

An Improved Approach for Stress Detection Using Physiological Signals

Kushagra Nigam, Kirti Godani, Deepshi Sharma, Shikha Jain*

Jaypee Institute of Information Technology, India

Abstract

Stress is a major problem in society. Prolonged stress can lead to ill-health and a decrease in self-confidence. It is necessary to detect stress at an early stage to prevent its adverse effects on our physical and psychological health. The paper presents a stress detection model using physiological signals. In this paper, WESAD (Wearable and Stress Affect Detection) dataset is used which consists of physiological data recorded from both the chest and wrist. Further, a Long Short-Term Memory (LSTM) based model is used to detect stress. The simulation results indicate that, indeed, Electrocardiograph (ECG), Electromyogram (EMG), and Respiration (RESP) signals may not be necessary for identifying stress. A three-way validation is carried out with an accuracy of 98%. The novelty of the paper is the way time-series data is handled to make it closer to real-time data captured from sensors. The work can be used widely in clinical practices to detect stress at an early stage.

Keywords: Stress Detection, Electro dermal Activity (EDA), 3-axis Acceleration (ACC), Body Temperature (TEMP), Long Short-Term Memory Network (LSTM), Wearable and Stress Affect Detection (WESAD).

Received on 12 February 2021, accepted on 01 May 2021, published on 14 May 2021

Copyright © 2021 Kushagra Nigam *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.14-5-2021.169919

*Corresponding author. Email: shi_81@rediffmail.com

1. Introduction

Nowadays, pandemic, natural disasters and economic recession are affecting our mental health negatively. It is leading to stress (Kaur et al, 2019; Gautam et al, 2020; Sharma et al, 2020; Ausin et el, 2021; Khalaf, 2020; Daniels et al, 2020) in our lives which is causing various psychological illness such as anxiety, depression, eating disorders, attention deficit disorder etc. (Olf et. al, 2020) and making the existing problems even worse. Moreover, it is affecting our physical health as well.

Stress can be defined as a response to situations that involve a high workload or demands high psychological thinking. In the current scenario, every other person can be seen in stress. If this problem is not dealt with, it might prompt real well-being issues at a later stage.

Furthermore, in most exceedingly bad situations, it may lead to even deaths. Recently, researchers observed a strong relationship between physical health and psychological health of an individual (Picard, 2016; Liapis et al., 2020). They observed that physiological signals are a good indicator of our emotional well-being. Hence, these signals can be used for periodic monitoring of our mental health including stress, anxiety, and depression. The applications of stress detection systems lie in various domains such as among employees, students, drivers, passengers, doctors, patients, job applicants, and many more.

The methods proposed in the past researches to detect stress depend on self-reporting diary as ground

truth that is to be filled by subjects. The drawback is stressed people don't fill up diaries with interest accurately. Thus, inconspicuous and accurate systems are required to identify stress in people for improving health. Researchers proposed different approaches in the past to recognize stress automatically (Carneiro, et al., 2012; Cho, et al., 2017; Gjoreski, et al., 2020; Lin, et al., 2017; Kurniawan, et al, 2013; Vizer, et al. 2009).

Though lots of research has been carried out in the field of automatic stress detection, the research is still far away from good accuracy while being close to real scenarios (Can, et al., 2019). In this paper, a physiological signal based dataset (WESAD) is used along with LSTM to design a stress detection model. To reduce the complexity of the model, first, a rigorous preprocessing is applied to clean the dataset and pick the most relevant features only.

The paper is organized as follows: Section 2 discusses the state-of-art. Proposed work is explained in detail in Section 3 followed by implementation and results in Section 4. Finally, the paper is discussed, concluded with future directions in Section 5 and 6.

2. Related Work

To detect stress in people, a considerable amount of work and researches have been done already in the past. Traditionally the researchers have done a considerable amount of work by anticipating the approach of stress detection which includes stress during day and night through smart devices (Muaremi, et al., 2013). The subject's data was recorded using mobiles. Using Smartphone sensors, variations in self-perception of the stress of people were recorded. Subjects were asked to fill questionnaires daily to have data. Voice recordings were also analyzed to measure stress. After experiments, HRV features were considered more important. However many more measures such as readings through biological sensors impact impeccably in recognizing stress-level. Considering this, there are several interesting works done in recognizing stress autonomously. Lin et al. (2019) used a deep fusion network for recognizing stress. LDA and Adaboost classifiers have also been used. Fusion schemes (early and late fusion) were framed and results show late fusion was superior to early fusion.

Some have used sensors available in commercial smartwatches to identify stress (Siirtola, 2019). A comparison was made between sensors in different operating systems available. Tizen OS was better than WearOS and WatchOS. The maximum accuracy achieved was 84% when the window size was 120s with an SVM classifier. The authors (Singh & Queyam, 2013) proposed a novel method for detecting stress in

