IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 30 November 2023, accepted 20 December 2023, date of publication 25 December 2023,
date of current version 4 January 2024.

*Digital Object Identifier 10.1109/ACCESS.2023.3346689*

RESEARCH ARTICLE

# Detection of Obstructive Sleep Apnoea Using Features Extracted From Segmented Time-Series ECG Signals With a One Dimensional Convolutional Neural Network

**STEVEN THOMPSON, DENIS REILLY, PAUL FERGUS, AND CARL CHALMERS**
Department of Computer Science, Liverpool John Moores University, L3 3AF Liverpool, U.K.
Corresponding author: Steven Thompson (s.r.thompson@ljmu.ac.uk)

**ABSTRACT** This paper reports on ongoing research, which aims to prove that features of Obstructed Sleep Apnoea (OSA) can be automatically identified from single-lead electrocardiogram (ECG) signals using a One-Dimensional Convolutional Neural Network (1DCNN) model. The 1DCNN is also compared against other machine learning (ML) classifier models, namely Support Vector Machine (SVM) and Random Forest Classifier (RFC). The 1DCNN architecture consists of 4 major parts, a Convolutional Layer, a Flattened Dense Layer, a Max Pooling Layer and a Fully Connected Multilayer Perceptron (MLP), with 1 Hidden Layer and a SoftMax output. The model repeatedly learns how to better extract prominent features from one-dimensional data and map it to the MLP for increased prediction. Training and validation are achieved using pre-processed time-series ECG signals captured from 35 ECG recordings. Using our unique windowing strategy, the data is shaped into 5 datasets of different window sizes. A total of 15 models (5 for each group, 1DCNNs, RFCs, SVMs) were evaluated using various metrics, with each being run over numerous experiments. Results show the 1DCNN-500 model delivered the greatest degree of accuracy and rapidity in comparison to the best producing RFC and SVM classifiers. 1DCNN-500 (Sensitivity 0.9743, Specificity 0.9708, Accuracy 0.9699); RFC-500 (Sensitivity/Recall (0) 0.90 / (1) 0.94, Precision (0) 0.94 / (1) 0.90, Accuracy 0.91); SVM-500 (Sensitivity (0) 0.94 / (1) 0.50, Precision (0) 0.65 / (1) 0.90, Accuracy 0.72). The model presents a novel approach that could provide support mechanisms in clinical practice to promptly diagnose patients suffering from OSA.

**INDEX TERMS** Apnoea–Hypopnoea index (AHI), electrocardiography (ECG), obstructed sleep Apnoea (OSA), one dimensional convolutional neural network (1DCNN), machine learning (ML), deep learning (DL), polysomnography (PSG), random forest classifier (RFC), support vector machine (SVM).

## I. INTRODUCTION

Obstructed Sleep Apnoea (OSA) is a sleep disorder that affects the breathing as you sleep. Severe apnoea sufferers can have up to 600 episodes of apnoea per night, with each episode lasting up to 40 seconds [1]. There is a range of symptoms that can indicate the presence of OSA, which include: chronic snoring, insomnia, gasping and breath holding, unrefreshing sleep and daytime sleepiness [2]. OSA is a common condition, with many estimates showing it is currently affecting approximately 1.5 billion people worldwide [3], although it is proven to be more prevalent amongst the age group 30 to 60 years. The Apnoea–Hypopnoea Index (AHI) is used to indicate the severity of OSA with an AHI value $<5$ classed as Normal, Mild AHI $\geq 5$, but $< 15$ per hour, Moderate AHI $\geq 15$, but $< 30$ per hour,

The associate editor coordinating the review of this manuscript and approving it for publication was Wai-Keung Fung.

with Severe AHI $\geq$ 30 per hour [3]. Estimates have shown that OSA affects 20% of the general population, where AHI is $\geq$5 [4], [5].

The effects of OSA can range from minor issues, such as daytime fatigue and tiredness to more life-threatening issues that include, heart failure and strokes. This not only puts a strain on health services, but also on the global economy, and it is estimated that direct and indirect costs of OSA, such as health care costs, accidents, decreased productivity and sickness reach into the billions annually [6]. One of the biggest challenges to OSA is the correct diagnosis of the condition.

The diagnosis of OSA dates back over a century when in 1913 French scientist Henri Pieron examined the physiological impact of the sleep disorder. Since then, many giant steps and major advances have been made in the diagnosis of OSA, the development of sleep societies, organisations and bodies have been formed and there are now thousands of accredited sleep specialists and hundreds of clinical sleep laboratories worldwide. However, there is still much evidence that shows the diagnosis of OSA is still not speedy or precise enough to keep up with demand. It is suggested that many OSA sufferers go undiagnosed, with estimates showing that over 80% of OSA patients also remain incorrectly diagnosed [5], [6]. Consequently, OSA represents a major public health concern and left untreated can lead to numerous negative health-related consequences and in some cases mortality [7], [8].

A major factor to this problem of diagnosis is through the many drawbacks and limitations of the traditional and existing diagnostic techniques and systems, which range from simple form filling and information gathering, such as the Epworth Sleepiness Scale (ESS) [9], Berlin [10] and STOP-Bang Questionnaires [11], to physical examinations and overnight clinical stays and high-tech monitoring systems, such as Polysomnography (PSG). All of which can be either time-consuming, expensive, complex and intrusive, often meaning OSA sufferers don't get adequate treatment in good time and sometimes never.

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 are lightweight, portable and in some cases, wearable, which means a reduction in physiological sensors and clinicians than that required for a standard PSGs [12]. Since these systems are better accessible, they are now used for first line diagnosis of OSA, which has 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 [13], this is primarily since HSATs find it difficult to compute the Respiratory Event Index (REI) [14].

In more recent years, the introduction of machine learning systems began to emerge, these intelligent machines brought a whole new approach to the diagnosis of OSA. Many studies show this approach not only improved diagnosis, but just as importantly, it brought a reduction in the required equipment, time and costs [15], [16]. Nevertheless, akin to the more traditional diagnosis systems, this approach also has its drawbacks, chiefly, the required domain expertise and consumed time.

This study presents a novel system that builds on recent advancements in the field of machine learning. Using deep learning neural networks for the automatic and early detection of OSA, could provide mechanisms in clinical practice to help diagnose patients suffering from OSA. This study also presents the results and findings from alternative ML classifier models, namely RFC and SVM, when compared against the IDCNN model.

Unique contributions of the paper include:

- Deliver a support apparatus to diagnose patients suffering from OSA using a 1DCNN deep learning model to overcome the requirement to manually extract features.
- Introduce 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
- Comparison of different classifiers (1DCNN, SVM, RFC) when utilising PhysioNet (Apnoea-ECG) dataset. The 1DCNN and the automated feature extraction associated with deep learning models proves significantly better than traditional machine learning models.

