ICIC Express Letters
Volume **17**, Number **1**, January **2023**

pp. **31–39**

DETECTING STRESS WITH ELECTROCARDIOGRAM SIGNAL USING THE CONVOLUTIONAL NEURAL NETWORK AND LONG SHORT-TERM MEMORY

Arvin Melvillo and Sani Muhamad Isa

Computer Science Department, BINUS Graduate Program – Master of Computer Science
Bina Nusantara University
JL. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
{ arvin.melvillo; sani.m.isa }@binus.ac.id

Received January 2022; accepted April 2022

Abstract. *Stress is something common in our life nowadays. Because physiological signals may identify stress, it is a great way to keep an eye on it. One of the most common signals used in the medical field was Electrocardiogram (ECG). That means ECG can be used to detect stress. ECG can be extracted to Heart Rate Variability (HRV) and HRV is also proved that can detect stress effectively. ECG signal will be acquired from the WESAD dataset which contains another physiological signal. This study proposes a simpler Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) model to detect stress using HRV since HRV was extracted from an ECG signal. The model was trained using the HRV time domain and frequency domain. The model also will be compared to the CNN model and LSTM model while using other types of data such as time domain, frequency domain, and raw ECG data. The model's highest accuracy achieved 97.07% with time domain and frequency domain data. The result demonstrated that a simpler deep learning model with extracted data can achieve a higher result than a more complex deep learning model.*
Keywords: Stress, Electrocardiogram, Convolutional neural network, Long short-term memory, Heart rate variability

1. **Introduction.** Stress is something common in our life nowadays. Most of the people become under pressure from their job that makes them more stressed. As an example, in 2020 when COVID-19 was spreading worldwide and became pandemic, many people became more stressed than usual. In the USA, about 78% of adults became stressed during the pandemic while 67% of adults said it was more stressful during the pandemic [1]. Stress can cause someone's productivity and other negative impacts on those around him. The impact itself can make other people become stressed and give another negative impact to others.

In *Stress: Concept, Definition and History* written by Fink, the most generic definition of stress proposed by Hans Selye is the non-specific response of the body to any demand [2]. The definition itself said that there is a response from our body to stress. In 2019, there is a study that stress response can be detected by physiological signals [3]. Therefore, stress will give a response such as a physiological signal and can be detected from those responses. This study used an ECG signal as the physiological signal. ECG signal acquired from the heart. Stress can trigger arrhythmia in the atrium and ventricle [4]; therefore, it is possible to detect stress using ECG. While ECG itself can be processed into HRV. HRV was acquired from a group of ECG signals over a period. HRV also can directly affect stress in an individual [5].

Stress detection carried out several stages of research starting from experiments to obtain ECG datasets that indicate mental stress, selecting features for several methods that

DOI: 10.24507/icicel.17.01.31

do not use deep learning, and classifying data into stress and non-stress. Since most of the studies used deep learning, they did not have to select any features or do feature extraction [6-11]. Several studies use feature extraction to do stress detection using machine learning techniques [12,13].

However, using the deep learning method without doing any feature extraction tends to make a complex model for stress detection. Therefore, this study aims to create a simpler deep learning model that can detect stress using HRV data without dropping an accuracy from a previous study. The proposed model will use the combination of CNN and LSTM model for stress detection. Finally, the proposed model performance will be compared based on several experiments based on different ECG data (HRV and raw data) and different deep learning models.

The remaining sections from this paper are organized as follows: Section 2 briefly explains the list of works related to the study; Section 3 explains the proposed system and methodology for this study; Section 4 discusses the evaluation result in this study; Section 5 concludes the study and offers suggestions for future works.

2. **Related Works.** Several studies detect stress using deep learning with raw ECG data [6-11], while some studies use HRV data and machine learning to detect stress [12,13]. One of the deep learning studies proposed a model named Deep ECGNet [6]. Deep ECGNet consists of 1D CNN and Recurrent Neural Network (RNN) which the input will be accepted by 1D convolution layer followed by activation layer, pooling layer, and dropout layer that will be continued with two of RNN layer and batch normalization in a row. Then Deep ECGNet used softmax as their classification activation. Deep ECGNet was evaluated on two datasets and achieved 87.39% for the first dataset and 73.96% for the second dataset.