automobile drivers where stress function was used as a classifier. They took ECG, EMG, RR, GSR physiological data of 10 drivers for building the system. The problem was dealt with as 3- class classification with classes as high, moderate, and low stress. The highest accuracy of 88% was achieved when the ANN classifier was employed for the task. Eye blinks and brain activities (Haak et al., 2009) were used in determining stress. Eye blinks were recorded while moving on streets with straight roads, curved roads, and crashes on a curved track. It was concluded that blink in eyes and spontaneous activity in certain events point towards stress. Authors (Padmaja et al, 2018) made use of wireless physical activity trackers to detect stress. 10 user's in IT Company's data were recorded using the FITBIT device and an online questionnaire. Their sleeping patterns, BMI, Heart rate data, and physical activity were analyzed using a mobile app. The output will be as the class low, moderate, and high level of stress. Build models were compared based on AIC. Sociometric and physiological sensors were used by Mozos, et al. (2017) for the task. These sensors gave data of EDA, PPG, HRV, social activity, daily movement, and voice. The problem was dealt with as a binary classification task. For ground truth, TSST, STAI questionnaires were employed. KNN, SVM, and Adaboost were used as classifiers (Dayalin & Anuja, 2014; Schmidt, et al, 2018; Zhai, et al., 2005; Wang, et al. 2013; Patil & Hansen, 2010). Accuracy differs in a range from 84% to 94%. Also, Sun, et al. (2010) detect stress in car drivers. From ECG data of drivers, QRS was extracted; the breathing rate was taken from Respiration, etc. ECG, RESP, and EEG were taken to identify stress in car drivers. Local Binary Pattern was used for preprocessing before features can be used for the task. The only SVM was employed for binary-classification with the class as abnormal/normal. Rostgoo et al., (2019) detected stress for drivers using LSTM and CNN and achieved an accuracy of 92.8% only. Umematsu et al. (2019) predicted the next day stress level among students based upon previous 7 days data and achieved accuracy of 83.6%.

Though, lots of research has been carried out in the field of automatic stress detection, the research is still far away from good accuracy while being close to real scenario.

3. Methods

This paper proposes an LSTM based stress detection model using physiological signals. As discussed in the previous section, the stress cannot be detected based upon the values of physiological signals

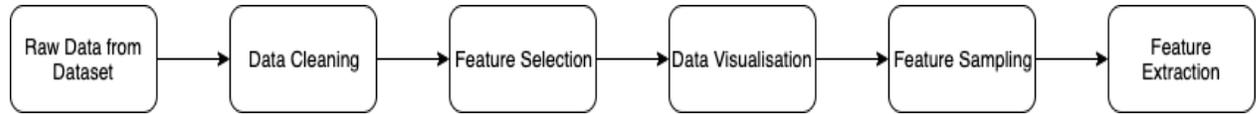


Fig. 1. Steps to get data fit for training

captured at an instance of time. Rather, continuous monitoring of these values is required to detect the stress accurately. Hence, the proposed model is a continuous model that takes the sequential physiological signals while detecting stress. Hence, LSTM is used for this purpose. The steps followed to make data fit for training is depicted in Figure 1.

3.1 Dataset: WESAD-Wearable and Stress Affect Detection

WESAD (Schmidt, et al., 2018) is a freely accessible labeled dataset consists of physiological data recorded from both the chest and wrist. It consists of data of 15 subjects (individuals) who were part of the experiment. Each of the subject data is provided with their gender, weight, height, and some relevant common questions which were asked before the experiment. In total, the study was conducted for about two hours. The WESAD dataset deals with a study of an individual's stressful state by making subjects proceed with several procedures. Based on the sequence of the procedure, subjects are segregated into 2 groups: version A and version B (Figure 2). Version A needs to follow the pattern of Baseline-Amusement- Meditation- Stress- Rest- Meditation and version B comprising Baseline- Stress- Rest- Meditation-Amusement- Meditation. Accordingly, samples are labeled as baseline, stress, amusement, and meditation.

For the data collection, they have used both a chest-worn device (RespiBAN) and a wrist-worn device (Empatica E4). A RespiBAN Professional and an Empatica E4, respectively. The RespiBAN is equipped with sensors to measure electrocardiogram (ECG), electro dermal activity (EDA), electromyogram (EMG), respiration, body temperature, and three-axis acceleration (ACCx, ACCy, ACCz). All signals are sampled at 700 Hz. The RespiBan was placed around the subject's chest. The RESP is recorded via a respiration inductive plethysmograph sensor. The ECG data was recorded via a standard three-point ECG. To allow the subject to move as freely as possible, the EDA signal was recorded on the rectus and the TEMP sensor was placed on the sternum. The EMG data were recorded on the upper trapezius muscle on both sides of the spine. To avoid wireless packet loss, the recorded data was stored locally and transferred to a computer for further processing after the experiment. All subjects wore the Empatica E4 on their non-

dominant hand. The E4 records blood volume pulse (BVP, 64 Hz), electro dermal activity (EDA, 4 Hz), body temperature (4 Hz), and three-axis acceleration (32 Hz).

3.2 Data Cleaning

Data cleaning is the way towards distinguishing and adjusting (or removing) off base records from a dataset. It includes recognizing inadequate, wrong, erroneous, or unimportant sections of the data and erasing it. Here, a two-level data cleaning process is carried out. In literature, it is found that the physiological signals obtained from the chest-worn device are more accurate compared to the physiological signals obtained from the wrist-worn device (Shi, et al., 2010). Therefore, at the first level, in the proposed model, the data captured from the chest-worn device i.e. RespiBAN is considered. Hence, the dataset with 6 modalities (EMG, ECG, RESP, TEMP, EDA, 3 axis ACC) is taken care of.