The remainder of this paper is structured as follows. Section II describes related work, Section III presents the methodology, which includes details of the data, test subject trial and system components. Section IV presents the experimental results. Section V provides a discussion of the results and Section VI concludes this paper, including any future work.

## II. RELATED WORK

Over the past 20yrs, various physiological signals that include, ECG, blood oxygen saturation (SpO2) and snoring have been regularly used by supervised machine learning algorithms for the detection of OSA. Such algorithms that include; Random Forests (RF) [17], Support Vector Machines (SVM) [18], K-nearest neighbor (KNN) [19], Naïve-Bayesian (NB) [20], Artificial Neural Networks (ANN) [21] and Convolutional Neural Networks (CNN) [22].

The following section looks at previous related studies, describing how some of these supervised ML algorithms have performed when used to classify a condition through the use of single-lead electrocardiogram (ECG) time-series data.

### A. MACHINE LEARNING METHODS USING ECG TIME-SERIES DATA

In [23], the authors endeavour was to automatically diagnose Obstructed Sleep Apnoea using a novel approach, which was based on the transformation of the Cepstral domain using the statistical model, Hidden Markov Model (HMM) with SVM classifiers. Results suggested that Cepstrum and HM model classifier were not enough, therefore their next approach was to use the HHM kerne, whilst also introducing an SVM model to classify the data. This approach showed great improvements with excellent results for accuracy.

The approach in [24] was to develop a stacked SAE (sparse auto-encoder) based deep neural network (DNN) combined with Hidden Markov model (HMM) using classifiers SVM and ANN. The author realised that combining HMM and DNN, with a Confidence Score-based Decision Fusion method, improved both the classification accuracy and classification performance, along with a discriminating balance of sensitivity and specificity. They also found that further classification accuracy was achieved with the addition of an extra hidden layer, where 2 hidden layers empirically provided their best results.

In [25] the authors used the signal processing technique Tunable-Q Factor Wavelet Transform (TQWT) to extract specific apnoea features from sample ECG signals. OSA Classification was then demonstrated using Random Under Sampling Boosting (RUSBoost). Using this method provided well balanced results. Further to this, an evaluation of RUSBoost proved its superiority when compared to eight commonly used classifiers, being; extreme learning machine (ELM), Prazen's probabilistic neural network (Prazen PNN), bootstrap aggregating (Bagging), k-Nearest Neighbors (kNN), support vector machine (SVM), least-square SVM (LS-SVM), random forest (RF), and adaptive boosting (AdaBoost).

Reference [26] considered a more state-of-the-art approach. Here, they developed a 1DCNN model consisting of an architecture that included, a rectified linear unit (ReLU) activation, a max pooling and dropout layers. All 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. Excellent results were achieved using three-layer architecture all the way up to nine-layer, 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 out performed each of these models; SVM, Fuzzy reasoning module, LDA, QDA AdaBoost, Bagging REPTree and Kernel density classifier.

In [27] detection of sleep apnoea (SA) was achieved using a multimodal approach that included the combined-channel feature analysis of ECG and SpO2. Classification was performed using RFC with excellent results; sensitivity 95.9, specificity 98.4 and accuracy 97.5. To better understand these results, other traditional classifiers where used, but with less success; SVM, KNN, and Linear Regression (LR). The authors found interesting results when testing both the stand-alone ECG and SpO2 signals, with SpO2 feature set providing a better accuracy, sensitivity, and specificity. However, SA results using SpO2 alone, have been shown to be imitated by other breathing conditions, namely chronic obstructive pulmonary disease or alveolar hypoventilation.

The authors in [28] used IHR (Instantaneous Heart Rate) signals for detection of SA. The experiment was performed using two network topologies, Single and Stacked LSTM (long short-term memory), with varying parameters. Training and testing were performed using various LSTM-RNN (recurrent neural network). Results ranged from very good to excellent; however, the authors did acknowledge training and testing was performed on small portions of the dataset.

Reference [29] proposed a snoring-based obstruction site detection model to identify the site of a collapse in the upper airway. They first processed the audio signal using VAD (voice activity detection) and mixed Gaussian distribution model. Features were then obtained from the audio signal using the popular method of MFCC (Mel-Frequency Cepstrum Coefficient) also known as Meyer cepstrum coefficient characteristic. They then ran 24 classification experiments using different feature vector dimensions each time across 3 separate classifiers, KNN, SVM and Gaussian NB producing satisfactory results. The KNN was seen to outperform the SVM, since the data set is much larger than the number of features and the Naive Bayes algorithm is often used in smaller feature sets with fewer outliers. The authors also claim that their model performed better than similar previous studies, since their data included the variables age, gender and BMI.

Reference [30] presents our previous study, a 1DCNN model, designed for the automated detection of OSA captured from single-lead (ECG) signals. The dataset was acquired from PhysioNet, used in this current study. The data was preprocessed into 5 exclusive datasets before being trained. The model consists of Convolutional Layer, Flattened Dense Layer, Max Pooling Layer and Fully Connected Multilayer Perceptron (MLP), with Hidden Layer and SoftMax output. The model was evaluated using various metrics. Results showed the model produced high classification, (Sensitivity 0.9705%, Specificity 0.9725%, F1_Score 0.9717%, Accuracy 0.9377%, ROCAUC 0.9945%).

Using digitised ECG signals [31] looked to compare thirteen classic ML models and four DL models for automatic detection of OSA. Preprocessing involved removing unwanted frequency noises using a digital IIR notch filter. Feature extraction codes were then applied to capture

nine specific features from ECG signals, this helps reduce the data's high dimension and improve the overall performance. The results showed the 4 DL models outperformed the 13 classical ML models, with the hybrid model CNNLSTM network producing a best performance of accuracy 86.25%, sensitivity 88.8% and AUC 95.1%, when compared to other previous studies.

In [32] a CNNLSTM hybrid model was developed to automatically detect OSA using ECG signals. Their model is split into 4 blocks. Blocks 1 & 2 consist of 1DCNN in each for feature extraction, block 3 consists of two LSTM networks for gradient vanishing problem and long-term dependency, and block 4 consists of two separate classifiers (Sigmoid and SVM). The model using the SVM classifier produced the best results. The model also achieves excellent scores of ACC 90.92%, SE 91.24% SP 90.36% F1 92.76% when compared to other studies, where they mainly used feature engineering techniques.

