

# A Multi-Layered LSTM for Predicting Physician Stress During an ED Shift

Vishnunarayan Girishan Prabhu (✉ [vgirish@clemsun.edu](mailto:vgirish@clemsun.edu))

Clemson University <https://orcid.org/0000-0001-5410-9894>

Kevin Taaffe

Clemson University

Ronald Pirralo

PRISMA Health- Upstate <https://orcid.org/0000-0002-8321-3866>

---

## Research Article

**Keywords:** Physician wellbeing, emergency department, physiological data, stress management, neural networks

**Posted Date:** March 11th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-318589/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# **A Multi-Layered LSTM for Predicting Physician Stress During an ED Shift**

**Vishnunarayan Girishan Prabhu, Kevin Taaffe**  
**Dept. of Industrial Engineering, Clemson University**  
**Clemson, SC 29634**

**Ronald Pirrallo**  
**Dept. of Emergency Medicine, PRISMA Health - Upstate**  
**Greenville, SC 29605**

## **Abstract**

Emergency Department (ED) physicians are faced with complex care settings, including a high level of uncertainty and intensity. Burnout among physicians is increasing every year, and ED physicians are one of the groups most prone to burnouts and work-related stress in the US. This research focused on developing a supervised Long Short-Term Memory (LSTM) artificial neural network model to predict a physician's stress level based on their physiological data. Twelve attending physicians working a 3:00 pm - 11:00 pm shift at Greenville Memorial Hospital (GMH) in Greenville, SC, participated in the study. Stress levels were estimated using physiological measures, including heart rate and electrodermal activity. Over 100 hours of physiological data were collected from 12 eight-hour shifts. Initially, an 80:20 split was used on the 12 individual datasets for training and testing the model. Further, to develop a generalizable model, the data were merged, and a 60:20:20 split was used for training, validating, and testing the model. On the test set, the model achieved values of 0.98, 0.17, and 0.005 as the R-squared, RMSE, and loss for EDA data and 0.99, 0.41, and 0.002 for HR data.

## **Keywords**

Physician wellbeing, emergency department, physiological data, stress management, neural networks

## **1. Introduction**

Emergency Department (ED) is an essential patient entry point into the healthcare system that contributes approximately 50% of the hospital admission [1]. As society's healthcare safety net, patients with no other options for medical care access the ED because the federal government mandates an ED to provide screening and stabilizing care to all patients regardless of their ability to pay [2]. The total number of patients visiting EDs is increasing annually, and according to the 2016 report, over 145 million visits are made to the US EDs annually [3]. ED physicians are often exposed to high levels of stress due to the diverse nature of patient conditions and the overwhelming volume of patient visits. Prior studies have reported that stress and fatigue are important factors that contribute to medical errors and patient fatalities. Additionally, frequent and prolonged exposure to stress often leads to burnout.

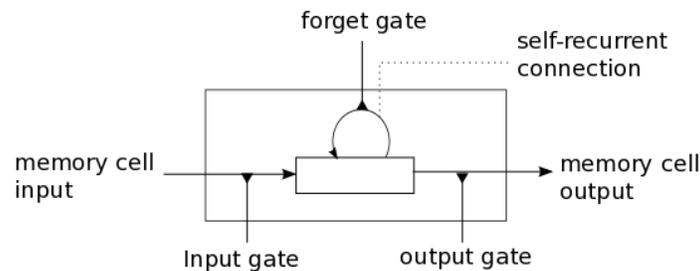
Burnout is defined as a condition of high emotional exhaustion, high depersonalization, and low personal accomplishment. The burnout rates among physicians are increasing every year, and it has been observed that physicians have a 10% higher chance of burnout compared to other general working adults [4, 5]. A 2017 survey which included 14,000 physicians from 30 different specializations, reported a 25% increase in burnout scores compared to the 2013 report [4]. Moreover, in the same survey, it was observed that 59% of the ED physicians exhibited higher levels of burnout [4]. Along with ED overcrowding, physicians have reported lengthy working hours, psychological demand, increasing implementation of EHRs, prolonged stress, and poor support as the crucial factors leading to burnout [6]–[8]. Physician burnout has a negative impact on their wellbeing, patient safety, as well as on hospital management. A study that specifically investigated the effect of burnout on physician wellbeing reported that physicians with higher levels of burnout had disturbed sleep and lower sleep quality [9]. Moreover, a survey among the ED physicians, including residents and attendings, reported that higher levels of burnout were positively correlated

with frequencies of self-reported suboptimal care, depression, and lower career satisfaction [10]. Further, hospitals have observed a higher turnover among the ED staff due to work-related stress and fatigue [11].