Later, Deep ECGNet will be compared by other models such as the CNN model proposed by Cho et al. [7]. Cho et al.'s model was a CNN model with 8 convolution blocks. The convolution block consists of 2 convolution layers, 1 pooling layer, 1 concatenate layer, 1 batch normalization layer, 1 relu activation layer, and 1 0.3 dropout layer. Each convolution layer had a filter width of 16 and the number of filters was $8 \times 2k$ where k started from 0 and will increase each time he advances to the next step. The CNN model got 90.19% accuracy with the ECG stress dataset.

Another 1D CNN model was proposed to detect stress with an ECG signal called 1-Dimensional Deep Wide Convolutional Neural Network (DWNet1D) [8]. The model consists of 3 modules with each of them followed by the pooling layer and batch normalization layer. The dataset from the experimental results will be divided into 16 data for training, 4 for validation, and 4 for testing. The model managed to get the highest accuracy up to 99.1% with an average accuracy of 89.8%.

Another model that compared Deep ECGNet was DeepER Net which was proposed by Seo et al. DeepER Net consists of CNN and LSTM model and receives ECG and RESP as the input to classify stress [9]. The input will be received by the batch-normalization layer to normalize each signal with the aim that the network can learn to normalize the signal based on the data itself, followed by the 1D convolution layer and the max-pooling layer and using the relu activation function. Then follow again with a 1D convolutional layer and with several LSTM layers and combine ECG and RESP signals with a concatenate layer. To avoid overfitting, Seo et al. added batch-normalization and dropout layers. At the end of the model will be fully connected by using the sigmoid activation function to detect stress or not. DeepER Net got an average accuracy of 83.9%, F1 0.81, and AUC 0.92.

Rastgoo et al. used CNN-LSTM model to classify stress for a driver using ECG signal as their physiological signal and driver simulation data [10]. They compared the classification handcraft feature method with CNN-LSTM algorithm. The CNN-LSTM model they proposed uses a conventional layer, an ELU (Exponential Linear Unit) activation function, a batch normalization layer, and dropout and max-pooling sequentially twice then followed by two LSTM layers and a dense layer and closed with a softmax layer to present the multimodal features. The results obtained from CNN-LSTM model have better evaluation results than the handcraft feature method with the best average accuracy in the 30s feature window, namely 81.4% on the ECG signal, 77.2% on dynamic vehicle data, 58.2% on environmental data, and a combined total of 96.3%.

Another CNN-LSTM model was used to detect stress with ECG signal by image data [11]. Kang et al. used 28 ECG signals from ST Change dataset and 30 ECG signals from WESAD dataset to detect stress. They used FFT (Fast Fourier Transform) and spectrogram to gather 58 time domain data and 58 frequency domain data. The CNN-LSTM model began with $124 \times 124 \times 3$ sequential image data entered into the sequential input layer. Then it was put into a sequential fold layer to be converted into an array and passed to a 2D convolutional layer with 6 filters measuring 5×5. The first 2D convolutional layer produced an output of $124 \times 124 \times 6$. Then it was continued with the BatchNormalization layer and forwarded to the MaxPooling layer with 2×2 filter and 2 steps so that it produces an output size of $62 \times 62 \times 12$. The second convolutional layer had 12 filters measuring 3×3 and finally produces an output size of $31 \times 31 \times 12$. Then normalization was carried out with sequential opening layers and the feature vector is obtained with a flatten layer. Then the data from the flatten layer will be received by the LSTM which has a value of 800×11532. Then it was continued with a fully connected layer measuring 2 and then classified with softmax to determine stress or non-stress conditions. The model achieved 94.8% accuracy for time domain and 98.3% accuracy for frequency domain.

