Proceedings Paper:
Yang, D., Zhang, J., Bhargava, E. et al. (3 more authors) (Accepted: 2025) Deep learning methods for apnoea detection based on pulse and oximetry data. In: Proceedings of the 11th IEEE International Conference on Data Science and Systems 2025 (DSS-2025). 11th IEEE International Conference on Data Science and Systems 2025 (DSS-2025), 13 Aug - 15 Jul 2025, Exeter, UK. Institute of Electrical and Electronics Engineers (IEEE) (In Press)

Deep Learning Methods for Apnoea Detection Based on Pulse and Oximetry Data

1st Dongjin Yang
School of Electrical & Electronic Eng.
University of Sheffield
Sheffield, United Kingdom
dyang34@sheffield.ac.uk

2nd Jingqiong Zhang
School of Electrical & Electronic Eng.
University of Sheffield
Sheffield, United Kingdom
jingqiong.zhang@sheffield.ac.uk

3rd Eishaan Bhargava
Dept. of Paediatric Otolaryngology
Sheffield Children's Hospital
Sheffield, United Kingdom
e.bhargava@nhs.net

4th Heather Elphick
Dept. of Paediatric Respirator & Sleep Medicine, Sheffield Children's Hospital
Sheffield, United Kingdom
h.elphick@nhs.net

5th Sanja Dogramadzi
School of Electrical & Electronic Eng.
University of Sheffield
Sheffield, United Kingdom
s.dogramadzi@sheffield.ac.uk

6th Lyudmila S. Mihaylova
School of Electrical & Electronic Eng.
University of Sheffield
Sheffield, United Kingdom
l.s.mihaylov@sheffield.ac.uk

***Abstract*—This paper proposes a deep learning approach for contactless detection of sleep apnoea using pulse and blood oxygen saturation (SPO2) data. Three convolutional neural network architectures are adopted for apnoea classification purposes by fusing different features of the available time series signals. A conventional convolutional neural network (CNN), a CNN with a support vector machine (CNN-SVM), and a CNN combined with a recurrent neural network (CNN-RNN) are compared. The RNN includes Gated Recurrent Units (GRU) and Bidirectional GRU (BiGRU). The CNN is utilised to extract features, whilst the SVM and RNN are used for classification. In addition, we compare two different fusion methods, signal-level and feature-level fusion. The performance is validated and evaluated on a public dataset obtain from St. Vincent University Hospital. The results show that the concatenation of SPO2 and pulse signal at the signal level enhances the classification performance compared to using the individual signal. In addition, the classification sensitivity with signal-level fusion is higher than that with feature-level fusion. Overall, the proposed CNN-RNN with GRU (CNN-GRU) architecture gives the best performance with an accuracy of 85.4%, a sensitivity of 61.5%, a specificity of 91.9%, an F_1 score of 0.64, and a κ score of 0.551 with a dropout rate of 0.5 and a 20-second overlap. The results demonstrate that the proposed deep learning approach offers a promising solution for non-invasive detection of sleep apnoea using affordable physiological signals.**

***Index Terms*—Sleep Apnoea, Deep Learning, Data Fusion, Convolutional Neural Network, Recurrent Neural Network**

I. Introduction

Sleep apnoea syndrome is a common sleep disorder characterized by repeated breathing interruptions during sleep [1]. It significantly impacts the cardiovascular, nervous systems and the quality of life. Current clinical diagnosis mainly relies on polysomnography (PSG) [2]. However, this method has cost, comfort, and scalability limitations, and it is challenging to meet the needs of early screening and home monitoring. In recent years, advancements in deep learning technology have prompted researchers to construct automated apnoea detection models utilising physiological inputs, including blood oxygen saturation (SPO2) [3], pulse, electroencephalogram (EEG), and electrocardiogram (ECG) [4].

A. Related Deep Learning Approaches

Convolutional neural networks (CNN) [5], recurrent neural networks (RNN) [6] and their hybrid models have shown promising potential in time series signal analysis [7]. However, existing methods generally face problems such as model robustness, strong dependence on a single signal, and difficulty handling class imbalance, single and ensemble outliers. With the development of Artificial Intelligence (AI) and automation, deep learning has gradually been proven to improve the accuracy of sleep analysis. By analysing research published between 2008 and 2018, Mostafa et al. [8] provided insights into the effectiveness, advantages, and potential future directions of deep learning for sleep apnoea detection. Same as [9], although CNN, RNN, and hybrid models have shown high performance, they believe that ongoing research needs to focus more on overcoming data challenges and ensuring clinical implementation. Most sleep analysis applications nowadays concentrate on feature extraction and signal processing. Sillaparaya et al. [10] proposed a deep-learning approach to classify OSA using snoring sounds. They obtained an accuracy of 85.25% by using Mel-frequency cepstral coefficients (MFCC) and a three-layer fully connected network. However, there is still more work to be done, as demonstrated by the difficulties presented by feature overlap, class imbalance, and small datasets. Barnes and his colleagues [11], unlike Sillaparaya and his group, used a single-channel EEG and created a CNN architecture with three convolutional layers to detect sleep apnoea. Their framework was more understandable than others because they trained a CNN classifier using raw EEG data. This provided inspiration and evidence for our direct use of signals.

Since this study uses two data types, a two-dimensional (2D) CNN is the model basis. Jiménez-García et al. [12] published a 2D CNN model to detect sleep apnoea in children using airflow and oximetry. They used a CNN model with two convolutional blocks to classify the severity of apnoea in children. Due to the extreme data imbalance, the classification results were not perfect. Later, they improved the model by adding Bidirectional Gated Recurrent Units (BiGRU) as a prediction model [13]. The authors of [13] pointed out that BiGRU is used to analyze the temporal patterns of data in two directions. However, it was not compared with the Gated Recurrent Units (GRU) model [14]. Our study changes this model to a GRU model and found that the overall ability to predict apnoea was better than BiGRU.

