

Table 15: Apnoea-ECG Database Subject Information

5.1.9 MATERIALS

- **SOFTWARE R AND PACKAGES**

A selection of software was used for the creation of the dataset and algorithm, that included, Anaconda, R-programming, Python, TensorFlow and Keras.

Anaconda presented a comprehensive environment for managing Python and R packages, ensuring a smooth installation of necessary libraries and updates. R's statistical capabilities provided valuable preprocessing data analysis of the ECG time-series data, whilst Python offered a comprehensive set of libraries and simplicity for programming the model. TensorFlow and Keras combined to offer a robust framework and user-friendly API for building, training, and evaluating the model. Finally, Jupyter Labs served as an integrated development environment for code execution, visualisation, and model deployment.

5.2 DATA PRE-PROCESSING AND WINDOWING STRATEGY

Data pre-processing is a crucial step to ensuring the data is in a workable state that can be computed in a practical manner. The pre-processing path required for purging the database consists of several important and essential stages that includes, Data Profiling, Data Cleaning, Data Reduction and Data Transforming. Figure 15 provides a graphical view of this process.

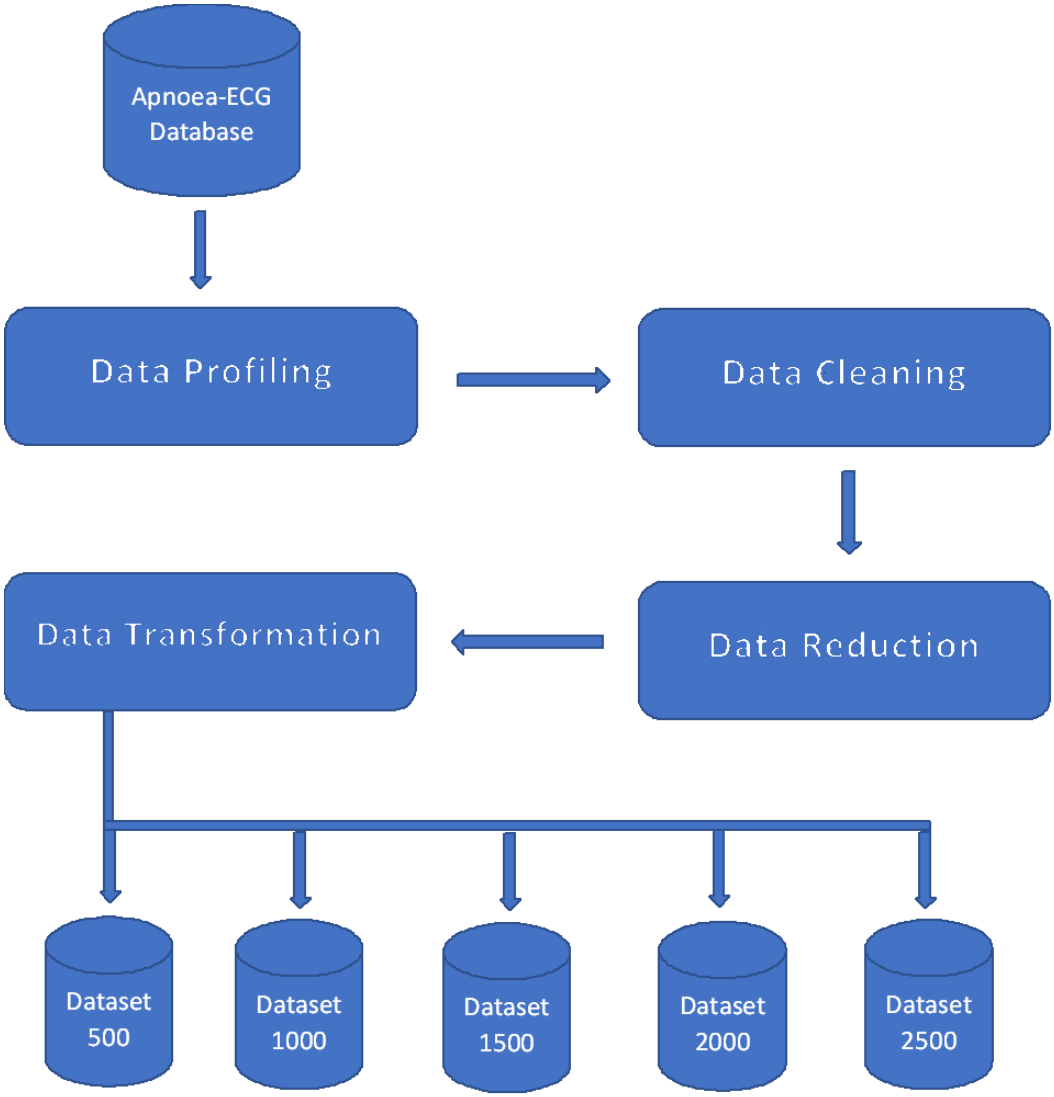

Figure 15: Data Preprocessing into Datasets

5.2.1 DATA PROFILING

The process starts by using data profiling techniques to understand and summarise the characteristics of each file's viability for the construction of the dataset. Five major types of files (Table 16) existed in the dataset. After opening and analysing each files attributes and consideration of what's required for the construction, it was decided that the ECG signal files (rnn.dat) and annotations files (.apn) would be the most suitable options. These two files held the unique identifiers, values and format required for the approach.

File Type	File Description
.hea files	• Short text header files that describe the contents of the associated signal files. Specifying the names and formats
QRS annotations files	• Machine-generated (binary) annotation files, made using the single-channel QRS detector software (sqrs125)
rnn.dat files	• **Contain the digitised ECGs (16 bits per sample, least significant byte first in each pair, 100 samples per second, nominally 200 A/D units per millivolt).**
.apn files	• **Binary annotation files, containing an annotation for each minute of each recording indicating either the presence or absence of apnoea**
XWS files	• WAVEScript file associated with the WAVE application (Linux). To view each record

Table 16: Apnoea-ECG Database Files

5.2.2 DATA CLEANING

Through the process of cleaning the chosen data, approximately 75 million samples (cells) were calculated and assessed. Elimination of unwanted variables and null values provided a solid source of data. Additional information that included variables such as Age, Sex, Height, and Weight had no value and were regarded has surplus. Within the signal files only 14 empty cells were identified and ultimately filled with (0) zeros values. No invalid data was identified. As the process advances through the merging stages, more surplus columns and data are removed.

5.2.3 DATA REDUCTION

This process is designed to extract the necessary amount of ECG signal data for the construction of the datasets for training. At the centre of this process are two significant files: annotation.txt

and signal.csv. There was a total of 35 ECG signal files (a01 – C10) with 35 corresponding annotation files (a01 to c10).

Annotation Reduction (Automated process)

The extraction of specific data was achieved using the R-programming, python scripting and Excel software.

- Using Microsoft Excel software, open the Annotation record (example a01.txt) from '***Annotation***' folder.
- Convert time by formatting cells in (column B) to custom hh.mm.ss.
- Copy columns B (Elapsed times) & C (Numbered Sectors) into the spreadsheet, (example a01).

Figure 16 presents the first record to be analysed, a01.txt (columns 1 – 7 and rows 1 – 489). The first column shows the row no; the next two columns indicate the elapsed time (i.e. the interval) from the beginning of the record to the sample marked by the annotation. (Column 2 gives this in hours, minutes, and seconds, and column 3 gives the same information in sample intervals. In these records, one sample interval equals 10 milliseconds.) The fourth column contains a mnemonic code for the annotation type, A for "Apnoea" and N for "Normal (Non-Apnoea)". The remaining three columns contained no data and could be discounted. This sample output (a01.txt) shows that during the onset of sleep (initial 17mins): the first 12 minutes (beginning at 0:00.000 to 12:00.000 or sample number 0 to 72000) shows apnoea is not present, and denoted by the character (N); the following 5 minutes (beginning at 13:00.000 to 17:00.000 or sample number 78000 to 102000), shows apnoea is present, and denoted by the character (A).

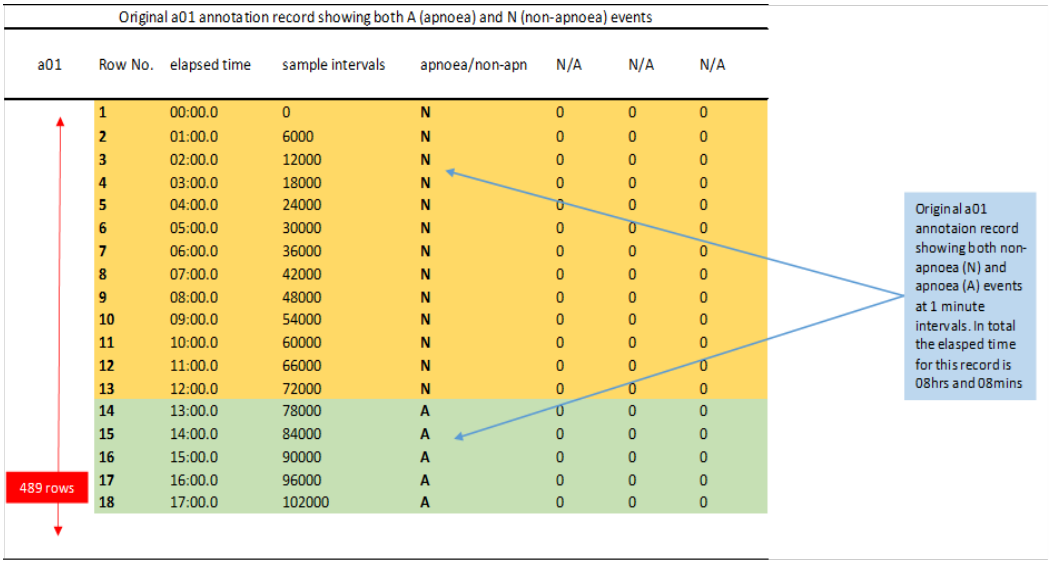

Figure 16: Annotation file a01 - showing both Non-Apnoea and Apnoea events

Using the same annotation file (a01.txt), the next stage (Figure 17) in the process was to run the code **a01N <-a01[which(a01$N=="N"),]** and **View(a01N)** to show "only" non-apnoea (N) minutes/events only (Figure 18). This reveals a total of 19 recorded minutes (3 unbroken periods) of non-apnoea over an 8hr 08min period. (01:00.0 – 12:00.0 or sample number 78000 – 72000); (3:47:00.0 – 3:50:00.0 or sample number 1362000 – 1380000) and (5:52:00.0 – 5:53:00.0 or sample number 2112000 – 2118000).

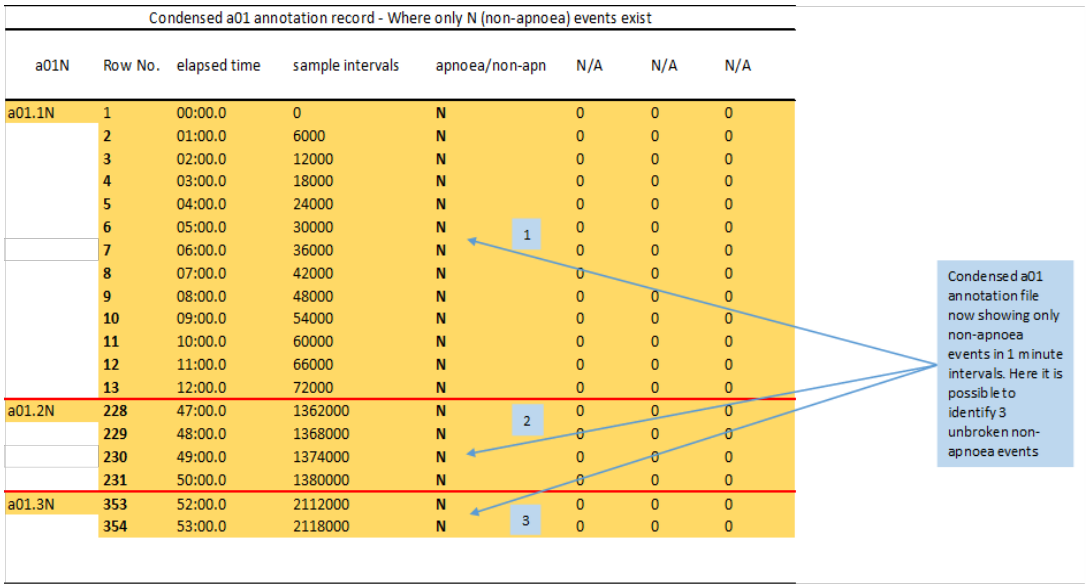

Figure 17: Annotation file a01 - showing non-Apnoea events only.

The next step was to open its corresponding ECG signal file (a01s.csv) in Figure 18. This file has 3 columns and 2957001 rows. The first column shows the row number (Row No.). The second column indicates the elapsed time in 100th/s. (hours, minutes, seconds and 100/s) and the third column shows the ECG-signals in millivolts (mV).

Original a01s ECG record containing apnoea & non-apnoea

a01s	Row No.	'hh:mm:ss.mmm'	'mV'
	1	'0:00.000'	-0.06
	2	'0:00.010'	-0.065
	3	'0:00.020'	-0.06
	4	'0:00.030'	-0.075
	5	'0:00.040'	-0.065

Figure 18: Signal file a01.txt - **showing the first 5 rows**

The next process was to match up the annotation times/events with the ECG-signal times/events, each time saving the signal data out into separate files. Using the code **a01.1N <- a01s[1:72002,]** and **write.table(a01.1N, file = "a01.1N.txt", sep = "\t")** we were able to capture and save out our first identified non-apnoea cluster as a TXT file (a01N.txt), viewable in Figure 19. This first non-apnoea cluster started at row 1 – 72002 and Zero – 12:00:000 minutes. (The RED ARROWS indicate the file was continuous with no break).

Newly created a01.1N non-apnoea signal file

a01.1N	Row No.	'hh:mm:ss.mmm'	'mV'
	1	'0:00.000'	-0.06
	2	'0:00.010'	-0.065
	3	'0:00.020'	-0.06
	4	'0:00.030'	-0.075
	5	'0:00.040'	-0.065
	6	'0:00.050'	-0.07
	7	'0:00.060'	-0.07
	71996	'11:59.940'	0.09
	71997	'11:59.950'	0.125
	71998	'11:59.960'	0.17
	71999	'11:59.970'	0.205
	72000	'11:59.980'	0.23
	72001	'11:59.990'	0.255
	72002	'12:00.000'	0.24

a01.1N. This file is the first portion of unbroken (1 - 72002) non-apnoea events to be extracted and saved out from the original a01 signal file

Figure 19: Shows the first Non-Apnoea cluster, row 1 – 72002 and 0 – 12:00:000 minutes

Using the same annotation file (a01.txt), but this time to show apnoea (A) events only (Figure 20). After running the code **a01N <-a01[which(a01$N=="N"),]** and **View(a01N)** the a01 annotation file displayed a total of 399 individual recorded minutes/events of apnoea over an 8hr 08min sleep period. Studying this closer we can see the 399 minutes of apnoea, grouped in 3 unbroken periods; (13:00.0 – 3:46.00.0 or sample number 78002 – 1356000); (3:51:00.0 – 5:51:00.0 or sample number 1386002 – 2106000) and (5:54:00.0 – 8:08:00.0 or sample number 2124002 – 2928000).

(The **RED ARROWS** indicate the file was continuous with no break).

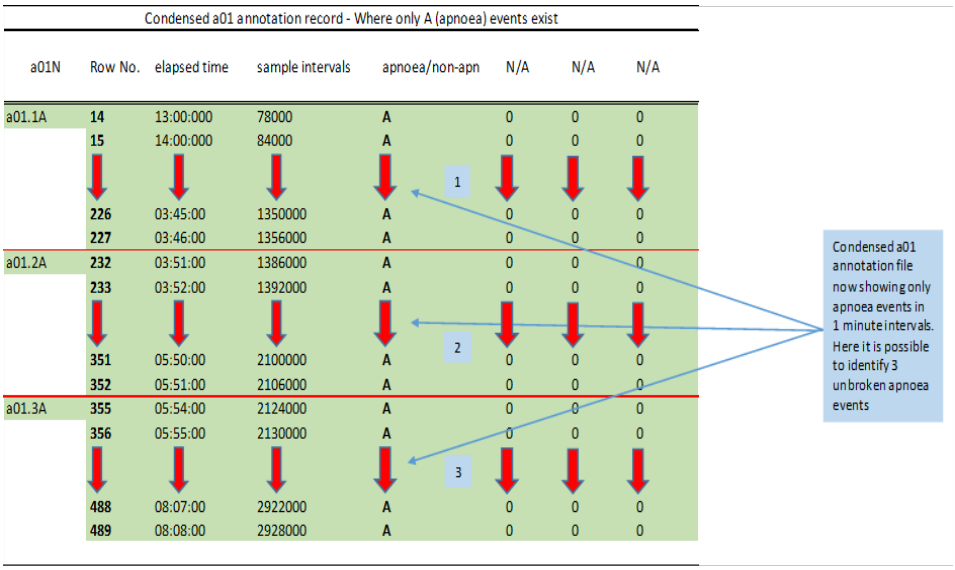

Figure 20: Shows the 3 broken periods of Apnoea

Using the corresponding ECG signal file and code **a01.1A <- a01s[1:72002,]** and **write.table(a01.1A, file = "a01.1A.txt", sep = "\t")** saved out all apnoea events as txt files. Figure 21 shows the first apnoea cluster started at row 78002 – 1356002 or 13:00.000 – 3:46:00.000. A total of 3 unbroken groups of apnoea were saved out as text files (13:00.0 – 3:46.00.0 or sample number 78002 – 1356000); (3:51:00.0 – 5:51:00.0 or sample number 1386002 – 2106000) and (5:54:00.0 – 8:08:00.0 or sample number 2124002 – 2928000).

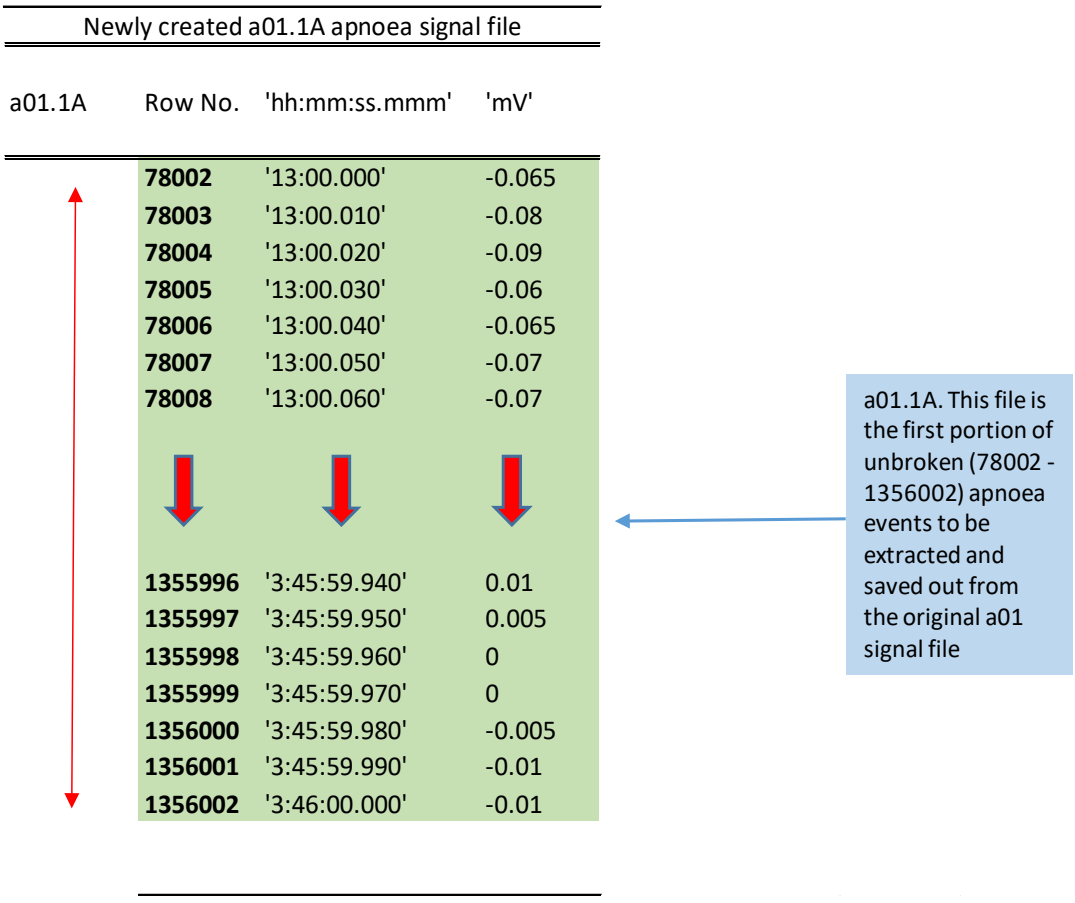

Figure 21: Shows the extracted Apnoea event, taken from the signal file

5.2.4 DATA TRANSFORMATION

The Transformation stage concentrates on the construction of the datasets. The extracted data segments are firstly introduced to 3-stage merging process. Following this, the merged data is squared, before finally moving through the windowing strategy process.