This research focused on predicting physician stress levels to alert the physicians and prevent them from working under prolonged stress. Stress can be estimated using various techniques, including subjective and objective measures. Few validated subjective measures include the Perceived Stress Scale (PSS) and other questionnaires, and objective measures include heart rate (HR), heart rate variability (HRV), electrodermal activity (EDA), and cortisol levels in endocrine stress response [12]–[17]. Although PSS and other questionnaires are validated methods to estimate stress levels, to monitor the stress levels continuously without interrupting the user objective measures are primarily adopted.

The current developments in Machine Learning methods provide a great opportunity to use these stress response data to predict future stress levels and prevent risks. Deep learning neural networks has been applied in various research, including image detection in healthcare, natural language processing, detection of health conditions from electronic medical records, etc. with high accuracy and outperforming the current practices [18]–[22]. Deep learning is a type of machine learning where a model is trained to predict outputs based on the inputs with the help of multiple hidden layers. Deep learning is highly efficient compared to other traditional methods because the multiple hidden layers enhance the model performance by calculating the probability of each output and updating the weights. A recent study implemented deep learning to predict in-hospital cardiac arrest, and this model significantly outperformed other methods, including the random forest algorithm and logistic regression [23].

Deep learning can be incorporated with different techniques depending on the type of input data. One of the most common approaches used in predicting temporal sequence data is Recurrent Neural Network (RNN). In RNN, unlike a typical feedforward NN, it uses their internal memory to hold the temporal behavior of the input data to predict the output. A few common RNN architectures currently used for speech recognition, time-series data prediction is Long Short-term Memory networks (LSTMs), and Gated Recurrent Unit (GRUs). LSTM is a type of RNN which can keep track of temporal behavior of the sequence without losing the long-term dependencies. The main advantage of LSTM over a traditional RNN is its ability to address the vanishing gradient problem. Vanishing gradients occur in stochastic gradient descent or any gradient-based learning methods where the NN weights are not updated as the gradient values diminish. The gradient value decreases during the backpropagation through time as the gradient values are computed by chain rule during the backpropagation. In a few cases, the vanishing gradients stop a NN from further training. Most of the time, the NN keeps training slowly but may leave out critical information from the previous sequences resulting in developing an incorrect model for prediction.



**Figure 1:** Fundamental architecture of an LSTM cell.

LSTMs addresses this issue with the help of a memory cell with gates that regulate the flow of information. **Figure 1** above shows the fundamental design of an LSTM cell without focusing on the underlying activations and mathematical complexities. An LSTM has multiple gates and cell states which manage to pass the critical information without loss. The three gates in an LSTM cell are input gate, output gate, and forget gate. The first gate in the LSTM cell is the forget gate as this gate decides how much information from the past and new input should be allowed to the input gate. The input gate is used to update the cell state where the data from the previous hidden state and new input is transferred. The cell state, which is multiplied by the forget vector forgets values close to zero, and the remaining values are added to the data from the input gate. The last gate in an LSTM cell is the output gate, which passes the new hidden state to the next LSTM cell where this process is repeated. An LSTM cell has a self-recurrent connection, as seen in **Figure 1** above. This research developed a deep learning supervised LSTM to predict the physician HR and EDA based on their current HR and EDA to help them better manage an ED shift.

