

Quantitative Assessment of Respiratory Distress using Convolutional Neural Network for Multivariate Time Series Segmentation

Rohit Pardasani¹, Rupanjali Chaudhuri², Navchetan Awasthi³, Sheetal Chaurasia⁴, Sushma Maya⁵

¹ General Electric Healthcare, Bangalore, Karnataka, India

² Cerner Corporation, Bangalore, Karnataka, India

³ Massachusetts General Hospital, Harvard University, Boston, Massachusetts, USA

⁴ Manipal Hospital Whitefield, Bangalore, Karnataka, India

⁵ Publicis Sapient, Bangalore, Karnataka, India

Abstract

Respiratory Distress (RD) is typically associated with many critical conditions, particularly in aged population. The present-day monitoring solutions do not track the severity of RD distinctly and this obscures accurate prognosis. Our method quantifies RD condition to a ‘Severity Index (SI)’ in a real time monitoring setup. We achieve this by tracking pattern of respiratory rate (RR), oxygen saturation (SpO₂) and heart rate (HR) over time using Convolutional Neural Network (CNN) based model for multivariate time series segmentation. The CNN based approach is then compared with other two different approaches viz. Long-Short Term Memory (LSTM) model and Bayesian Inference model based on Symbolic Aggregate approximation (SAX). We work on segments extracted from 180 records of MIMIC-III Clinical and Waveform Database, after getting them clinically annotated. Our CNN, LSTM and SAX models achieve RMSE (Root Mean Square Error) of 0.288, 0.301 and 0.305 respectively with respect to the clinical annotations. All methods achieve AUC (Area Under Curve for Receiver Operating Characteristic(ROC)) close to 0.95. The dataset, code and trained models are available at https://github.com/rohit-pardasani/RDsegmentation_v1.

1. Introduction

Respiratory Distress is forewarning of many pathologies particularly related to failure of cardiac or respiratory system [1, 2]. It is a prevalent condition among aged population admitted to intensive care units (ICUs). A coarse characterization of RD is laboured breathing or signs that patient is not getting enough oxygen. It is difficult to gauge the acuteness of RD by short term trend or value of single vital parameter. Hence, present threshold based monitoring alarms fail to capture long term interplay of multiple

vitals that is prerequisite for accurate assessment of respiratory abnormality [2]. Most of the prior works in domain are inclined towards predicting Acute Respiratory Distress Syndrome (ARDS) [3] or reducing false alarms linked to respiratory conditions.

We propose a method for quantitative assessment of RD severity in real time using three non-invasively acquired parameters viz. respiratory rate (RR), peripheral capillary oxygen saturation (SpO₂) and heart rate (HR). The proposed system quantifies anomalous respiratory behaviour to a single value, that we denote as ‘Respiratory Distress Severity Index (RDSI)’ and ranges from 0 to 1. The system captures the entire retrospective pattern of vitals using a Convolutional Neural Network (CNN), that is trained to segment multivariate time series into RD and non-RD segments. We also use Long-Short Term Memory (LSTM) model and Bayesian Inference on Symbolic Aggregate approximation (SAX) representation of time series [4] for the same task. We compare the results of three models and illustrate how they can be used for continuous real time assessment of severity of RD.

2. Approach and Methods

Our approach maps ‘the problem of determining severity of RD in non-ventilated patients’ onto a ‘multi-variate time series segmentation problem’. Section 2.1 describes the selection of records for the dataset and the annotation process for creating the ground truth. Section 2.2, 2.3 and 2.4 describes the CNN, LSTM and Bayesian Inference on SAX model approaches respectively for quantification of severity of RD. We draw comparison among these three models and in section 2.5 describe how we can use these models in real time monitoring systems, even though they have been trained and tested on stationary data.

2.1. Data Selection and Annotation

We randomly selected 180 records from the MIMIC- III [5] Waveform Database of patients who were not mechanically ventilated throughout their ICU stay. Following are the attributes for 180 records selected for study:

- Age: 19 yrs - 88 yrs (Average = 66 yrs)
- Gender: 90 females, 90 males
- ICU Stay: 70 hrs - 120 hrs (Average= 93 hrs)
- Respiratory Diseases: 124 with some resp. disease

We extracted the raw data (denoted as T_{raw} in Eq. 1) of three clinical vitals i.e. RR, SpO₂ and HR. The extracted numerics data is noisy owing to artifacts related to probe fitting, patient movement, optical interference etc. and it needs to be pre-processed for noise reduction.