- **LIST OF ALL SUBJECT SEGMENTED SIGNAL FILES**

Table 17 presents a list of the 35 Subjects and their associated apnoea and non-apnoea recordings (ECG signal files). A total of 650 files (314 Apnoea & 336 Non-apnoea) were generated. These files were copied to guarantee enough data is available to construct the 5 datasets. At this time, the amounts of apnoea and non-apnoea were unbalanced. This overhang of non-apnoea was adjusted and balanced during the stages of squaring and merging.

35 ECG Recording	Apnoea events (segmented signal files)	Non-Apnoea events (segmented signal files)
A01	3	3
A02	11	11
A03	11	11
A04	3	3
A05	15	16
A06	10	11
A07	23	23
A08	32	33
A09	14	14
A10	18	18
A11	7	7
A12	7	7
A13	20	20
A14	8	9
A15	16	17
A16	13	14
A17	14	15
A18	6	6
A19	16	16
A20	16	17
B01	6	7
B02	13	14
B03	11	12
B04	3	4
B05	7	7
C01	0	1
C02	1	2
C03	0	1
C04	0	1
C05	2	3
C06	1	2
C07	4	5
C08	0	1
C09	2	3
C10	1	2
Total Segmented	314	336

Table 17: Shows the extracted Apnoea and non-Apnoea events

Presented in Table 18 are the Python scripts used to merge and square the selected files into the 5 unique window sizes.

FileMerge.py	• Steps through each subjects A01-C10 • Merges their Apnoea events (segmented signal files) into 1 file • Merges their Non-apnoea events (segmented signal files) into 1 file • Generates 70 files (35 Apnoea and 35 Non-apnoea)
SleepDataAggregator.py	• Strip out 4 unwanted columns (X.Elapsed.time, NA, NA, NA) • Transpose vertical running ECG signal data in the column 'X.ECG.' to run in horizontal rows. • Run these rows until a set (user input) column limit was reached.
RemoveRows.py	• Check the bottom row length against the inputted column number. • If this the row length is less than the stated number, the row will be deleted
FileMergeBlocks.py (FileMergeWithoutElse)	With all overhang removed from the window blocks of data, the following script was instructed to merge all these newly created file blocks (175 apnoea and 175 non-apnoea) into 10 combined file blocks, 5 for each condition and block size.

Table 18: Shows the python scripts and their functionality

The 3 merging stages to constructing the 5 datasets. Each dataset is built using the "non-apnoea" signals located at the bottom half of the dataset, labelled with the number "0" (zero) and apnoea signals located at the top half at the dataset, labelled with the number "1". At this stage, to better balance the dataset the overhang of non-apnoea rows was stripped out from the bottom of the dataset, ensuring the dataset was perfectly balanced. Each of the 5 finished datasets contains approximately 75 million samples, 37.5 million apnoea and 37.5 million non-apnoea.

First stage merging

Each of the 650 separate files (apnoea and non-apnoea) associated to each subject are merged into 70 files (35 for each type) This is performed 5 times for each window size.

Window Sizes	Apnoea Merged Files	Non-Apnoea Merged Files
500	35	35
1000	35	35
1500	35	35
2000	35	35
2500	35	35

Table 19 – First stage of merging

Second stage merging

Each of the 35 clusters of files are merged into 1 single file for each section.

Window Sizes	Apnoea Merged Files	Non-Apnoea Merged Files
500	1	1
1000	1	1
1500	1	1
2000	1	1
2500	1	1

Table 20 – Second stage of merging

Third stage merging

Each apnoea file and non-apnoea file associated to their window size are merged forming 1 file for each window. The 5 newly formed datasets were named W=500 through to W=2500

Window Sizes	Fully merged datasets	Dimensions columns x rows	Dataset Name
500	1	500 x 15038	W=500
1000	1	1000 x 75190	W=1000
1500	1	1500 x 50126	W=1500
2000	1	2000 x 37592	W=2000
2500	1	2500 x 30060	W=2500

Table 21 – Third stage of merging

- **WINDOW STRATEGY**

The goal of this process is to construct 5 unique datasets of specific window sizes (500, 1000, 1500, 2000, 2500), with observations of 5s, 10s, 15s, 20s, and 25s. Developing datasets in this manner is beneficial for the following reasons.

- Allows the reducing of the signal to capture more OSA events.
- Enables better training of more observations using smaller time series windows to achieve improved classification results.
- Addresses the dataset class imbalance using real data, thus avoiding the use of synthetic data.

Medical datasets can be particularly messy and suffer with severe class imbalance. This imbalance can cause real performance issues when training a model [296]. Generally, the fix to an imbalance dataset can be corrected using one of three options. Collect more data from subjects, but this can be problematic, lengthy and including red tape regulations [297]. The second option is to remove data from the majority, however, if the dataset is already small, reducing the number of observations further will hinder the models training and performance [298]. The final option is to use oversampling methods by adding synthetic data, however, this method comes with many drawbacks, that can also cause problems when training the model [296]. The approach of our Window Strategy avoids all these options. By taking each observation and transposes it into multiple observations, this reduces the time-series length, making them smaller, but increasing the overall observations and allows better balancing of the classes.

5.3 CLASSIFICATION MODEL COMPARISON STUDY

This section (A, B & C) presents the three classification ML models (1DCNN. RFC, SVM) used in this comparison study. It first investigates how they were constructed, the major parts to their architecture, their parameter configurations and functionality.

A. One Dimensional Convolutional Neural Network Model

- **Architecture**

The 1DCNN architecture in this study (Figure 22) is constructed with several layers which include a convolutional feature layer, max pooling layer, activation layer and a fully connected Multilayer Perceptron (MLP), comprising of a hidden layers and a SoftMax output for classification..

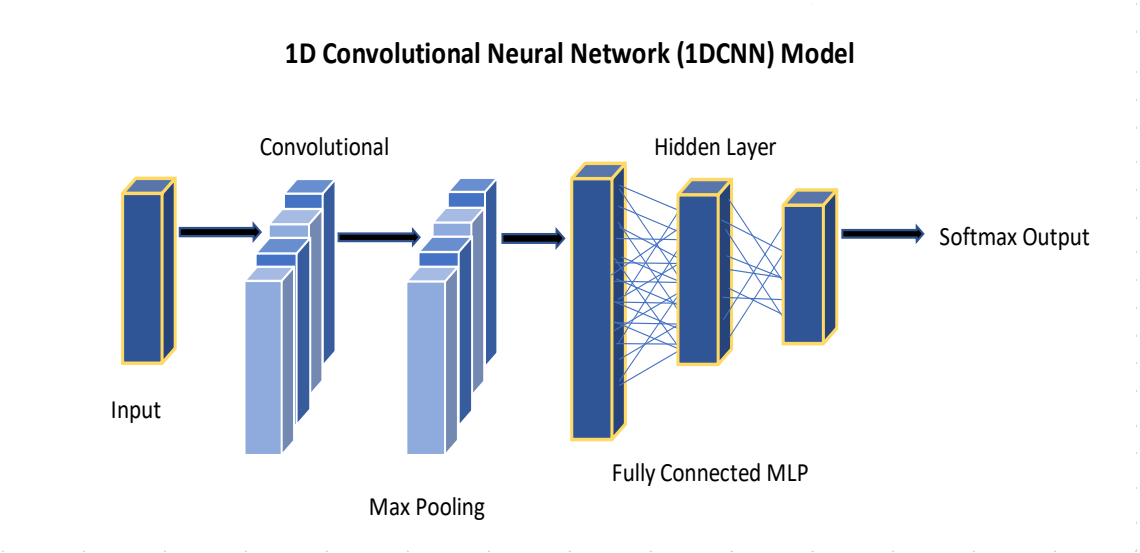

Figure 22: Typical Architecture of a One-Dimensional Convolutional Neural Network.

The model architecture in this PhD study was developed using the Python programming language combined with the high-level APIs Keras and the open-source platform TensorFlow as its backend. The model is constructed using several key layers and important functions, including an Input layer, a Convolutional layer, a Max Pooling layer and a Fully Connected Multilayer Perceptron (MLP) consisting of hidden layers, a Softmax output layer, a ReLU activation function and an ADAM activation function with Back Propagation. At the input layer a set of neurons, dictated by the batch size, feeds in and passes through the pre-processed time-series single-lead ECG-signal data. The convolutional layer, pre-set by the kernel size (matrix) hyperparameter, then slides across the input data extracting the most prominent features. These features are then built into a feature map and captured by the overriding filter hyperparameters. To further assist the convolutional process at this stage, a Max Pooling layer is incorporated. This layer skilfully summarises any

captured features, thus reducing overfitting and computation, whilst increasing overall performance of the model. The penultimate section of the 1DCNN is the Fully Connected Multilayer Perceptron (MLP). Here the newly formed output is firstly received by an input layer before being propagated forwarded through hidden layers. The hidden layers contain the activation function (ReLU), which transforms the input before passing through to the final layer, SoftMax output. The purpose of the SoftMax function is to improve classification by using probability sums. The final process within the MLP is controlled by the method 'back propagation of error', using the ADAM optimiser algorithm, this method performs calculation iterations of the layers, continuously training the network by updating the neuron weights, thus minimising the errors, and making the difference between the predicted output and actual output.

- **1DCNN MATHEMATICAL FUNCTIONS**

The following describes the mathematical function to many of the essential components within the 1DCNN and a representation of how their own specific equations can be defined.

The IDCNN layers of the convolutional function is represented in (1) by ***y*** = *conv*1*d*(***x,w,b***). The input to this function is denoted by ***x***, filters are shown as ***w***, bias as ***b*** and the final output of the convolutional layers is presented as ***y*** [299].

$$yi'k' = \sum_{ik} w_{ikk'} x_{i+i',k} + b_{k'} \qquad (1)$$

The Max Pooling layer function is a primary process within the CNN. In (2) the *MaxPooling* formula uses stride values ***sx, sy*** and a pooling window, defined by filter ***fx, fy*** and channel sizes ***k***. It operates by moving across the data capturing the highest valued features through input (***X***), where the values are summed and outputted ***i, j***. By reducing overfitting and computation, this increases the overall performance of the model and can be defined as below [300][301].

$$\text{Max Pooling}\,(X)_{i,j,k} = \max_{m,n} X_{i\cdot s_x+m,\,j\cdot s_y+n,k} \qquad (2)$$

The Fully Connected Multilayer Perception is applied to continuously recalculate and adjust the weight parameters through each layer and at each convolution. In (3) the operation of this function is shown. By using ***y*** = *fullyCon*(***x,w,b***), where ***x*** denotes the input, ***w*** = weights, ***b*** = bias and ***y*** = outputs [299].

$$yi' = \sum_i w_{ii'} x_i + b_{i'} \qquad (3)$$

The ReLU (Rectifier) activation function in (4) shows a mathematical representation of how it can be defined. This function is integral to the training and performance of the 1DCNN. Its role is to transform the weighted inputs *x* from each node and pass the outputted results *ReLU* (***x***) to the final layer [302].

$$\text{ReLU(x)} = \max\,(0, \text{x}) \qquad (4)$$

The softmax fuction denoted by ***S*** in (5) is the final activation function of the 1DCNN. Its purpose is to improve classification output for the number of classes ***n***. This is achieved by taking an input of vector numbers ***yi***, applying an exponential function to convert these real numbers into probability sums, using normalisation to ensure each value is between 0 and 1[299].

$$S(y)i = \frac{\exp(yi)}{\sum_{j=i}^{n} \exp(yi)} \qquad (5)$$

B. RANDOM FOREST CLASSIFIER ARCHITECTURE AND EVALUATION

The RFC built to train the datasets was constructed using the hyperparameters of *n_estimators*, *Random_State* and *Gini Index*. The *n_estimators* determine the number of decision trees to be used within a forest to predict an outcome. For the RFC, this is set to ***500***. The *Random_state* controls the randomness of the data for training and testing. Setting this hyperparameter to ***42***, ensures stability in the results. The *Gini Index* (7) is a measure of impurity of the sample sets ***S***. This is the probability ***Pi*** of the incorrectly labelling of a randomly selected class ***k*** [303]. The Gini Index improves classification by decreasing the numerical value of feature importance at each node within a decision tree. Further to this, the Gini Index assists to provide quicker computations [270].

$$\{h(x, \theta k), k = 1, \dots\} \tag{6}$$

$$Gini(S) = 1 - \sum_{i=1}^{k} p_i^2 \tag{7}$$

C. SUPPORT VECTOR MACHINE ARCHITECTURE AND EVALUATION

The SVM built to compare against the 1DCNN and RFC was constructed to train the datasets using the hyperparameters *Random_State* and the *RBF_Kernel* (Radial Basis Function). The *RBF_kernel* assists to make better classification decisions when training on non-linear data. Based on the Gaussian Distribution kernel, which calculates the similarity or closeness of two fixed points. In (8) the fixed points $\{\boldsymbol{X}\_\mathbf{1}, \boldsymbol{X}\_\mathbf{2}\}$, are calculated using the decision boundary parameter ***Y***, the *RBF_kernel* ***K (., .)*** maps the input data into a high-dimensional space, thus enabling the SVM to find the best position of the hyperplane for classification [304].

$$\mathcal{K}(\mathcal{X}_i, \mathcal{X}_j) = e^{-\gamma \|\mathcal{X}_i - \mathcal{X}_i\|^2} \tag{8}$$

5.3.1 PERFORMANCE METRICS

Performance metrics are critical gauges to evaluating how well the any ML algorithm model is working. This section briefly describes all of the metrics used in the evaluation of the three ML models (1DCNN, RFC & SVM)

A. CONFUSION MATRIX

Confusion Matrix is a visualisation tool used to measure the performance of a classification model. Represented as a table of predicted and true classes, it better summarises the performance and facilitates the calculation of other metrics, that includes, Recall, Precision, Accuracy, F1 score and AUC-ROC curve.

		Actual Class	
		Positive	Negative
Predicted Class	Positive	TPs	FPs
	Negative	FNs	TNs

(9)

True Positives (TP) when the actual value is Positive and predicted is also Positive, True Negatives (TN), when the actual value is Negative, and prediction is also Negative, False Positives (FP), when the actual is negative, but prediction is Positive and False Negatives (FN), when the actual is Positive but the prediction is Negative.

B. VALIDATION LOSS AND VALIDATION ACCURACY METRICS

Validation loss and Validation Accuracy metrics function in a similar way to the loss and accuracy metric by evaluating the quality and performance of the model. However, the validation loss metric is measured after each iteration of epoch. Furthermore, the validation loss metric does not signal the model to updates the weights at each passing.

C. SENSITIVITY AND SPECIFICITY METRICS

The function of the Sensitivity and Specificity metrics is to demonstrate the accuracy of a classification test. This is calculated by the presence or absence of an instant. Sensitivity measures the true-positive rate, what the model has correctly predicted, and Specificity measures the true-negative rate, again what the model has correctly predicted [305].

$$\text{Sensitivity} = \frac{True\ Positives}{True\ Positives\ + False\ Negatives} \tag{10}$$

$$\text{Specificity} = \frac{True\ Negaties}{True\ Negatives\ + False\ Positives} \tag{11}$$

D. PRECISION AND RECALL

Precision and Recall are used to measure the model's performance when predicting binary classification. Precision measures how many correct predictions the model has correctly predicted out of all the predictions made. Recall works to measure all the positives are correctly identified out of all the predictions.

$$\text{Precision} = \frac{True\ Positives}{True\ Positives\ + False\ Positives} \tag{12}$$

$$\text{Recall} = \frac{True\ Positives}{True\ Positives\ + False\ Negatives} \tag{13}$$

E. F1_SCORE

F1_Score is a measurement of the model's accuracy. This measurement is performed by calculating by the means of both the precision and recall (classification values)

$$\text{F1} = 2 * \frac{precision * recall}{precision + recall} \tag{14}$$

F. ROC AUC (Receiver Operator Characteristic, Area Under the Curve)

The Area under the ROC Curve (AUC) is a visual representation of a models performance and accuracy. ROC measures the probability of the model by plotting sensitivity (True positive rate) against specificity (False positive rate) and the AUC measures the ability of a model to distinguish between the two classes. This measurement is achieved by using a ranking system, which scores the separate classes on a scale of 0 to 1. The higher the AUC, the better the model is at prediction and class separability.

$$\text{True Positive Rate (TPR)} = \frac{TP}{TP + FN} \tag{15}$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP + TN}$$

G. Loss & Accuracy Metrics

These two metrics are calculated very differently, however they both indicate how well the model is learning through the progression of training. On each batch iteration of the training set, the loss metric calculates the sum of error/bad predictions and then presents how good/bad the model is performing. Through calculation of these sums the model will continually attempt to improve its performance by altering the neuron weights (cost function) at each passing. The lower the loss, the better the model. The function of the Accuracy metric is to evaluate the model's performance in an interpretable way. It calculates and presents the number of correctly classified predictions against the actual number of true predictions, it can be defined as follows [306]

(16)

$$\text{Accuracy} = \frac{True\ Positives + True\ Negatives}{True\ Positives + True\ Negatives + \ False\ Positives\ + Flase\ Negatives}$$

H. KAPPA_SCORE

Another accuracy indicator is Kappa score. This metric demonstrates the level of agreement between two raters on a classification problem. The closer the score is to 1, the better the agreement between the raters and the better the model is at classification. The sum of Kappa score is achieved by calculating Po (accuracy), Pe (expected accuracy) and 1 – Pe (Value range). Po, being the amount of observed agreement in relation to the total number, Pe, being the amount of observed probability of chance agreement and 1 – Pe, being the kappa value range, -1 no agreement to +1 complete agreement [307].

$$\mathrm{k} = \frac{\mathrm{Po} - \mathrm{P\,e}}{1 - \mathrm{P\,e})} \tag{17}$$

I. LOG_LOSS

A further accuracy indicator is Log Loss or Binary Cross Entropy Loss. Based on probabilities, it measures the accuracy of a classification model, where the output is a value between 0 and 1. It achieves this by comparing the prediction probability result to the actual result. The closer these two sums are, the smaller the log loss becomes and the more accurate the model is at classification. In this formula p is the probability of class 1, and (1 – $\hat{p}$) is the probability of class 0 [308].

$$CE\ (p, \hat{p}) = -\ (p * log\ (\hat{p}) + (1 - p) log\ (1 - \hat{p}) \tag{18}$$

J. MACROAVERAGE

Macro averaging reduces the multiclass predictions down to multiple sets of binary predictions. It then calculates the corresponding metric for each of the binary cases before averaging the results together.