## **2. Methods**

### **2.1. Participants**

Participants for this study included 12 emergency physicians (8 male, 4 female) working a 3:00 pm - 11:00 pm shift at Greenville Memorial Hospital (GMH) in Greenville, SC. The Greenville Health System (GHS), now called PRISMA Health, is the largest healthcare provider in South Carolina and serves as a tertiary referral center for the entire Upstate region. The flagship GMH academic Department of Emergency Medicine is integral to GHS patient care services as the Adult Level 1 and Pediatric Level 2 Trauma Center, Stroke, and ST Elevation Myocardial Infarction (STEMI) Comprehensive Center seeing over 106,000 patients annually. Six participants (mean age =  $26.8 \pm 1.5$  years, 4 male, 2 female) were first-year resident physicians, and the other six (mean age =  $42.66 \pm 2.8$  years, 4 male, 2 female) were attending physicians with an average experience of 8 years of practice. This particular sample set was selected to represent the diverse population of ED physicians. Consent was obtained from physicians before the shift, and the study was approved by GHS IRB Pro00058516.

### **2.2. Apparatus**

The Empatica E4 watch is a wearable research device that allows real-time physiological data acquisition. This wrist band is equipped with four types of sensors: two metallic electrodes for measuring galvanic skin response (GSR), a three-axis accelerometer, optical thermometer, and photoplethysmogram (PPG) sensors for recording the heart rate (HR). Prior researchers have validated the effectiveness of this device, and one study which specifically compared it to the medical devices used in the hospital reported that Empatica E4 echoed the data collected from the medical devices [24]. Additionally, multiple research studies have used this research device for computing stress, emotional arousal, sleep quality, and arterial fibrillation [25]–[28]. Empatica E4 collects the EDA data at a sampling rate of 4 Hz and HR data at 1 Hz.

### **2.3. Procedure and Data Processing**

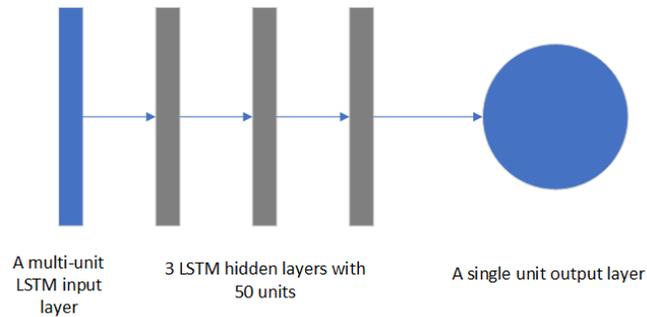
First, we identified the resident and attending physicians working an eight-hour shift in the ED. Prior to the shift, each physician was handed the consent form. The physicians were then asked to put on the Empatica E4 wristwatch at least five minutes prior to the start of their shift to obtain the baseline data. As mentioned above, Empatica E4 collects various physiological measures, including HR and EDA. Data collected using Empatica were first preprocessed for each physician separately. Initially, the data was visualized to remove the outliers and incorrect data points. Further, the missing values were interpolated using cubic spline interpolation. [29]. Following the initial data preparation, the data was standardized to address the variations in the HR and EDA data. Later, each dataset was split into an 80:20 ratio for training and testing purposes, which roughly converts to 23,040 data points for training and 5,760 data points for testing for each physician. This split was adopted for two reasons: 1) to obtain a robust training model and 2) prior studies have proved that a physician's productivity decreases as the shift progress and increases the chances of errors [29, 30]. We aimed at predicting the last 1.5 hours of the shift for each physician, which can thus help in managing the stress/fatigue experienced during the end of the shift.

Finally, after training and testing each dataset individually, the hyperparameters were further tuned to improve the accuracy of the model. To evaluate if a general model with data from multiple physicians could improve the model, the individual datasets were merged, resulting in a dataset with 345,600 data points. As each data point represented HR and EDA values for a second, two consecutive data points were averaged to reduce the dataset by half. Further, to validate and test the new model, the data were randomly split into a train, validation, and test set with 60:20:20 split. A validation set approach was adopted to address the model overfitting issue. Following the training, the model was initially fit on the validation set, and hyperparameters were further tuned and tested on the random validation set. Finally, the model was evaluated on the test set.