In addition to CNN-RNN, the CNN with a support vector machine (CNN-SVM) hybrid model is a popular classification model nowadays. Baresary and his colleague apply CNN-SVM to classify sleep apnoea by using PSG [15]. The performance is outstanding. However, they did not give the dataset conditions, and the details and parameters of the model are unknown. They simulated noise and added it to the raw data to get closer to reality. The current trend in apnoea detection is to use less and less expensive signal data to detect the disease. Although their research was highly accurate, it was time-consuming and costly.

B. Main Contributions

The main contributions of this paper are as follows: (1) This study proposes a deep learning approach for detecting sleep apnoea using pulse and SPO2 signals. We adopt and compare three deep learning methods: CNN, a CNN-SVM model, and a CNN-RNN model. The RNN block includes GRU and BiGRU structures. (2) The effects of different signal fusion strategies, dropout probability settings, and window overlap lengths on detection performance are evaluated. Specifically, two fusion strategies are compared, signal-level fusion and feature-level fusion. The signal-level fusion performs better than feature-level fusion, yielding a higher sensitivity of 61.2% for apnoea detection. (3) The performance of the proposed approach is evaluated on a public dataset, St.Vincent University Hospital. The validation results show that combining the pulse and SPO2 signals using the proposed CNN-GRU architecture outperforms the single signal model, yielding satisfactory performance.

The rest of the paper is as follows: Section II describes the adopted deep learning algorithms. Section III gives implementation details of the deep learning architecture. Section IV presents the validation and performance analysis. Finally, Section V summarises the conclusions and future work.

II. The Proposed Hybrid CNN-SVM and CNN-RNN Architectures for Sleep Apnoea Detection

Starting with the well-known CNNs [5] and RNNs [6] architectures, this paper proposes hybrid approaches benefitting from a prediction model CNN-BiGRU [13]. This section describes the CNN-SVM and CNN-RNN architectures.

A. The CNN-SVM architecture

A hybrid classification approach, CNN-SVM combines convolutional neural networks (CNN) with support vector machines (SVM). This method [16] uses CNN to extract feature representations from raw input data. The extracted features are then used as input for the SVM classifier. CNN enables the automated extraction of features in a hierarchical structure. The SVM method maximises the margins between various classes to generate appropriate decision limits for classification tasks. The CNN feature extraction combined with SVM classification could raise the general performance of the model for different applications. This hybrid model efficiently leverages the representational capabilities of CNN in conjunction with the robust generalisation properties of SVM.

Fig. 1 presents the basic architecture of CNN-SVM, which is also the architecture used in this experiment. Because 2D CNN is used, the processed signal needs to be reshaped to meet the high-dimensional output requirements of the CNN. The detailed setting of the CNN block is in Section III. The output of the fully connected layer of the CNN block will be used as the input of the SVM model. SVM can be changed to any other machine learning method.

B. The CNN-RNN architecture

Recurrent neural networks (RNNs) [6] are deep learning models that process and convert sequential data inputs into specific sequential data outputs. Sequential data comprises sequential components interrelated by complex semantics and syntax constraints. Examples of sequential data include words, sentences, and time-series data. A combination of CNN and RNN models is designed and trained using a dataset of 1-minute segments of Pulse and SPO2 signals, labelled with apnoea episodes.

Fig. 2a presents a deep learning architecture for the concatenation in the signal stage and the overall CNN+RNN idea. The data is divided into 1-minute segments. Assuming the the data with sampling frequency of 8 Hz and the signal length of 480. If the concatenation is performed at the signal stage, the input size is 480×2. If concatenation is performed at the CNN feature stage, the architecture should have two input sizes of 480×1 (see Fig. 2b). Different from the basic CNN model [5], in the CNN hybrid model, the dropout layer is followed by a flatten layer instead of a fully connected layer. This is because the flattening layer is used to convert a multi-dimensional tensor into a one-dimensional vector without doing any additional processing [17]. This operation is a reshaping function with no learnable parameters.

In a CNN+RNN architecture, features extracted from the CNN block are usually directly converted into a sequence format suitable for RNN processing without further mixing or transformation. The flattening layer translates the multi-dimensional feature map into a one-dimensional vector free of parameters, which preserves the original feature information obtained by CNN. A fully connected layer, on the other hand, will add more weights and biases and combine and change features in linear and nonlinear ways. This makes it more difficult for RNN to get time series information because it

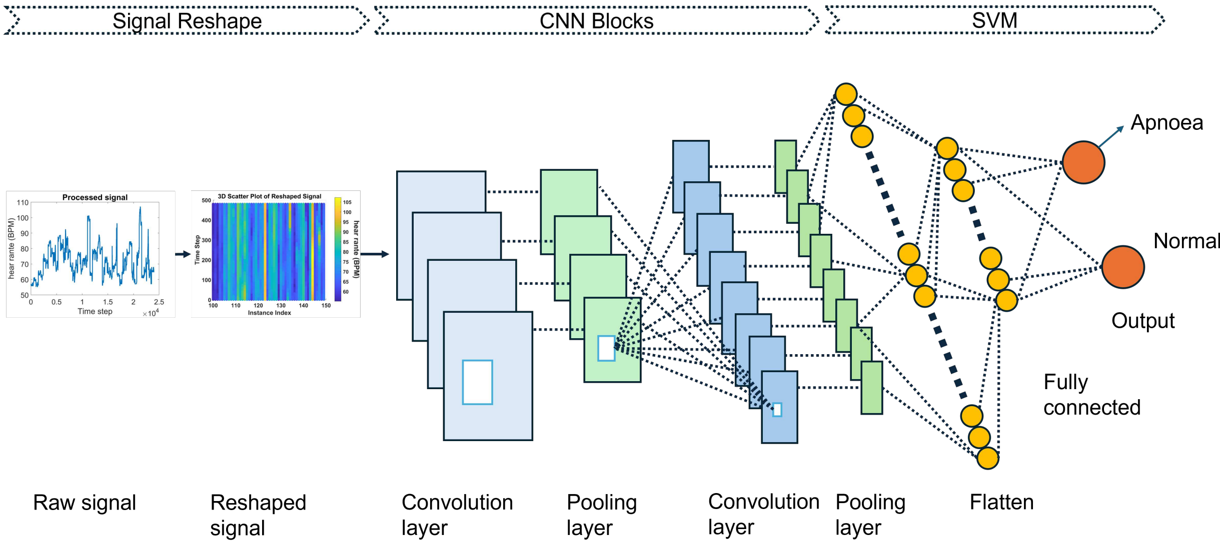


Fig. 1 CNN+SVM architecture

adds more model factors and computational work. It may also change the original feature structure.