K. WEIGHTED AVERAGE

Weighted average is a calculation that takes into account the varying degrees of importance of the numbers in a dataset. In calculating a weighted average, each number in the data set is multiplied by a predetermined weight before the final calculation is made.

5.3.2 HYPERPARAMETER EXPERIMENTS FOR FINE TUNING THE 1DCNN MODEL

To thoroughly investigate the efficacy of the proposed model, a series of rigorous experiments were conducted that systematically explored the available method space. This involved tuning a multitude of parameters for machine optimisation. Key parameters under consideration included Learning rate, Window sizes, Training Sample size, Validation Sample size, Filter size, Batch size, Kernel size, length of Epochs and depth of network. Over a thousand experiments were conducted across the 5 1DCNN models (1DCNN-500 through to 1DCNN-2500) and windows dataset sizes W=500 through to W=2500). However, for the purpose of this thesis, only the most significant experiments, using the 1DCNN-500 window dataset, were recorded for this section.

The earliest experiments started with low value inputs across all available hyperparameters. Table 22 shows a selection (No.1 – No.10) of these experiments that were performed on the dataset 1DCNN-500. Viewable in this table is the “n_filters”, “k_size”, “batch_size”, “Acc”, “Loss”, “Val_Acc”, “Val_Loss”. The training and validation samples were fixed at 108,276 and 12,048 respectively, for all experiments performed using the W=500 dataset. Due to the inadequate scoring in the experimental output, the calculation of some key performance metrics, including Sensitivity (Recall), Specificity, F1_score, Kappa_score, log_loss and ROCAUC, were not sufficient to record in the experimental spreadsheet and therefore not available to view in this table. Furthermore, all experiments shown here were run using a single layer 1DCNN. Other network depths were tried across all models, without any improvement.

Given the modest size of the datasets, an initial learning rate of 0.001 was chosen to balance convergence and speed. Fine tuning began with gradual increases in batch size

(2^n) and epoch lengths of 25, 50 and 100, with each experiment yielding slight improvement in both accuracy and loss.

Inputted Hyperparameters						Outputted Results			
Number	Learning Rate	n_filters	k_size	batch_size	epochs	Acc	Loss	Val_Acc	Val_Loss
1	0.001	32	32	32	25	0.4021	0.7102	0.4018	0.7106
2	0.001	32	32	32	50	0.4033	0.6941	0.4031	0.6948
3	0.001	32	32	32	100	0.4041	0.6891	0.4037	0.6898
4	0.001	32	32	64	25	0.5010	0.6502	0.5005	0.6507
5	0.001	32	32	64	50	0.4024	0.7122	0.4021	0.7123
6	0.001	32	32	64	100	0.5002	0.6511	0.5001	0.6515
7	0.001	32	32	128	25	0.5010	0.6502	0.5007	0.6505
8	0.001	32	32	128	50	0.5011	0.6491	0.4999	0.6533
9	0.001	32	32	128	100	0.5010	0.6502	0.5001	0.6571
10	0.001	32	32	256	25	0.4024	0.7122	0.4011	0.7191

Table 22 – 1DCNN-500 Experiments No.1 – 10

Subsequent experiments revealed that further increases in batch size (beyond 256 or 512) did not lead to any substantial improvements. Similarly, adjustments to the learning rate, including increases to 0.01 and 0.1, resulted in erratic behaviour of the model's performance, affecting the loss, causing oscillations, and rapidly fluctuating without any convergence. Therefore, from experiment 35 (Table 23) onwards, the focus shifted to optimising the filter and kernel configurations. Initially these two hyperparameters had been set to 32, then proceeding with (2^n) increases throughout the experiments. To balance computational efficiency, interpretability and understanding of patterns and overfitting, the filter and kernel sizes were set to identical, low, odd-numbered values. At first, implementing his new approach did not lead to any performance gains. However, starting with experiment 48, a configuration of 50 filters and kernels, began to produce some significant results. However, despite the notable advancements in accuracy and loss reduction, the model's performance in terms of sensitivity and specificity, key indicators of balanced classification, were still falling someway short.

Inputted Hyperparameters						Outputted Results			
No.	Learning Rate	n_filters	k_size	batch_size	epochs	Acc	Loss	Val_Acc	Val_Loss
35	0.001	15	15	32	25	0.5033	0.6915	0.5002	0.6943
36	0.001	15	15	32	50	0.5390	0.6891	0.5000	0.6935

37	0.001	15	15	32	100	0.5412	0.6885	0.5010	0.6901
38	0.001	25	25	32	25	0.5049	0.6899	0.5001	0.6943
39	0.001	25	25	32	50	0.5028	0.6901	0.4999	0.6934
40	0.001	25	25	32	100	0.5043	0.6887	0.4998	0.6935
41	0.001	35	35	32	25	0.5001	0.6931	0.5000	0.6934
42	0.001	35	35	32	50	0.5000	0.6934	0.5000	0.6934
43	0.001	35	35	32	100	0.5008	0.6931	0.5003	0.6930
45	0.001	45	45	32	25	0.5019	0.6923	0.4981	0.6935
46	0.001	45	45	32	50	0.5033	0.6915	0.5002	0.6943
47	0.001	45	45	32	100	0.5390	0.6891	0.5000	0.6935
48	0.001	55	55	32	25	0.9283	0.1777	0.8782	0.3421
49	0.001	55	55	32	50	0.9472	0.1311	0.8962	0.3271
50	0.001	55	55	32	100	0.9441	0.1378	0.8952	0.3212

Table 23 – 1DCNN-500 Experiments No.35 – 50

Continuing to build on this method and results, experiment 124 (Table 24), marked more significant improvements in consistency and a more balanced sensitivity and specificity. Key changes to the hyperparameters included substantial increases to the filter, kernel, and batch sizes. Additionally, the filter and kernel sizes were no longer using identical values, with a shift towards alternating the kernel sizes between odd and even numbers. Finally, the parameter of 25 epochs was abandoned, to allow better convergence.

Inputted Hyperparameters						Outputted Results					
No.	Learning Rate	n_filters	k_size	batch_ size	epochs	Acc	Loss	Val_Acc	Val_Loss	Sen	Spec
124	0.001	100	200	256	50	0.9602	0.0987	0.9554	0.1116	95.8	94.5
125	0.001	100	200	256	100	0.9565	0.1079	0.9509	0.1235	95.8	95.5
126	0.001	100	225	256	50	0.9531	0.1172	0.9517	0.1191	95.3	95.4
127	0.001	100	225	256	100	0.9557	0.1077	0.9520	0.1200	95.7	96.1
128	0.001	100	230	256	50	0.9560	0.1081	0.9506	0.1240	93.8	97.7
129	0.001	100	230	256	100	0.9531	0.1172	0.9517	0.1191	95.9	95.6
130	0.001	150	235	256	50	0.9552	0.1081	0.9515	0.1202	95.1	95.6
131	0.001	150	235	256	100	0.9595	0.0984	0.9557	0.1155	95.9	96.3
132	0.001	150	240	256	50	0.9558	0.1066	0.9515	0.1216	96.7	95.5
133	0.001	150	240	256	100	0.9561	0.1060	0.9534	0.1142	95.3	95.5
134	0.001	150	250	256	50	0.9516	0.1149	0.9491	0.1239	95.0	95.7
135	0.001	150	250	256	100	0.9518	0.1142	0.9495	0.1234	93.6	97.8
136	0.001	150	260	256	50	0.9562	0.1055	0.9494	0.1273	94.1	97.2
137	0.001	100	260	256	100	0.9565	0.1068	0.9518	0.1230	93.8	96.9
138	0.001	100	270	256	50	0.9559	0.1063	0.9518	0.1214	96.8	95.7
139	0.001	100	270	256	100	0.9562	0.1058	0.9539	0.1140	95.4	95.8

Table 24 – 1DCNN-500 Experiments No.124 – 139

Subsequent experiments further built upon these promising results. Table 25 showcasing experiments No.162 onwards, demonstrates continued improvements across all outputs. A substantial increase in batch size led to a more balanced sensitivity and specificity. This trend was further optimised in experiments 236 onwards, where reverting back to identical filter and kernel sizes yielded the best results. Specifically applying a configuration of 150 filters, 150 kernels, a batch size of 8192, run over 50 epochs was seen to produce the best performance.

Inputted Hyperparameters						Outputted Results					
No.	Learning Rate	n_filters	k_size	batch_ size	epochs	Acc	Loss	Val_Acc	Val_Loss	Sen	Spec
162	0.001	150	400	8192	50	0.9334	0.1750	0.9352	0.1786	96.0	97.0
163	0.001	150	400	8192	100	0.8994	0.2464	0.9011	0.2469	96.0	95.1
164	0.001	150	425	8192	50	0.9496	0.1353	0.9390	0.1629	93.0	97.0
165	0.001	150	425	256	100	0.9670	0.0823	0.9518	0.1370	95.5	95.6
166	0.001	200	450	256	50	0.9691	0.0762	0.9515	0.1474	94.6	95.5
167	0.001	200	450	8192	100	0.9390	0.1573	0.9281	0.1846	95.2	97.2
168	0.001	250	475	8192	50	0.9340	0.1753	0.9181	0.2172	95.2	94.6
169	0.001	250	480	2048	100	0.9531	0.1237	0.9377	0.1764	95.2	94.6
170	0.001	300	480	2048	50	0.9353	0.1660	0.9212	0.2040	94.5	95.5
171	0.001	300	475	8192	100	0.9071	0.2382	0.9030	0.2532	94.0	95.0
172	0.001	350	475	8192	50	0.8521	0.3244	0.8470	0.3365	95.2	95.1
173	0.001	350	475	8192	100	0.9601	0.1094	0.9344	0.1890	94.8	95.1
⇩	⇩	⇩	⇩	⇩	⇩	⇩	⇩	⇩	⇩	⇩	⇩
236	0.001	100	150	8192	50	0.9635	0.1050	0.9570	0.1214	97.0	97.4
237	0.001	100	150	8192	50	0.9480	0.1440	0.9482	0.1497	96.1	97.4
238	0.001	100	150	8192	50	0.9421	0.1541	0.9433	0.1568	96.6	96.5
239	0.001	100	150	8192	50	0.9377	0.1641	0.9403	0.1640	97.0	97.2
240	0.001	150	150	8192	50	0.9699	0.0814	0.9662	0.0942	97.4	97.0
241	0.001	150	150	8192	50	0.9699	0.0814	0.9662	0.0942	97.4	97.0
242	0.001	150	150	8192	50	0.9699	0.0814	0.9662	0.0942	97.4	97.0
243	0.001	150	150	8192	50	0.9699	0.0814	0.9662	0.0942	97.4	97.0

Table 25 – 1DCNN-500 Experiments No.162 – 173 and 236 – 243

This methodology was applied to the 4 remaining models and datasets (1DCNN-1000, 1DCNN-1500, 1DCNN-2000, and 1DCNN-2500), with excellent results.

6. TRAINING AND TEST RESULTS

This chapter evaluates the effectiveness of all models by presenting their results through training and testing. The chapter is split into areas (6.1, 6.2 & 6.3). In subsection 6.1, the training and validation phase, firstly results (Tables 26 - 28) for the best performing model from each group (1DCNN, RFC, SVM), this also includes (Tables 29 - 31), which shows the results for all 15 models. Subsection 6.2 presents the testing phase and results for the best performing 1DCNN models (Table 32). Following this in 6.3 is a set of results (Table 33) for traditional ML models when trained with feature engineering methods. Finally, presented in subsection 6.4 (Table 34) is a results comparison of the proposed model against other OSA studies, discussed earlier in the Chapter 4 Related Work section.

6.1 TRAINING AND VALIDATION OF THE MODELS

The training and validation phase involves the optimization of each model's hyperparameters to keep minimising the loss function when running over the selected datasets. This process aims to find the best configuration that maps the input data to the desired output labels. A total of 15 models, 5 for each group (1DCNNs, RFCs, SVMs) were part of this process. Each model was run numerous times through training and validation. The first models to be assessed was the 1DCNNs. This was conducted by running separate experiments using the 5 pre-built balanced datasets (W=500 through to W=2500). For each of these training experiments the data was split into 80% training and 20% validation. These sizes are calculated based on the amount of data contained within each dataset. The same experiments, using the same datasets, were again performed on the RFCs and SVMs models. Performance of each experiment was measured using a variety of common metrics, presented earlier in Chapter 5, Performance Metrics.

Each experiment was run and executed on the same computer and specifications: Intel i7 processor, Nvidia GTX 1080 and 16GB Ram. The main objective of these experiments is to find the

model that frequently produces the best performances, using the least computational power and in the quickest times.

- **BEST PERFORMING MODEL FOR EACH TRAINING GROUP**

Tables 26 - 28 and Figures 23 - 27, show the best performing model of each group (1DCNN, RFC, SVM) their optimum configuration (inputs) and results (outputs), and where available, a confusion matrix measurement is presented. It firstly presents the 1DCNN model, followed by the RFC model and finally the SVM model.

1. One-Dimensional Convolutional Neural Network Configuration and Results

This section assesses the best performing 1DCNN model (1DCNN-500) after training and validation. It shows the model's hyperparameter configuration (Table 26), along with graphical representations of *accuracy*, *loss* and *ROCAUC* results (Figures 23 – 25).

Configuration		Results	
Window Size	500	Accuracy	0.9699%
Train on Samples	108276	Loss	0.0814%
Validate on Samples	12031	Validation Accy	0.9662%
n_Filters	150	Validation Loss	0.0942%
k_Size	150	Sensitivity (Recall)	0.9743
Batch_Size	8192	Specificity	0.9708
Epochs	50	F1_Score	0.9726
---		Kappa_Score	0.9451
---		Log_Loss	0.0759
---		ROCAUC	0.9966

Table 26: 1DCNN-500 Configuration and results

Table 26 presents both the inputs and outputs for the 1DCNN-500 model when running the W=500 dataset. Applying inputs of 150 Filters (*n_Filters*), with a Kernel size (*k_Size*) of 150, a peak threshold *Batch_size* of 8192, when run over 50 epochs, was empirically found to return the best results. These results are listed in the 'Results' column, which shows exceptional high scores across all metrics. Especially when analysing the results of *Accuracy, Sensitivity and Specificity*. The configuration and results show this to be the best performing 1DCNN model.

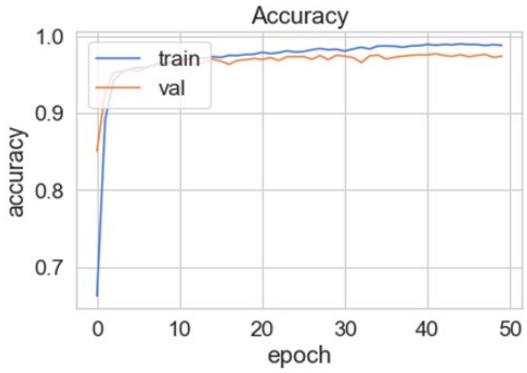

Figure 23. Graphical output results from the 1DCNN-500 model using dataset W=500. Showing Training and Validation Accuracy

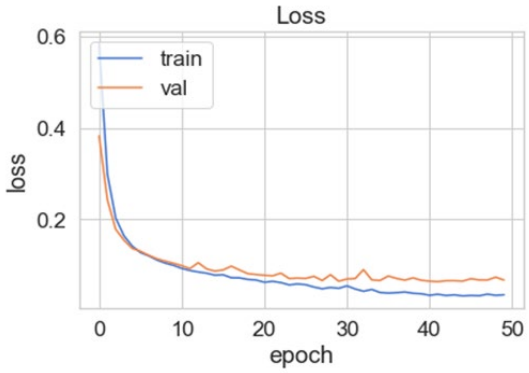

Figure 24. Graphical output results from the 1DCNN-500 model using dataset W=500. Showing Training and Validation Loss

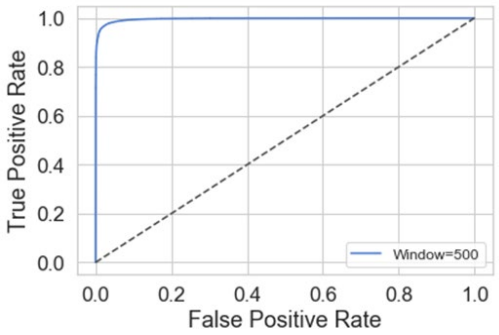

Figure 25. Graphical output results from the 1DCNN-500 model using dataset W=500. Showing ROCAUC plot.

Figures 23 – 25, shows three images for the 1DCNN-500 model results in Table 26 (training and validation accuracy). It is clear to see almost instant merging and high accuracy scoring at 10 epochs. Additionally, it shows the model is still producing a steady increase in accuracy through the 50-epoch marker with no signs of over-fitting. The bottom graph presents the true-positive results of the model in a ROCAUC plot. The tightness of the curve to the top-left hand corner, along with the very high AUC scoring at almost 1.0, demonstrates how well this model is at predicting between the two separate classes.

2. Random Forest Classifier Configuration and Results

The same 5 experiments were then performed and assessed on the best performing RFC model (RFC-500)

Configuration		Results	
Window Size	500	Accuracy	0.91
Random State	42	Precision (0)	0.94
Forest Size	500	Precision (1)	0.90
Support	30077	Recall (0)	0.90
		Recall (1)	0.94
		F1_Score (0)	0.92
		F1_Score (1)	0.92
		Support (0)	15088
Confusion Matrix		Support (1)	14989
	Predicted Label	Macro Avg.	92,92,92
True Label	(TN) 13517 1571 (FP)	Weighted Avg.	92,92,92
		Support	30077
	(FN) 901 14088 (TP)		

Table 27: RFC-500 Configuration & Results and Confusion Matrix

Table 27, Configuration column presents the optimal configuration for the RFC-500 model. This model was the best performing of the RFC models. Using a *n_estimator* size (amount of decision trees) of 500 and a *random_state* of 42 was found to return the best results. When examining the Confusion Matrix values, which represent the number of correct classification data predictions over the total amount of classification predictions, as well as calculated scores for Precision, Recall (Sensitivity) and Accuracy. This gives a good measure of the performance of the classification model's performance by providing a measure of misclassified instances. Misclassifications are typically the result of noise in the dataset. Results are pretty good; FP is about 3% and FN about 5%. Overall, this model has performed to a very good standard. However, producing this level of performance incurred drawbacks, notably time consumption. The higher the value of the decision tree value (*n_estimator*), the higher the accuracy, but increasing this value meant the longer the duration of the experiment took to complete. Moreover, inputted values above 500 decision trees didn't show any further improvements.

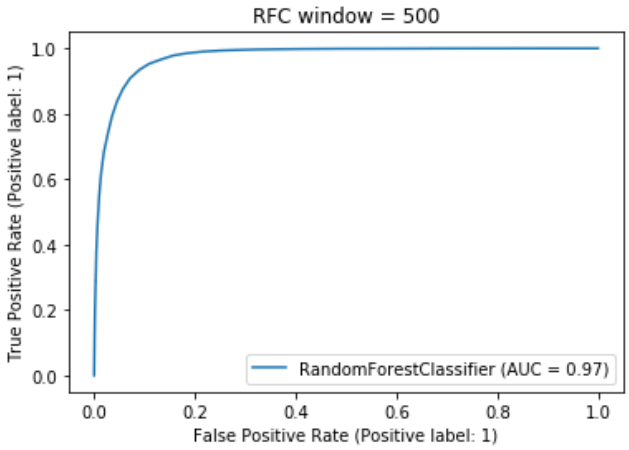

Figure 26. Graphical output results from the RFC-500 model using dataset W=500. Showing AUC plot.