### **2.4. Model Architecture**

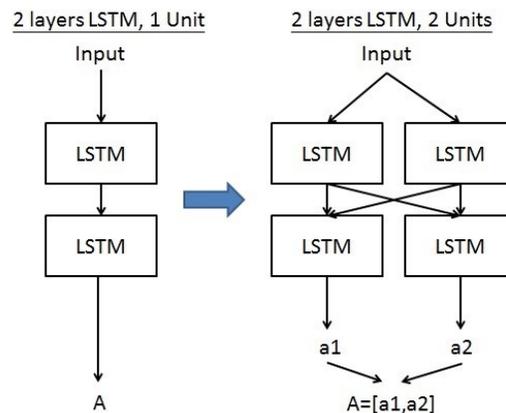
A deep learning neural network with a single input layer, three hidden layers, and a single output layer was developed. The input was a multi-unit LSTM with input channel shape similar to the training data shape, i.e., the LSTM can hold  $t-n$  steps of data in the input layer, where  $t$  denotes a data point at time  $t$  and  $n$  denotes the `look_back(n)` function. It equips the model to learn from the past  $n$  data points as input variables to predict the output variable. The output layer was designed to hold one output value. Between the input and output layer, there were three bidirectional multi-unit LSTMs. Each layer contains 50 units (25 in each direction). A dropout rate of 0.2 was applied to the final layer, and a tanh (hyperbolic tangent function) activation was used, which resulted in the outputs range from -1 to 1. The output

was later inverse transformed for deriving the HR and EDA. These values were selected from multiple model iterations, and testing on the validation set, testing set, and prior research, which built an LSTM model for similar input data [32]. **Figure 2** below shows the underlying architecture of the final model, which was used to predict the physician's HR and EDA.



**Figure 2:** Model Architecture.

Initially, we used a multi-layer single unit LSTM; however, this resulted in under fitting where the model did not capture the temporal dependencies. To address this, the hidden layers were stacked with multiple LSTM units. Although this architecture resulted in more computational time, the multiple connections between the units assured consideration of all dependencies and improved the robustness of the model resulting in a better model fit. **Figure 3** below shows the difference between a single unit LSTM and multi-unit LSTM cells and their computational differences. In this research, we used a 50-unit multi-unit LSTM with three hidden layers.



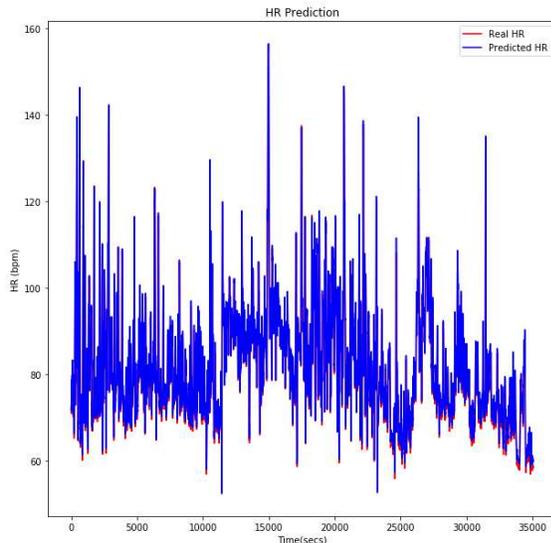
**Figure 3:** A single unit v/s multi-unit LSTM.

Additionally, in this model, we used a mean squared error method from the Keras library to compute the loss and a stochastic gradient descent algorithm: Adam. Adam is a combination of Adaptive Gradient Algorithm (AdaGrad), which maintains a per-parameter learning rate that improves performance on problems with sparse gradients, and Root Mean Square Propagation (RMSProp), which uses the same learning rate technique that is adapted based on the average of recent magnitudes of the gradients for the weight. Adam uses the benefits of both methods to result in a better algorithm. Additionally, a dropout with a probability of 0.2 was used to prevent overfitting. This rate was derived from multiple model iterations and prior study that used HR to predict cardiovascular risk [32]. Lastly, return sequences were used in this model so that the hidden state output for each input time was used for developing the model.