A 1D deep CNN model for the automatic detection of OSA using single-lead ECG signals was developed in [33]. The 1DCNN used 10 identical convolutional layers, 5 Fully-Connected layers and 4 identical classification layers. Preprocessing was achieved using Butterworth bandpass filtering and z-score normalization. Compared to several studies, this model had the best accuracy 87.9%, specificity 92.0%, sensitivity 81.1% and AUC of 94 for per-minute apnoea detection.

A final study in [34] looked to address the limitations of feature extraction using traditional ML models, 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 using a bandpass filter to find R-peak from 2-minute ECG signal segments to detect OSA. Results of segment detection showed a sensitivity 87.6%, specificity 91.9% and accuracy 90.3%.

All but two of the studies (26, 29) in this section, used the same Apnoea-ECG database, but with varying formats, techniques and methods. Further to this, some of the approaches discussed are depending on 3rd party signal processing applications to prepare their data, whilst others are using traditional ML methods with hand-crafted features, all of which can be time-consuming and expensive. Some are using LSTM and hybrid approaches, 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 other classification models discussed, our results are more than comparable, with a much simpler workflow.

## III. DATA ACQUISITION, SUBJECT INFORMATION, AND PRE-PROCESSING

This first part of this section describes the dataset (Apnoea-ECG Database) used to train and test the 3 models, the proposed 1DCNN model, the RFC model and the SVM model. It shows the value of the dataset, the information held within the dataset and how the dataset was carefully pre-processed, and feature engineered. The latter part of this section looks at the construction of the three machine learning algorithms (1DCNN, RFC, SVM), including their architectural makeup, how they were evaluated, the metrics used to gauge their performances and their produced results.

**TABLE 1. The Apnoea-ECG database properties.**

| | | |
|---|---|---|
| Subjects | 32 | 35 annotated recording |
| 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 |

### A. APNOEA-ECG DATABASE

The Apnoea-ECG database (Table 1) was acquired from the publicly renowned on-line database website, PhysioNet. Researching this database showed it has been actively used and extensively published in previous high-quality publications and journals. The Apnoea-ECG database was constructed through the observation and merger of data taken from two separate studies, in 1993 and 1999, that involved the recordings of ECG signals from patients suffering with obstructive sleep apnoea [1].

A total of 70 night-time ECG/EEG recordings were observed in the database, the 35 annotated recoding were used for this study. Breaking down the data as presented in Table 2; Group A (Apnoea-Set) contained 20 subject recordings with 6250 mins of Apnoea and 3811 mins of Normal (Non-Apnoea). Group B (Borderline-Set) had 5 subject recordings, with 252 mins of Apnoea and 2060 mins Non-Apnoea. Group C (Normal-Set) had 10 subject recordings with 12 mins of Apnoea and 4740 mins of Non-Apnoea.

**TABLE 2. Breakdown of the 35 subjects recoding.**

| Subject Recordings | EGC Files | Group Type | Apnoea Events (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 |
| | | | 6514 | 10611 |

### 1) BUILDING THE DATASETS

Each Apnoea observation was scored and annotated by an expert sleep clinician. Using a feature selection process, all unwanted characteristics were identified and removed, leaving only the required variables and features for the dataset. The next step was to cross-reference and separate all the annotated files that resided in groups A, B & C, into two sample recording groups (Apnoea and Non-Apnoea). This resulted in 650 segmented files. A total of 314 files contained Apnoea and a total of 336 files contained Non-Apnoea, as shown in Table 3.

### 2) MERGING OF SEGMENTED SAMPLE FILES STAGE I

With the segmentation of Apnoea and Non-Apnoea files fully completed (*Table 3*), the construction of the dataset could now begin. This process involved the merging together of each individual segmented sample file within their own specific code and group. Completing this task resulted in 35 newly formed files for each of the two groups (35 Apnoea files and 35 Non-Apnoea files), shown in *Table 4*.

### 3) MERGING OF SEGMENTED SAMPLE FILES STAGE II

The processes to build the dataset were to firstly merge each of the 35 newly formed sample files within their own specific section within their specific group. i.e. the 20 newly formed files in the 'A' section under the 'Apnoea' group, were merged together to form one file, likewise the same process was applied to the files in the B and C sections of their specific groups, shown in *Table 5* .

Using the same technique and keeping the separate groups (Apnoea and Non-Apnoea), the process of merging the newly formed files together continued, thus leaving two files, 1 file for Apnoea and 1 file for Non-Apnoea (*Table 6*). The final step before moving onto the windowing strategy involved two parts, one was to firstly remove any overhang, in this case ''Non-apnoea'' overhang rows where removed to match the ''Apnoea'' row size, and secondly to merge these two files together to create a balanced dataset.

**TABLE 3. Breakdown of the annotated files into two groups (Apnoea and Non-Apnoea).**

| 35 ECG Recording | Apnoea events (segmented files) | Non-Apnoea events (segmented 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 |

### 4) DATASET WINDOWING STRATEGY

At this stage, the newly built dataset was reshaped into 5 separately balanced datasets of specific window sizes; 500, 1,000, 1,500, 2,000 and 2,500. Each newly formed window

**TABLE 4.** Merging of the segmented sample files into the two groups.

| Subject Recordings | Apnoea events (merged files) | Non-Apnoea events (merged files) |
|---|---|---|
| A01 – A20 | 20 | 20 |
| B01 – B05 | 5 | 5 |
| C01 – C10 | 10 | 10 |
| Total | 35 | 35 |

**TABLE 5.** Merging of the segmented sample files into six files.

| Subject Recordings | Apnoea events (merged files) | Non-Apnoea events (merged files) |
|---|---|---|
| A01 – A20 | 1 | 1 |
| B01 – B05 | 1 | 1 |
| C01 – C10 | 1 | 1 |
| Total | 3 | 3 |

**TABLE 6.** Merging of the segmented sample files into two files.

| Subject Recordings | Apnoea events (merged files) | Non-Apnoea events (merged files) |
|---|---|---|
| A01 – C10 | 1 | 1 |
| Total | 1 | 1 |

**TABLE 7.** 5 separately balanced datasets of specific window sizes.

| Dataset | Columns (Window size) | Rows | Apnoea & Non-Apnoea Samples (balanced) |
|---|---|---|---|
| W=500 | 500 | 150,384 | 75,192,000 |
| W=1000 | 1,000 | 75,190 | 75,190,000 |
| W=1500 | 1,500 | 50,126 | 75,189,000 |
| W=2000 | 2,000 | 37,592 | 75,184,000 |
| W=2500 | 2,500 | 30,060 | 75,150,000 |

contained the same amount of Apnoea and Non-Apnoea, approx. 37 million samples per group (approx. 75 million samples). The Non-Apnoea rows, were labelled as '0', and filled the bottom half of the dataset and the Apnoea rows, labelled as '1' filled the top half of the dataset. *Table 7* provides a view of the structure for each of the newly constructed 5 balanced datasets. It shows the number of columns (window size), n/rows (samples), n/Apnoea samples and n/Non-Apnoea samples.