Figure 26, presents the AUC graph for the RFC-500 model results in Table 27. Although not as tight to the top-left hand corner as Figure 25, 1DCNN AUC, this is still a very good scoring and demonstrates this model is good at predicting between the two separate classes.

3. Support Vector Machine Configuration and Results

The same 5 experiments were then performed and assessed on the best performing SVM model.

Configuration		Results	
Window Size	500	Accuracy	0.72
Random State	42	Precision (0)	0.65
Kernel	RBF	Precision (1)	0.90
Support Size	30077	Recall (0)	0.94
		Recall (1)	0.50
		F1_Score (0)	0.77
		F1_Score (1)	0.64
		Support (0)	15013
Confusion Matrix		Support (1)	15064
	Predicted Label	Macro Avg.	77,72,71
True Label	(TN) 14142 871 (FP)	Weighted Avg.	77,72,71
		Support	30077
	(FN) 7555 7509(TP)		

Table 28: SVM-500 Configuration & Results and Confusion Matrix

In Table 28, the Configuration column presents the optimal configuration for model SVM-500. For this model, using the Radial Basis Function (RBF) Kernel and a Random State of 42 was empirically found to return the best results using this dataset. The overall performance of this model (Results

column) is moderate. Actual classification within the confusion matrix is unbalanced, particularly when examining the large number of False Negatives that have been produced.

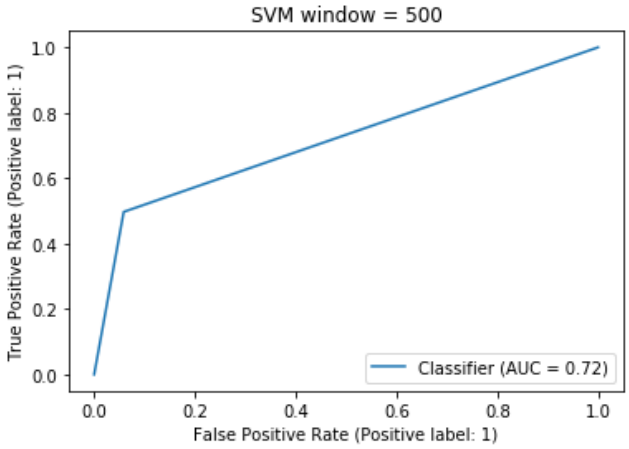

Figure 27. Graphical output results from the SVM-500 model using dataset W=500. Showing AUC plot.

Figure 27 presents the AUC graph for the RFC-500 model results in Table 28. When compared to 1DCNN and RFC, this model performed quite poor, in both classification of the two separate classes and duration of time to complete the task.

- **COMPLETE LIST OF TRAINING RESULTS**

The three tables (Table 29 - 31) below present the training and validation results for the execution of the 15 models averaged over 50 runs. They show the optimal architecture for each model for the automatic detection of OSA using single-lead ECG signals. The top table (Table 29) shows the 5 1DCNN configuration hyperparameters and results, the middle table (Table 30), shows the 5 RFC configurations and results and the bottom table (Table 31), shows the 5 SVM configuration hyperparameters and results.

When evaluating the whole set of results across the three tables, it is clear to that the 1DCNN models outperforms both the RFC and SVM models. Particularly notable, is the high performance

of Sensitivity & Specificity across the 1DCNN models, where results range from very good to excellent with well-balanced classification.

Further comparison analysis shows the 1DCNN models produce results in significantly quicker time and using less computational power. Experiment times alter significantly between 1DCNN models (3 to 4 minutes) and RFC models, (1+hrs – up to 1.40hrs) and even more so when analysing the SVM models (10+hrs – up to 17hrs). Moreover, looking at the 1DCNN results, from the bottom (No.5) to the top (No.1), it is possible to see performance slightly increases each time. This pattern coincides with the novel dataset windowing strategy. Windows with more rows and fewer columns, shows a gradual increase in performance results. This same pattern is also evident in both the RFC and SVM experiments. Additional window dimension testing showed the minimum threshold was at around 500 columns, after this point results didn't improve significantly.

Training and learning of the 1DCNN models were shaped by hyperparameter influences. The importance of these hyperparameters is evident when looking at the wide variation of configurations between each model. For the best performing model, 1DCNN-500, reducing and balancing both the k_size and n_filter dimensions for convolving and output, scaling up the Batch_size for training, and minimizing epoch iterations for updating learning values, was empirically found to return the best results. However, for the 1DCNN-2500 model, which still performed very well, but with slight signs of overfitting, using a small dimensional kernel and a large filter output with a scaled-up batch size, was empirically found to return the best results.

Training and configuration hyperparameters of the RFC models were more straight-forward. The focal hyperparameter setting was the input value of the n_estimators. This value indicates the amount of decision trees to be used within the random forests when running the model. The amount of decision trees dictates both performance and duration of an experiment. At this stage, the influences of the Random_state hyperparameter controls the randomness of the data for training, testing and stability in the results. Experiments were set from 100 estimators, with short

durations, through to 2000 estimators, that took many hours. To reduce lengthy testing durations and without dropping model performance, the implementation of Gini Index was chosen over Entropy. Gini provides quicker results and less computational power. Further attempts to improve results included initiating the hyperparameter, max_depth, however, this only succeeded to increase overfitting. The sweet spot for this model was using 500 estimators, anything higher than this value only increased the testing duration, but not the results.

Of the 3 groups, the table shows SVMs was the worst performing models for both results and duration of testing, taking up to 17hrs to complete. Finding the optimum performance of the SVM included the hyperparameter RBF_Kernel, for classification and Random_State to control the randomness of the data for training and testing along with stability of the results. Furthermore, attempts to try and improve performance results for some SVM experiments were influenced by the hyperparameter Gamma, however, with no positive effects.

Configuration						Results			
No	Model	Dataset	n_Filters	k_Size	Batch_Size	Time h/m	Sensitivity (Recall)	Specificity	Accuracy
1	1DCNN-500	W=500	150	150	8192	0.04ms	0.9743	0.9708	0.9699%
2	1DCNN-1000	W=1000	250	250	4096	0.04ms	0.9612	0.9730	0.9528%
3	1DCNN-1500	W=1500	100	1000	4096	0.03ms	0.9592	0.9472	0.9161%
4	1DCNN-2000	W=2000	100	500	4069	0.03ms	0.9575	0.9702	0.9086%
5	1DCNN-2500	W=2500	100	800	4096	0.03ms	0.9414	0.9545	0.9046%

Table 29: 1DCNN Training and Validation Experiments - Configuration and Results

Configuration						Results			
No	Model	Dataset	Estimators	Random state	Classification	Time h/m	Sensitivity (Recall)	precision	Accuracy
1	RFC-500	W=500	500	42	0/1	1+hrs	0.90/0.94	0.94/0.90	0.91
2	RFC-1000	W=1000	500	42	0/1	1+hrs	0.84/0.85	0.85/0.85	0.85
3	RFC-1500	W=1500	500	42	0/1	1+hrs	0.85/0.87	0.87/0.84	0.86
4	RFC-2000	W=2000	500	42	0/1	1+hrs	0.84/0.88	0.87/0.85	0.86
5	RFC-2500	W=2500	500	42	0/1	1+hrs	0.84/0.85	0.85/0.85	0.85

Table 30: RFC Training and Validation Experiments - Configuration and Results

Configuration						Results			
No	Model	Dataset	Kernel	Random state	Classification	Time h/m	Sensitivity (Recall)	precision	Accuracy
1	SVM-500	W=500	RBF	42	0/1	10+hrs	0.94/0.50	0.65/0.90	0.72
2	SVM-1000	W=1000	RBF	42	0/1	10+hrs	0.95/0.44	0.63/0.89	0.69
3	SVM-1500	W=1500	RBF	42	0/1	10+hrs	0.92/0.44	0.62/0.84	0.68
4	SVM-2000	W=2000	RBF	42	0/1	10+hrs	0.89/0.44	0.62/0.79	0.67
5	SVM-2500	W=2500	RBF	42	0/1	10+hrs	0.88/0.44	0.61/0.79	0.66

Table 31: SVM Training and Validation Experiments - Configuration and Results

6.2 TESTING THE MODELS

Following the training of the model, the next phase was to test the model. Due to the lower training scores of the RFC and SVM model's, it was decided the testing would just concentrate on the group of 1DCNN models. This involves evaluating each model's performance using a separate dataset that was not used for training. This is done to assess the model's ability to generalise to new and unseen data and identify any potential overfitting or underfitting issues. The performance of each model was measured using the most common metrics; Sensitivity, Specificity, Accuracy, F1 Score and ROCAUC. These metrics were discussed earlier in Chapter 5, Performance Metrics. Presented in table 32 are the averaged results of the 5 models after testing. Each scores presented is taken from an average of 20 runs.

The steps involved in the testing stage:

- Prepare and load the test data: The datasets were formatted, transposed, and shaped using the same windowing strategy as the training datasets.
- Predict the labels: Load the newly trained model, with learned parameters to predict the class labels for the new datapoints.
- Analyse the results: Analyse the metrics results to identify all areas of performance, good or bad.
- Consider model improvements: If the model shows signs of underperforming. Adjust the model architecture, hyperparameters and training configurations.

Model	Sensitivity	Specificity	Accuracy	F1 Score	ROCAUC
1DCNN-500	97.42	97.15	97.28	97.26	99.63
1DCNN-1000	96.26	96.21	96.24	96.23	99.33
1DCNN-1500	95.38	94.74	95.06	95.11	98.52

1DCNN-2000	95.15	97.04	96.09	96.06	99.00
1DCNN-2500	94.62	95.63	95.13	95.08	98.49

Table 32: Test results for the five 1DCNN models averaged over 20 runs

The overall position of results at the testing phase were very similar to those of the training and validation phase, in that the best performing model was the 1DCNN-500. However, the individual results showed increases of Accuracy across all the models, particularly when examining the models 1DCNN-1500, 1DCNN-2000 and 1DCNN-2500, which had rose by approximately 5 points, bringing them slightly more in-line with the two best performing models, 1DCNN-500 and 1DCNN-1000. The most important metrics of Sensitivity and Specificity have stayed high and stable across all models, showing an almost perfect balance for each model, and producing excellent classification. Further high performance and excellent measures of accuracy are presented in the results for F1_Score and ROCAUC.

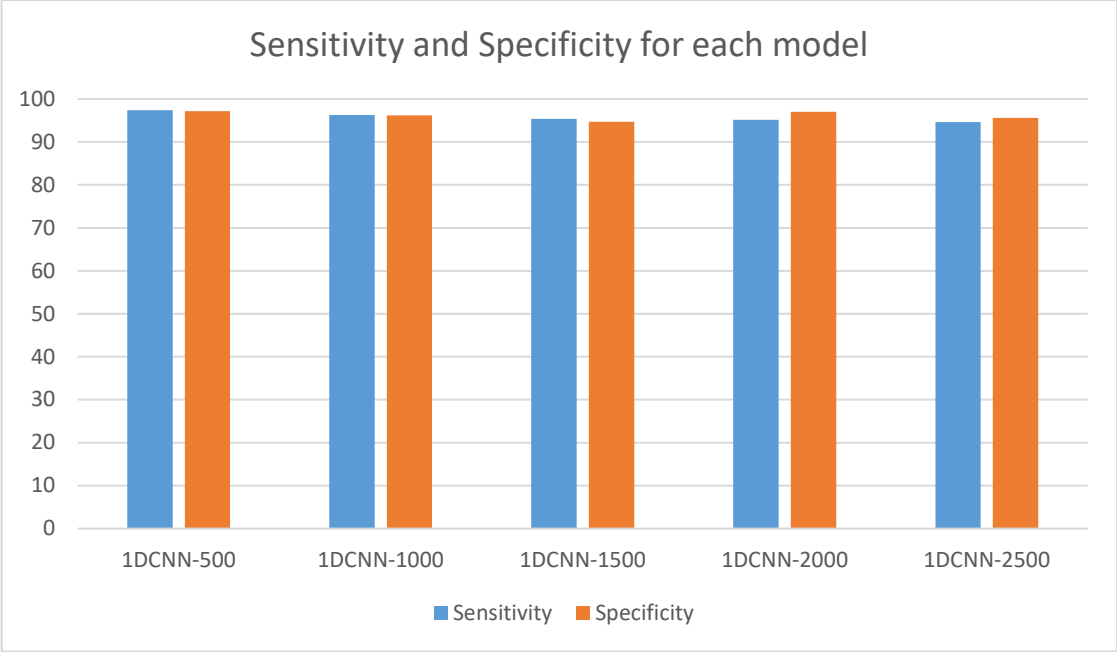

Figure 28: The average scores (Sensitivity and Specificity) of the 5 models after testing

6.3 RESULTS FOR RFC AND SVM USING FEATURE ENGINEERING METHODS

The previous section of this thesis demonstrated the superior performance of a 1DCNN in the diagnosing of OSA using raw ECG time-series data, when compared to both the RF and SCM classifiers without feature engineering. Whilst these results are promising, it is crucial to acknowledge that feature engineering can significantly enhance the performance of traditional ML models. To provide a comprehensive evaluation, Table 33 provides a comparative analysis of RF and SVM classifiers, that have employed feature engineering techniques in the diagnosing of OSA. All these models were trained and tested on the same Apnoea-ECG database, which served as the primary dataset throughout this PhD study.

Model	Feature Engineering Method	Sens/Recall	Spec/Prec(P)	Acc	Author
RFC	Statistical Feature Engineering and Time Series Feature Engineering	1.00	93.02(P)	96.25	A. Jezzini et al.
RFC	Time Series Feature Engineering (Time-Domain Feature Engineering)	92.17	94.79	95.01	AP. Razi et al.
RFC	Domain-Specific Feature Engineering and Transform-Based Feature Engineering	88.75	92.44	91.51	AI.Sharaf et al.
RFC	Time Series Feature Engineering (Time-Domain Feature Engineering and Frequency-Domain Feature Engineering)	84.50	91.20	88.60	J.Zhu et al.
SVM	Statistical Feature Engineering and Time Series Feature Engineering	97.50	97.50	97.50	A. Jezzini et al.
SVM	Time Series Feature Engineering (Time-Domain Feature Engineering)	91.18	93.64	93.69	AP. Razi et al.
SVM	Statistical feature extraction	89.70	94.67	92.59	B.Fatimah et al.
SVM	Time-Domain Feature Engineering, Frequency-Domain Feature Engineering, Nonlinear Feature Engineering	88.84	83.29	85.07	C.Varon et al.

Table 33: Comparison of proposed model V's other OSA ML and DL models

A total of 8 published studies were evaluated for their high quality, high performance and the varied application of feature engineering applied. Each of these studies employed either a RF or SVM classifier. The feature engineering techniques utilised in these studies encompassed a variety of approaches, including Domain-Specific, Transform-Based, Statistical and Time-Series methods (both Time-Domain and Frequency-Domain). It's fair to say that The best performing model is seen in the study by A.Jezzini et al, which employed an SVM using a statical and time-series feature engineering approach based on RR-interval measurements. This study produced a high score of 97.50 as well as a perfectly balanced Sensitivity and Specificity.

6.4 COMPARISON OF RESULTS AGAINST OTHER OSA DETECTION METHODS

Table 34 provides a comparison of result using the proposed method in this current PhD study against other state-of-the-art methods discussed earlier in Chapter 4, Related Work for the Diagnosis of OSA. These studies were chosen based on factors that considered, the quality of paper and the year of publish. Other influences were the type of ML or DL model used and the achieved results. The selected studies ensured a balance of traditional ML models and DL models that fitted the criteria and had produced good to excellent results. Further characteristics were the type of database employed, its rating, and how it was applied in the training and testing phases. Of the chosen studies, 70% employed the same proven database as used in this PhD study. Also, 70% reported their testing was performed on the withheld part of the chosen database. Some limitations to the chosen studies where 60% hadn't described if their presented results were from a single test run or averaged over numerous runs. Further to this, 70% didn't publish the duration of each validation or tests runs.

Model	Sens	Spec/ Prec	Acc	Author
SVM+HMM	82.6	88.4	86.2	C. Song et al
HMMK+SVM	-	-	99.23	C. M. Travieso
RFC	88.75	92.44	91.51	I. Sharaf

RFC	92.17	94.79	95.01	A. P. Razi et al
RFC	95.9	98.4	97.5	J. Zhu et al
RFC	82.4	82.1 (Pr)	82.43	E. Tuncer
RFC	89.0	-	86.0	X. Zhao et al
RUSBoost	87.58	91.49	88.88	A. R. Hassan et al
MFCC-KNN	-	-	85.55	Y. Liu et al
1DCNN+LSTM+SVM	91.24	90.36	90.92	H. Almutairi et al
1DCNNLSTM	88.7	86.6 (Pr)	86.25	A. Sheta et al
1DCNN-AE	90.32	91.63	91.13	K. Cao et al
1DCNN	81.1	98.5	96.6	S. H. Choi et al
1D-SEResGNet	87.6	91.9	90.3	Q. Yang et al
1DCNN	81.1	92.0	87.9	H.-Y. Chang et al
1DCNN	96.0	96.0 (Pr)	NA	E. Urtnasan, J. et al
W-500 1DCNN	97.05	97.25	93.77	Our Previous Study
1DCNN-500	97.42	97.15	97.28	This Model

Table 34: Comparison of proposed model V's other OSA ML and DL models

Observing the results, it is clear to see the 1DCNN approach produces better results in comparison to the other types of approaches and suggests an excellent method for the detection of OSA. Of these 1DCNNs, this current model and our previous model show excellent results. The high accuracy scoring as well as the perfectly balanced sensitivity and specificity are crucial metrics when performing medical diagnostic testing. Having unbalanced sensitivity and specificity will cause trade-offs where people without the condition will be flagged or people with the condition will be missed. Other standout 1DCNN results are seen in *E. Urtnasan, J. et al* and *S. H. Choi et al* studies, however, a highlighted limitation in *E. Urtnasan, J. et al* study was the small dataset size used for training and testing the model, furthermore this was using a 6 layer CNN. In *S. H. Choi* et al model, the concern was the imbalance of sensitivity and specificity, which would cause the misidentification of apnoea events as normal. Looking at the more traditional approaches, there

are two standout results produced by *C. M. Travieso et al* and *J. Zhu et al*. Assessing the *M. Travieso et al* study, a main limitation was that it focused only on the accuracy of the model and doesn't show or produce scores for sensitivity and specificity. Using the metric of accuracy alone can be misleading, particularly with unbalanced data. The highlighted limitation of the *J. Zhu et al* study, was the small sample size used for training and testing, which can lead to skewed results in training phase and poor generalisation on unseen data at the testing phase.

7. DISCUSSION

This study set out with two main objectives. Firstly, to evaluate the 1DCNN model and secondly, to compare it against other classification models. The 1DCNN model was constructed using the state-of-the-art techniques in 1DCNNs, consisting of a Convolutional Layer, a Max Pooling Layer, a Flattening layer, a Fully Connected Multilayer Perceptron (MLP) with hidden dense layers (FC layers) and a SoftMax output for classification. For training the network, a combination of an Adam optimiser and Cross Entropy loss function is applied.