The RNN block contains an RNN layer, a fully connected layer and a softmax layer. The last two layers are for classification. Since GRU has the same performance as LSTM and lower computational cost [13], two layers related to GRU are selected for the RNN layer, GRU layer and BiGRU layer.

The BiGRU layer integrates the capabilities of GRU with bidirectional processing, enabling the model to learn past and future details about the input sequence. The BiGRU layer has two GRU layers (see Fig. 3), each of which concurrently processes the input sequence in both forward and backwards directions [18]. In the forward pass, the LSTM layer captures information from previous time steps, while the backwards pass acquires information from subsequent time steps. This bidirectional processing allows the model to precisely capture long-term dependencies in the input sequence. Finally, the outputs are contacted and sent into a fully connected layer and a softmax layer for classification.

III. Implementation

A. Implementation Details

The parameter settings for the CNN model of each task are shown in Table I. Adam optimiser is chosen for its good binary classification. The size of the mini-batch to use for each training iteration is 128. The maximum number of epochs to use for training is set to 500. The initial learning rate used for training is set to 0.001. The experiments are performed using the Matlab2024b version.

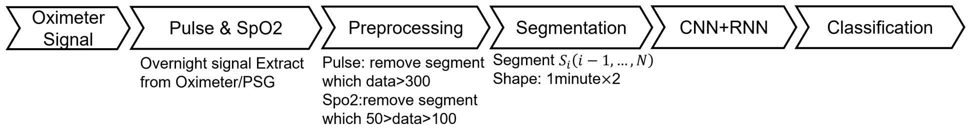
B. Data Preparation

The dataset from St. Vincent's University Hospital [19] are used in this experiment. This dataset has 25 cases, and all the data are split into overlapping one-minute segments. SPO2 and pulse readings below 50 and above 300 are considered artefacts and removed from further considerations. The segments

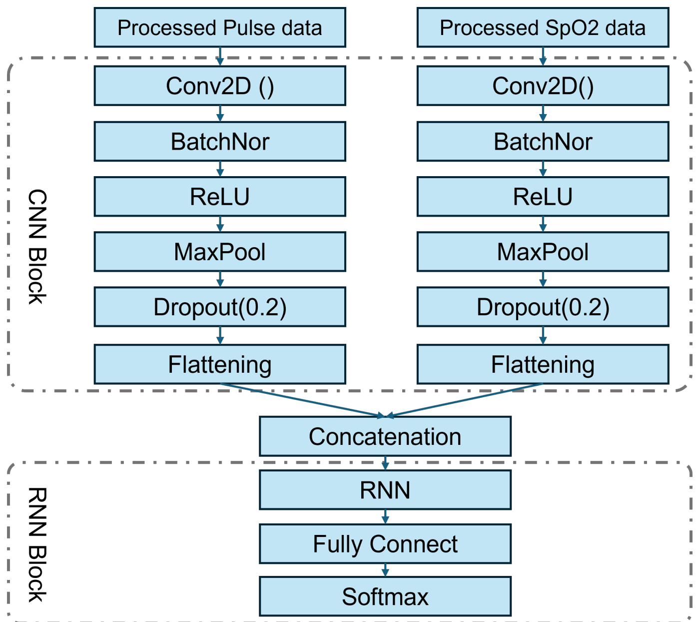
TABLE I
Parameters setting for the CNNs

CNN	Conv2D_1	number	32
		size	[16,1]
		stride	[1,1]
	Conv2D_2	number	64
		size	[16,1]
		stride	[1,1]
	Maxpooling_1	size	[2,1]
		stride	[1,1]
	Maxpooling_2	size	[2,1]
		stride	[1,1]
	dropout	probability	0.3
CNN+RNN	Conv2D_1	number	32
		size	[16,1]
		stride	[1,1]
	Maxpooling_1	size	[2,1]
		stride	[1,1]
	dropout	probability	0.2
CNN+SVM	Conv2D_1	number	32
		size	[16,1]
		stride	[1,1]
	Conv2D_2	number	64
		size	[16,1]
		stride	[1,1]
	Maxpooling_1	size	[2,1]
		stride	[1,1]
	dropout	probability	0.2

containing apnoea events are classified as 'apnoea' while those without respiratory disturbances are labelled 'normal'. When an apnoea event spanned two consecutive segments, fine-scale classification was applied. Respiratory disturbances lasting less than five seconds in any segment are classified as 'normal,' as such brief interruptions do not significantly affect the overall respiratory pattern. If disturbances surpass this duration in any segment, it is classified as 'apnoea', signifying a substantial interruption in respiratory function. This method guarantees accurate classification of each minute according to the intensity and length of interruptions. The detail of the data in different overlaps (ovlp) is shown in Table II. The data is