### B. MACHINE LEARNING ALGORITHMS

This section presents the three machine learning algorithms used in this study. It first talks about the proposed algorithm in this study, namely the 1DCNN classification model, in terms of how it was constructed, its architecture and how it works. To further evaluate the 1DCNN model, the section then describes two alternative classification models to be used for comparison experiments, namely Random Forest Classifier (RFC) and Support Vector Machine (SVM), The same datasets, training and validation were applied to all the models.

#### 1) 1DCNN ARCHITECTURE AND EVALUATION

Over the last decade Convolutional Neural Networks (CNNs) have become highly recognised and a very popular means of performing machine learning tasks. This is mainly because CNN's are decisively more powerful and accurate when compared to traditional machine learning algorithms. Similar to Artificial Neural Networks (ANNs), CNNs use the feed-forwarding technique. The most common types of CNN's are the 2 and 3-dimentional models, which have very high accuracy when dealing with complex image processing tasks. However, in recent years 'the state of the art' 1DCNN has become a desirable choice for classification tasks, particularly where time-series data is used. They are also proven to work well with one-dimensional arrays, providing excellent feature extraction capabilities, thus avoiding the need for domain expertise.

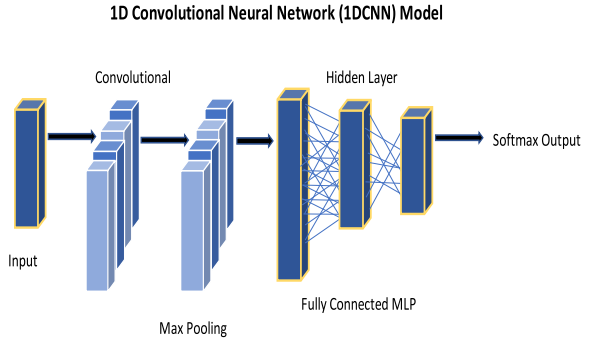


**FIGURE 1.** Architecture of the one-dimensional convolutional neural network.

Figure.1 shows the architecture of the 1DCNN. The model 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 a number of 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 1 Hidden layer, 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) hyper-parameter, 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 hyper-parameters. To further assist the convolutional process at

this stage, a Max Pooling layer is incorporated. This layer skillfully 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 to the hidden layer. The hidden layer contains the activation function (ReLU), which transforms the input before passing through to the final layer, 'softmax output'. The purpose of the softmax activation 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.

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.

Below in (1) the IDCNN layers of the convolutional function is represented by $\boldsymbol{y} = conv1d(\boldsymbol{x},\boldsymbol{w},\boldsymbol{b})$. The input to this function is denoted by $\boldsymbol{x}$, filters are shown as $\boldsymbol{w}$, bias as $\boldsymbol{b}$ and the final output of the convolutional layers is presented as $\boldsymbol{y}$ [35].

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

The *Max Pooling layer* function is a primary process within the CNN. In (2) the *Max Pooling* formula uses stride values $\boldsymbol{sx}, \boldsymbol{sy}$ and a pooling window, defined by filter $\boldsymbol{fx}, \boldsymbol{fy}$ and channel sizes $\boldsymbol{k}$. It operates by moving across the data capturing the highest valued features through input ($\boldsymbol{X}$), where the values are summed and outputted $\boldsymbol{i}, \boldsymbol{j}$. By reducing overfitting and computation, this increases the overall performance of the model and can be defined as below [36].

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

The aim of a *Fully Connected Multilayer Perception* is 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 $\boldsymbol{y}$= *fullyCon*($\boldsymbol{x},\boldsymbol{w},\boldsymbol{b}$), where $\boldsymbol{x}$ denotes the input, $\boldsymbol{w}$ = weights, $\boldsymbol{b}$ = bias and $\boldsymbol{y}$ = outputs [35].

$$yi' = \sum_{i} w_{ii'} x_i + b_{i'} \tag{3}$$

(4) shows a mathematical representation of how the *ReLU*(Rectifier) activation function 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 ($\boldsymbol{x}$) to the final layer [37].

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

In (5) the *softmax fuction* $\boldsymbol{S}$ is the final activation function of the 1DCNN. Its purpose is to improve classification output for the number of classes $\boldsymbol{n}$. This is achieved by taking an input of vector numbers $\boldsymbol{yi}$, applying an exponential function to convert these real numbers into probability sums, using normalisation to ensure each value is between 0 and 1 [35].

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

### 2) RANDOM FOREST CLASSIFIER ARCHITECTURE AND EVALUATION

The Random Forest Classifier first came into prominence approx. 20yrs ago. It is a supervised machine learning algorithm, primarily used with non-linear classification tasks. Random Forests are constructed using an ensemble of decision tree classifiers in (6) $\{\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\theta k}), \boldsymbol{k} = 1, \ldots\}$, where each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest $\{\boldsymbol{\theta k}\}$, before a vote is casted for the most popular class $x$. This vote is achieved by taking the mean number from the output of the previous tree at each stage before making a final decision [38].

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

The RFC built to train the datasets was constructed using the hyperparameters of *n_estimators*, *Random_State* and *Gini Index*. The *n_estimators* determines 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 $\boldsymbol{S}$. This is the probability $\boldsymbol{Pi}$ of the incorrectly labelling of a randomly selected class $\boldsymbol{k}$ [39]. 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 [40].

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

### 3) SUPPORT VECTOR MACHINE ARCHITECTURE AND EVALUATION

Support Vector Machine has been developed into a very robust and well-established supervised machine learning algorithm, which is primarily associated with classification tasks. Through the mathematical functionality of support kernels and the calculations of margins using plotted datapoints and hyperspace, the SVM finds the most meaningful hyperplane that enables it to separate one class from another class [41].

The SVM built to compare against the 1DCNN 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 $\{\boldsymbol{X_1}, \boldsymbol{X_2}\}$, are calculated using the decision boundary parameter $\boldsymbol{Y}$, the *RBF_kernel* $\boldsymbol{K}$, maps the input data into a high-dimensional space, thus enabling the SVM to find the best position of the hyperplane for classification [42].

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

### C. TRAINING THE 1DCNN MODEL

The model was trained using the 5 uniquely designed datasets with different window sizes (W=500 through to W=2500). This included running large volumes of detailed experiments using various numbers of layers and hyperparameters (n_Filters k_Size, Batch_Size, Epochs) to find the optimum performance of each model. Discussed in more detailed in ''Experiments and Results'' section.