An extensive and rigorous experimental training process was conducted, systematically exploring the method space to achieve optimal model performance. This involved over a thousand experiments, each time fine-tuning hyperparameters, including Learning rate, Window sizes, Training Sample size, Validation Sample size, Filter size, Batch size, Kernel size, and length of Epochs. While the application of this method to all models with datasets was sometimes tedious, each group of experiments provided both valuable insights and strong reference points. This in turn, facilitated the optimisation of subsequent models by establishing a solid foundation for hyperparameter fine-tuning. Additionally, it was also found that applying varying depths of convolutional layers provided no improvement to the 1DCNN model. Using one layer decreased both the computational complexity and load, thus producing excellent classification performances that empirically provided the overall best results.

To further evaluate the 1DCNN, it was decided to perform comparison experiments against other more traditional ML algorithms, namely RFC and SVM, since these are well-known for their binary classification problem-solving. The main objective of this comparison testing was to find the model that frequently produces the best performances, in the quickest times and using the least computational power.

All the models were evaluated using a well-received dataset, containing approx. 285hrs of segmented ECG single-lead time-series signals obtained from 35 subjects. Over 75hr of non-

apnoea segments were initially removed to balance the dataset at approx. 108hrs for each group, apnoea/non-apnoea. To ensure fairness of training and validation, the segments were grouped into a single balanced dataset containing approx. 37.5 million samples of apnoea and 37.5 million samples of non-apnoea.

Changes and limitations to the acquired dataset. In the data of the original evaluation study, the scoring of apnoea's and hypopneas was done according to standard criteria, where the number of apnoea's and hypopnoea's were marked and scored separately using the values Apnoea Index (AI) and Hypopnoea Index (AHI). For the dataset used in this study all the marking and scoring was done by an expert sleep specialist in a different way. Here the new marking and scoring method did not differentiate between apnoea and hypopnoea. The result of the scoring were markings for the beginning and the end of episodes of disordered breathing. The disordered breathing may contain one single apnoea or hypopnoea or may contain a longer sequence of apnoea's and hypopnoea's. The markings were mapped to time with a resolution of one minute. Therefore, it is unknown exactly how much of each scored minute is accommodated with apnoea and/or hypopnoea, whether this is fully or partial. The final result of the scoring was a binary outcome for each minute of the recording being coded as either "normal breathing" (N) which the study refers to as non-apnoea or "disordered breathing" (A), which this study refers to a apnoea. The total number of minutes spent in apnoea's or hypopnoea's was determined for each recording. All scoring was assessed against the Apnoea–Hypopnoea Index (AHI).

The novel idea of reshaping the dataset into 5 different window sizes provided the opportunity to improve training and evaluation of the models. The results showed that using different sizes, slightly impacts on the performances of each model. Results appear to coincide with window sizes, where more rows with less columns (smaller, but more observations) generally produced increased performance, however, this increase seemed to plateau at approx. 500 columns.

Of the 3 groups of models (1DCNN, RFC, SVM) evaluated in the training and validation experiments, the 1DCNN group was shown to be the strongest group when using the W-500 dataset. The 1DCNN-500 model, Sensitivity 0.9743 Specificity 0.9708 Accuracy 0.9699 ROCAUC 0.9966 produced the best results. The W-500 windowed dataset, also produced the best performing models of the other two groups; RFC-500 model, Recall (0)/(1) 0.90/0.94, Precision (0)/(1) 0.94/0.90, Accuracy 0.91 and SVM-500 model, Recall (0)/(1) 0.94/0.50, Precision (0)/(1) 0.65/0.90, Accuracy 0.72.

Figure 29 presents comparable ROCAUC curve scores for the best performing model from each group (1DCNN, RFC, SVM) using the W-500 windowed dataset. 1DCNN-500 produces the best performance.

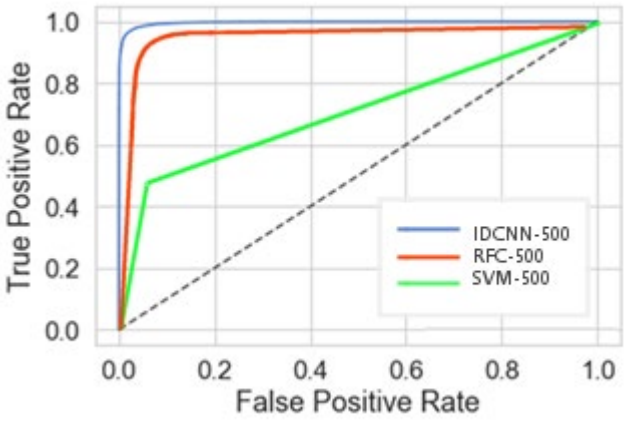

Figure 29: Comparable ROCAUC plot results for the best performing model from each group, using dataset size W=500.

The overall two best performing models, (1DCNN-500 and 1DCNN-1000), produced excellent classification results. Interestingly, the other 3 CNN models (1DCNN-1500, 2000, & 2500), which also performed to a very good standard, with only slight signs of overfitting, found their optimum performances using almost polar-opposite hyperparameter configurations to the two best performing models. Adding dropout layers to each of these 3 models could improve performance.

The overall performance of the RFC models produced very good results with some overfitting. However, the main drawback to this model was the duration of experiments, sometimes taking

over 1hr to complete. Attempts to speed up this process using hyperparameter influences and reducing estimator values, was very limited before results started to dramatically decrease.

The overall presentation of them SVMs was poor, both in terms of performance and results. Classification was very unbalanced, and the duration of the experiments was extremely slow, with some experiments taking almost 17hrs to complete.

The limitations and drawbacks shown by some RFC and SVM results could be associated to the type of data used for the experiments. RFCs and SVMs are often better suited to text analysis, also in the case of SVMs, smaller datasets. Another point is the small number of variables used in these experiments. Both RFC and SVM respond better to higher amounts of variables. However, it is important to note that a separate evaluation of the RFC and SVM was conducted in Chapter 6.3, utilising the same Apnoea-ECG database in this thesis. This time employing feature engineering methods and techniques, some demonstrated excellent performances and scores.

For the testing phase, it was decided to only test the five 1DCNN models. All of the results showed excellent scores across all 5 models with the 1DCNN-500 again displaying the best performance with a sensitivity Sensitivity 0.9742, Specificity 0.9715, Accuracy 0.9728. These scores were averaged over 20 runs. All the results presented in this study have demonstrated the complexity and value of the hyperparameter selection required to achieve an optimal performance in the automated detection of OSA.

A final evaluation was seen in the comparison study of this 1DCNN approach against other OSA detection approaches discussed in section *4.4 Related Work for the Diagnosis of OSA*. The study revealed the 1DCNN model outperformed other models. All but four of the studies used the same Apnoea-ECG database, but with varying formats, techniques, and methods. Further to this, many of the approaches discussed were dependent on third-party signal processing applications to prepare their data, whilst others used traditional ML methods with hand-crafted features, all of which can be time-consuming and expensive. Some limitations were highlighted through the

chosen databases for training and testing, with some using small datasets and another using partial data, all of which can skew training results and cause future generalisability issues when testing. Examining the applied models, some are using traditional methods, whilst others are using LSTM and hybrid approaches, which are based on predictive measures that rely on both accurate data and how well missing values can be guessed. Our model is based on classification, which simply identifies or determines an observations class. Comparing our model to the other classification models discussed, our results are better than comparable, with a more efficient and automated workflow.

A limitation to the 1DCCN model in this PhD study is its inability to determine between apnoea's and hypopneas, also the model has been trained to extract features from ECG signals that are produced from similar devices and constructed using comparable sample and bit rates. Using signals constructed from different variations of sample and bit rates may diminish the produced results. Moreover, the model hasn't been trained to detect features from other types of physiological signals. Further considerations are constraints that include the simplicity of the model and the diminutive size of the dataset used. The current dataset is constructed using 75 million samples, however, 1DCNNs tend to produce their best performances using much larger datasets and multiple convolutional layers. A final consideration would be to further develop the model to detect OSA using a multimodal approach that combines two or more physiological signals.

7.1 SCOPE

This study involves the research of machine learning (ML), which is a very current and wide-ranging topic within the field of computer science. ML approaches (new and current) are starting to make significant contributions towards the diagnosis and classification of sleep disorders, thereby supporting traditional clinical methods. Significantly, the study uses ML technology, One-Dimensional Convolutional Neural Network (1DCNN), this study examines the relationship

between the disease Obstructed Sleep Apnoea (OSA) and time-series ECG data captured from subjects who suffer with OSA.

Boundaries of the study

- The Apnoea-ECG database was the only data used in this study.
- 35 subject (70 recordings) night-time sleep recordings (Time-series data).
- No clinical trial was involved.
- Covid-19 restrictions limited and changed the pattern of work.

8. CONCLUSIONS AND FUTURE WORK

8.1 CONCLUSIONS

Presented in Chapter 1.1 is a list of aims and objectives. These have all been addressed through the development and evaluation of a 1DCNN model for the automated detection of OSA using single-channel ECG signals. The model has been rigorously trained and tested on a carefully selected preprocessed ECG database, demonstrating superior performance compared to other models. By systematically exploring different window sizes and hyperparameters, the model was optimised to achieve high accuracy with a perfect balance of sensitivity and specificity.

OSA, is a prevalent and costly health condition, that affects a significant portion of the global population [4], [5]. Traditional diagnostic methods often fall short, leading to underdiagnosis and serious health consequences [11], [12]. While ML techniques have shown promise in OSA detection, they require substantial expertise and resources. DL models, specifically 1DCNN, offer a more effect and accurate solution [34].

The research in this PhD thesis demonstrates the effectiveness of a 1DCNN in automatically detecting OSA from a single-lead ECG signals. By incorporating a windowing strategy and a well-defined experimental procedure, our model not only surpasses traditional ML models, but also

contemporary DL models, in terms of accuracy, speed and reliability. This innovated approach has the potential to significantly improve the efficiency and accuracy of OSA diagnosis, leading to earlier intervention and better patient outcomes. This research contributes to the advancements of OSA research by providing novel and effective method for automated OSA detection.

8.2 Future Work

This study has provided a view where the design and implementation of the 1DCNN system could deliver a support mechanism in clinical practice for the diagnosis of patients suffering with OSA. However, whilst the approach gives us confidence to perform such tasks, the work could be further expanded applying the following purposes.

- **Further evaluation and testing using alternative ECG signal datasets.**

  To better evaluate the model, further testing using alternative ECG signal datasets will be necessary. When exploring and assessing datasets, the University College Dublin (UCD) Dataset, appears to be the most suitable. A comprehensive testing plan for this dataset is currently in progress.

- **Attain clinical approval to better evaluate this study in a clinical setting.**

  We are in the early stages of dialog with NHS Sleep Medicine clinic, Aintree hospital, Liverpool. This can lead to a collaboration where thorough investigations involving specific subjects from the sleep clinic can take place and clinical approval is granted. This exceeds the scope of the current work and will require ethical consent.

- **Development of a frontend system to interact with the 1DCNN model when classifying uploaded ECG signals.**

  This will involve the creation of a user-facing part such as a website or application where interactions can be performed via API's (Application Programming Interface) with the backend of the system to allow the exchanging of information.

- **Further application in various domains.**

  Akin to its uses in healthcare, the development this system could be applied to a wide variety of domains, such as finance, manufacturing, transport, engineering, energy or anywhere where anomalies can be detected through time-series data, thus avoiding helping with safety and prevent costly problems. Two specific examples could be:

  1. **Motor Bearing Fault Detection**: The system could analyse vibration data to detect early signs of bearing faults.
  2. **Electrical Current Fault Detection**: The system could be used to analyse electrical current data to identify anomalies that may indicate equipment faults.

9. REFERENCES

[1] D. P. White, “Sleep-related breathing disorder.2. Pathophysiology of obstructive sleep apnoea.,” *Thorax*, vol. 50, no. 7, pp. 797–804, 1995, doi: 10.1136/thx.50.7.797.

[2] B. S. Young T, Palta M, Dempsey J, Skatrud J, Weber S, “The occurrence of sleep-disordered breathing among middle-aged adults,” *N Engl J Med 1993;328:1230-5.*, vol. 3, no. 3, p. 4, 1993, doi: 10.1145/1077246.1077257.

[3] R. Cartwright, “Obstructive sleep apnea: A sleep disorder with major effects on health,” *Disease-a-Month*, vol. 47, no. 4, pp. 105–147, 2001, doi: 10.1016/S0011-5029(01)90002-1.

[4] S. M. Ejaz, I. S. Khawaja, S. Bhatia, and T. D. Hurwitz, “Obstructive sleep apnea and depression: A review,” *Innov Clin Neurosci*, vol. 8, no. 8, pp. 17–25, 2011.

[5] B. Lechat *et al.*, “Multinight prevalence, variability, and diagnostic misclassification of obstructive sleep apnea,” *Am J Respir Crit Care Med*, vol. 205, no. 5, pp. 563–569, 2022.

[6] N. M. Punjabi, “The epidemiology of adult obstructive sleep apnea,” Feb. 2008. doi: 10.1513/pats.200709-155MG.

[7] B. S. Young T, Palta M, Dempsey J, Skatrud J, Weber S, “The occurrence of sleep-disordered breathing among middle-aged adults,” *N Engl J Med 1993;328:1230-5.*, vol. 3, no. 3, p. 4, 1993, doi: 10.1145/1077246.1077257.

[8] T. Young, L. Evans, L. Finn, and M. Palta, “Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women,” *Sleep*, vol. 20, no. 9, pp. 705–706, 1997, doi: 10.1093/sleep/20.9.705.

[9] T. Young, L. Evans, L. Finn, and M. Palta, “Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women,” *Sleep*, vol. 20, no. 9, pp. 705–706, 1997, doi: 10.1093/sleep/20.9.705.

[10] Frost & Sullivan, “Hidden Health Crisis Costing America Billions: Underdiagnosing and Undertreating Obstructive Sleep Apnea Draining Healthcare System,” *American Academy of Sleep Medicine, 2016.*, 2016.

[11] N. M. Punjabi *et al.*, “Sleep-disordered breathing and mortality: A prospective cohort study,” *PLoS Med*, vol. 6, no. 8, 2009, doi: 10.1371/journal.pmed.1000132.

[12] T. Young, “Epidemiological insights into the public health burden of sleep disordered breathing: Sex differences in survival among sleep clinic patients,” *Thorax*, vol. 53, no. SUPPL. 3, pp. 16–19, 1998.

[13] J. C.-G. J. TERÁN-SANTOS, A. JIMÉNEZ-GÓMEZ, “Association Between Sleep Apnea and the Risk of Traffic Accidents,” *N Engl J Med*, vol. 340, pp. 847–851, 1999.

[14] Frost & Sullivan, “Hidden Health Crisis Costing America Billions: Underdiagnosing and Undertreating Obstructive Sleep Apnea Draining Healthcare System,” *American Academy of Sleep Medicine, 2016.*, 2016.

[15] W. T. McNicholas, “Diagnosis of Obstructive Sleep Apnea in Adults,” *Proc Am Thorac Soc*, vol. 5, no. 2, pp. 154–160, 2008, doi: 10.1513/pats.200708-118MG.

[16] F. Chung, H. R. Abdullah, and P. Liao, “STOP-bang questionnaire a practical approach to screen for obstructive sleep apnea,” *Chest*, vol. 149, no. 3, pp. 631–638, 2016, doi: 10.1378/chest.15-0903.

[17] N. C. Netzer, R. A. Stoohs, C. M. Netzer, K. Clark, and K. P. Strohl, “Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome,” 1999. doi: 10.7326/0003-4819-131-7-199910050-00002.

[18] Johns M. W., “A new method for measuring daytime sleepiness: the Epworth sleepiness scale,” *Sleep*, vol. 14, no. 6, pp. 540–545, 1991, doi: 10.1016/j.sleep.2007.08.004.

[19] A. L. Chesson *et al.*, “The indications for polysomnography and realted procedures. An American sleep disorders association review,” *Pneumologie*, vol. 52, no. 3, p. 154, 1998.

[20] W. W. Flemons, N. J. Douglas, S. T. Kuna, D. O. Rodenstein, and J. Wheatley, “Access to Diagnosis and Treatment of Patients with Suspected Sleep Apnea,” *Am J Respir Crit Care Med*, vol. 169, no. 6, pp. 668–672, 2004, doi: 10.1164/rccm.200308-1124pp.

[21] W. W. Flemons, N. J. Douglas, S. T. Kuna, D. O. Rodenstein, and J. Wheatley, “Access to Diagnosis and Treatment of Patients with Suspected Sleep Apnea,” *Am J Respir Crit Care Med*, vol. 169, no. 6, pp. 668–672, 2004, doi: 10.1164/rccm.200308-1124pp.

[22] N. A. Collop *et al.*, “Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Portable Monitoring Task Force of the American Academy of Sleep Medicine.,” *J Clin Sleep Med*, vol. 3, no. 7, pp. 737–47, 2007.

[23] L. Abrahamyan *et al.*, “Diagnostic accuracy of level IV portable sleep monitors versus polysomnography for obstructive sleep apnea: a systematic review and meta-analysis,” *Sleep and Breathing*, vol. 22, no. 3, pp. 593–611, 2018, doi: 10.1007/s11325-017-1615-1.

[24] V. K. Kapur *et al.*, “Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: An American academy of sleep medicine clinical practice guideline,” *Journal of Clinical Sleep Medicine*, vol. 13, no. 3, pp. 479–504, 2017, doi: 10.5664/jcsm.6506.

[25] L. Almazaydeh, K. Elleithy, and M. Faezipour, “Detection of obstructive sleep apnea through ECG signal features,” *IEEE International Conference on Electro Information Technology*, pp. 1–6, 2012, doi: 10.1109/EIT.2012.6220730.

[26] S. G. Jones *et al.*, “Regional Reductions in Sleep Electroencephalography Power in Obstructive Sleep Apnea: A High-Density EEG Study,” *Sleep*, vol. 37, no. 2, pp. 399–407, 2014, doi: 10.5665/sleep.3424.

[27] J. V. Marcos, R. Hornero, D. Álvarez, M. Aboy, and F. Del Campo, “Automated prediction of the apnea-hypopnea index from nocturnal oximetry recordings,” *IEEE Trans Biomed Eng*, vol. 59, no. 1, pp. 141–149, 2012, doi: 10.1109/TBME.2011.2167971.

[28] N. Selvaraj and R. Narasimhan, “Detection of sleep apnea on a per-second basis using respiratory signals,” *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 2124–2127, 2013, doi: 10.1109/EMBC.2013.6609953.

[29] B. L. Koley and D. Dey, “Automatic detection of sleep apnea and hypopnea events from single channel measurement of respiration signal employing ensemble binary SVM classifiers,”

Measurement (Lond), vol. 46, no. 7, pp. 2082–2092, 2013, doi: 10.1016/j.measurement.2013.03.016.

[30] A. Azarbarzin and Z. Moussavi, "Snoring sounds variability as a signature of obstructive sleep apnea," *Med Eng Phys*, vol. 35, no. 4, pp. 479–485, 2013, doi: 10.1016/j.medengphy.2012.06.013.

[31] E. Dafna, A. Tarasiuk, and Y. Zigel, "Automatic detection of whole night snoring events using non-contact microphone," *PLoS One*, vol. 8, no. 12, 2013, doi: 10.1371/journal.pone.0084139.

[32] J. Solà-Soler, J. A. Fiz, J. Morera, and R. Jané, "Multiclass classification of subjects with sleep apnoea-hypopnoea syndrome through snoring analysis," *Med Eng Phys*, vol. 34, no. 9, pp. 1213–1220, 2012, doi: 10.1016/j.medengphy.2011.12.008.

[33] L. Baier, F. Jöhren, and S. Seebacher, "Challenges in the Deployment and Operation of Machine Learning in Practice.," in *ECIS*, 2019.

[34] S. Asif *et al.*, "Advancements and Prospects of Machine Learning in Medical Diagnostics: Unveiling the Future of Diagnostic Precision," *Archives of Computational Methods in Engineering*, 2024, doi: 10.1007/s11831-024-10148-w.