However, the deep learning model has become more complex to achieve better accuracy with ECG signals. Therefore, one of the ways to achieve a simpler model without dropping accuracy is to extract ECG signal to HRV data and make a simpler deep learning model with it. Chandrasekaran et al. achieved 93% accuracy with HRV data using reservoir computing [12]. Their study results with HRV data achieve 3% higher accuracy than deep learning models with raw ECG signals.

Another HRV study also compared five different models for stress detection such as k-NN, General Linear Model (GLM), Naïve Bayes, Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), and random forest classifier [13]. The highest accuracy was obtained by the random forest classifier of 75.1% and the highest accuracy after feature selection was obtained by the SVM classification of 84.4%. The study claimed that the usage of one ECG channel made stress detection with HRV appropriate. Therefore, stress detection with HRV will be made a simpler model and easier to collect data through wearable devices.

This study focused more on building a simpler deep learning model. To achieve that, this study does feature extraction while using a deep learning model. Previous deep learning studies did not do feature extraction and made a model more complex and the previous HRV studies used machine learning to detect stress with ECG signals [6-11]. While some studies used HRV data and still achieve a good result [12,13]. Therefore, this study will be focused on building a deep learning model using HRV as the trained data.

3. **Research Methodology.** This study proposed a deep learning model that used HRV as the input to detect stress. The model combined two methods such as CNN and LSTM. HRV data was achieved from two domains namely frequency domain and time domain which are already mentioned in Table 1. HRV data duration for this study would be 5 minutes per data which means one data consists of 5 minutes long of ECG signal processing. Later this study also will use raw ECG data as the input to compare model

Table 1. HRV features explanation

	Variable	Units	Description
Time domain	SDNN	ms	The standard deviation of all NN intervals
	RMSSD	ms	The square root mean of the sum of the differences between adjacent NN intervals
	SDSD	ms	The standard deviation of the differences between adjacent NN intervals
	pNN50	%	Number of pairs adjacent NN intervals differing by more than 50 ms divided by the total number of all NN intervals in the entire recording
Frequency domain	Total power	ms^2	The variance of all NN intervals ($\approx\leq$ 0.4 Hz)
	VLF	ms^2	Power in range less than 0.04 Hz
	LF	ms^2	Power in a range between 0.04 Hz and 0.15 Hz
	HF	ms^2	Power in a range between 0.15 Hz and 0.4 Hz
	LF/HF		The ratio of LF/HF

results that use HRV data. The data will be split into 80% for training and 20% for testing. The evaluation will consist of precision, recall, f-measure, and accuracy.

3.1. **Dataset.** This study uses a dataset called Wearable Stress and Affect Detection (WESAD) which was gathered from 15 subjects. Subjects are specifically 17 graduate students that were not pregnant, heavy smokers, mental disorder, chronic and cardiovascular disease. However, two of the subjects had to be discarded because of malfunction so the data consist of 15 subjects. The remaining 15 subjects had a mean age of 27.5 $\pm$ 2.4 years with twelve of them being male and the other was female.

WESAD dataset contains high resolution physiological (ECG, EDA, EMG, RESP, and TEMP) and motion (ACC) data sampled at 700 Hz from a chest-worn device and lower resolution data from a wrist-worn device [14]. This study uses ECG signals only to detect stress. WESAD consists of three classes namely amusement, baseline, and stress. However, this study only uses baseline and stress signals since the baseline class has more data than stress and amusement data combined. Later the baseline class will be called non-stress and the stress class will be called stress.

ECG signals from WESAD data can be extracted into HRV data. HRV data from WESAD was acquired from Kaggle with 5 minutes extraction [15]. Data from Kaggle have EDA data and HRV data from the SWELL dataset and WESAD dataset. However, this study only uses HRV from the WESAD dataset. HRV extracted according to the standard proposed by the Task Force of the European Society of Cardiology [16]. The HRV data contain many features, but this study did not use every feature from the HRV dataset. HRV features that were used for this study can be seen in Table 1.