### D. PERFORMANCE METRICS

Performance metrics are critical gauges to evaluating how well the 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).

#### 1) 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 | Actual Class: 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.

#### 2) 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 update the weights at each passing.

#### 3) 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 [43].

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

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

#### 4) 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.

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

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

#### 5) 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)

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

#### 6) ROC AUC (RECEIVER OPERATOR CHARACTERIC, AREA UNDER THE CURVE)

The Area under the ROC Curve (AUC) is a visual representation of a model's 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.

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

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP + TN} \tag{15}$$

#### 7) 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 (16), shown at the bottom of the page, [44].

#### 8) 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 [45].

$$\mathrm{k} = \frac{\mathrm{Po} - \mathrm{Pe}}{1 - \mathrm{Pe})} \tag{17}$$

#### 9) 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 - $p^{\wedge}$**) is the probability of class 0 [46].

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

#### 10) 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.

#### 11) 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.

## IV. EXPERIMENTS AND RESULTS

This section evaluates the effectiveness of all models by presenting their results for training and validation. The section looks at three main areas. Firstly, *Subsections A, B and C*, present the best performing model from each group (1DCNN, RFC and SVM). Following this, *Subsection D* (*Tables 11 – 13*) show the results for all 15 models. Finally, presented in *Subsection E*. (*table 14*) is a classification results comparison study of the proposed model against our previous study and also other OSA studies, discussed earlier in the *II Related Works* section. 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. A total of 15 models, 5 for each group (1DCNNs, RFCs, SVMs) were part of this experiment. 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 experiments the data was split into 72% training, 20% testing and 8% 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 *Performance Metrics*.

Tables 8 through to X and (figures 2 – 6), 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.

### A. 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 8*), along with graphical representations of accuracy, loss and ROCAUC results (Figures 2 – 4).

#### 1) 1DCNN CONFIGURATION AND RESULTS

*Table 8* 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

$$\text{Accuracy} = \frac{\textit{True Positives} + \textit{True Negatives}}{\textit{True Positives} + \textit{True Negatives} + \textit{False Positives} + \textit{False Negatives}} \tag{16}$$

**TABLE 8.** 1DCNN-500 configuration and results.

| 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 |

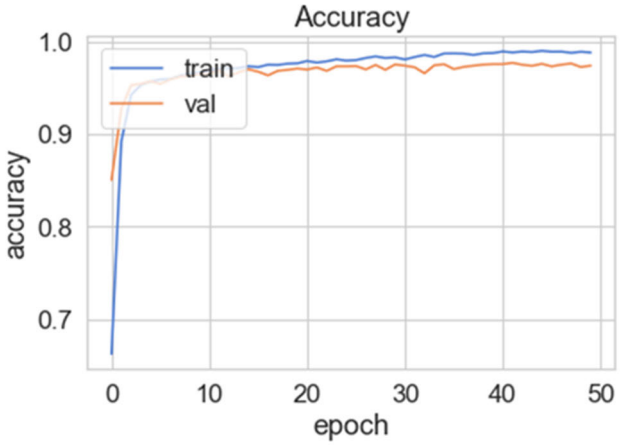


**FIGURE 2.** Graphical output results from the 1DCNN-500 model using dataset W=500. Showing training and validation accuracy.

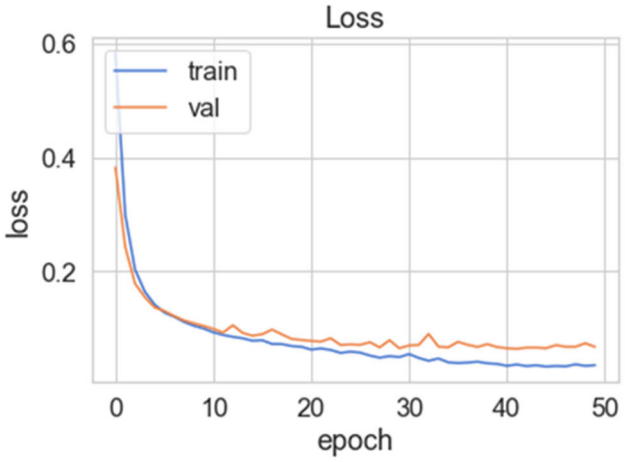


**FIGURE 3.** Graphical output results from the 1DCNN-500 model using dataset W=500. Showing training and validation loss.

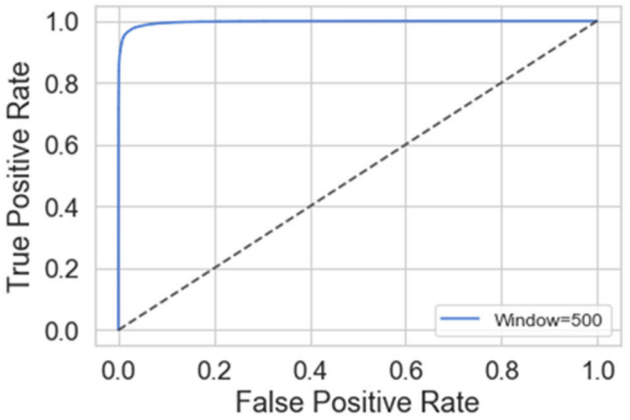


**FIGURE 4.** Graphical output results from the 1DCNN-500 model using dataset W=500. Showing ROCAUC plot.

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.

*Figures.2 – 4*, shows three images for the 1DCNN-500 model results in *Table 8* (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.

### B. RANDOM FOREST CONFIGURATION AND RESULTS

This section assesses the best performing RFC model (RFC-500) after training and validation. It shows the model's hyper-parameter configuration and Confusion Matrix (*Table 9*), along with graphical representations of ROCAUC results (Figures 5).

#### 1) RFC CONFIGURATION AND RESULTS

*Table 9*, Configuration column presents the optimal configuration for the RFC-500 model. This model was the best performing of the RFC models. Using an 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

**TABLE 9.** RFC-500 configuration & results and confusion matrix.

| 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 |
| | FN 901 14088 TP | Support | 30077 |

the duration of the experiment took to complete. Moreover, inputted values above 500 decision trees didn't show any further improvements.

*Figure.5* presents the AUC graph for the RFC-500 model results in *Table 9* . Although not as tight to the top-left hand corner as *figure.4* 1DCNN AUC, this is still a very good scoring and demonstrates this model is good at predicting between the two separate classes.

### C. SUPPORT VECTOR MACHINE CONFIGURATION AND RESULTS

This section assesses the best performing SVM model (SVM-500) after training and validation. It shows the model's hyper-parameter configuration and Confusion Matrix (*Table 10)*, along with graphical representations of ROCAUC results (Figures 6).