[35] M. Subramanian, K. Shanmugavadivel, and P. S. Nandhini, "On fine-tuning deep learning models using transfer learning and hyper-parameters optimization for disease identification in maize leaves," *Neural Comput Appl*, vol. 34, no. 16, pp. 13951–13968, 2022, doi: 10.1007/s00521-022-07246-w.

[36] A. V Benjafield *et al.*, "Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis," *Lancet Respir Med*, vol. 7, no. 8, pp. 687–698, 2019.

[37] A. Broström, O. Sunnergren, P. Nilsen, B. Fridlund, M. Ulander, and E. Svanborg, "Gender differences in respiratory disturbance, sleep and daytime sleepiness in hypertensive patients with different degrees of obesity," *European Journal of Cardiovascular Nursing*, vol. 12, no. 2, pp. 140–149, 2013, doi: 10.1177/1474515112438163.

[38] M. M. Lyons, N. Y. Bhatt, A. I. Pack, and U. J. Magalang, "Global burden of sleep-disordered breathing and its implications," *Respirology*, vol. 25, no. 7, pp. 690–702, 2020.

[39] "EMBARGOED UNTIL 12:00 GMT ON THURSDAY 2 MARCH 2023 PRESS RELEASE Economic impact of overweight and obesity to surpass $4 trillion by 2035." [Online]. Available: https://cdn.who.int/media/docs/default-source/obesity/who-accelertaion-plan-to-stop-

[40] R. Bódizs, B. Schneider, P. P. Ujma, C. G. Horváth, M. Dresler, and Y. Rosenblum, "Fundamentals of sleep regulation: Model and benchmark values for fractal and oscillatory neurodynamics," *Prog Neurobiol*, vol. 234, p. 102589, Mar. 2024, doi: 10.1016/J.PNEUROBIO.2024.102589.

[41] Z. Mohamad Noor, A. J. Smith, S. S. Smith, and L. M. Nissen, "A study protocol: A community pharmacy-based intervention for improving the management of sleep disorders in the community settings," *BMC Health Serv Res*, vol. 14, p. 74, 2014, doi: 10.1186/1472-6963-14-74.

[42] A. P. A. Vorster, E. J. W. van Someren, A. I. Pack, R. Huber, M. H. Schmidt, and C. L. A. Bassetti, “Sleep Health,” *Clinical and Translational Neuroscience*, vol. 8, no. 1, 2024, doi: 10.3390/ctn8010008.

[43] J. C. Carberry, J. Amatoury, and D. J. Eckert, “Personalized Management Approach for OSA,” *Chest*, vol. 153, no. 3, pp. 744–755, 2017, doi: 10.1016/j.chest.2017.06.011.

[44] K. Harding and M. Feldman, *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*, vol. 47, no. 4. 2008. doi: 10.1097/01.CHI.0000270812.55636.3b.

[45] M. J. Sateia, “International Classification of Sleep Disorders-Third Edition,” *Chest*, vol. 146, no. 5, pp. 1387–1394, 2014, doi: https://doi.org/10.1378/chest.14-0970.

[46] Y. C. G. Feti CETİN, “CLASSIFICATION OF SLEEP DISORDERS,” *Modern Medical and Health Sciences*, p. 57, 2023.

[47] M. J. Sateia, “International classification of sleep disorders-third edition highlights and modifications,” *Chest*, vol. 146, no. 5, pp. 1387–1394, 2014, doi: 10.1378/chest.14-0970.

[48] X. Wang, Y. Ouyang, Z. Wang, G. Zhao, L. Liu, and Y. Bi, “Obstructive sleep apnea and risk of cardiovascular disease and all-cause mortality: A meta-analysis of prospective cohort studies,” *Int J Cardiol*, vol. 169, no. 3, pp. 207–214, 2013, doi: 10.1016/j.ijcard.2013.08.088.

[49] S. Javaheri and M. S. Badr, “Central sleep apnea: pathophysiologic classification,” *Sleep*, vol. 46, no. 3, p. zsac113, Mar. 2023, doi: 10.1093/sleep/zsac113.

[50] S. Javaheri and S. Redline, “Insomnia and Risk of Cardiovascular Disease,” *Chest*, vol. 152, no. 2, pp. 435–444, 2017, doi: 10.1016/j.chest.2017.01.026.

[51] T. Zdanowicz, K. Turowski, J. Celej-Szuster, R. Lorencowicz, and E. Przychodzka, “Insomnia, Sleepiness, and Fatigue Among Polish Nurses,” *Workplace Health Saf*, vol. 68, no. 6, pp. 272–278, Feb. 2020, doi: 10.1177/2165079920901534.

[52] Z. Khan and L. M. Trotti, “Central disorders of hypersomnolence: Focus on the narcolepsies and idiopathic hypersomnia,” *Chest*, vol. 148, no. 1, pp. 262–273, 2015, doi: 10.1378/chest.14-1304.

[53] A. B. Williams, J. M. Dzierzewski, S. C. Griffin, M. J. Lind, D. Dick, and B. D. Rybarczyk, “Insomnia Disorder and Behaviorally Induced Insufficient Sleep Syndrome: Prevalence and Relationship to Depression in College Students,” *Behavioral Sleep Medicine*, vol. 18, no. 2, pp. 275–286, Mar. 2020, doi: 10.1080/15402002.2019.1578772.

[54] R. R. Auger, “Circadian Rhythm Sleep-Wake Disorders,” *An Evidence-Based Guide for Clinicians and Investigators. Cham: Springer Nature Switzerland AG*, 2020.

[55] M. J. Sateia, “International classification of sleep disorders-third edition highlights and modifications,” *Chest*, vol. 146, no. 5, pp. 1387–1394, 2014, doi: 10.1378/chest.14-0970.

[56] S. Tale, A. Kumar, L. K. Saini, S. P. Meitei, and R. Gupta, “Sleepwalking and prolonged partial sleep paralysis in a case of obstructive sleep apnea,” *Sleep Science*, vol. 14, no. 2, p. 186, 2021.

[57] F. Gemignani, F. Brindani, F. Vitetta, A. Marbini, and S. Calzetti, “Restless legs syndrome in diabetic neuropathy: A frequent manifestation of small fiber neuropathy,” *Journal of the*

Peripheral Nervous System, vol. 12, no. 1, pp. 50–53, 2007, doi: 10.1111/j.1529-8027.2007.00116.x.

[58] R. Cartwright, "Obstructive sleep apnea: a sleep disorder with major effects on health," *Disease-a-Month*, vol. 47, no. 4, pp. 105–147, 2001.

[59] N. Selvaraj and R. Narasimhan, "Detection of sleep apnea on a per-second basis using respiratory signals," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2013, pp. 2124–2127.

[60] J. V. Marcos, R. Hornero, D. Alvarez, M. Aboy, and F. Del Campo, "Automated prediction of the apnea-hypopnea index from nocturnal oximetry recordings," *IEEE Trans Biomed Eng*, vol. 59, no. 1, pp. 141–149, 2011.

[61] F. Hadavandsiri *et al.*, "Non-communicable disease comorbidities in HIV patients: diabetes, hypertension, heart disease, and obstructive sleep apnea as a neglected issue," *Sci Rep*, vol. 13, no. 1, p. 12730, 2023.

[62] T. Young, "The occurrence of steep-disordered breathing among middle-aged adults," *N Engl J Med*, vol. 328, pp. 230–235, 1993.

[63] N. Selvaraj and R. Narasimhan, "Detection of sleep apnea on a per-second basis using respiratory signals," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2013, pp. 2124–2127.

[64] B. L. Koley and D. Dey, "Automatic detection of sleep apnea and hypopnea events from single channel measurement of respiration signal employing ensemble binary SVM classifiers," *Measurement*, vol. 46, no. 7, pp. 2082–2092, 2013, doi: https://doi.org/10.1016/j.measurement.2013.03.016.

[65] K. A. Carter, N. E. Hathaway, and C. F. Lettieri, "Common sleep disorders in children," *Am Fam Physician*, vol. 89, no. 5, pp. 368–377, 2014.

[66] J.-M. YOON, H.-R. LEE, and K.-H. LEE, "Characteristics of sleep disorders in childhood and adolescence," *Journal of the Korean Child Neurology Society*, pp. 185–191, 2007.

[67] B. N. Mandrell *et al.*, "Predictors of narcolepsy and hypersomnia due to medical disorder in pediatric craniopharyngioma," *J Neurooncol*, vol. 148, pp. 307–316, 2020.

[68] I.-H. Chung, W.-C. Chin, Y.-S. Huang, and C.-H. Wang, "Pediatric narcolepsy—A practical review," *Children*, vol. 9, no. 7, p. 974, 2022.

[69] M. Gradisar *et al.*, "Sleep's role in the development and resolution of adolescent depression," *Nature Reviews Psychology*, vol. 1, no. 9, pp. 512–523, 2022.

[70] "SleepAdolescents.pdf."

[71] X. Liu, "Sleep and adolescent suicidal behavior.," *Sleep*, vol. 27, no. 7, pp. 1351–1358, 2004, doi: 10.1093/sleep/27.7.1351.

[72] "SleepHabitsChildren.pdf."

[73] J. H. Gilmore, "NIH Public Access," *North*, vol. 29, no. 10, pp. 1883–1889, 2008, doi: 10.3174/ajnr.A1256.Functional.

[74] “National Sleep Foundation.,” 2010.

[75] L. Palagini, E. Hertenstein, D. Riemann, and C. Nissen, “Sleep, insomnia and mental health,” *J Sleep Res*, vol. 31, no. 4, p. e13628, 2022.

[76] C. A. Webb, R. Cui, C. Titus, A. Fiske, and M. R. Nadorff, “Sleep Disturbance, Activities of Daily Living, and Depressive Symptoms among Older Adults,” *Clin Gerontol*, vol. 00, no. 00, pp. 1–9, 2017, doi: 10.1080/07317115.2017.1408733.

[77] N. Tsuno, A. Besset, and K. Ritchie, “Sleep and depression,” 2005. doi: 10.4088/JCP.v66n1008.

[78] D. Wajszilber, J. A. Santiseban, and R. Gruber, “Sleep disorders in patients with ADHD: impact and management challenges,” *Nat Sci Sleep*, pp. 453–480, 2018.

[79] J. L. Paterson, A. C. Reynolds, S. A. Ferguson, and D. Dawson, “Sleep and obsessive-compulsive disorder (OCD),” *Sleep Med Rev*, vol. 17, no. 6, pp. 465–474, 2013.

[80] L. Basquin *et al.*, “Study of the different sleep disturbances during the prodromal phase of depression and mania in bipolar disorders,” *Bipolar Disord*, 2024.

[81] W. Xu, C.-C. Tan, J.-J. Zou, X.-P. Cao, and L. Tan, “Sleep problems and risk of all-cause cognitive decline or dementia: an updated systematic review and meta-analysis,” *J Neurol Neurosurg Psychiatry*, vol. 91, no. 3, pp. 236–244, 2020.

[82] S. Garbarino, P. Lanteri, N. L. Bragazzi, N. Magnavita, and E. Scoditti, “Role of sleep deprivation in immune-related disease risk and outcomes,” *Commun Biol*, vol. 4, no. 1, p. 1304, 2021.

[83] S. Liu, X. Wang, Q. Zheng, L. Gao, and Q. Sun, “Sleep deprivation and central appetite regulation,” *Nutrients*, vol. 14, no. 24, p. 5196, 2022.

[84] J. W. Cho and J. F. Duffy, “Sleep, sleep disorders, and sexual dysfunction,” *World J Mens Health*, vol. 37, no. 3, pp. 261–275, 2019.

[85] S. C. Liew and T. Aung, “Sleep deprivation and its association with diseases-a review,” *Sleep Med*, vol. 77, pp. 192–204, 2021.

[86] B. Jurado-Gámez, O. Guglielmi, F. Gude, and G. Buela-Casal, “Workplace Accidents, Absenteeism and Productivity in Patients With Sleep Apnea,” *Archivos de Bronconeumología (English Edition)*, vol. 51, no. 5, pp. 213–218, 2015, doi: 10.1016/j.arbr.2014.12.002.

[87] “Brigham and Women ’ S Hospital,” 2013.

[88] “American Academy of Sleep Medicine.”

[89] “Brigham and Women ’ S Hospital,” 2013.

[90] N. T. Ayas and C. A. Marra, “Continuous positive airway pressure therapy for obstructive sleep apnea syndrome: do the dollars make sense?,” *Sleep*, vol. 28, no. 10, pp. 1211–3, 2005.

[91] M. Karimi, J. Hedner, H. Habel, O. Nermann, and L. Grote, *A sleep apnea related risk of vehicle accident is reduced by CPAP - Swedish traffic accident data acquisition (STRADA) registry*. 2014.

[92] O. Schlafer, V. Wenzel, and B. Högl, “Schlafstörungen bei Ärzten im Schichtdienst,” *Anaesthesist*, vol. 63, no. 11, pp. 844–851, 2014, doi: 10.1007/s00101-014-2374-z.

[93] R. Powell and A. Copping, “Construction worker sleep deprivation and its effects on personal safety,” no. September, pp. 203–211, 2010.

[94] K. Suzuki, T. Ohida, Y. Kaneita, E. Yokoyama, and M. Uchiyama, “Daytime sleepiness, sleep habits and occupational accidents among hospital nurses,” *J Adv Nurs*, vol. 52, no. 4, pp. 445–453, 2005, doi: 10.1111/j.1365-2648.2005.03610.x.

[95] O. Schlafer, V. Wenzel, and B. Högl, “Schlafstörungen bei Ärzten im Schichtdienst,” *Anaesthesist*, vol. 63, no. 11, pp. 844–851, 2014, doi: 10.1007/s00101-014-2374-z.

[96] M. T. De Mello *et al.*, “Sleep disorders as a cause of motor vehicle collisions,” *Int J Prev Med*, vol. 4, no. 3, pp. 246–257, 2013.

[97] R. Gupta and S. R. Pandi-perumal, “Hypersomnolence and Traff ic Saf ety Drivers Drowsiness Sleepiness Narcolepsy Obstructive sleep apnea,” p. 225503, 2017.

[98] W. T. McNicholas, “Diagnosis of Obstructive Sleep Apnea in Adults,” *Proc Am Thorac Soc*, vol. 5, no. 2, pp. 154–160, 2008, doi: 10.1513/pats.200708-118MG.

[99] M. T. De Mello *et al.*, “Sleep disorders as a cause of motor vehicle collisions,” *Int J Prev Med*, vol. 4, no. 3, pp. 246–257, 2013.

[100] M. Karimi, J. Hedner, H. Habel, O. Nermann, and L. Grote, *A sleep apnea related risk of vehicle accident is reduced by CPAP - Swedish traffic accident data acquisition (STRADA) registry*. 2014.

[101] N. T. Ayas and C. A. Marra, “Continuous positive airway pressure therapy for obstructive sleep apnea syndrome: do the dollars make sense?,” *Sleep*, vol. 28, no. 10, pp. 1211–3, 2005.

[102] S. Garbarino *et al.*, “Insomnia is associated with road accidents. Further evidence from a study on truck drivers,” *PLoS One*, vol. 12, no. 10, pp. 1–10, 2017, doi: 10.1371/journal.pone.0187256.

[103] D. Fischer, E. B. Klerman, and A. J. K. Phillips, “Measuring sleep regularity: theoretical properties and practical usage of existing metrics,” *Sleep*, vol. 44, no. 10, p. zsab103, 2021.

[104] L. Q. Cerri, M. C. Justo, V. Clemente, A. A. Gomes, A. S. Pereira, and D. R. Marques, “Insomnia Severity Index: A reliability generalisation meta-analysis,” *J Sleep Res*, vol. 32, no. 4, p. e13835, 2023.

[105] T. Roth *et al.*, “A new questionnaire to detect sleep disorders,” *Sleep Med*, vol. 3, no. 2, pp. 99–108, 2002.

[106] A. N. Vgontzas, T. L. Tan, E. O. Bixler, L. F. Martin, D. Shubert, and A. Kales, “Sleep apnea and sleep disruption in obese patients,” *Arch Intern Med*, vol. 154, no. 15, pp. 1705–1711, 1994.

[107] N. A. Collop *et al.*, “Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Portable Monitoring Task Force of the American Academy of Sleep Medicine.,” *J Clin Sleep Med*, vol. 3, no. 7, pp. 737–47, 2007.

[108] M. Elbaz, G. M. Roue, F. Lofaso, and M. a Quera Salva, “Utility of actigraphy in the diagnosis of obstructive sleep apnea.,” *Sleep*, vol. 25, no. 5, pp. 527–531, 2002, doi: 10.1093/sleep/25.5.525.

[109] S. D. Kyle, C. B. Miller, Z. Rogers, A. N. Siriwardena, K. M. MacMahon, and C. A. Espie, “Sleep Restriction Therapy for Insomnia is Associated with Reduced Objective Total Sleep Time, Increased Daytime Somnolence, and Objectively Impaired Vigilance: Implications for the Clinical Management of Insomnia Disorder,” *Sleep*, vol. 37, no. 2, pp. 229–237, 2014, doi: 10.5665/sleep.3386.

[110] N. Baranwal, K. Y. Phoebe, and N. S. Siegel, “Sleep physiology, pathophysiology, and sleep hygiene,” *Prog Cardiovasc Dis*, vol. 77, pp. 59–69, 2023.

[111] V. C. Abad and C. Guilleminault, “New developments in the management of narcolepsy,” *Nat Sci Sleep*, vol. 9, pp. 39–57, 2017, doi: 10.2147/NSS.S103467.

[112] A. T. Siebern, S. Suh, and S. Nowakowski, “Non-Pharmacological Treatment of Insomnia,” *Neurotherapeutics*, vol. 9, no. 4, pp. 717–727, 2012, doi: 10.1007/s13311-012-0142-9.

[113] J. D. Lie, K. N. Tu, D. D. Shen, and B. M. Wong, “Pharmacological Treatment of Insomnia.,” *P T*, vol. 40, no. 11, pp. 759–71, 2015, doi: 10.1007/978-1-60327-042-7_23.

[114] H. Chang, Y. Chen, and J. Du, “Obstructive sleep apnea treatment in adults,” *Kaohsiung J Med Sci*, vol. 36, no. 1, pp. 7–12, 2020.

[115] T. B. Huedo-Medina, I. Kirsch, J. Middlemass, M. Klonizakis, and A. N. Siriwardena, “Effectiveness of non-benzodiazepine hypnotics in treatment of adult insomnia: Meta-analysis of data submitted to the Food and Drug Administration,” *BMJ (Online)*, vol. 346, no. 7889, pp. 1–13, 2013, doi: 10.1136/bmj.e8343.

[116] M. J. Sateia, D. J. Buysse, A. D. Krystal, D. N. Neubauer, and J. L. Heald, “Clinical Practice Guideline for the Pharmacologic Treatment of Chronic Insomnia in Adults: An American academy of sleep medicine clinical practice guideline,” *Journal of Clinical Sleep Medicine*, vol. 13, no. 2, pp. 307–349, 2017, doi: 10.5664/jcsm.6470.

[117] U. Kallweit and C. L. Bassetti, “Pharmacological management of narcolepsy with and without cataplexy,” *Expert Opin Pharmacother*, vol. 18, no. 8, pp. 809–817, May 2017, doi: 10.1080/14656566.2017.1323877.

[118] P. J. Hauri, “Sleep disorders,” in *Handbook of Psychology and Health, Volume I*, Routledge, 2021, pp. 211–260.

[119] C. Positive and A. Pressure, “Oral appliances for sleep apnoea”.