Moreover, there are four classes in the WESAD dataset namely: baseline, stress, amusement, and meditation. Since the objective of the proposed work is to detect stress, we used two-class data from the dataset, i.e., baseline (neutral) and stress. Hence, all the records which correspond to amusement and meditation have been eliminated.

Then, the data is segregated into 2 parts:

- Records of each person when he/she is in stress.
- Records of each person where he/she is not in stress.

It is found that there are more records of data when the person is in stress compared to a person when he/she is not in stress for each subject.

3.3 Feature Selection

Feature selection is the process of selecting the most relevant subset of features that can be used in further processing. Having superfluous features in the data can diminish the exactness of the models and cause the models to learn dependent on unessential features.

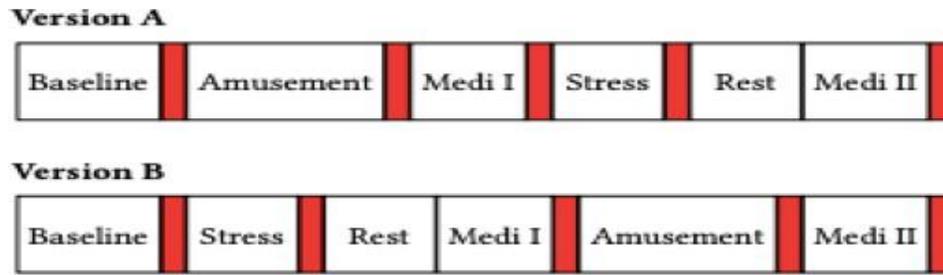


Fig. 2. The two different versions of the study protocol (Adapted from (Schmidt, et al, 2018))

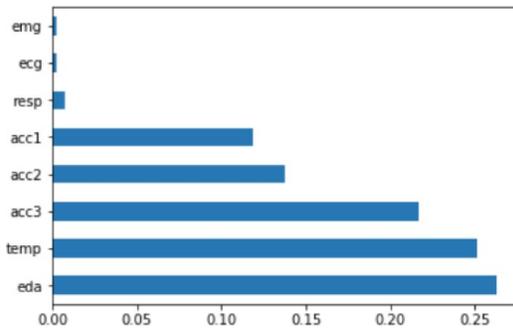


Fig. 3. Features Importance.

Therefore, it is necessary to select only those features which are most accountable for the prediction. For feature selection, we have used Extra Tree Classifier (ETC). The features taken into account by ETC are ECG, EDA, EMG, TEMP, RESP, ACCx, ACCy, ACCz. It ranked the features according to their importance as shown in Figure 3. Here, X-axis represents the various features and Y-axis represents the importance parameter. It is observed that EMG, ECG, and RESP are the least important features and hence can safely be discarded.

Therefore, the selected modalities for feature extraction are:

- ACC (ACCx, ACCy, ACCz)
- TEMP
- EDA

3.4 Data Visualisation

Data visualization deals with graphical interpretation of data by plotting graphs, visualizing charts, and mapping data point elements. It provides an approachable way to understand patterns and depict outliers in data. The graphs have been plotted for each subject data corresponding to time which helps to classify subjects and their versions (whether the subject belongs to version A or version B) as mentioned in Figure 2. All 15 subjects have been plotted in Figure 4 to identify version A and version B subjects. Each graph corresponds to one subject.

The plot is between class label that is the baseline (1), stress (2), amusement (3), and meditation (4) with respect to time (Y-axis). It is necessary to know which subject belongs to which version as it is not stated in [16] so that while taking a cluster of subjects for training and testing, not all subjects should belong to a single version. If this becomes the case, then the testing accuracy can be affected.

3.5 Feature Generation

Feature generation deals with dimensional reduction by which raw data is reduced to more segments for further processing. Large datasets contain enormous data that require techniques for computing such a huge dataset. Aggregating the features used in signal segments and combining them for classification serves as an initial step for pre-processing. Extracted features are grouped in several possible ways like time- or frequency-domain features, linear or non-linear features, uni-modal or multi-modal features, etc. Considering statistical features (like dynamic range, mean, standard deviation, etc.) are feature dependent on complex features.

In the proposed model, the following features are used for ACC as there are 3-axis X, Y, and Z:

- *maxX* – Maximum of X-axis
- *maxY* – Maximum of Y-axis
- *maxZ* – Maximum of Z-axis
- *mean* – the combined mean of x, y, and z-axis that is $(\text{mean } x + \text{mean } y + \text{mean } z)/3$.
- *std* – Standard deviation

For TEMP and EDA, we include the following features:

- *max* – maximum of the feature
- *min* – minimum of the feature
- *mean* – mean of the feature
- *range* – Dynamic range of the feature (max-min)
- *std* – Standard deviation

So, it resulted in a dataset with 5 attributes each for EDA, RESP, and ACC. Also, samples are from 15 subjects and labeled with two classes: stress and baseline (neutral).