#### 1) SVM CONFIGURATION AND RESULTS

In *Table 10*, 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.6* presents the AUC graph for the RFC-500 model results in *Table 10*. 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.

### D. COMPLETE LIST OF RESULTS

The three tables (*Table 11, Table 12, Table 13*) 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 11*) shows the 5 1DCNN configuration hyperparameters and results, the middle table (*Table 12*), shows the 5 RFC configurations and results and the bottom table (*Table 13*),

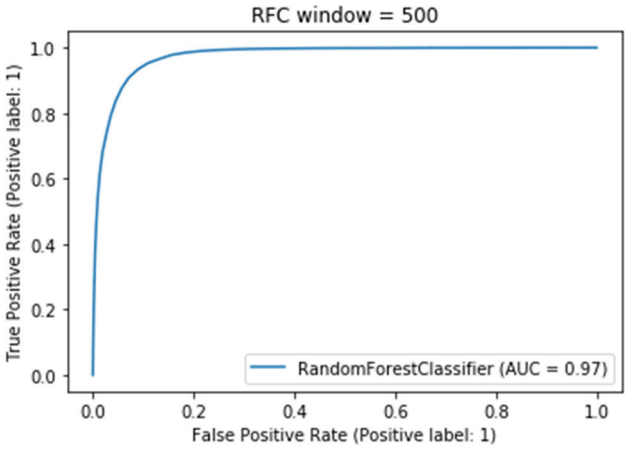


**FIGURE 5.** Graphical output results from the RFC-500 model using dataset W=500. Showing AUC plot.

**TABLE 10.** SVM-500 configuration & results and confusion matrix.

| 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 |
| | FN 7555 7509 TP | Support | 30077 |

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

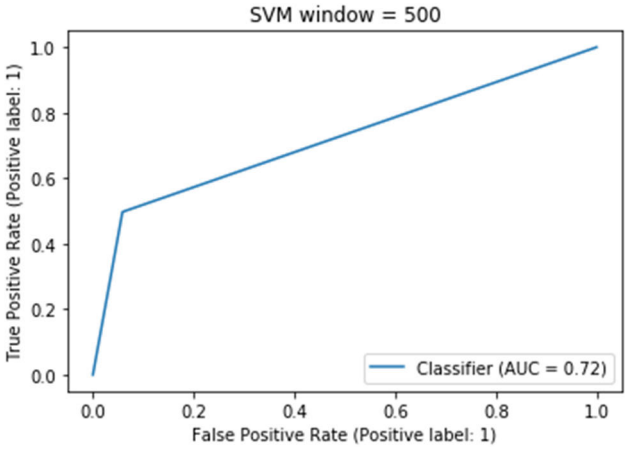


**FIGURE 6.** Graphical output results from the SVM-500 model using dataset W=500. Showing AUC plot.

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 dimension's 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 fairly 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.

### E. COMPARISON STUDY AGAINST OTHER OSA DETECTION METHODS

*Table 14* provides a comparison of results, using this proposed model against other state-of-the-art classification methods, discussed earlier in section *II Related Works*. These studies present a range of different ML and DL approaches. The results for all 1DCNN models including our current and previous models performed better in comparison to others, suggesting this is an excellent approach for detection of OSA. R. Pathinarupothi et al LSTM-RNN model does show slightly higher scores, however as previously mentioned, the authors acknowledge training and testing was performed on small portions of the Apnoea-ECG database.

## V. DISCUSSION

This study set out with two main objectives. Firstly, to evaluate the 1DCNN model and secondly, to compare it against other classification models (RFC and SVM). The 1DCNN model was constructed using the state-of-the-art techniques in 1DCNNs, consisting of a Convolutional and a Max Pooling Layer and a fully connected Multilayer Perceptron (MLP) which included a hidden layer and SoftMax output for classification. It was found that using multiple convolutional layers showed no improvement to the 1DCNN model. Using one layer decreased both complexity and load and produced excellent classification performance that empirically provided the best results.

It was decided to perform some comparison experiments against other more traditional ML algorithms, RFC and SVM, since they 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. 216hrs of segmented ECG single-lead time-series signals obtained from 35 subjects. Over 70hr of non-apnoea segments were initially removed to balance the dataset at approx. 108hrs for each group, apnoea/non-apnoea. To ensure fairness of testing, the segments were grouped into a single balanced dataset containing approx. 35 million samples of Apnoea and 35 million samples of Non-apnoea or Normal.

Changes and limitations to the acquired dataset. In the data of the original evaluation study, the scoring of apnoeas and hypopneas was done according to standard criteria, where the number of apnoeas and hypopnoeas were marked and scored separately using the values Apnoea Index (AI) and

**TABLE 11. 1DCNN experiments - configuration and results.**

| | 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 12. RFC 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 13. SVM 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 |

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. This new marking and scoring method did not differentiate between apnoea (AI) 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 apnoeas and hypopnoeas. 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) or ''disordered breathing'' (A). The total number of minutes spent in apnoeas or hypopnoeas 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 impacted the performances of each model. Results appear to coincide with window sizes, more rows with less columns generally produced increased performance, however, this increase seemed to plateau at reduction of approx. 500 columns.

Of the 3 groups of models (1DCNN, RFC, SVM) evaluated in the experiment, the 1DCNN group was shown to be the strongest group and 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.

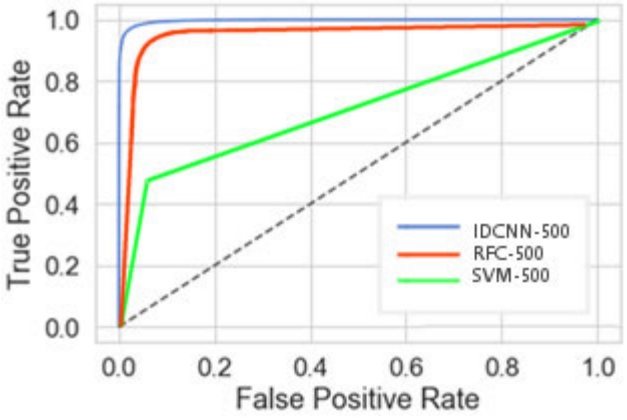


**FIGURE 7. Comparable ROCAUC plot results for the best performing model from each group, using dataset size W=500.**

Figure.7 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.