3.2. **Proposed model.** The proposed model referenced CNN and LSTM which both models mostly used for recent stress detection studies. The model will consist of a convolution layer as the input layer followed by the LSTM layer and fully connected with the sigmoid to classify the stress. The proposed model did not use any pooling layer to make every feature from the convolution layer received by the LSTM layer. The proposed model had simpler architecture than other deep learning models. For instance, the proposed model had 7.216 parameters while Deep ECGNet had 39.772 parameters [6]. The total parameter itself can conclude that the proposed model has simpler architecture than Deep ECGNet.

Later the proposed model will be compared to the CNN model without pooling layer and LSTM model. CNN model will consist of two convolution layers followed by fully

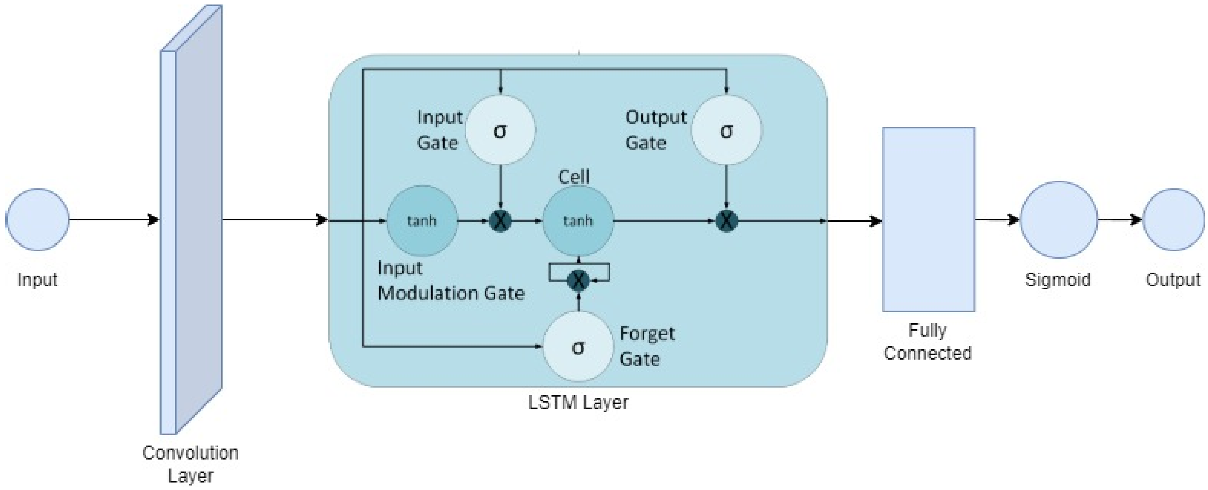

Figure 1. Proposed model architecture

Table 2. Detail of the proposed model architecture

Layer	Detail	Parameter
Input shape		
Conv1D	filter = 45, kernel_size = 1, activation = 'relu'	90
LSTM	filter = 25, activation = 'relu'	7100
Flatten		0
Dense	filter = 1, activation = 'sigmoid'	26
Total parameter		7216

connected with sigmoid to classify stress. Meanwhile, the LSTM model consists of one LSTM layer followed by a normal neural network, and a fully connected layer with a sigmoid to classify stress. Figure 1 shows the proposed model architecture and Table 2 shows the detail of the proposed model architecture.

Other than comparing every model, every model will use four types of input to achieve the best results to detect stress. Four type inputs consist of raw data, time domain, frequency domain, both time domain, and frequency domain. Every type of input will be compared to see which input got the best result to detect stress. However, raw ECG data will get a lower result if we compare it with HRV data with the same model. Since HRV is the extraction from ECG signal, there is a huge possibility that HRV got a lower result. To overcome this issue, another model customized from Deep ECGNet [6] will be used to detect stress with raw ECG signals.