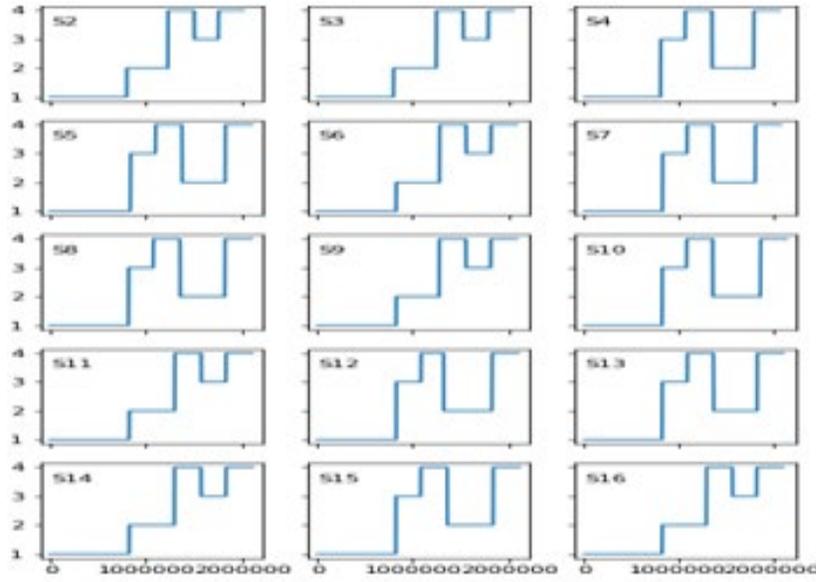


Fig. 4. Baseline (1), Stress (2), Meditation (3) and Amusement (4) vs Time plot for each subject.

According to Schmidt, et al., (2018), each subject is subjected to 20 minutes of baseline and 10 minutes of stressed tasks. Therefore:

- Number of records for class Baseline: 800 thousand approximately.
- Number of records for class Stress: 400 thousand approximately.

The concept of window size and the shift are depicted in Figure 5. Since, we have unequal number of data in both the classes, keeping the window size and the shift same for both class may lead to more records of baseline data than the records of stressed data which can create further unbalancing. To balance the classes, the window size of 42000 (as all samples are recorded at 700Hz, hence $(1/700) * 42000$ gives a time window of 1 minute) with a window shift of 175 (0.25 seconds) is considered for computing the features corresponding to the records of baseline and similarly, for class stress, a window size of 42000 and window shift of 105 (0.15 seconds) was considered. There is a difference in the window shifts between the two to get an almost equal number of samples from baseline and stress data. The concept of window size and the shift are depicted in Figure 5. If window size and the shift are kept constant, then definitely records of baseline data will be more than the records of stressed data and it may lead to outliers in the model. So, the number of samples considered for each subject for class baseline as well as the class stress is approximately 4500. However, by approximation, it can be in the range 3500–4800. Hence, feature sampling was applied so that data from each of the 15 subjects may get an equal number of samples.

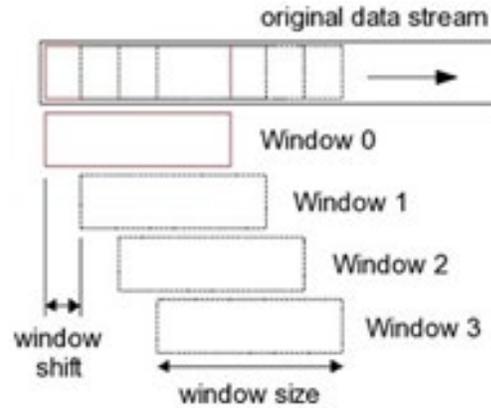


Fig. 5. Concept of window size and window shift.

3.6 FEATURE SAMPLING

As mentioned in the previous section, approximately 4800 samples were taken for each subject. Hence feature sampling is used. The values are up-sampled according to the demand of the dataset. 4800 samples were chosen as it is closest to all the records computed by each subject. If the data is lesser than 4800 the data samples were up-sampled and substituted by mean of each column. Bringing the uniformity in the number of columns is achieved for better interpretation of data points.

After feature sampling, the data is ready for analysis. Figure 6 depicts the procedure followed for training and testing the model.

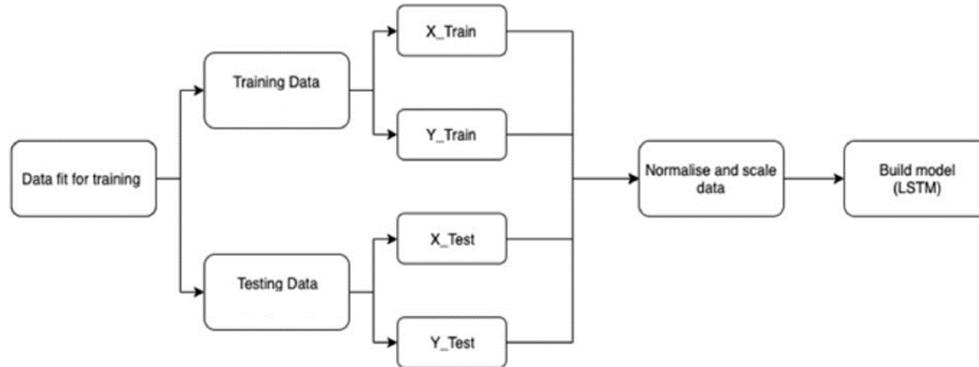


Fig. 6. Steps for model training

4. Implementation and Results

After preprocessing, we have created a single data-frame consisting of 72,000 (4800*15) baseline records followed by 72,000 (4800*15) stress records. This gives a total of 1,44,000 records. The baseline and stress labels are assigned 0 and 1 respectively. Figure 6 depicts the procedure followed for training and testing the model. The data-frame is divided into two parts-X and Y, where X contains an array of data-points and Y contains a single dimension array of 0(Baseline) and 1(Stress). X is further divided into two parts X Train and X Test for training and testing respectively. Similarly, Y is also divided into Y Train and Y Test. More emphasis is given to consider the samples sequentially which preserves the dependency of current value over the previous value. To ensure this, special care has been taken to train the model subject-wise, i.e., feed the data of one subject (means 4800 sample) at once for training. By training the model subject-wise, it is meant that first the subject 1 is trained and carried forward then the second subject is trained and so on for all the subjects.