$$T_{raw} \xrightarrow[\text{using interpolation}]{\text{fill zeros and NaNs}} T_{filled} \quad (1)$$

$$T_{filled} \xrightarrow[\text{Chauvenet's Criterion}]{\text{outlier rejection}} T_{filtered} \quad (2)$$

$$T_{filtered} \xrightarrow[\text{LOESS}]{\text{smoothing}} T_{smoothed} \quad (3)$$

Eq. 1, Eq. 2 and Eq. 3 show the pre-processing steps performed on the data to make it suitable for annotation. The numerics extracted have a sampling rate of 1 sample/minute and all processes ensure that sampling frequency remains preserved. The regions of RD are given labels '1' and those without RD are marked as '0' by clinician using $T_{smoothed}$. Fig. 1 shows variation of RR, SpO₂ and HR in a record at T_{filled} (scatter plot) and $T_{smoothed}$ (dotted plot) stage. Fig. 1 shows a rectangular waveform (indicating 0 and 1 labelling), which is visual representation of ground truth labels. We use $T_{smoothed}$ data for marking clinical ground truth, while T_{filled} (after normalization) is used as input to CNN, LSTM and SAX based models detailed below.

2.2. Convolutional Neural Network (CNN) Segmentation Model

Quantifying severity of RD requires capturing information from current as well as retrospective trend of vital parameters. It has been proved that CNNs are quite efficient in identifying patterns, thus seemed a viable choice for this problem. Our architecture of 1D-CNN is inspired from single image super resolution 2D-CNN networks such as DnCNN [6]. For our task we map multivariate time series of a given length with equal length segmentation map. Our network consists of 12 convolutional layers as shown in Fig. 2. During training, the input to the network is standard score normalized multivariate (RR, SpO₂, HR) time-series and the target is time-series of labels (or segmentation map of time-series). The dataset of 180 records is

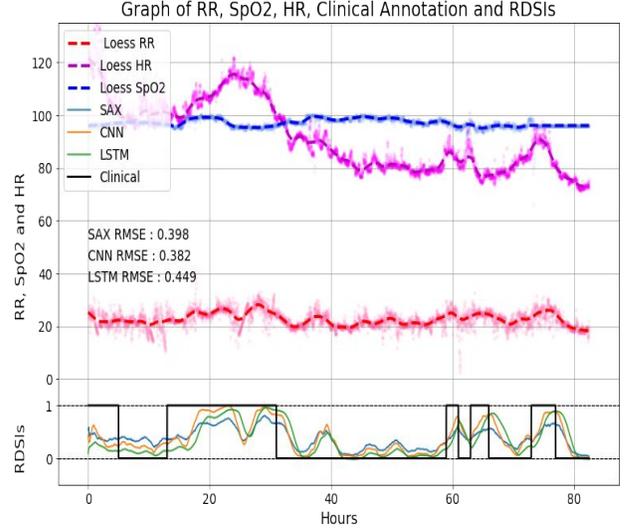


Figure 1. Graphs of T_{filled} and $T_{smoothed}$ vitals (RR, SpO₂ and HR) for a record from test set, shown as scatter plot and dotted plot respectively. The waveforms in the region below $y = 0$ represent clinical labels and time-series of RDSI obtained through CNN, Bayesian Inference SAX model and LSTM.

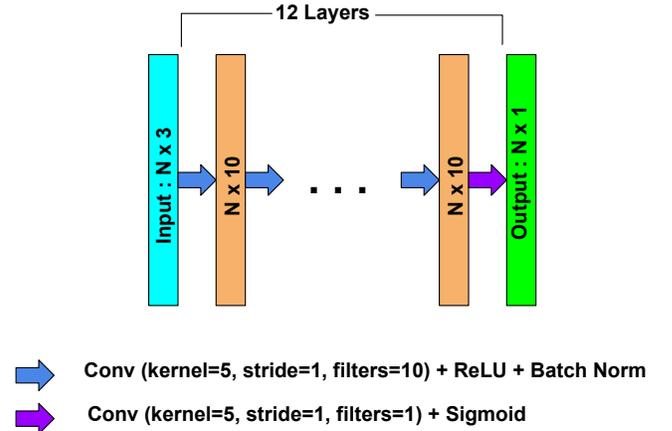


Figure 2. Proposed CNN Model for multivariate time series segmentation where N denotes number of time samples in time series. Three input channels correspond to RR, SpO₂ and HR respectively.