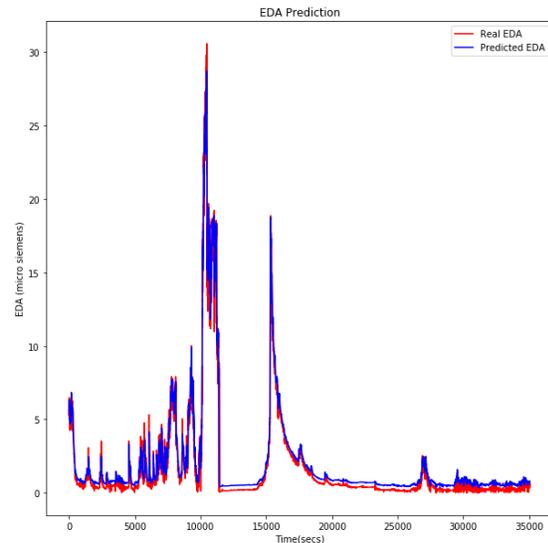
### 3. Results

This research focused on developing a multi-unit deep learning LSTM model to predict physician's HR and EDA during an ED shift. To test the model, initially, we used  $n = 60$  data points to predict the next data point for each physician. The model was run for ten epochs, using both return sequences and a dropout rate of 0.2. The computation time was around 15 minutes per dataset, and R-squared, Root Mean Square Error (RMSE), and loss were computed

post-run. The average observed R-squared, RMSE, and loss for the HR data of twelve physicians were 0.90, 0.97, and 0.004 and 0.89, 1.04, and 0.003, for the EDA data. Further, to develop and evaluate a general model, all 12 datasets were merged. Following the training, the model was validated against the validation set and evaluated on the test set. A validation set approach was adopted to address the issue of model overfitting commonly observed in the machine learning model. On the validation set, the model achieved the average values of 0.97, 0.31, and 0.004 for the R-squared, RMSE, and loss for EDA, and values of 0.99, 0.44, and 0.002 for HR. On the test set, the model achieved values of 0.98, 0.17, and 0.005 for the R-squared, RMSE, and loss for EDA, and values of 0.99, 0.41, and .002 for HR. Finally, the predicted HR and EDA values were plotted against the real HR and EDA values, as represented in **Figure 4** and **Figure 5** below. The model was able to predict with high accuracy, as seen in **Figure 4** and **Figure 5** below on the test data because of the model validation and hyperparameters tuning.



**Figure 4:** Predicted HR v/s Real HR.



**Figure 5:** Predicted EDA v/s Real EDA.

#### 4. Conclusions and Future Work

This research observed that a multi-unit deep learning LSTM could be used to develop general models for predicting heart rate and electrodermal activity that can assist physicians in better managing their shift. We observed that training the model with more participants could develop a much more generalizable model that can better estimate the HR and EDA values. Our next step is to predict heart rate variability (HRV), which provides more precise information regarding the stress. Further, we plan to utilize a questionnaire to interpret the physician's perceived stress and develop a score: stress score, which will be a function of HR, HRV, EDA, and the questionnaire. This stress score can be used to inform physicians and help them manage their shifts by taking short breaks or assigning less severe patients when stress levels are high.

#### References

- [1] Committee on the Future of Emergency Care in the United States Health System and Institute of Medicine, "Hospital-Based Emergency Care: At the Breaking Point," National Academies Press, Washington, D.C., May 2006.
- [2] Centers for Medicare & Medicaid Services (CMS), *Emergency Medical Treatment and Labor Act*. Department of Health and Human Services, 1986, p. 262.
- [3] Centers for Disease Control and Prevention, "FastStats - Emergency Department Visits," 2016. [Online]. Available: <https://www.cdc.gov/nchs/fastats/emergency-department.htm>. [Accessed: 29-Oct-2019].
- [4] C. Peckham, "Medscape Lifestyle Report 2017: Race and Ethnicity, Bias and Burnout," New York, 2017.
- [5] T. D. Shanafelt *et al.*, "Burnout and Satisfaction With Work-Life Balance Among US Physicians Relative to the General US Population," *Arch. Intern. Med.*, vol. 172, no. 18, p. 1377, Oct. 2012.
- [6] I. Bragard, G. Dupuis, and R. Fleet, "Quality of work life, burnout, and stress in emergency department physicians: A qualitative review," *Eur. J. Emerg. Med.*, vol. 22, no. 4, pp. 227–234, 2015.
- [7] R. S. Patel, R. Bachu, A. Adikay, M. Malik, and M. Shah, "Factors Related to Physician Burnout and Its