4.1 Recurrent Neural Networks

Recurrent Neural Networks are applied to sequential data. They have nodes similar to neurons organized into a network of layers. They store what has been processed before, so they are said to possess 'memory'. The input in the current state is the output from the previous state in the network. They are called recurrent because they recurrently process each item in the sequence of data. They pose a problem of exploding and vanishing gradient which is solved by LSTMs by introducing some modifications in the basic structure of RNN.

The Problem of Long-Term Dependencies: It may be possible that due to RNN being used, past information is connected to current information. Sometimes, it is required to examine recent information to perform the present task which is exactly in our case.

And therefore, Long Short-Term Memory networks (LSTM) will be suitable for training our dataset.

4.2 LSTM Networks

LSTM networks are built to overcome the problem of vanishing and exploding gradient posed by RNNs during their training. It solves the problem as it allows the gradients to flow without change. This type of neural network architecture has the ability not only to process single data points but large sequences of data. One single unit of LSTM includes a forget gate, an input gate, an output gate, and a cell as shown in Figure 7. These gates altogether help in regulating the flow of data in and out of the cell. Input gate checks what information is going as input to the cell i.e. how much new values pass into the cell, forget gate is to check which information is not to be kept from previous cell state, and output gate is to take decision what data is to be passed to next state. Forget gate uses a sigmoid activation function.

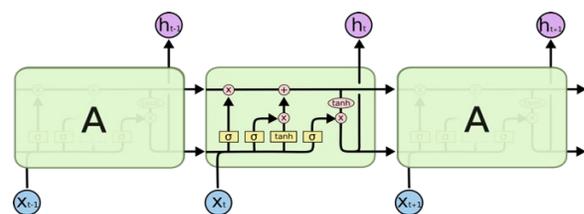


Fig. 7. The repeating module in an LSTM contains four interacting layers (Adapted from (Zhai et al., 2005)).

4.3 Training and testing of the model

For training, we designed a sequential LSTM network with 1 hidden layer and 15 neurons (as there are 15 attributes in our dataset). We applied 3 types of classification:

1. Randomly picking 80% samples for the training set and remaining for the testing set.
2. Training the model using 14 subjects (which includes 8 persons from version A and 6 persons from version B) and testing it with the samples of 15th subject (who belong to version B).
3. Training the model using 13 subjects (which includes 7 persons from version A and 6 persons from version B) and testing it with the samples of the remaining 2 persons (one from version A and the other from version B category).

The first type of classification is the traditional approach which is used in (Carneiro, et al., 2012; Vizer, et al. 2009; Cho, et al., 2017; Lin, et al., 2017; Kurniawan, et al, 2013). In this type of approach, samples are randomly picked from the dataset. In this case, the accuracy of training and testing can be high for such a scenario, however, it treats each sample as independent data. Though the accuracy is high, its behavior does not match the real-world scenario.

The second and third types of classification are novel and necessary. Here, instead of picking up the samples randomly, the samples are considered subject-wise. All the samples of a subject are fed into the model in a sequence. As a result, the model is trained first with a sequence of data of one subject. Then, the next subject is considered for training, and so on. In the second type, 14 subjects are considered for training, and the remaining one is used for testing. In the third type, 13 subjects are considered for training and the remaining two are used for testing. The two test subjects are selected such that one belongs to version A and second belongs to version B.

Further, experimentation is performed by varying the hyper-parameters like batch size and number of epochs to observe their effect on training. It is carried out for all 3 types of classification.

- Batch size = 4, Number of Epochs = 1
- Batch size = 32, Number of Epochs = 5
- Batch size = 32, Early stopping used for epochs

Early Stopping is used to avoid over-training of the neural networks. Following some criteria, it restricts the number of epochs after some time. The criterion that has been used for early stopping for model training is to minimize the validation loss while training in each epoch.

Batch size of 32 or 4 is chosen as it is a factor 4800, the number of stress and baseline samples in each subject. By batch size of 32, the algorithm takes the first 32 samples (from 1st to 32) from the training dataset and trains the network. Next, it takes the second 32 samples (from 33rd to 64th) and trains the network again. The

same procedure is repeated until all samples have been propagated through our network. It helps in avoiding the mixing of samples of different subjects. Further, the sigmoid activation function is used for the output layer using loss function as binary cross-entropy.

4.4 Results

A detailed evaluation of testing is demonstrated in Table 1. All the types of classification that are discussed in the previous sub-section give promising results. This is by far the best accuracy that has been achieved in the binary classification of stress vs baseline. Each of the types of classification gives the best training accuracy of 98%, 99%, and 99% for type 1, type 2, and type 3 respectively.