(a) Signal-level fusion architecture



(b) Feature-level fusion architecture

Fig. 2 An overview of CNN+RNN architecture

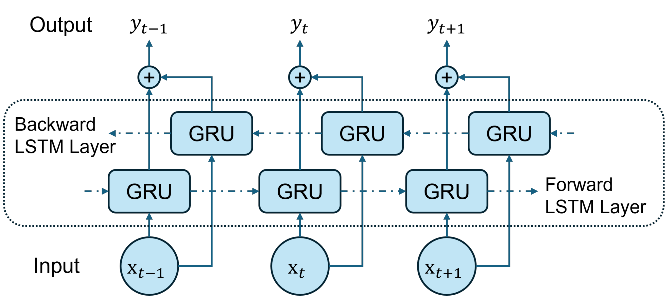


Fig. 3 The BiGRU architecture

TABLE II
DETAIL OF THE DATA IN DIFFERENT OVERLAPS(OVLP)

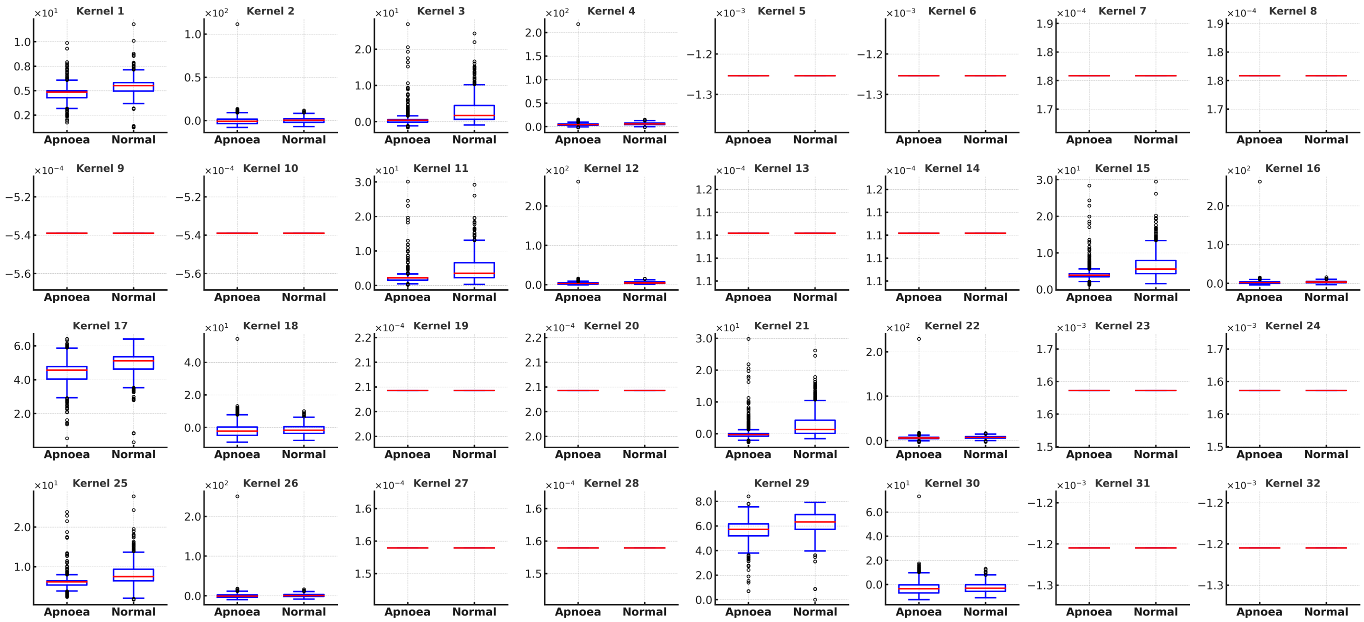
	apnoea	normal
0_ovlp	2569	7406
10_ovlp	2882	9094
20_ovlp	3289	11674
30_ovlp	3820	16137

divided into a training set and a test set in a ratio of 8:2. Since the convolutional layer used is a convolutional 2D layer, the data must be reshaped into four-dimensional data to meet the model input requirements. The training data is divided into 2000 segments and one segment has 480 samples. The input data can be expressed as [480 2000]. In order to meet the input conditions of CNN, the input matrix is reshaped in the order of $[S\ C\ B\ T]$ to become 4-D data $[480\ 1\ 1\ 2000]$. These four numbers represent samples, channels, batches, and time, respectively.

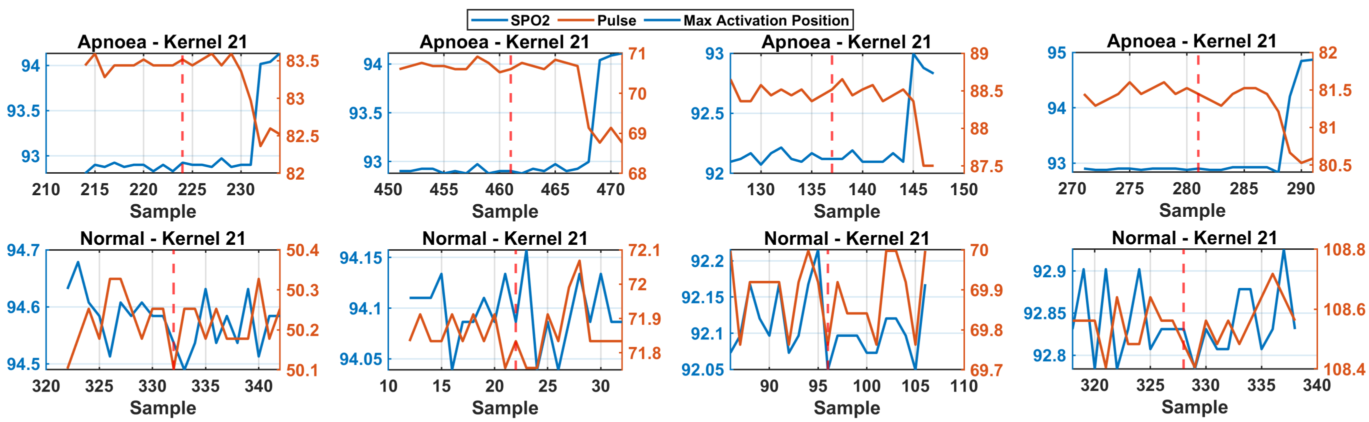
C. Feature Map from CNN

Figure 4a shows the maximum activation values of the 32 convolution kernels of the CNN model for the apnoea and normal classes. Each box plot shows the statistical characteristics of the maximum response value of the convolution kernel to the input sequence of the two classes. The central line in the boxplot means the median, while the two edges of the box mean the interquartile range. Distinct disparities exist in the activation distribution for some convolution kernels between the two classes, which means that these kernels effectively capture the temporal structural attributes associated with apnoea and significantly influence the classification decision of the model.