- Consequences: A Review.,” *Behav. Sci. (Basel)*, vol. 8, no. 11, pp. 1–7, Oct. 2018.
- [8] L. Kane, “Medscape National Physician Burnout, Depression & Suicide Report,” 2019. [Online]. <https://www.medscape.com/slideshow/2019-lifestyle-burnout-depression-6011056>. [Accessed: 08-Jul-2019].
- [9] A. Vela-Bueno *et al.*, “Insomnia and sleep quality among primary care physicians with low and high burnout levels,” *J. Psychosom. Res.*, vol. 64, no. 4, pp. 435–442, Apr. 2008.
- [10] D. W. Lu, S. Dresden, C. McCloskey, J. Branzetti, and M. A. Gisondi, “Impact of burnout on self-reported patient care among emergency physicians,” *West. J. Emerg. Med.*, vol. 16, no. 7, pp. 996–1001, Dec. 2015.
- [11] J.-A. V. Sawatzky and C. L. Enns, “Exploring the key predictors of retention in emergency nurses,” *J. Nurs. Manag.*, vol. 20, no. 5, pp. 696–707, Jul. 2012.
- [12] M. Birhanu, B. Gebrekidan, G. Tesefa, and M. Tareke, “Workload Determines Workplace Stress among Health Professionals Working in Felege-Hiwot Referral Hospital, Bahir Dar, Northwest Ethiopia,” *J. Environ. Public Health*, vol. 2018, 2018.
- [13] N. Gördeles Beşer, “Nursing Studies and Practice International Perceived Stress Levels of Physicians and Other Health Personnel Working in the 112 Emergency Service and Associated Factors,” 2018.
- [14] F.-X. Lesage, S. Berjot, and F. Deschamps, “Clinical stress assessment using a visual analogue scale,” *Occup. Med. (Chic. Ill)*, vol. 62, no. 8, pp. 600–605, Dec. 2012.
- [15] S. W. Porges, “Cardiac vagal tone: A physiological index of stress,” *Neurosci. Biobehav. Rev.*, vol. 19, no. 2, pp. 225–233, Jun. 1995.
- [16] R. M. Rose, “Endocrine Responses to Stressful Psychological Events,” *Psychiatr. Clin. North Am.*, vol. 3, no. 2, pp. 251–276, Aug. 1980.
- [17] V. Girishan Prabhu, K. Taaffe, R. Pirrallo, and D. Shvorin, “Stress and burnout among attending and resident physicians in the ED: a comparative study,” *IISE Trans. Healthc. Syst. Eng.*, vol. 11, no. 1, pp. 1–10, 2020.
- [18] Z. Che, D. Kale, W. Li, M. T. Bahadori, and Y. Liu, “Deep Computational Phenotyping,” in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15*, 2015, pp. 507–516.
- [19] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, “Doctor AI: Predicting Clinical Events via Recurrent Neural Networks,” in *Proceedings of the 1st Machine Learning for Healthcare Conference*, 2016, pp. 301–318.
- [20] L. Deng and D. Yu, “Deep Learning: Methods and Applications,” *Found. Trends Signal Process.*, vol. 7, no. 3, pp. 197–387, 2013.
- [21] Z. C. Lipton, D. C. Kale, C. Elkan, and R. Wetzel, “Learning to Diagnose with LSTM Recurrent Neural Networks,” *arXiv Prepr. arXiv1511.03677*, 2018.
- [22] N. Razavian, J. Marcus, and D. Sontag, “Multi-task Prediction of Disease Onsets from Longitudinal Laboratory Tests,” in *Proceedings of Machine Learning Research*, 2016, pp. 73–100.
- [23] J. M. Kwon, Y. Lee, Y. Lee, S. Lee, and J. Park, “An algorithm based on deep learning for predicting in-hospital cardiac arrest,” *J. Am. Heart Assoc.*, vol. 7, no. 13, pp. 1–11, 2018.
- [24] C. McCarthy, N. Pradhan, C. Redpath, and A. Adler, “Validation of the Empatica E4 wristband,” in *Proceedings of IEEE EMBS International Student Conference (ISC)*, 2016, pp. 1–4.
- [25] V. D. A. Corino, R. Laureanti, L. Ferranti, G. Scarpini, F. Lombardi, and L. T. Mainardi, “Detection of atrial fibrillation episodes using a wristband device,” *Physiol. Meas.*, vol. 38, no. 5, pp. 787–799, May 2017.
- [26] J. A. Onton, D. Y. Kang, and T. P. Coleman, “Visualization of whole-night sleep EEG from 2-channel mobile recording device reveals distinct deep sleep stages with differential electrodermal activity,” *Front. Hum. Neurosci.*, vol. 10, pp. 605–621, Nov. 2016.
- [27] D. Lutscher, “The relationship between skin conductance and self-reported stress : does the relationship exist and, if so, does it differ across different types of stressors?,” 2016.
- [28] M. Matsubara, O. Augereau, C. L. Sanches, and K. Kise, “Emotional arousal estimation while reading comics based on physiological signal analysis,” in *Proceedings of the 1st International Workshop on coMics ANalysis, Processing and Understanding - MANPU '16*, 2016, pp. 1–4.
- [29] M. A. Peltola, “Role of editing of R-R intervals in the analysis of heart rate variability,” *Front. Physiol.*, vol. 3, no. 148, pp. 1–10, 2012.
- [30] R. Jeanmonod, D. Jeanmonod, and R. Ngiam, “Resident productivity: does shift length matter?,” *Am. J. Emerg. Med.*, vol. 26, no. 7, pp. 789–791, Sep. 2008.
- [31] M. Silverman, “Shifting Shift Lengths: Is Shorter Better? -,” *Emergency Physicians Monthly*, 2011. .
- [32] B. Ballinger *et al.*, “DeepHeart: Semi-Supervised Sequence Learning for Cardiovascular Risk Prediction,” in *Proceedings of The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, 2018.