- The best testing accuracy for type 1 classification is 98% with a batch size of 32 and the number of epochs equal to 31(early stopped).
- The best testing accuracy for type 2 classifications is 98% with a batch size of 32 and the number of epochs equal to 8(early stopped).
- The best testing accuracy for type 3 classifications is 93% with a batch size of 32 and the number of epochs equal to 40 (early stopped).

The plots of training and testing accuracy vs. the number of epochs and training and testing loss vs. the number of epochs for type1, type 2, and type 3 classifications are depicted in Figure 8, Figure 9, and Figure 10 respectively. It can be observed that type 2 (Figure 9) is the best among all others in terms of accuracy, loss and number of epochs as compared to type 1 (Figure 8) and type 3 (Figure 10). Moreover, it is closer to the human pattern of stress. Just by looking at one instance of values of different parameters, one cannot be declared in stress. Rather, monitoring is to be done over a period of time and the collective analysis of all those values helps in concluding whether a person is in stress or not. This is what is happening exactly in type 2 and type 3 classification.

5. Discussion

In this section, we compared the existing models with the proposed one (Table 2). It is observed that in the base paper (Schmidt, et al., 2018), number of experiments are performed and maximum accuracy achieved is 92.83% with LDA. Later, in 2019, another experiment (Gjoreski, et al., 2019) is performed which reported accuracy of 97.55% with EMI-Fast GRNN. Then, Indikawati and Winiarti (2020) designed a stress detection model and

achieved an accuracy of 88% to 99% with random forest. However, none of them considers the data as time-series data.

Table 1.
EVALUATION OF THE 3 TYPES

| Type of Classification | Criteria | | Training Accuracy | Testing Accuracy | Recall |
|------------------------|------------|--------------------------|-------------------|------------------|-------------|
| | Batch Size | Epochs | | | |
| Type 1 | 4 | 1 | 92% | 90% | 0.92 |
| | 32 | 5 | 94% | 90% | 0.89 |
| | 32 | 31 (with early stopping) | 98% | 98% | 0.99 |
| Type 2 | 4 | 1 | 99% | 97% | 0.97 |
| | 32 | 5 | 98% | 96% | 0.96 |
| | 32 | 8 (with early stopping) | 99% | 98% | 0.98 |
| Type 3 | 4 | 1 | 84% | 84% | 0.84 |
| | 32 | 5 | 90% | 75% | 0.74 |
| | 32 | 40 (with early stopping) | 99% | 93% | 0.87 |

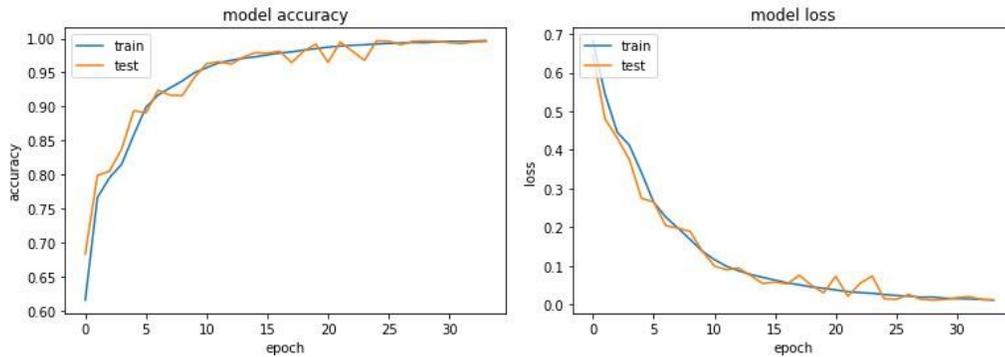


Fig. 8. Accuracy vs Epochs and Loss vs Epochs for training and testing of type 1.

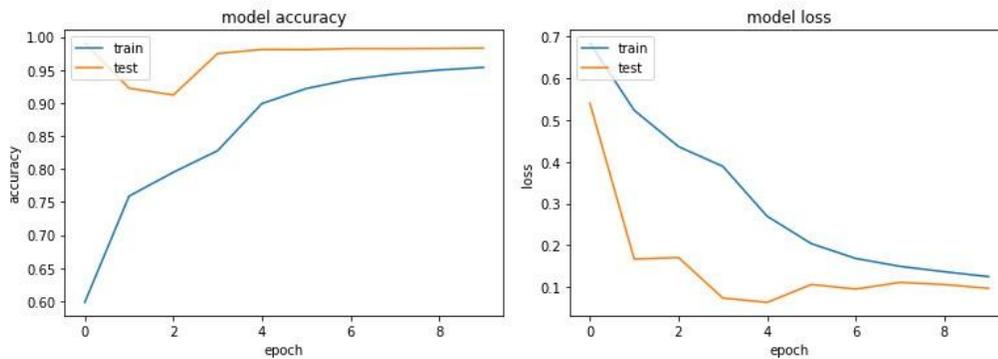


Fig. 9. Accuracy vs Epochs and Loss vs Epochs for training and testing of type 2.