divided into training, validation and test sets of size 110, 30 and 40 respectively. Partially overlapping segments of 1024 samples (1024 minutes = ~ 17 hours) are extracted from them for training, validating and testing the model. The model is trained till 500 epochs using Stochastic Gradient Descent (SGD) optimizer with learning rate of 0.1 and momentum set to zero. Mean Square Error (MSE) is used as loss function and we retained the model that gave

best validation loss. The sigmoid activation layer after last convolution ensures that output of the network will be between 0 and 1 for all samples of time series and this output can be interpreted as ‘time series of Respiratory Distress Severity Index (RDSI)’. However, the output (RDSI series) is not smooth and has lot of fluctuations acting as noise, and hence difficult to interpret in a condition like RD, because its progression or decline is a gradual process [2]. These fluctuations are attributed to the fact that model was applied on T_{filled} (sampled per minute) that itself was quite fluctuating due to noise. Thus, to capture the actual long term trend of the RDSI, we run a 2 hours (120 sample) moving average on output and finally get the desired SI behaviour as shown in Fig. 1. It is clear from the figure that RDSI so obtained varies smoothly and follows the pattern of annotation.

2.3. Long-Short Term Memory (LSTM) Sequence to Sequence Mapping Model

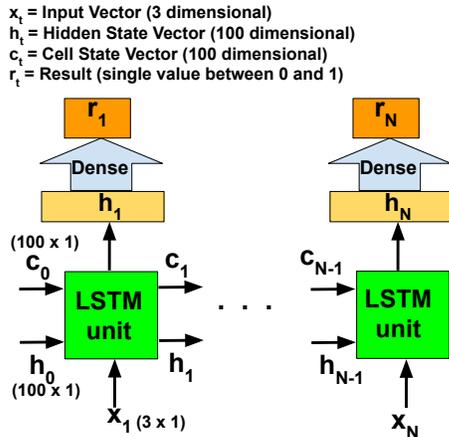


Figure 3. Proposed LSTM Model for sequence to sequence mapping where N denotes time samples in time series and 3-D input vector corresponds to RR, SpO₂, HR

LSTMs have been proved quite effective for time series data. We figure out that the most suitable architecture of LSTM for our case is synced sequence input and output model (many-to-many mapping). For each sample in time-series we need a RDSI, thus corresponding to our input time series we have a ‘time series of RDSI’ of same length. The network consists of LSTM unit followed by a dense layer, as shown in Fig. 3. We use exactly the same datasets for training, validation and testing that we used in CNN. The model is trained till 100 epochs using Adam optimizer (with learning rate = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$) and MSE as loss function. The fluctuations in the output pattern when model is used for inference are again dealt by applying 2 hours moving average to get time series of

RDSI. As can be seen from Fig. 1, RDSI mimics behaviour of annotation but smoothly.

2.4. Bayesian Inference on Symbolic Aggregate approximation (SAX)

Table 1. Symbols used in SAX representation along with normalized and actual ranges. *SpO₂ cannot exceed 100, but ranges are mentioned to maintain symmetry of table

Sym	Norm	RR	SpO ₂	HR
A	$-\infty$ to -1.5	$-\infty$ - 10	$-\infty$ - 89	$-\infty$ - 59
B	-1.5 to -1	10 - 13	89 - 91	59 - 68
C	-1 to -0.5	13 - 16	91 - 93	68 - 77
D	-0.5 to 0	16 - 20	93 - 96	77 - 86
E	0 to 0.5	20 - 23	96 - 98	86 - 95
F	0.5 to 1	23 - 26	98 - 100	95 - 104
G	1 to 1.5	26 - 29	100 - 102*	104 - 113
H	1.5 to ∞	29 - ∞	102 - ∞ *	113 - ∞

Table 2. Example showing SAX representation with corresponding labels and probability outputs after applying Bayesian Inference

Time	...	t_k	t_{k+1}	t_{k+2}	...
RR, RR _{sym}	...	15.8, C	26.3, F	27.6, G	...
SpO ₂ , SpO _{2sym}	...	96.0, D	93.0, C	97.0, E	...
HR, HR _{sym}	...	73.8, C	128.2, H	127.2, H	...
SAX	...	CDC	FCH	GEH	...
Labels	...	0	1	1	...
$P(X = 1)$...	0.014	0.545	0.782	...