Figure 4b shows how the kernel 21 captures the local structure of the original input signal at its maximum response position. The figure shows the signal pattern within a specific range before and after the response position. It can be observed from the figure that in the apnoea sample, the kernel tends to detect local structures with downward trends in SPO2 and upward trend patterns in pulse. These patterns may correspond to the sudden change in flow before apnoea.



(a) Max responses of all kernels/filters (apnoea vs no apnoea). The red line means median. The blue box is the data range.



(b) Data around the strongest activation location in the input

Fig. 4 Feature map visualisation from CNN block

IV. Results and Discussion

This section mainly gives the results of different models based on CNN. The left side of the table represents different inputs, for example, 'SPO2_30' means the input data is SPO2 signal in 30 seconds overlap. 'Feature' here means feature concatenation. 'SigConcate' means signal concatenation. The difference between these two is the location of the concatenation. Concatenation at the full-connection layer is feature concatenation, while signal concatenation is concatenated at the signal stage. 'F1-N' and 'F1-A' mean the F_1 score of the normal class and the F_1 score of the apnoea class. Since this experiment is a medical classification, the sensitivity (sens) is mainly used to evaluate the model.

1) Results with CNN architecture: Table III shows the results of the CNN model. In this table, the signal concatenation data with a 20-second overlap performs well. The feature

TABLE III
Results with CNN Model

	acc	sens	spec	F_1-N	F_1-A	κ
SPO2_0	82.1%	53.7%	91.4%	0.88	0.60	0.48
SPO2_10	82.3%	56.1%	90.2%	0.89	0.59	0.48
SPO2_20	84.3%	57.9%	91.8%	0.90	0.61	0.51
SPO2_30	85.6%	52.4%	93.3%	0.91	0.58	0.49
Pulse_0	69.4%	30.7%	84.0%	0.80	0.36	0.16
Pulse_10	71.8%	30.0%	85.0%	0.82	0.34	0.16
Pulse_20	74.2%	28.3%	87.6%	0.84	0.33	0.18
Pulse_30	78.2%	25.8%	91.8%	0.87	0.33	0.21
Featrue_0	80.5%	60.2%	87.2%	0.87	0.61	0.48
Feature_10	81.4%	56.2%	89.1%	0.88	0.59	0.47
Feature_20	84.0%	55.7%	92.0%	0.90	0.61	0.51
Feature_30	84.7%	53.5%	92.4%	0.91	0.58	0.49
SigConcate_0	80.5%	56.7%	88.7%	0.87	0.60	0.47
SigConcate_10	82.0%	58.3%	89.7%	0.88	0.61	0.50
SigConcate_20	**84.0%**	**60.1%**	**90.4%**	**0.90**	**0.61**	**0.51**
SigConcate_30	85.0%	53.4%	92.3%	0.91	0.57	0.48

concatenation with no overlap has 60.2% sensitivity, but the κ value is not as good as signal concatenation. Compared to the same 20-second overlap input, the sensitivity of the connection is higher than that of the single signal result. However, there is still room for improvement in the overall performance, which can be achieved by optimising network parameters or enhancing data preprocessing.

TABLE IV
RESULTS WITH CNN-SVM MODEL

	acc	sens	spec	F_1-N	F_1-A	κ
SPO2_0	79.0%	45.1%	90.9%	0.87	0.53	0.40
SPO2_10	83.4%	47.9%	94.3%	0.90	0.58	0.48
SPO2_20	84.1%	45.6%	95.5%	0.90	0.57	0.48
SPO2_30	85.0%	40.4%	96.2%	0.91	0.52	0.44
Pulse_0	68.0%	28.0%	82.0%	0.79	0.31	0.11
Pulse_10	74.8%	18.8%	92.3%	0.85	0.26	0.14
Pulse_20	76.8%	16.2%	93.4%	0.86	0.23	0.12
Pulse_30	79.0%	14.9%	95.0%	0.88	0.22	0.13
SigConcate_0	81.4%	50.7%	92.2%	0.88	0.59	0.47
SigConcate_10	83.8%	54.7%	92.9%	0.90	0.62	0.52
SigConcate_20	83.5%	58.8%	90.2%	0.90	0.60	0.50
SigConcate_30	84.7%	51.1%	93.1%	0.91	0.57	0.48

2) Results with CNN+SVM : Table IV shows the results of the CNN-SVM model. In this model, using a single signal as input is unsatisfactory, but the fusion of two signals performs relatively well. Based on previous studies, SPO2 has always performed well in apnoea classification. This is because apnoea directly affects oxygen intake, which means the oxygen content in the blood [20]. However, in this CNN-SVM model, the classification sensitivity of the SPO2 signal is less than 50%. This may be due to the unbalanced signal and may also be due to the parameter setting.

3) Results with CNN+RNN: Based on different RNN models, this experiment designed two CNN-RNN models, the CNN-GRU and the CNN-BiGRU models. Table V shows the results of the CNN-GRU model. The results show that based on the CNN-GRU model, the comprehensive performance of the signal concatenation fusion input with a 20-second overlap is better than that of other inputs. However, due to the randomness of the experiment (such as the existence of the dropout layer), the experimental parameters can be further optimized to obtain better performance.