[120] J. C. Carberry, J. Amatoury, and D. J. Eckert, “Personalized Management Approach for OSA,” *Chest*, vol. 153, no. 3, pp. 744–755, 2017, doi: 10.1016/j.chest.2017.06.011.

[121] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter, “Apnea-ECG database,” *Comput Cardiol*, pp. 255–258, 2000, doi: 10.1109/cic.2000.898505.

[122] T. R. M. team in collaboration with the C. and F. I. Group, “Obstructive sleep apnea, Great Ormond St Hospital,” NHS. Accessed: Oct. 01, 2023. [Online]. Available: https://www.gosh.nhs.uk/conditions-and-treatments/conditions-we-treat/obstructive-sleep-apnoea

[123] J. M. Slowik, A. Sankari, and J. F. Collen, “Obstructive Sleep Apnea.,” Treasure Island (FL), 2023.

[124] M. Suzuki, “Obstructive sleep apnea-consideration of its pathogenesis,” *Auris Nasus Larynx*, vol. 49, no. 3, pp. 313–321, 2022.

[125] J. L. Carroll and D. F. Donnelly, “Respiratory physiology and pathophysiology during sleep,” *Principles and Practice of Pediatric Sleep Medicine. 2nd edn. Philadelphia, Elsevier*, pp. 179–194, 2014.

[126] B. A. Edwards and D. P. White, “Control of the pharyngeal musculature during wakefulness and sleep: implications in normal controls and sleep apnea.,” *Head Neck*, vol. 33 Suppl 1, no. Suppl 1, pp. S37-45, Oct. 2011, doi: 10.1002/hed.21841.

[127] P. Collard, P. Rombaux, and D. O. Rodenstein, “Why should we enlarge the pharynx in obstructive sleep apnea?,” *Sleep*, vol. 19, no. suppl_9, pp. S85–S87, 1996.

[128] Columbiadoctors, “Obstructive sleep apnea: Blocked upper airway,” Healthwise. Accessed: Oct. 15, 2023. [Online]. Available: https://www.columbiadoctors.org/health-library/multimedia/obstructive-sleep-apnea-blocked-upper-airway/

[129] A. Patwa and A. Shah, “Anatomy and physiology of respiratory system relevant to anaesthesia.,” *Indian J Anaesth*, vol. 59, no. 9, pp. 533–541, Sep. 2015, doi: 10.4103/0019-5049.165849.

[130] E. van Lunteren, “Muscles of the pharynx: structural and contractile properties,” *Ear Nose Throat J*, vol. 72, no. 1, pp. 27–33, 1993.

[131] GettyImages, “Anatomical structure of the Upper Respiratory Tract,” iStock. Accessed: Oct. 15, 2023. [Online]. Available: https://www.istockphoto.com/vector/anatomy-of-the-mouth-and-tongue-medical-vector-illustration-on-white-background-gm1141280498-305696760

[132] M. H. Kryger, T. Roth, and W. C. Dement, *Principles and Practice of Sleep Medicine E-Book: Expert Consult-Online and Print*. Elsevier Health Sciences, 2010.

[133] A. R. Schwartz *et al.*, “The hypotonic upper airway in obstructive sleep apnea: role of structures and neuromuscular activity,” *Am J Respir Crit Care Med*, vol. 157, no. 4, pp. 1051–1057, 1998.

[134] L. Pereira *et al.*, “Prevalence of adenoid hypertrophy: A systematic review and meta-analysis,” *Sleep Med Rev*, vol. 38, pp. 101–112, 2018.

[135] W. C. Orr and R. J. Martin, “Obstructive sleep apnea associated with tonsillar hypertrophy in adults,” *Arch Intern Med*, vol. 141, no. 8, pp. 990–992, 1981.

[136] L. Chi *et al.*, “Identification of craniofacial risk factors for obstructive sleep apnoea using three-dimensional MRI,” *European Respiratory Journal*, vol. 38, no. 2, pp. 348–358, 2011.

[137] S. N. Tan and B. Abdullah, “The association between obstructive sleep apnea and allergic rhinitis: current literature review,” *Curr Respir Med Rev*, vol. 17, no. 1, pp. 13–19, 2021.

[138] P. C. Deegan and W. T. McNicholas, “Pathophysiology of obstructive sleep apnoea,” *European Respiratory Journal*, vol. 8, no. 7, pp. 1161–1178, 1995.

[139] A. Carneiro-Barrera *et al.*, “Interdisciplinary weight loss and lifestyle intervention for obstructive sleep apnoea in adults: Rationale, design and methodology of the INTERAPNEA study,” *Nutrients*, vol. 11, no. 9, p. 2227, 2019.

[140] National Diabetes Information Clearinghouse, “What I need to know about Carbohydrate Counting and Diabetes,” *NIH Publication Nº 14–7496*, pp. 1–22, 2013, doi: 10.1196/annals.1417.033.Associations.

[141] M. Sands-Lincoln *et al.*, “Sleep Duration, Insomnia, and Coronary Heart Disease Among Postmenopausal Women in the Women’s Health Initiative,” *J Womens Health*, vol. 22, no. 6, pp. 477–486, 2013, doi: 10.1089/jwh.2012.3918.

[142] A. Hansford, “Thirty years of CPAP: A brief history of OSA,” *ResMedica Clinical Newsletter*, vol. 14, p. 16, 2011.

[143] J. H. Auchincloss, E. Cook, and A. D. Renzetti, “Clinical and physiological aspects of a case of obesity, polycythemia and alveolar hypoventilation,” *J Clin Invest*, vol. 34, no. 10, pp. 1537–1545, 1955.

[144] C. S. Burwell, E. D. Robin, R. D. Whaley, and A. G. Bickelmann, “Extreme obesity associated with alveolar hypoventilation—a Pickwickian syndrome,” *Am J Med*, vol. 21, no. 5, pp. 811–818, 1956.

[145] J. Tambling, “Pickwick Papers,” in *The Oxford Handbook of Charles Dickens*, Oxford University Press, 2018, p. 35.

[146] M. Kryger, “Charles Dickens: impact on medicine and society,” *Journal of Clinical Sleep Medicine*, vol. 8, no. 3, pp. 333–338, 2012.

[147] J. H. Auchincloss, E. Cook, and A. D. Renzetti, “Clinical and physiological aspects of a case of obesity, polycythemia and alveolar hypoventilation,” *J Clin Invest*, vol. 34, no. 10, pp. 1537–1545, 1955.

[148] H. Gastaut, C. A. Tassinari, and B. Duron, “Polygraphic study of the episodic diurnal and nocturnal (hypnic and respiratory) manifestations of the Pickwick syndrome,” *Brain Res*, vol. 1, no. 2, pp. 167–186, 1966.

[149] D. M. Rapoport, S. M. Garay, H. Epstein, and R. M. Goldring, “Hypercapnia in the obstructive sleep apnea syndrome: a reevaluation of the ‘Pickwickian syndrome,’” *Chest*, vol. 89, no. 5, pp. 627–635, 1986.

[150] J. W. Shepard *et al.*, “History of the development of sleep medicine in the United States,” *Journal of clinical sleep medicine*, vol. 1, no. 01, pp. 61–82, 2005.

[151] A. W. De Weerd, “Actigraphy, the alternative way?,” *Front Psychiatry*, vol. 5, p. 155, 2014.

[152] Z. Pouliot, M. Peters, H. Neufeld, and M. H. Kryger, “Using self-reported questionnaire data to prioritize OSA patients for polysomnography,” *Sleep*, vol. 20, no. 3, pp. 232–236, 1997.

[153] F. Kapsimalis and M. H. Kryger, “Gender and obstructive sleep apnea syndrome, part 1: clinical features,” *Sleep*, vol. 25, no. 4, pp. 409–416, 2002.

[154] N. F. Watson, “Health care savings: the economic value of diagnostic and therapeutic care for obstructive sleep apnea,” *Journal of Clinical Sleep Medicine*, vol. 12, no. 8, pp. 1075–1077, 2016.

[155] I. M. Rosen *et al.*, “Clinical use of a home sleep apnea test: an American Academy of Sleep Medicine position statement,” *Journal of Clinical Sleep Medicine*, vol. 13, no. 10, pp. 1205–1207, 2017.

[156] M. T. Bianchi and B. Goparaju, “Potential underestimation of sleep apnea severity by at-home kits: rescoring in-laboratory polysomnography without sleep staging,” *Journal of clinical sleep medicine*, vol. 13, no. 4, pp. 551–555, 2017.

[157] I. Kononenko, “Machine learning for medical diagnosis: history, state of the art and perspective,” *Artif Intell Med*, vol. 23, no. 1, pp. 89–109, 2001.

[158] D. A. Pisner and D. M. Schnyer, “Support vector machine,” in *Machine learning*, Elsevier, 2020, pp. 101–121.

[159] M. Ahmed, R. Seraj, and S. M. S. Islam, “The k-means algorithm: A comprehensive survey and performance evaluation,” *Electronics (Basel)*, vol. 9, no. 8, p. 1295, 2020.

[160] I. Wickramasinghe and H. Kalutarage, “Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation,” *Soft comput*, vol. 25, no. 3, pp. 2277–2293, 2021.

[161] S. Rangineni and D. Marupaka, “Data Mining Techniques Appropriate for the Evaluation of Procedure Information,” *International Journal of Management, IT & Engineering*, vol. 13, no. 9, pp. 12–25, 2023.

[162] R. Genuer, J.-M. Poggi, R. Genuer, and J.-M. Poggi, *Random forests*. Springer, 2020.

[163] S. Suryadevara and A. K. Y. Yanamala, “Fundamentals of Artificial Neural Networks: Applications in Neuroscientific Research,” *Revista de Inteligencia Artificial en Medicina*, vol. 11, no. 1, pp. 38–54, 2020.

[164] K. A. Bhavsar, A. Abugabah, J. Singla, A. A. AlZubi, and A. K. Bashir, “A comprehensive review on medical diagnosis using machine learning,” *Computers, Materials and Continua*, vol. 67, no. 2, p. 1997, 2021.

[165] H. Chang, Y. Chen, and J. Du, “Obstructive sleep apnea treatment in adults,” *Kaohsiung J Med Sci*, vol. 36, no. 1, pp. 7–12, 2020.

[166] J. Patel, J. Baldwin, P. Bunting, and S. Laha, “The effect of a multicomponent multidisciplinary bundle of interventions on sleep and delirium in medical and surgical intensive care patients,” *Anaesthesia*, vol. 69, no. 6, pp. 540–549, 2014.

[167] A. Malhotra *et al.*, “Metrics of sleep apnea severity: beyond the apnea-hypopnea index,” *Sleep*, vol. 44, no. 7, p. zsab030, 2021.

[168] J. White, C. Cates, and J. Wright, “Continuous positive airways pressure for obstructive sleep apnoea.,” *Cochrane Database Syst Rev*, no. 2, pp. CD001106–CD001106, 2002.

[169] S. Craig *et al.*, “Effect of CPAP on cardiac function in minimally symptomatic patients with OSA: results from a subset of the MOSAIC randomized trial,” *Journal of Clinical Sleep Medicine*, vol. 11, no. 9, pp. 967–973, 2015.

[170] C. A. Kushida *et al.*, “Effects of continuous positive airway pressure on neurocognitive function in obstructive sleep apnea patients: the Apnea Positive Pressure Long-term Efficacy Study (APPLES),” *Sleep*, vol. 35, no. 12, pp. 1593–1602, 2012.

[171] S. ok Hong, Y.-F. Chen, J. Jung, Y.-D. Kwon, and S. Y. C. Liu, “Hypoglossal nerve stimulation for treatment of obstructive sleep apnea (OSA): a primer for oral and maxillofacial surgeons,” *Maxillofac Plast Reconstr Surg*, vol. 39, pp. 1–5, 2017.

[172] M. Barnes *et al.*, “Efficacy of positive airway pressure and oral appliance in mild to moderate obstructive sleep apnea,” *Am J Respir Crit Care Med*, vol. 170, no. 6, pp. 656–664, 2004.

[173] P. G. Catcheside, “Predictors of continuous positive airway pressure adherence,” *F1000 Med Rep*, vol. 2, 2010.

[174] R. Varghese, N. G. Adams, N. L. Slocumb, C. F. Viozzi, K. Ramar, and E. J. Olson, “Maxillomandibular advancement in the management of obstructive sleep apnea,” *Int J Otolaryngol*, vol. 2012, no. 1, p. 373025, 2012.

[175] B. W. Rotenberg, D. Murariu, and K. P. Pang, “Trends in CPAP adherence over twenty years of data collection: a flattened curve,” *Journal of Otolaryngology-Head & Neck Surgery*, vol. 45, no. 1, p. 43, 2016.

[176] M. M. Hamoda, Y. Kohzuka, and F. R. Almeida, “Oral appliances for the management of OSA: an updated review of the literature,” *Chest*, vol. 153, no. 2, pp. 544–553, 2018.

[177] D. Lavery, I. Szollosi, S. Czyniewski, F. Beer, K. McCloy, and C. Hart, “Safety and efficacy of a novel oral appliance in the treatment of obstructive sleep apnea,” *J Dent Sleep Med*, vol. 4, no. 3, pp. 57–63, 2017.

[178] J. C. Carberry, J. Amatoury, and D. J. Eckert, “Personalized management approach for OSA,” *Chest*, vol. 153, no. 3, pp. 744–755, 2018.

[179] D. P. White and S. Shafazand, “Mandibular advancement device vs CPAP in the treatment of obstructive sleep apnea: are they equally effective in short term health outcomes?,” *Journal of Clinical Sleep Medicine*, vol. 9, no. 9, pp. 971–972, 2013.

[180] G. R. Diercks *et al.*, “Hypoglossal nerve stimulation in adolescents with down syndrome and obstructive sleep apnea,” *JAMA Otolaryngology–Head & Neck Surgery*, vol. 144, no. 1, pp. 37–42, 2018.

[181] “Upper airway stimulation for obstructive sleep apnea: durability of the treatment effect at 18 months,” *Sleep*, vol. 38, no. 10, pp. 1593–1598, 2015.

[182] G. R. Diercks *et al.*, “Hypoglossal nerve stimulation in adolescents with down syndrome and obstructive sleep apnea,” *JAMA Otolaryngology–Head & Neck Surgery*, vol. 144, no. 1, pp. 37–42, 2018.

[183] A. Kumar, M. Camacho, and R. Capasso, “Quantitative assessment of an obstructive sleep apnea patient before and after tracheostomy: a case study,” *J Otol Rhinol*, vol. 2, p. 2, 2013.

[184] M. Giarda, M. Brucoli, F. Arcuri, R. Benech, A. Braghiroli, and A. Benech, “Efficacy and safety of maxillomandibular advancement in treatment of obstructive sleep apnoea syndrome,” *Acta Otorhinolaryngologica Italica*, vol. 33, no. 1, p. 43, 2013.

[185] P. Yalamanchi and P. T. Hoff, “Hypopharyngeal Surgery for OSA Patients,” *Management of Obstructive Sleep Apnea: An Evidence-Based, Multidisciplinary Textbook*, pp. 269–280, 2021.

[186] J. Cambi *et al.*, “Outcomes in single-stage multilevel surgery for obstructive sleep apnea: Transoral robotic surgery, expansion sphincter pharyngoplasty and septoplasty,” *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 15, no. 6, p. e2034, 2019.

[187] S. Zaghi *et al.*, “Maxillomandibular advancement for treatment of obstructive sleep apnea: a meta-analysis,” *JAMA Otolaryngology–Head & Neck Surgery*, vol. 142, no. 1, pp. 58–66, 2016.

[188] S. M. Werz, M. Pfeifle, F. Schrader, P. Jurgens, M. Briel, and B. Berg, “Surgery for obstructive sleep apnoea in adults,” *Cochrane Database Syst Rev*, vol. 2017, no. 9, 2017.

[189] G. D. Singh, S. E. Cress, M. K. McGuire, and R. Chandrashekhar, “Case presentation: effect of mandibular tori removal on obstructive sleep apnea parameters,” *Dialogue Can Philos Assoc*, vol. 1, pp. 22–24, 2012.

[190] K. C. Guimarães, L. F. Drager, P. R. Genta, B. F. Marcondes, and G. Lorenzi-Filho, “Effects of oropharyngeal exercises on patients with moderate obstructive sleep apnea syndrome,” *Am J Respir Crit Care Med*, vol. 179, no. 10, pp. 962–966, 2009.

[191] M. Camacho *et al.*, “Myofunctional therapy to treat obstructive sleep apnea: a systematic review and meta-analysis,” *Sleep*, vol. 38, no. 5, pp. 669–675, 2015.

[192] T. Fleury Curado *et al.*, “Chemogenetic stimulation of the hypoglossal neurons improves upper airway patency,” *Sci Rep*, vol. 7, no. 1, p. 44392, 2017.

[193] S. S. Barold, “Willem Einthoven and the birth of clinical electrocardiography a hundred years ago.,” *Card Electrophysiol Rev*, vol. 7, no. 1, pp. 99–104, Jan. 2003, doi: 10.1023/a:1023667812925.

[194] E. A. Ashley, E. Ashley, and J. Niebauer, “Cardiology explained,” 2004.

[195] Vetronic, “Understanding the terminology used when ECG Monitoring,” *Understanding the terminology used when ECG Monitoring*.

[196] H. Hilmisson, N. Lange, and S. P. Duntley, “Sleep apnea detection: accuracy of using automated ECG analysis compared to manually scored polysomnography (apnea hypopnea index),” *Sleep and Breathing*, vol. 23, pp. 125–133, 2019.

[197] S. Ishaque, N. Khan, and S. Krishnan, “Trends in heart-rate variability signal analysis,” *Front Digit Health*, vol. 3, p. 639444, 2021.

[198] D. Urbanik *et al.*, “Obstructive sleep apnea as a predictor of reduced heart rate variability,” *Sleep Med*, vol. 54, pp. 8–15, 2019.

[199] D. S. Desai and S. Hajouli, “Arrhythmias,” in *StatPearls [Internet]*, StatPearls Publishing, 2023.

[200] Y. Sattar and L. Chhabra, “Electrocardiogram,” in *StatPearls [Internet]*, StatPearls Publishing, 2023.

[201] Anthony Atkielski, “Basic ECG waveforms, intervals and segments.” Accessed: Oct. 16, 2023. [Online]. Available: https://www.wikilectures.eu/w/Electrocardiogram#/media/File:ECG-PQRST+popis.svg

[202] A. Sheta *et al.*, “Diagnosis of obstructive sleep apnea from ECG signals using machine learning and deep learning classifiers,” *Applied Sciences*, vol. 11, no. 14, p. 6622, 2021.

[203] S. H. Jambukia, V. K. Dabhi, and H. B. Prajapati, “Classification of ECG signals using machine learning techniques: A survey,” in *2015 International Conference on Advances in Computer Engineering and Applications*, IEEE, 2015, pp. 714–721.

[204] V. Dunjko and H. J. Briegel, “Machine learning & artificial intelligence in the quantum domain: a review of recent progress,” *Reports on Progress in Physics*, vol. 81, no. 7, p. 74001, 2018.

[205] J. M. Bishop, “Artificial intelligence is stupid and causal reasoning will not fix it,” *Front Psychol*, vol. 11, p. 2603, 2021.

[206] M. Wilson, M. Mugnai, and H. van Ruler, *Leibniz: Dissertation on Combinatorial Art*. Oxford University Press, USA, 2020.

[207] J. Mittelstrass, “The Philosopher’s conception of Mathesis Universalis from Descartes to Leibniz,” *Ann Sci*, vol. 36, no. 6, pp. 593–610, Nov. 1979, doi: 10.1080/00033797900200401.