4. **Results and Discussion.** The proposed model is trained by the WESAD dataset while the dataset is pre-processed as described in Section 3.1. The proposed model will receive input from the convolution layer and be forwarded to the LSTM layer. Lastly, the model will be fully connected and classified as stress or non-stress with the sigmoid. The proposed model will be compared with the CNN model and LSTM model with four types of data, i.e., raw ECG data, time domain data, frequency domain data, and both time domain and frequency domain data. The result will be displayed in Table 4. The result shows that the proposed model with both time domain and frequency domain had the best result with 97.07% accuracy. While the other type of model gets a good result with the same type of data.

Despite being trained by a different type of data, the time domain and frequency domain got a good result with the proposed model. While another type of model trained with the time domain data and frequency domain data got an accuracy between 75% to 84%.

TABLE 3. Hyperparameter values in the experiment

Hyperparameter	Values
batch_size	[10, 10.000] batch_size 10.000 only used for raw ECG data
activation_function	[relu, sigmoid]
epochs	50
loss	binary_crossentropy
optimizer	Adam (learning_rate = 0.0001)
metrics	accuracy

TABLE 4. Experiment result with WESAD ECG data

Feature	Model	Precision	Recall	F1-score	Accuracy
Time domain + Frequency domain	CNN + LSTM	**96.68**	**96.98**	**96.82**	**97.07**
	CNN	88.58	88.74	88.66	89.52
	LSTM	94.25	95.87	95.05	95.42
Time domain	CNN + LSTM	**78.87**	**81.44**	**80.14**	**82.01**
	CNN	71.02	74.09	72.52	75.44
	LSTM	78.56	81.06	79.79	81.70
Frequency domain	CNN + LSTM	**82.30**	**83.38**	**82.84**	**84.31**
	CNN	75.93	77.84	76.87	79.07
	LSTM	81.08	82.08	81.57	83.15
Raw data	CNN + LSTM	**53.12**	59.97	56.34	64.54
	CNN	52.05	59.67	55.60	64.29
	LSTM	52.56	60.08	56.07	64.46
	CNN + LSTM & Batch normalization	52.62	**61.58**	**56.75**	**64.76**
	CNN & Pool + LSTM & Batch normalization	52.62	61.37	56.65	64.72

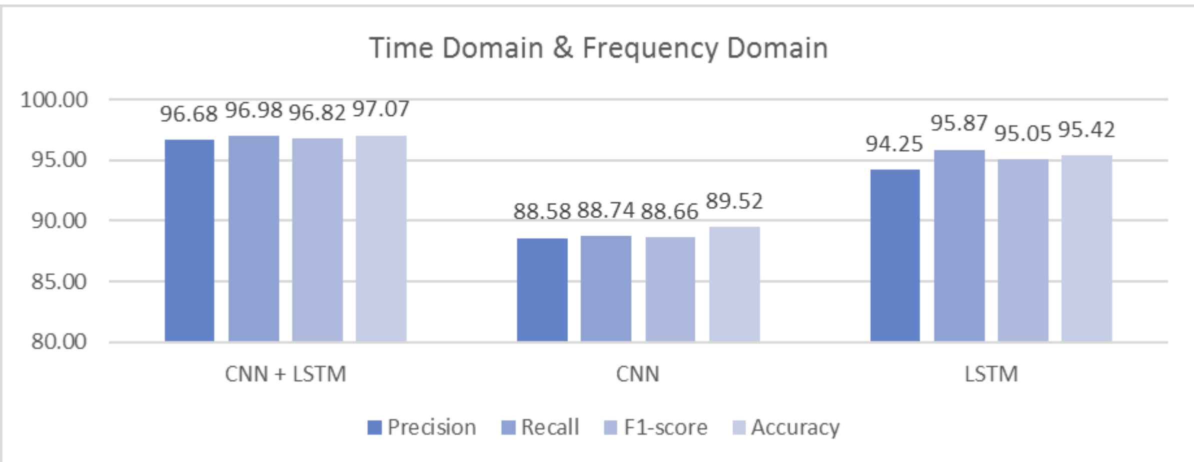

FIGURE 2. Graphic result for HRV time domain and frequency domain data

The result between the time domain and frequency domain did not have many differences. However, the time and the frequency domain type of data had 14% differences between frequency domain data which show a better result than the time domain data.