Table VI shows the results of the CNN-BiGRU model. The overall trend of this result is similar to that of CNN-GRU. However, a 10-second overlap is more suitable for this model. By comparing the detection results after fusion of the two stages, the results of fusion in the signal stage are slightly higher than the others. These four tables show that the blood oxygen signal can directly detect apnoea, while the pulse signal is slightly insufficient. This may be due to the limitation of the pulse signal for apnoea classification. Although the pulse signal also changes when apnoea occurs, the pulse signal still has limitations in diagnosing apnoea. Sleep apnoea is mainly caused by airway obstruction or central nervous system abnormalities. However, the pulse signal does not directly indicate the occurrence of apnoea like the airflow sensor but is indirectly inferred through the secondary effects

TABLE V
RESULTS WITH CNN-GRU MODEL

	acc	sens	spec	F_1-N	F_1-A	κ
SPO2_0	81.1%	55.3%	90.2%	0.88	0.60	0.48
SPO2_10	81.3%	56.0%	89.2%	0.88	0.59	0.47
SPO2_20	83.4%	58.3%	91.5%	0.90	0.61	0.51
SPO2_30	85.2%	51.6%	93.0%	0.91	0.57	0.48
Pulse_0	70.5%	24.8%	86.1%	0.81	0.30	0.12
Pulse_10	72.7%	20.5%	88.6%	0.83	0.26	0.11
Pulse_20	73.5%	19.7%	88.6%	0.84	0.25	0.10
Pulse_30	76.1%	20.0%	90.0%	0.86	0.25	0.12
Featrue_0	81.3%	54.5%	90.6%	0.88	0.60	0.48
Feature_10	82.3%	55.9%	90.7%	0.89	0.60	0.49
Feature_20	85.4%	57.7%	92.9%	0.91	0.63	0.54
Feature_30	83.7%	53.8%	91.0%	0.90	0.56	0.46
SigConcate_0	80.2%	50.3%	90.7%	0.87	0.57	0.44
SigConcate_10	82.9%	57.0%	91.0%	0.89	0.61	0.51
SigConcate_20	**83.6%**	**61.2%**	**89.7%**	**0.90**	**0.61**	**0.51**
SigConcate_30	85.2%	50.5%	93.0%	0.91	0.56	0.47

TABLE VI
RESULTS WITH CNN-BIGRU MODEL

	acc	sens	spec	F_1-N	F_1-A	κ
SPO2_0	80.6%	51.0%	90.1%	0.88	0.56	0.44
SPO2_10	81.6%	55.0%	90.6%	0.88	0.60	0.48
SPO2_20	83.6%	56.6%	91.1%	0.90	0.60	0.50
SPO2_30	84.6%	52.7%	92.2%	0.91	0.57	0.48
Pulse_0	70.3%	26.8%	84.8%	0.81	0.31	0.13
Pulse_10	71.2%	23.0%	86.4%	0.82	0.28	0.11
Pulse_20	73.2%	27.5%	86.7%	0.83	0.32	0.16
Pulse_30	76.4%	24.1%	88.5%	0.86	0.28	0.14
Featrue_0	80.9%	56.7%	89.3%	0.87	0.61	0.48
Feature_10	83.2%	56.1%	91.6%	0.89	0.61	0.51
Feature_20	82.4%	54.9%	89.8%	0.89	0.57	0.46
Feature_30	83.7%	55.5%	90.7%	0.90	0.58	0.48
SigConcate_0	79.8%	54.3%	88.8%	0.87	0.58	0.45
SigConcate_10	81.7%	56.2%	89.0%	0.88	0.60	0.49
SigConcate_20	**83.6%**	**58.4%**	**91.8%**	**0.90**	**0.61**	**0.51**
SigConcate_30	84.1%	57.4%	90.6%	0.90	0.58	0.49

of the cardiovascular system [21]. This means that if some short or mild apnoea does not cause an obvious heart rate response, the pulse signal may not have an apparent change, which may cause missed events. The result after signal fusion is still better than that of a single signal, which shows that pulse signal is still helpful in indirectly detecting apnoea.

A. Performance Comparison and Discussion

Table VII presents results from different models based on signal concatenation with 20-second overlap. In addition to the CNN model introduced in this paper, there is also the CUSUM algorithm for change point detection [22]. CNN-GRU shows the best performance among these models. Given

TABLE VII
CNN RESULTS BASED ON CONCATENATED DATA

	acc	sens	spec	F_1-N	F_1-A	κ
CNN	84.0%	60.1%	90.4%	0.90	0.61	0.51
CNN+SVM	83.5%	58.8%	90.2%	0.90	0.60	0.50
CNN+GRU	**83.6%**	**61.2%**	**89.7%**	**0.90**	**0.61**	**0.51**
CNN+BiGRU	83.6%	58.4%	91.8%	0.90	0.61	0.51
CUSUM [22]	51.1%	72.3%	45.1%	0.59	0.39	0.11

that the outcomes of the CNN model closely resemble those of

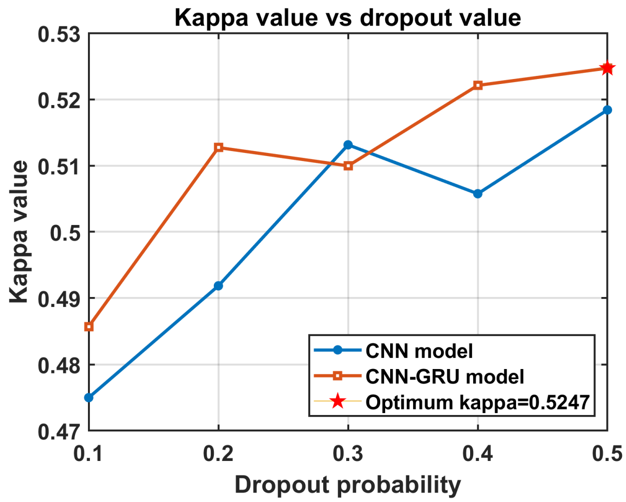


Fig. 5 Diagnostic performance of CNN and CNN + RNN architectures for different numbers of dropout probabilities.