[208] A. G. Bromley, “Charles babbage’s analytical engine, 1838,” *Ann Hist Comput*, vol. 4, no. 3, pp. 196–217, 1982.

[209] C. E. Shannon, “A universal Turing machine with two internal states,” *Automata studies*, vol. 34, pp. 157–165, 1956.

[210] R. Rojas, “Konrad Zuse’s legacy: the architecture of the Z1 and Z3,” *IEEE Annals of the History of Computing*, vol. 19, no. 2, pp. 5–16, 1997.

[211] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine learning: An artificial intelligence approach*. Springer Science & Business Media, 2013.

[212] J. G. Carbonell, R. S. Michalski, and T. M. Mitchell, “An overview of machine learning,” *Mach Learn*, pp. 3–23, 1983.

[213] T. O. Ayodele, “Machine learning overview,” *New Advances in Machine Learning*, vol. 2, pp. 9–18, 2010.

[214] B. J. Copeland, “The Turing Test*,” *Minds Mach (Dordr)*, vol. 10, no. 4, pp. 519–539, 2000, doi: 10.1023/A:1011285919106.

[215] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *Bull Math Biophys*, vol. 5, no. 4, pp. 115–133, 1943, doi: 10.1007/BF02478259.

[216] D. O. Hebb, *The organization of behavior: A neuropsychological theory*. Psychology press, 2005.

[217] S. C. Shapiro, “The Turing Test and the economist,” *ACM SIGART Bulletin*, vol. 3, no. 4, pp. 10–11, 1992.

[218] A. L. Samuel, “Some studies in machine learning using the game of checkers,” *IBM J Res Dev*, vol. 3, no. 3, pp. 210–229, 1959.

[219] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain,” *Psychol Rev*, vol. 65, no. 6, pp. 386–408, 1958, doi: 10.1037/h0042519.

[220] J. Orbach, “Principles of neurodynamics. Perceptrons and the theory of brain mechanisms.,” *Arch Gen Psychiatry*, vol. 7, no. 3, pp. 218–219, 1962.

[221] M. Minsky and S. Papert, “An introduction to computational geometry,” *Cambridge tiass., HIT*, vol. 479, no. 480, p. 104, 1969.

[222] L. E. Peterson, “K-nearest neighbor,” *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.

[223] E. Fix and J. L. Hodges, “Discriminatory analysis: Nonparametric discrimination: Small sample performance,” 1952.

[224] B. W. Silverman and M. C. Jones, “E. fix and jl hodges (1951): An important contribution to nonparametric discriminant analysis and density estimation: Commentary on fix and hodges (1951),” *International Statistical Review/Revue Internationale de Statistique*, pp. 233–238, 1989.

[225] S. Hochreiter and J. Schmidhuber, “Flat minima,” *Neural Comput*, vol. 9, no. 1, pp. 1–42, 1997.

[226] O. Chapelle, P. Haffner, and V. N. Vapnik, “Support vector machines for histogram-based image classification,” *IEEE Trans Neural Netw*, vol. 10, no. 5, pp. 1055–1064, 1999.

[227] R. E. Schapire and Y. Singer, “Improved boosting algorithms using confidence-rated predictions,” in *Proceedings of the eleventh annual conference on Computational learning theory*, 1998, pp. 80–91.

[228] O. Sagi and L. Rokach, “Ensemble learning: A survey,” *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 8, no. 4, p. e1249, 2018.

[229] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning.,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[230] A. Adadi, “A survey on data-efficient algorithms in big data era. J. Big Data (2021).”

[231] I. H. Sarker, “Machine learning: Algorithms, real-world applications and research directions,” *SN Comput Sci*, vol. 2, no. 3, p. 160, 2021.

[232] M. F. Dixon, I. Halperin, and P. Bilokon, *Machine learning in finance*, vol. 1170. Springer, 2020.

[233] A. Alanazi, “Using machine learning for healthcare challenges and opportunities,” *Inform Med Unlocked*, vol. 30, p. 100924, 2022.

[234] N. Anantrasirichai and D. Bull, “Artificial intelligence in the creative industries: a review,” *Artif Intell Rev*, pp. 1–68, 2022.

[235] S. P. Adam, S.-A. N. Alexandropoulos, P. M. Pardalos, and M. N. Vrahatis, “No free lunch theorem: A review,” *Approximation and optimization: Algorithms, complexity and applications*, pp. 57–82, 2019.

[236] G. Bonaccorso, *Machine learning algorithms*. Packt Publishing Ltd, 2017.

[237] J. R. Rabunal and J. Dorado, *Artificial neural networks in real-life applications*. IGI Global, 2006.

[238] G. Dreyfus, *Neural networks: methodology and applications*. Springer Science & Business Media, 2005.

[239] M. I. Jordan and T. M. Mitchell, “Machine learning: Trends, perspectives, and prospects,” *Science (1979)*, vol. 349, no. 6245, pp. 255–260, 2015.

[240] W. Z. T. Tareq and M. Davud, “Classification and clustering,” in *Decision-Making Models*, Elsevier, 2024, pp. 351–359.

[241] M. A. Nielsen, *Neural networks and deep learning*, vol. 25. Determination press San Francisco, CA, USA, 2015.

[242] S. Haykin, *Neural networks and learning machines, 3/E*. Pearson Education India, 2009.

[243] R. Nagyfi, “The differences between artificial and biological neural networks,” *Towards Data Science [online]. Sep*, vol. 4, 2018.

[244] D. J. Livingstone, *Artificial neural networks: methods and applications*. Springer, 2008.

[245] R. Hecht-Nielsen, “Theory of the backpropagation neural network,” in *Neural networks for perception*, Elsevier, 1992, pp. 65–93.

[246] J. Karhunen, T. Raiko, and K. Cho, “Unsupervised deep learning: A short review,” *Advances in independent component analysis and learning machines*, pp. 125–142, 2015.

[247] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Adv Neural Inf Process Syst*, vol. 25, 2012.

[248] T. Guo, J. Dong, H. Li, and Y. Gao, “Simple convolutional neural network on image classification,” in *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, IEEE, 2017, pp. 721–724.

[249] S. Hershey *et al.*, “CNN architectures for large-scale audio classification,” in *2017 ieee international conference on acoustics, speech and signal processing (icassp)*, IEEE, 2017, pp. 131–135.

[250] G. Mittag, B. Naderi, A. Chehadi, and S. Möller, “NISQA: A deep CNN-self-attention model for multidimensional speech quality prediction with crowdsourced datasets,” *arXiv preprint arXiv:2104.09494*, 2021.

[251] W. H. L. Pinaya, S. Vieira, R. Garcia-Dias, and A. Mechelli, “Convolutional neural networks,” in *Machine learning*, Elsevier, 2020, pp. 173–191.

[252] N. Ketkar, J. Moolayil, N. Ketkar, and J. Moolayil, “Convolutional neural networks,” *Deep learning with Python: learn best practices of deep learning models with PyTorch*, pp. 197–242, 2021.

[253] D. Li, J. Zhang, Q. Zhang, and X. Wei, “Classification of ECG signals based on 1D convolution neural network,” in *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, IEEE, 2017, pp. 1–6.

[254] Ö. Yıldırım, U. B. Baloglu, and U. R. Acharya, “A deep convolutional neural network model for automated identification of abnormal EEG signals,” *Neural Comput Appl*, vol. 32, pp. 15857–15868, 2020.

[255] K. K. Lella and A. Pja, “Automatic COVID-19 disease diagnosis using 1D convolutional neural network and augmentation with human respiratory sound based on parameters: Cough, breath, and voice,” *AIMS Public Health*, vol. 8, no. 2, p. 240, 2021.

[256] J. Yoo *et al.*, “Residual one-dimensional convolutional neural network for neuromuscular disorder classification from needle electromyography signals with explainability,” *Comput Methods Programs Biomed*, vol. 226, p. 107079, 2022.

[257] D. Moitra and R. K. Mandal, “Classification of non-small cell lung cancer using one-dimensional convolutional neural network,” *Expert Syst Appl*, vol. 159, p. 113564, 2020.

[258] Uk.mathworks.com, “Analysis of CNN-based speech recognition system using raw speech as input,” Towards Data Science [online]. Sep. Accessed: Oct. 02, 2023. [Online]. Available: https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939

[259] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, R. Tibshirani, and J. Friedman, “Random forests,” *The elements of statistical learning: Data mining, inference, and prediction*, pp. 587–604, 2009.

[260] A. Liaw and M. Wiener, “Classification and regression by randomForest,” *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[261] M. Madaan, A. Kumar, C. Keshri, R. Jain, and P. Nagrath, “Loan default prediction using decision trees and random forest: A comparative study,” in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 2021, p. 12042.

[262] C. Liu, Y. Chan, S. H. Alam Kazmi, and H. Fu, “Financial fraud detection model: Based on random forest,” *Int J Econ Finance*, vol. 7, no. 7, 2015.

[263] P. Song and Y. Liu, “An XGBoost algorithm for predicting purchasing behaviour on E-commerce platforms,” *Tehnički vjesnik*, vol. 27, no. 5, pp. 1467–1471, 2020.

[264] M. Ferdous, J. Debnath, and N. R. Chakraborty, “Machine learning algorithms in healthcare: A literature survey,” in *2020 11th International conference on computing, communication and networking technologies (ICCCNT)*, IEEE, 2020, pp. 1–6.

[265] P. Mekha and N. Teeyasuksaet, “Image classification of rice leaf diseases using random forest algorithm,” in *2021 joint international conference on digital arts, media and technology with ECTI northern section conference on electrical, electronics, computer and telecommunication engineering*, IEEE, 2021, pp. 165–169.

[266] S. Du, X. Wang, C.-C. Feng, and X. Zhang, “Classifying natural-language spatial relation terms with random forest algorithm,” *International Journal of Geographical Information Science*, vol. 31, no. 3, pp. 542–568, 2017.

[267] A. P. Razi, Z. Einalou, and M. Manthouri, “Sleep apnea classification using random forest via ECG,” *Sleep Vigil*, vol. 5, pp. 141–146, 2021.

[268] L. Breiman, “Random Forests,” *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[269] J. Egbert and L. Plonsky, “Bootstrapping techniques,” in *A practical handbook of corpus linguistics*, Springer, 2021, pp. 593–610.

[270] M. Kaur, “An Approach for Sentiment Analysis Using Gini Index with Random Forest Classification BT - Computational Vision and Bio-Inspired Computing,” S. Smys, J. M. R. S. Tavares, V. E. Balas, and A. M. Iliyasu, Eds., Cham: Springer International Publishing, 2020, pp. 541–554.

[271] Spotfire, “A typical Random Forest.” Accessed: Oct. 16, 2023. [Online]. Available: https://www.spotfire.com/glossary/what-is-a-random-forest

[272] Y. Xu, S. Zomer, and R. G. Brereton, “Support vector machines: a recent method for classification in chemometrics,” *Crit Rev Anal Chem*, vol. 36, no. 3–4, pp. 177–188, 2006.

[273] S. Ramaswamy *et al.*, “Multiclass cancer diagnosis using tumor gene expression signatures,” *Proceedings of the National Academy of Sciences*, vol. 98, no. 26, pp. 15149–15154, 2001.

[274] M. S. Manerkar, U. Snekhalatha, S. Harsh, J. Saxena, S. P. Sarma, and M. Anburajan, “Automated skin disease segmentation and classification using multi-class SVM classifier,” 2016.

[275] Y. Jiang *et al.*, “Immunomarker support vector machine classifier for prediction of gastric cancer survival and adjuvant chemotherapeutic benefit,” *Clinical Cancer Research*, vol. 24, no. 22, pp. 5574–5584, 2018.

[276] R. Rodríguez-Pérez and J. Bajorath, “Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery,” *J Comput Aided Mol Des*, vol. 36, no. 5, pp. 355–362, 2022.

[277] A. Ben-Hur and J. Weston, “A user’s guide to support vector machines,” *Data mining techniques for the life sciences*, pp. 223–239, 2010.

[278] F. Mendonca, S. S. Mostafa, A. G. Ravelo-Garcia, F. Morgado-Dias, and T. Penzel, “A review of obstructive sleep apnea detection approaches,” *IEEE J Biomed Health Inform*, vol. 23, no. 2, pp. 825–837, 2018.

[279] C. Song, K. Liu, X. Zhang, L. Chen, and X. Xian, “An obstructive sleep apnea detection approach using a discriminative hidden Markov model from ECG signals,” *IEEE Trans Biomed Eng*, vol. 63, no. 7, pp. 1532–1542, 2015.

[280] C. M. Travieso, J. B. Alonso, M. del Pozo, J. R. Ticay, and G. Castellanos-Dominguez, “Building a Cepstrum-HMM kernel for Apnea identification,” *Neurocomputing*, vol. 132, pp. 159–165, 2014, doi: https://doi.org/10.1016/j.neucom.2013.04.048.

[281] A. I. Sharaf, “Sleep Apnea Detection Using Wavelet Scattering Transformation and Random Forest Classifier,” *Entropy*, vol. 25, no. 3, p. 399, 2023.

[282] A. P. Razi, Z. Einalou, and M. Manthouri, “Sleep apnea classification using random forest via ECG,” *Sleep Vigil*, vol. 5, pp. 141–146, 2021.

[283] A. Jezzini, M. Ayache, L. Elkhansa, and Z. al abidin Ibrahim, “ECG classification for sleep apnea detection,” in *2015 international conference on advances in biomedical engineering (ICABME)*, IEEE, 2015, pp. 301–304.

[284] J. Zhu, A. Zhou, Q. Gong, Y. Zhou, J. Huang, and Z. Chen, “Detection of Sleep Apnea from Electrocardiogram and Pulse Oximetry Signals Using Random Forest,” 2022. doi: 10.3390/app12094218.

[285] E. Tuncer, “Development of New Machine Learning Based Algorithm for the Diagnosis of Obstructive Sleep Apnea from ECG Data,” *Journal of Computer Science Research*, vol. 5, no. 3, pp. 15–21, Jul. 2023, doi: 10.30564/JCSR.V5I3.5762.

[286] X. Zhao *et al.*, “Classification of sleep apnea based on EEG sub-band signal characteristics,” *Sci Rep*, vol. 11, no. 1, p. 5824, 2021.

[287] A. R. Hassan and Md. A. Haque, “An expert system for automated identification of obstructive sleep apnea from single-lead ECG using random under sampling boosting,” *Neurocomputing*, vol. 235, pp. 122–130, 2017, doi: https://doi.org/10.1016/j.neucom.2016.12.062.

[288] Y. Liu, Y. Feng, Y. Li, W. Xu, X. Wang, and D. Han, “Automatic classification of the obstruction site in obstructive sleep apnea based on snoring sounds,” *Am J Otolaryngol*, vol. 43, no. 6, p. 103584, 2022, doi: https://doi.org/10.1016/j.amjoto.2022.103584.

[289] H. Almutairi, G. M. Hassan, and A. Datta, “Classification of Obstructive Sleep Apnoea from single-lead ECG signals using convolutional neural and Long Short Term Memory networks,” *Biomed Signal Process Control*, vol. 69, p. 102906, 2021.

[290] K. Cao and X. Lv, “Multi-task feature fusion network for Obstructive Sleep Apnea detection using single-lead ECG signal,” *Measurement*, vol. 202, p. 111787, 2022.

[291] S. H. Choi *et al.*, “Real-time apnea-hypopnea event detection during sleep by convolutional neural networks,” *Comput Biol Med*, vol. 100, pp. 123–131, 2018.

[292] Q. Yang, L. Zou, K. Wei, and G. Liu, “Obstructive sleep apnea detection from single-lead electrocardiogram signals using one-dimensional squeeze-and-excitation residual group network,” *Comput Biol Med*, vol. 140, p. 105124, 2022.

[293] H.-Y. Chang, C.-Y. Yeh, C.-T. Lee, and C.-C. Lin, “A sleep apnea detection system based on a one-dimensional deep convolution neural network model using single-lead electrocardiogram,” *Sensors*, vol. 20, no. 15, p. 4157, 2020.

[294] E. Urtnasan, J.-U. Park, E.-Y. Joo, and K.-J. Lee, “Automated Detection of Obstructive Sleep Apnea Events from a Single-Lead Electrocardiogram Using a Convolutional Neural Network.,” *J Med Syst*, vol. 42, no. 6, p. 104, Apr. 2018, doi: 10.1007/s10916-018-0963-0.

[295] S. Thompson, P. Fergus, C. Chalmers, and D. Reilly, “Detection of Obstructive Sleep Apnoea Using Features Extracted from Segmented Time-Series ECG Signals Using a One Dimensional Convolutional Neural Network,” *Proceedings of the International Joint Conference on Neural Networks*, no. April, 2020, doi: 10.1109/IJCNN48605.2020.9207470.

[296] L. Gao, L. Zhang, C. Liu, and S. Wu, “Handling imbalanced medical image data: A deep-learning-based one-class classification approach,” *Artif Intell Med*, vol. 108, p. 101935, 2020.

[297] M. Greener, “The good, the bad and the ugly red tape of biomedical research: How could regulators lower bureaucratic hurdles in clinical research without compromising the safety of patients?,” *EMBO Rep*, vol. 10, no. 1, pp. 17–20, 2009.

[298] J.-J. Zhu, M. Yang, and Z. J. Ren, “Machine learning in environmental research: common pitfalls and best practices,” *Environ Sci Technol*, 2023.

[299] K. Liu, G. Kang, N. Zhang, and B. Hou, “Breast Cancer Classification Based on Fully-Connected Layer First Convolutional Neural Networks,” *IEEE Access*, vol. 6, pp. 23722–23732, 2018, doi: 10.1109/ACCESS.2018.2817593.

[300] V. Christlein, L. Spranger, M. Seuret, A. Nicolaou, P. Král, and A. Maier, “Deep Generalized Max Pooling,” in *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 2019, pp. 1090–1096. doi: 10.1109/ICDAR.2019.00177.

[301] “Neural Networks Pooling Layers_Accessed 15082023.”

[302] A. D. Rasamoelina, F. Adjailia, and P. Sinčák, “A Review of Activation Function for Artificial Neural Network,” in *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, 2020, pp. 281–286. doi: 10.1109/SAMI48414.2020.9108717.

[303] A. Wang, G. Wan, Z. Cheng, and S. Li, “An incremental extremely random forest classifier for online learning and tracking,” in *2009 16th IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 1449–1452. doi: 10.1109/ICIP.2009.5414559.

[304] M. Claesen, F. De Smet, J. Suykens, and B. De Moor, “Fast prediction with SVM models containing RBF kernels,” Mar. 2014.

[305] K. Chu, “An introduction to sensitivity, specificity, predictive values and likelihood ratios,” *Emerg Med (N Y)*, vol. 11, no. 3, pp. 175–181, Sep. 1999, doi: https://doi.org/10.1046/j.1442-2026.1999.00041.x.

[306] D. Chavarría-Bolaños, L. Rodríguez-Wong, D. Noguera-González, V. Esparza-Villalpando, M. Montero-Aguilar, and A. Pozos-Guillén, “Sensitivity, Specificity, Predictive Values, and Accuracy of Three Diagnostic Tests to Predict Inferior Alveolar Nerve Blockade Failure in Symptomatic Irreversible Pulpitis,” *Pain Res Manag*, vol. 2017, p. 3108940, 2017, doi: 10.1155/2017/3108940.

[307] M. Grandini, E. Bagli, and G. Visani, “Metrics for Multi-Class Classification: an Overview,” pp. 1–17, 2020.

[308] M. Shafiei Neyestanak, H. Jahani, M. Khodarahmi, J. Zahiri, and M. S. Yekaninejad, “A Quantitative Comparison between Focal Loss and Binary Cross-Entropy Loss in Brain Tumor Auto-Segmentation Using U-Net”.