**TABLE 14. Comparison of proposed model Vs other OSA ML and DL models.**

| Model | Sens | Spec/ Prec | Acc | Author |
|---|---|---|---|---|
| HMMK+SVM | NA | NA | 99.23 | C. M. Travieso |
| DNNHMM+SVM/ANN | 88.9 | 82.1 | 84.7 | K. Li, |
| RUSBoost | 87.58 | 91.49 | 88.88 | A. R. Hassan *et al* |
| 1DCNN | 96.0 | 96.0 (Pr) | NA | E. Urtnasan, J. *et al* |
| ECG + SpO2 RFC | 95.9 | 98.4 | 97.5 | J. Zhu *et al* |
| LSTM-RNN | 99.9 | 100 | 99.9 | R. Pathinarupothi *et al* |
| GDM/KNN,SVM | NA | NA | 87.98 | Y. Liu *et al* |
| CNNLSTM | 88.8 | 86.8 (Pr) | 86.25 | A. Sheta *et al* |
| CNNLSTM | 91.24 | 90.36 | 90.92 | H. Almutairi *et al* |
| 1DCNN | 81.1 | 92.0 | 87.9 | H.-Y. Chang *et al* |
| 1D-SEResGNet | 87.6 | 91.9 | 90.3 | Q. Yang *et al* |
| W-500 1DCNN | 97.05 | 97.25 | 93.77 | Our Previous Study |
| 1DCNN-500 | 97.43 | 97.08 | 96.99 | This Model |

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 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 the 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 RCF and SVM results could be associated to the type of data used for the experiments. RCFs and SVMs are often better suited to text analysis, also in the case of SVMs, small datasets. Another point is the number of variables used in these experiments. RFC and SVM respond better to higher amounts of variables.

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.

## VI. CONCLUSION

Obstructed Sleep Apnoea is a worldwide problem that will affect 1 in every 5 people at any one time and will affect 1 in every 2 people at some stage over their lifetime. It is a condition that can develop into serious health complications, both physically and mentally and can lead to mortality. The global economic impact of OSA costs billions of pounds per annum and is forecast to continue to grow year on year. Traditional diagnosis techniques are not enough. Over 80% of patients still remain incorrectly diagnosed. In more recent years newer OSA diagnostic solutions have emerged with some success, particularly in the area of ML algorithms, however, these innovative methods require extensive domain experience and time. Over the past decade there has been various supervised machine learning algorithm developed to better diagnose certain human conditions and illness.

The approach of a 1DCNN looked to address many of these issues and it has demonstrated the capability to automatically detect instances of OSA through captured single-lead ECG signals. The study has also shown that the 1DCNN model provides greater classification accuracy, rapidity and robustness when compared to the other traditional ML algorithms.

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, it will first require some important steps, to be published in future papers;

- Further evaluation and testing using alternative ECG signal dataset (University College Dublin (UCD), Dataset)
- Attain clinical approval to better evaluate this study in a clinical setting
- Tests using real-world data captured from test subjects
- Development of a frontend system to host and interact with the 1DCNN model when classifying uploaded ECG signals

## ACKNOWLEDGMENT

The authors would like to thank all those involved in making the dataset available to the general public. The dataset for this

study was sourced from Physiobank, which is a subdivision of the publicly accessible and well-renowned on-line data exchange site, Physionet. PhysioNet is a web-based library of physiological data, accompanied by analytic software https://archive.physionet.org/physiobank/.

## REFERENCES

[1] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter, "Apnea-ECG database," *Comput. Cardiol.*, vol. 27, pp. 255–258, Feb. 2000.

[2] D. P. White, "Sleep-related breathing disorder. 2. Pathophysiology of obstructive sleep apnoea," *Thorax*, vol. 50, no. 7, pp. 797–804, Jul. 1995.

[3] 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.

[4] T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, "The occurrence of sleep-disordered breathing among middle-aged adults," *New England J. Med.*, vol. 328, no. 17, pp. 1230–1235, Apr. 1993.

[5] 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, Sep. 1997.

[6] *Hidden Health Crisis Costing America Billions: Underdiagnosing and Undertreating Obstructive Sleep Apnea Draining Healthcare System*, Amer. Acad. Sleep Med., Frost & Sullivan, San Antonio, TX, USA, 2016.

[7] N. M. Punjabi, B. S. Caffo, J. L. Goodwin, D. J. Gottlieb, A. B. Newman, G. T. O'Connor, D. M. Rapoport, S. Redline, H. E. Resnick, J. A. Robbins, E. Shahar, M. L. Unruh, and J. M. Samet, "Sleep-disordered breathing and mortality: A prospective cohort study," *PLoS Med.*, vol. 6, no. 8, 2009, Art. no. e1000132.

[8] T. Young and L. Finn, "Epidemiological insights into the public health burden of sleep disordered breathing: Sex differences in survival among sleep clinic patients," *Thorax*, vol. 53, no. Supplement 3, pp. S16–S19, Oct. 1998.

[9] 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.

[10] 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," *Ann. Internal Med.*, vol. 131, no. 7, pp. 485–491, 1999.

[11] M. W. Johns, "A new method for measuring daytime sleepiness: The Epworth sleepiness scale," *Sleep*, vol. 14, no. 6, pp. 540–545, Nov. 1991.

[12] N. A. Collop, W. M. Anderson, B. Boehlecke, D. Claman, R. Goldberg, D. J. Gottlieb, D. Hudgel, M. Sateia, and R. Schwab, "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. 737–747, 2007.

[13] L. Abrahamyan, Y. Sahakyan, S. Chung, P. Pechlivanoglou, J. Bielecki, S. M. Carcone, V. E. Rac, M. Fitzpatrick, and M. Krahn, "Diagnostic accuracy of level IV portable sleep monitors versus polysomnography for obstructive sleep apnea: A systematic review and meta-analysis," *Sleep Breathing*, vol. 22, no. 3, pp. 593–611, Sep. 2018.

[14] V. K. Kapur, D. H. Auckley, S. Chowdhuri, D. C. Kuhlmann, R. Mehra, K. Ramar, and C. G. Harrod, "Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: An American academy of sleep medicine clinical practice guideline," *J. Clin. Sleep Med.*, vol. 13, no. 3, pp. 479–504, Mar. 2017.

[15] L. Almazaydeh, K. Elleithy, and M. Faezipour, "Detection of obstructive sleep apnea through ECG signal features," in *Proc. IEEE Int. Conf. Electro/Inf. Technol.*, May 2012, pp. 1–6.

[16] S. G. Jones, B. A. Riedner, R. F. Smith, F. Ferrarelli, G. Tononi, R. J. Davidson, and R. M. Benca, "Regional reductions in sleep electroencephalography power in obstructive sleep apnea: A high-density EEG study," *Sleep*, vol. 37, no. 2, pp. 399–407, Feb. 2014.