the CNN-GRU model, we choose to perform a more detailed evaluation based on different dropout values. Table VIII and

TABLE VIII
CNN MODEL RESULTS WITH DIFFERENT DROPOUT

Dropout	acc	sens	spec	F_1-N	F_1-A	κ
0.1	82.5%	58.7%	88.9%	0.89	0.59	0.4750
0.2	83.4%	57.9%	90.3%	0.90	0.60	0.4919
0.3	84.0%	60.1%	90.4%	0.90	0.61	0.5131
0.4	84.0%	57.9%	91.1%	0.90	0.61	0.5058
0.5	**84.7%**	**57.3%**	**92.1%**	**0.90**	**0.61**	**0.5184**

TABLE IX
CNN-GRU RESULTS WITH DIFFERENT DROPOUT

Dropout	acc	sens	spec	F_1-N	F_1-A	κ
0.1	83.7%	54.1%	91.7%	0.90	0.59	0.4857
0.2	84.5%	56.5%	92.1%	0.90	0.61	0.5127
0.3	83.6%	61.2%	89.7%	0.90	0.61	0.5100
0.4	84.6%	59.0%	91.5%	0.90	0.62	0.5221
0.5	**84.3%**	**61.0%**	**90.6%**	**0.90**	**0.62**	**0.5247**

Table IX show results from the CNN and the CNN-GRU models based on different dropout probabilities, using signal concatenation with a 20-second overlap as input. According to these results both models perform well when the probability is 0.5. CNN-GRU is the best among them, which can be shown in Fig. 5. figure 5 shows the κ values obtained on the validation set for different dropout values using CNN and CNN-GRU. The maximum performance on the validation set was $\kappa = 0.5247$ with $dropout\ probability = 0.5$ and the CNN-GRU model. The other configurations performed slightly lower, so the best model was ultimately chosen to continue evaluating the test data.

Fig. 6 shows the κ values obtained in the validation set using different numbers NG of neurons in the GRU layer for CNN and RNN. The maximum performance in the validation set is $\kappa = 0.5510$ with $NG = 4$, this is higher than previous research results, indicating that the CNN-GRU model has

TABLE X
CNN-GRU RESULTS BASED ON A DIFFERENT NG

NG	acc	sens	spec	F_1-N	F_1-A	κ
1	85.5%	53.4%	94.3%	0.91	0.61	0.5258
2	85.2%	56.3%	93.1%	0.91	0.62	0.5290
4	**85.4%**	**61.5%**	**91.9%**	**0.91**	**0.64**	**0.5510**
8	84.3%	60.2%	90.8%	0.90	0.62	0.5219
16	84.2%	56.6%	91.8%	0.90	0.61	0.5080
32	85.0%	60.7%	91.7%	0.91	0.63	0.5407
64	84.3%	61.0%	90.6%	0.90	0.62	0.5247

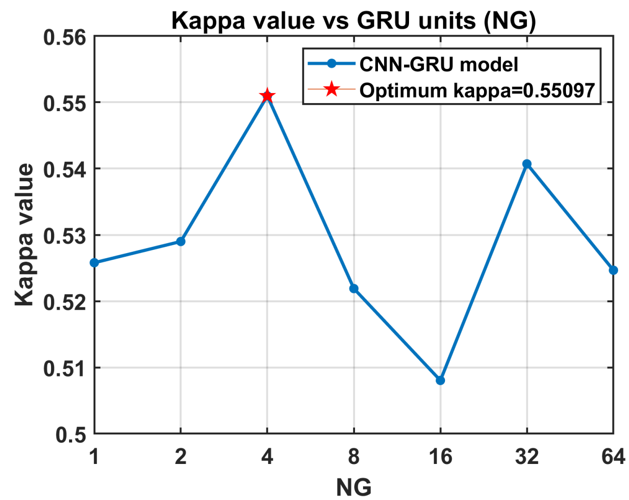


Fig. 6 Diagnostic performance of CNN + RNN architectures for different numbers of neurons in the GRU layer (NG).

higher consistency and reliability. The optimal $NG = 4$ was low, indicating that modelling the detection of apnoea in large clusters using 1-min segments does not require high complexity. Table X shows that with $NG = 4$, the model sensitivity reaches 61.5%, and F_1 score of the apnoea category is 0.64, which is the highest value in all experiments. The results obtained under different GRU unit sizes and dropout rates show minimal variation, indicating that the proposed CNN-GRU architecture exhibits stable performance across various parameter settings.

V. CONCLUSION AND FUTURE WORK

This study proposes a deep learning architecture for contactless detection of sleep apnoea using pulse and SPO2 signals. Three deep learning models, CNN, CNN-SVM, and CNN-RNN (with GRU and BiGRU variants), are designed, compared and evaluated on the public data from St. Vincent's University Hospital. The key design of the detection framework in this paper is that two feature fusion strategies are used, among which the result of signal-level fusion is better than that of feature-level fusion. The experimental results show that signal-level fusion of pulse and SPO2 signals can improve classification performance, compared to single-signal models. Among all model configurations, the CNN-GRU model with 20-second overlap and 0.5 dropout achieves the highest accuracy of 85.4%, sensitivity of 61.5%), specificity of 91.9%, and κ score of 0.551. In addition, the overall stability of the CNN-GRU architecture across different GRU units and dropout settings confirms its robustness and suitability for the classification task, without requiring extensive hyperparameter tuning. These results confirm the effectiveness of the proposed deep learning approach for low-cost, non-invasive sleep apnoea screening using physiological signals. The approach proposed in this paper is not only applicable to sleep apnoea detection, but can also be extended to other disease detection tasks based on physiological signals, such as arrhythmia detection, epileptic seizure prediction, and abnormal blood pressure monitoring. These tasks similarly involve the joint modelling of multimodal time series and are characterized by challenges such as non-linear feature relationships and strong temporal dependencies.