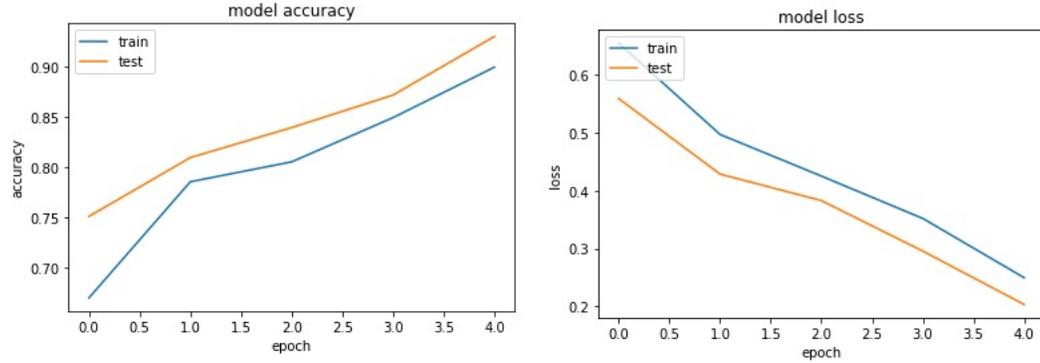


Fig. 10. Accuracy vs. Epochs and Loss vs. Epochs for training and testing of type 3.

Table 2. COMPARISON OF PROPOSED MODEL WITH EXISTING MODELS

| | Schmidt, et al., 2018 | Gjoreski, et al., 2019 | Indikawati and Winiarti, 2020 | Proposed Model |
|---|--|-----------------------------|---|----------------|
| Dataset | WESAD | WESAD, SWELL-KW and DREAMER | WESAD | WESAD |
| Considered Time-Series property? | No | No | No | Yes |
| Accuracy | 92.83% with Linear discriminant Analysis | 97.55% with EMI-Fast GRNN | 88% to 99% with random forest on 15 subjects that means accuracy is not consistent and varying with subjects. | 98% |

6. Conclusion

Nowadays stress is a part of our life. Dealing it at an early stage may avoid a lot many further problem. In this paper, we propose an automatic stress detection model using LSTM. The model used the WESAD dataset containing physiological signals of 15 subjects for training and testing purposes. Utmost care has been taken to preserve the human body pattern and hence preserved the sequence of the given dataset. Further, a thorough and vigorous preprocessing is applied followed by 3 types of classifications. The LSTM model provides a great possibility of exhibiting temporal dynamic behavior. All 3 types of classification using LSTM have shown promising results. Furthermore, type 2 is the best one having good testing accuracy of 98% with only 8 epoch. Although the results are promising, the LSTM model helps to categorize into stress and non-stress states using binary classification.

Future Directives: The scope of applying model as 4-class classifications will turn into a massive advantage

in terms of a better understanding of various emotions. Predicting Body Mass Index and differentiating the stress levels of men and women would add depth to the project. Working on different transition states of emotion will broaden the perspective of the study. Real-time testing of models with the available wrist sensors is still a challenge to cope with, but currently, biological sensors are proving to be suitable for analysis.

References

- [1]. Can, Y. S., Arnrich, B., & Ersoy, C. (2019). Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. *Journal of biomedical informatics*, 92, 103139.
- [2]. Carneiro, D., Castillo, J. C., Novais, P., Fernández-Caballero, A., & Neves, J. (2012). Multimodal behavioral analysis for non-invasive stress detection. *Expert Systems with Applications*, 39(18), 13376-13389.