The fundamental idea behind SAX representation is reducing an arbitrary length time series to a proportional length string [4]. Our approach closely follows the SAX approach discussed in [4], except that we do not use PAA (Piecewise Aggregate Approximation) of the T_{filled} data, instead we directly represent each time sample of data as a symbol triplet. This means that the length of the SAX representation is same as the length of time series. We map range of normalized values of vitals to symbols as shown in Table 1. The given range selection ensures that cut-points are close to clinically significant thresholds. Table 2 illustrates how a typical time-series (T_{filled}) is converted to series of symbol triplets using SAX representation. Based on the labels, the frequency of each symbol in ‘class 1’ (RD category) and ‘class 0’ (non-RD category) is calculated from the training set. The Bayesian rule (Eq. 4) is used to calculate the probability of a symbol belonging to ‘class 1’. We end up obtaining a time series of probabilities (example shown in last row of Table 2), which is inter-

puted as time-series of RDSI. Again, we encounter fluctuations similar to those present in previous models (Sec. 2.2 and Sec. 2.3) and we apply a 2 hours moving average to get a smoothly varying RDSI. Fig. 1 shows comparison of this smoothed RDSI (obtained via SAX representation method) with the clinical annotation. Below is equation representing Bayesian inference as applied to SAX:

N_{S1} = No. of occurrences of symbol S in ‘class 1’

N_S = Total no. of occurrences of symbol S

$$P(X = 1 | sym = S) = \frac{N_{S1}}{N_S} \quad (4)$$

2.5. Real Time Prediction

The versatility of all three models lies in the fact that their functioning is independent of the length of time series. The described CNN model contains only convolutional layers, thus can operate on varying length time series once trained. LSTM shows same characteristic because the number of weights learnt in the LSTM used are independent of length of time series. The Bayesian Inference on SAX representation also gives SI at each time sample irrespective of number of samples in input. So, all three models can be deployed easily for real time monitoring. Suppose, we want to update RDSI after n minutes interval and sampling rate is 1 sample/minute. Clearly, in this case we will run these models after every n minutes, on the samples we have accrued till that time. Therefore, our model will run on $n, 2n, 3n, \dots$ samples in subsequent updates of RDSI.

3. Results

The performance of the proposed models is compared in terms of RMSE (Root Mean Square Error), AUC (Area Under Curve for Receiver Operating Characteristics (ROC)), sensitivity, specificity and accuracy (Table 3). The RMSE is calculated between labels and RDSI values by treating this as a regression problem. On the other hand, while calculating AUC, sensitivity, specificity and accuracy, it is treated as a two class (‘RD’ and ‘non-RD’) classification problem. The models achieve AUC of about 0.95 on the test set. In terms of regression performance, these models achieve RMSE close to 0.3 when clinical time-series is compared with time-series of RDSI. Though the performance of all the models is almost at par, CNN model performs marginally better in terms of RMSE which we find is the most preferred metrics when comparing RDSI time series with annotation (Table 3).

4. Conclusion

CNN for time series segmentation was proposed and can be effectively used for detecting regions of respiratory dis-

Table 3. Results of classification and regression for the models on test set

Method	RMSE	AUC	Thresh.	Sens.	Spec.
CNN	0.288	0.94	0.31	0.83	0.88
LSTM	0.301	0.94	0.23	0.84	0.87
SAX	0.305	0.95	0.34	0.88	0.86

ress (RD) in a streaming setup. Our approach of quantifying severity RD can be integrated into a real time monitoring solution and can continuously track the intensity of respiratory abnormality. The approach is independent of amount of retrospective data available, so can be triggered just after the initiation of vital parameters monitoring. We have used three parameters for computing severity of RD, but the method can be easily extended to include more parameters if the annotations are done accordingly.

Acknowledgments

This research was funded by GE Healthcare.

References

- [1] Ravishankar H, Saha A, Swamy G, Genc S. An early respiratory distress detection method with markov models. In 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2014; 3438–3441.
- [2] Lynn LA, Curry JP. Patterns of unexpected in-hospital deaths: a root cause analysis. *Patient safety in surgery* 2011; 5(1):3.
- [3] Taoum A, Mourad-Chehade F, Amoud H. Evidence-based model for real-time surveillance of ards. *Biomedical Signal Processing and Control* 2019;50:83–91.
- [4] Lin J, Keogh E, Wei L, Lonardi S. Experiencing sax: a novel symbolic representation of time series. *Data Mining and knowledge discovery* 2007;15(2):107–144.
- [5] Johnson AE, Pollard TJ, Shen L, Li-wei HL, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG. MIMIC-iii, a freely accessible critical care database. *Scientific data* 2016;3:160035.
- [6] Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing* 2017; 26(7):3142–3155.

Address for correspondence:

Rohit Pardasani
 GE Healthcare,
 2nd Floor-Odyssey Building, JFWTC, Bangalore, India.
 rohit.r.pardasani@gmail.com