Future work will focus on multi-classification to include other sleep disorder categories, such as hypopnea. Moreover, the data imbalance problem can be addressed through techniques such as class-weighted loss function and targeted resampling to improve the detection sensitivity of all sleep disorder types.

Acknowledgement. We are grateful to the UK Engineering and Physical Sciences Research Council (EPSRC) for their support through the following projects: EP/V026747/1 (Trustworthy Autonomous Systems Node in Resilience) and EP/T013265/1 (NSF-EPSRC: ShiRAS. Towards Safe and Reliable Autonomy in Sensor Driven Systems). We also acknowledge the support of the National Science Foundation (NSF) under Grant No. ECCS 1903466 for the ShiRAS project. L. Mihaylova also acknowledges the support from the Insigneo Institute at the University of Sheffield.

References

[1] M. Emin Tagluk, M. Akin, and N. Sezgin, "Classıfication of sleep apnea by using wavelet transform and artificial neural networks," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1600–1607, 2010.

[2] J. V. Rundo and R. Downey, "Chapter 25 - Polysomnography," in *Clinical Neurophysiology: Basis and Technical Aspects*, ser. Handbook of Clinical Neurology, K. H. Levin and P. Chauvel, Eds. Elsevier, 2019, vol. 160, pp. 381–392.

[3] F. Mendonça, S. S. Mostafa, A. G. Ravelo-García, F. Morgado-Dias, and T. Penzel, "A review of obstructive sleep apnea detection approaches," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 825–837, 2019.

[4] S. Roomkham, B. Ploderer, S. Smith, and D. Perrin, *Technologies for Quantifying Sleep: Improved Quality of Life or Overwhelming Gadgets?* Springer International Publishing, 04 2022, pp. 151–164.

[5] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, 2022.

[6] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," *arXiv preprint arXiv:1506.00019*, 2015.

[7] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, pp. 5455–5516, 2020.

[8] S. S. Mostafa, F. Mendonça, A. G. Ravelo-García, and F. Morgado-Dias, "A systematic review of detecting sleep apnea using deep learning," *Sensors*, vol. 19, no. 22, 2019.

[9] G. Bazoukis, S. C. Bollepalli, C. T. Chung, X. Li, G. Tse, B. L. Bartley, S. Batool-Anwar, S. F. Quan, and A. A. Armoundas, "Application of artificial intelligence in the diagnosis of sleep apnea," *Journal of Clinical Sleep Medicine*, vol. 19, no. 7, pp. 1337–1363, 2023.

[10] A. Sillaparaya, A. Bhatranand, C. Sudthongkong, K. Chamnongthai, and Y. Jiraraksopakun, "Obstructive sleep apnea classification using snore sounds based on deep learning," in *Proceedings of 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2022, pp. 1152–1155.

[11] L. D. Barnes, K. Lee, A. W. Kempa-Liehr, and L. E. Hallum, "Detection of sleep apnea from single-channel electroencephalogram (eeg) using an explainable convolutional neural network (cnn)," *PLOS ONE*, vol. 17, no. 9, pp. 1–18, 09 2022.

[12] J. Jiménez-García, M. García, G. C. Gutiérrez-Tobal, L. Kheirandish-Gozal, F. Vaquerizo-Villar, D. Álvarez, F. del Campo, D. Gozal, and R. Hornero, "A 2D convolutional neural network to detect sleep apnea in children using airflow and oximetry," *Computers in Biology and Medicine*, vol. 147, p. 105784, 2022.

[13] J. Jiménez-García, M. García, G. C. Gutiérrez-Tobal, L. Kheirandish-Gozal, F. Vaquerizo-Villar, D. Álvarez, F. del Campo, D. Gozal, and R. Hornero, "An explainable deep-learning architecture for pediatric sleep apnea identification from overnight airflow and oximetry signals," *Biomedical Signal Processing and Control*, vol. 87, p. 105490, 2024.

[14] Y. Gao and D. Glowacka, "Deep gate recurrent neural network," in *Proceedings of Asian Conference on Machine Learning*. PMLR, 2016, pp. 350–365.

[15] D. Baresary, K. Kaur, T. Gera, S. K. Debnath, and S. Mehta, "A comparative study of cnn and svm for sleep apnea classification: Evaluating performance and efficiency," in *Proc. of Global Conference on Communications and Information Technologies (GCCIT)*, 2024, pp. 1–5.

[16] S. Ahlawat and A. Choudhary, "Hybrid cnn-svm classifier for handwritten digit recognition," *Procedia Computer Science*, vol. 167, pp. 2554–2560, 2020, International Conf. on Computational Intelligence and Data Science.

[17] J. Kreiser, M. Meuschke, G. Mistelbauer, B. Preim, and T. Ropinski, "A survey of flattening-based medical visualization techniques," *Computer Graphics Forum*, vol. 37, no. 3, pp. 597–624, 2018.

[18] Z. Hu, G. Liu, Y. Li, and S. Zhuang, "Sagb: self-attention with gate and bigru network for intrusion detection," *Complex & Intelligent Systems*, vol. 10, no. 6, pp. 8467–8479, 2024.

[19] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[20] A. R. Spector, D. Loriaux, and A. E. Farjat, "The clinical significance of apneas versus hypopneas: is there really a difference?" *Cureus*, vol. 11, no. 4, 2019.

[21] T. Penzel, "Is heart rate variability the simple solution to diagnose sleep apnoea?" *European Respiratory Journal*, vol. 22, no. 6, pp. 870–971, 2003.

[22] D. Yang, E. Bhargava, H. Elphick, and L. S. Mihaylova, "An adaptive cusum approach for automating sleep apnoea analysis based on pulse and oximetry data," in *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA)*, 2023, pp. 557–562.