[17] A. Cutler, D. R. Cutler, and J. R. Stevens, "Random forests," in *Ensemble Machine Learning: Methods and Applications*, C. Zhang and Y. Ma, Eds. Boston, MA, USA: Springer, 2012, pp. 157–175.

[18] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul./Aug. 1998.

[19] A. Kataria and M. D. Singh, "A review of data classification using $k$-nearest neighbour algorithm," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 3, no. 6, pp. 354–360, 2013.

[20] K. M. Leung, "Naive Bayesian classifier," *Polytech. Univ. Dept. Comput. Sci./Finance Risk Eng.*, vol. 2007, pp. 123–156, Nov. 2007.

[21] J. J. Hopfield, "Artificial neural networks," *IEEE Circuits Devices Mag.*, vol. MSD-4, no. 5, pp. 3–10, Sep. 1988.

[22] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015, *arXiv:1511.08458*.

[23] 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, May 2014.

[24] K. Li, W. Pan, Y. Li, Q. Jiang, and G. Liu, "A method to detect sleep apnea based on deep neural network and hidden Markov model using single-lead ECG signal," *Neurocomputing*, vol. 294, pp. 94–101, Jun. 2018.

[25] A. R. Hassan and M. 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, Apr. 2017.

[26] 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.

[27] 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," *Appl. Sci.*, vol. 12, no. 9, p. 4218, Apr. 2022.

[28] R. Pathinarupothi, V. Ravi, E. Rangan, E. A. Gopalakrishnan, and S. Kp, "Instantaneous heart rate as a robust feature for sleep apnea severity detection using deep learning," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat.*, Feb. 2017, pp. 293–296.

[29] 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," *Amer. J. Otolaryngol.*, vol. 43, no. 6, Nov. 2022, Art. no. 103584.

[30] 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," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.

[31] A. Sheta, H. Turabieh, T. Thaher, J. Too, M. Mafarja, M. S. Hossain, and S. R. Surani, "Diagnosis of obstructive sleep apnea from ECG signals using machine learning and deep learning classifiers," *Appl. Sci.*, vol. 11, no. 14, p. 6622, Jul. 2021.

[32] 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, Aug. 2021, Art. no. 102906.

[33] 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, Jul. 2020.

[34] 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, Jan. 2022, Art. no. 105124.

[35] 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.

[36] V. Christlein, L. Spranger, M. Seuret, A. Nicolaou, P. Král, and A. Maier, "Deep generalized max pooling," in *Proc. Int. Conf. Document Anal. Recognit. (ICDAR)*, Sep. 2019, pp. 1090–1096.

[37] A. D. Rasamoelina, F. Adjailia, and P. Sincák, "A review of activation function for artificial neural network," in *Proc. IEEE 18th World Symp. Appl. Mach. Intell. Informat. (SAMI)*, Jan. 2020, pp. 281–286.

[38] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001.

[39] A. Wang, G. Wan, Z. Cheng, and S. Li, "An incremental extremely random forest classifier for online learning and tracking," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 1449–1452.

[40] M. Kaur, "An approach for sentiment analysis using Gini index with random forest classification," in *Computational Vision and Bio-Inspired Computing*. Cham, Switzerland: Springer, 2020, pp. 541–554.

[41] S. Ding, X. Hua, and J. Yu, "An overview on nonparallel hyperplane support vector machine algorithms," *Neural Comput. Appl.*, vol. 25, no. 5, pp. 975–982, Oct. 2014.

[42] M. Claesen, F. De Smet, J. Suykens, and B. De Moor, "Fast prediction with SVM models containing RBF kernels," 2014, *arXiv:1403.0736*.
[43] K. Chu, "An introduction to sensitivity, specificity, predictive values and likelihood ratios," *Emergency Med.*, vol. 11, no. 3, pp. 175–181, Sep. 1999.
[44] 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. Manage.*, vol. 2017, Jun. 2017, Art. no. 3108940.
[45] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: An overview," 2020, *arXiv:2008.05756*.
[46] M. S. Neyestanak, H. Jahani, M. Khodarahmi, J. Zahiri, M. Hosseini, and M. S. Yekaninejad, "A quantitative comparison between focal loss and binary cross-entropy loss in brain tumor auto-segmentation using U-Net," *SSRN Electron. J.*, pp. 1–19, Jul. 2022.

**STEVEN THOMPSON** received the B.Sc. degree (Hons.) in information systems and the M.Sc. degree (Hons.) in wireless and mobile computing from Liverpool John Moores University, in 2008 and 2013, respectively, where he is currently pursuing the Ph.D. degree. He is also performing part-time research. He is also a Senior Technical Officer with the Department of Computer Science, Liverpool John Moores University. His research interests include data analytics, machine learning, deep learning, and neural networks.

**DENIS REILLY** received the B.Eng. degree (Hons.) in electrical and electronic engineering and the M.Sc. degree (Hons.) in computer science and software engineering from the University of Liverpool, in 1989 and 1991, respectively, and the Ph.D. degree in distributed system and middleware from Liverpool John Moores University, in 2003. He is currently a Principal Lecturer with the Department of Computer Science, Liverpool John Moores University. His research interests include machine learning, data analytics, distributed systems and middleware, medical informatics, and cloud computing and security. In addition to his research interests, he is a Leader for the B.Sc. degree (Hons.) in computer studies and the B.Sc. degree (Hons.) in computer networks.

**PAUL FERGUS** is currently a Professor in machine learning and the Head of the Data Science Research Centre. His main research interests include machine learning for detecting and predicting preterm births. He is also interested in the detection of foetal hypoxia, electroencephalogram seizure classification and bioinformatics. He is also conducting research with Mersey Care NHS Foundation Trust looking at the use of smart meters to detect activities of daily living in people living alone with Dementia by monitoring the use of home appliances to model habitual behaviors for early intervention practices and safe independent living at home. He has competitively won external grants to support his research from HEFCE, Royal Academy of Engineering, Innovate U.K., Knowledge Transfer Partnership, North West Regional Innovation Fund, and Bupa. He has published over 200 peer-reviewed papers in these areas.

**CARL CHALMERS** is currently a Senior Lecturer with the Department of Computer Science, Liverpool John Moores University. He is also leading a three-year project on smart energy data and dementia in collaboration with Mersey Care NHS Trust to monitor and model the behavior of dementia patients and facilitate safe independent living. In addition, he is also working in the area of high-performance computing and cloud computing to support and improve existing machine learning approaches, while facilitating application integration. His main research interests include advanced metering infrastructure, smart technologies, ambient assistive living, machine learning, high performance computing, cloud computing, and data visualization. His current research interests include remote patient monitoring and ICT-based healthcare.