- [3]. Chakraborty, S., Aich, S., Joo, M. I., Sain, M., & Kim, H. C. (2019). A Multichannel Convolutional Neural Network Architecture for the Detection of the State of Mind Using Physiological Signals from Wearable Devices. *Journal of healthcare engineering*, 2019.
- [4]. Cho, Y., Bianchi-Berthouze, N., & Julier, S. J. (2017, October). DeepBreath: Deep learning of breathing patterns for automatic stress recognition using low-cost thermal imaging in unconstrained settings. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 456-463). IEEE.
- [5]. Dayalin Subi, J., & Anuja, H. S. (2014). Detection of Automobile Drivers Stress from Physiological Signals.
- [6]. Gjoreski, M.; Kolenik, T.; Knez, T.; Luštrek, M.; Gams, M.; Gjoreski, H.; Pejović, V. Datasets for Cognitive Load Inference Using Wearable Sensors and Psychological Traits. *Appl. Sci.* 2020, *10*, 3843.
- [7]. Haak, M., Bos, S., Panic, S., & Rothkrantz, L. J. M. (2009). Detecting stress using eye blinks and brain activity from EEG signals. *Proceeding of the 1st driver car interaction and interface (DCII 2008)*, 35-60.
- [8]. <https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/>, December 2017.
- [9]. Indikawati FI, Winiarti S.(2020), Stress Detection from Multimodal Wearable Sensor Data. *IOP Conf. Series: Materials Science and Engineering* 771.
- [10]. Kurniawan, H., Maslov, A. V., & Pechenizkiy, M. (2013, June). Stress detection from speech and galvanic skin response signals. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems* (pp. 209-214). IEEE.
- [11]. Liapis, A., Katsanos, C., Karousos, N., Xenos, M., & Orphanoudakis, T. (2020). User Experience Evaluation: A Validation Study of a Tool-based Approach for Automatic Stress Detection Using Physiological Signals. *International Journal of Human-Computer Interaction*, 1-14.
- [12]. Lin, H., Jia, J., Qiu, J., Zhang, Y., Shen, G., Xie, L., & Chua, T. S. (2017). Detecting stress based on social interactions in social networks. *IEEE Transactions on Knowledge and Data Engineering*, 29(9), 1820-1833.
- [13]. Lin, J., Pan, S., Lee, C. S., & Oviatt, S. (2019, November). An Explainable Deep Fusion Network for Affect Recognition Using Physiological Signals. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 2069-2072).
- [14]. Mozos, O. M., Sandulescu, V., Andrews, S., Ellis, D., Bellotto, N., Dobrescu, R., & Ferrandez, J. M. (2017). Stress detection using wearable physiological and sociometric sensors. *International journal of neural systems*, 27(02), 1650041.
- [15]. Muaremi, A., Arnrich, B., & Tröster, G. (2013). Towards measuring stress with smartphones and wearable devices during workday and sleep. *BioNanoScience*, 3(2), 172-183.
- [16]. Padmaja, B., Prasad, V. R., & Sunitha, K. V. (2018). A Machine Learning Approach for Stress Detection using a Wireless Physical Activity Tracker. *Int. J. Mach. Learn. Comput*, 8, 33-38.
- [17]. Patil, S. A., & Hansen, J. H. (2010). The physiological microphone (PMIC): A competitive alternative for speaker assessment in stress detection and speaker verification. *Speech Communication*, 52(4), 327-340.
- [18]. Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. (2018). Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18)*. Association for Computing Machinery, New York, NY, USA, 400-408.
- [19]. Picard R.W. (2016), Automating the recognition of stress and emotion: from lab to real-world impact, *IEEE MultiMedia*, 23 (3), pp. 3-7
- [20]. Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., & Van Laerhoven, K. (2018, October). Introducing WESAD, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction* (pp. 400-408).
- [21]. Shi, Y., Nguyen, M. H., Blitz, P., French, B., Fisk, S., De la Torre, F., ... & Kamarck, T. (2010, June). Personalized stress detection from physiological measurements. In *International symposium on quality of life technology* (pp. 28-29).
- [22]. Siirtola, P. (2019, September). Continuous stress detection using the sensors of commercial smartwatch. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (pp. 1198-1201).
- [23]. Singh, M., & Queyam, A. B. (2013). A novel method of stress detection using physiological measurements of automobile drivers. *International Journal of Electronics Engineering*, 5(2), 13-20.

- [24]. Sun, F. T., Kuo, C., Cheng, H. T., Buthpitiya, S., Collins, P., & Griss, M. (2010, October). Activity-aware mental stress detection using physiological sensors. In International conference on Mobile computing, applications, and services (pp. 282-301). Springer, Berlin, Heidelberg.
- [25]. Vizer, L. M., Zhou, L., & Sears, A. (2009). Automated stress detection using keystroke and linguistic features: An exploratory study. *International Journal of Human-Computer Studies*, 67(10), 870-886.
- [26]. Wang, J. S., Lin, C. W., & Yang, Y. T. C. (2013). A k-nearest-neighbor classifier with heart rate variability feature-based transformation algorithm for driving stress recognition. *Neurocomputing*, 116, 136-143.
- [27]. Zhai, J., Barreto, A. B., Chin, C., & Li, C. (2005, April). Realization of stress detection using psychophysiological signals for improvement of human-computer interactions. In *Proceedings. IEEE SoutheastCon, 2005.* (pp. 415-420). IEEE.
- [28]. Rastgoo, M. N., Nakisa, B., Maire, F., Rakotonirainy, A., & Chandran, V. (2019). Automatic driver stress level classification using multimodal deep learning. *Expert Systems with Applications*, 138, 112793.
- [29]. Umematsu, T., Sano, A., Taylor, S., & Picard, R. W. (2019, May). Improving Students' Daily Life Stress Forecasting using LSTM Neural Networks. In *2019 IEEE EMBS IC on Biomedical & Health Informatics (BHI)* (pp. 1-4). IEEE.
- [30]. Kaur, Prableen, and Manik Sharma. "Diagnosis of human psychological disorders using supervised learning and nature-inspired computing techniques: a meta-analysis." *Journal of medical systems* 43.7 (2019): 1-30.
- [31]. Gautam, Ritu, and Manik Sharma. "2019-nCoV pandemic: A disruptive and stressful atmosphere for Indian academic fraternity." *Brain, behavior, and immunity* (2020).
- [32]. Sharma, Samriti, Manik Sharma, and Gurvinder Singh. "A chaotic and stressed environment for 2019-nCoV suspected, infected and other people in India: Fear of mass destruction and causality." *Asian journal of psychiatry* 51 (2020): 102049.
- [33]. Olf, Miranda, et al. "Screening for consequences of trauma—an update on the global collaboration on traumatic stress." *European journal of psychotraumatology* 11.1 (2020): 1752504.
- [34]. Ausín, Berta, et al. "Gender-related differences in the psychological impact of confinement as a consequence of COVID-19 in Spain." *Journal of Gender Studies* 30.1 (2021): 29-38.
- [35]. Khalaf, Mustafa Ali. "Psychological consequences of COVID-19 and challenges for post-traumatic interventions." *Journal of Psychology Research* 10 (2020): 24-29.
- [36]. Daniels, Teresa E., Elizabeth M. Olsen, and Audrey R. Tyrka. "Stress and psychiatric disorders: The role of mitochondria." *Annual review of clinical psychology* 16 (2020): 165-186.