# Figures

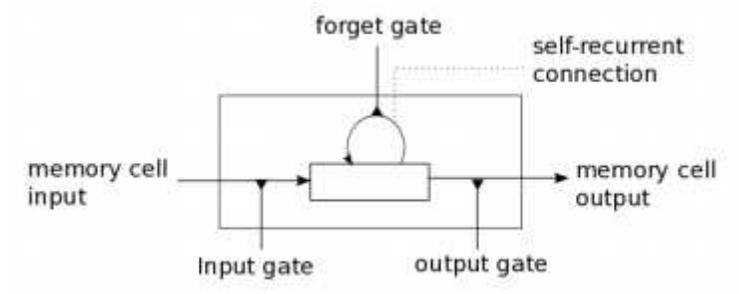


Figure 1

Fundamental architecture of an LSTM cell.

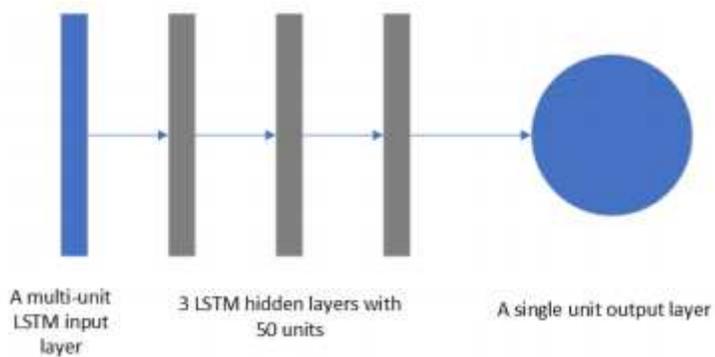


Figure 2

Model Architecture.

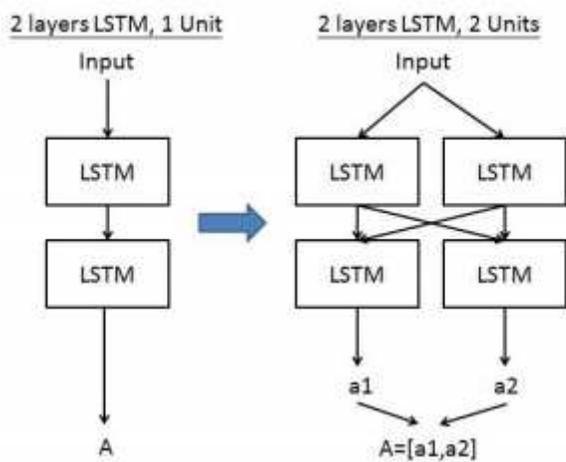
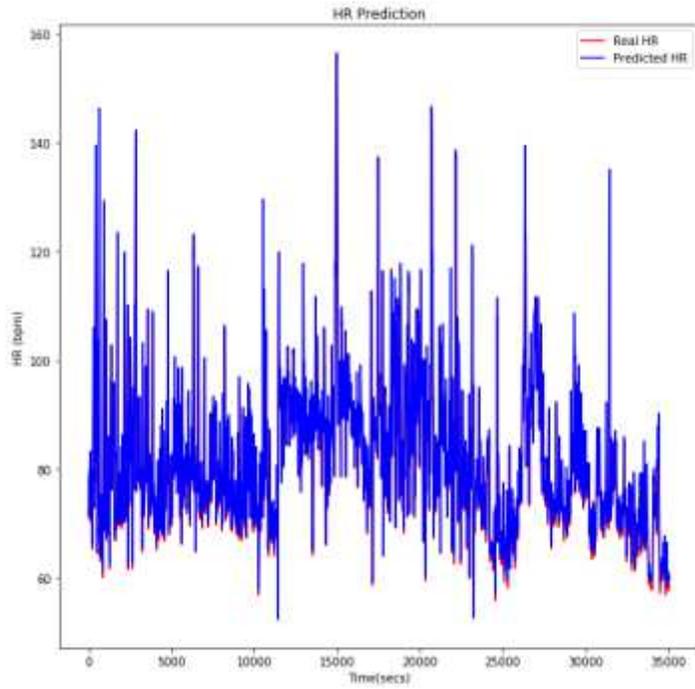


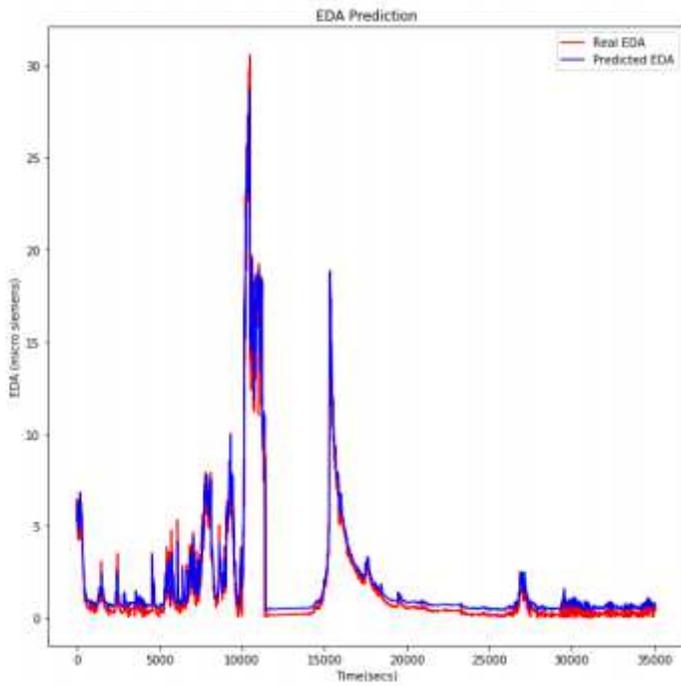
Figure 3

A single unit v/s multi-unit LSTM.



**Figure 4**

Predicted HR v/s Real HR.



**Figure 5**

Predicted EDA v/s Real EDA.