The proposed model got an increase in accuracy of about 4% from Chandrasekaran et al.'s model [12]. They used reservoir computing with WESAD as the dataset. They

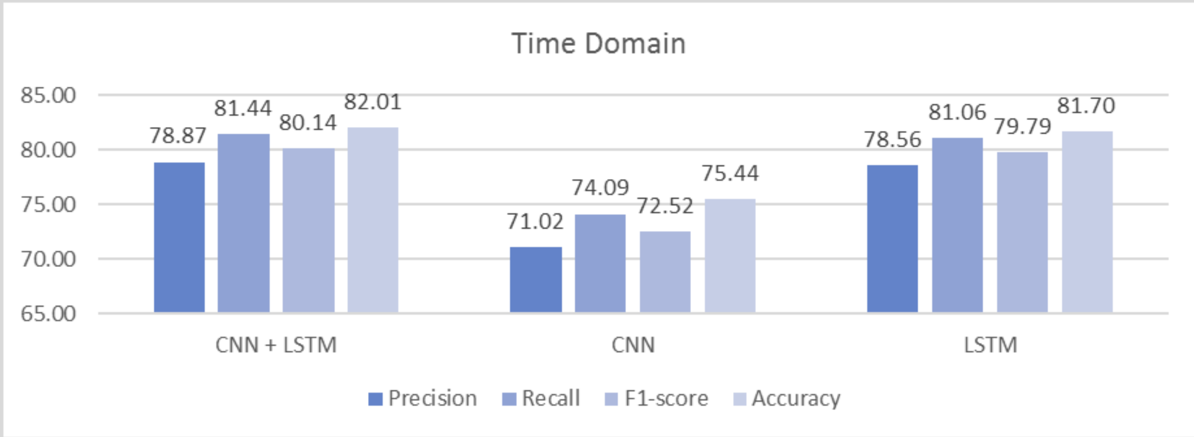

FIGURE 3. Graphic result for HRV time domain data

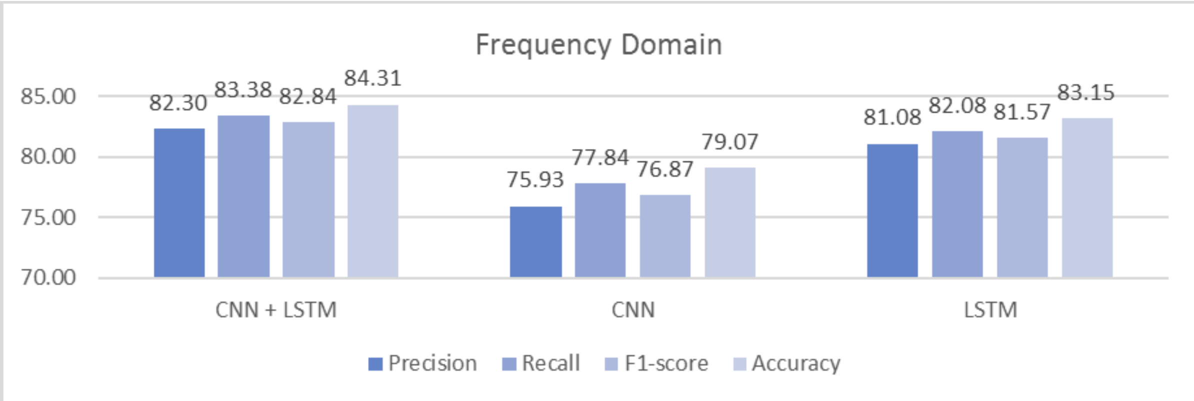

FIGURE 4. Graphic result for HRV frequency domain data

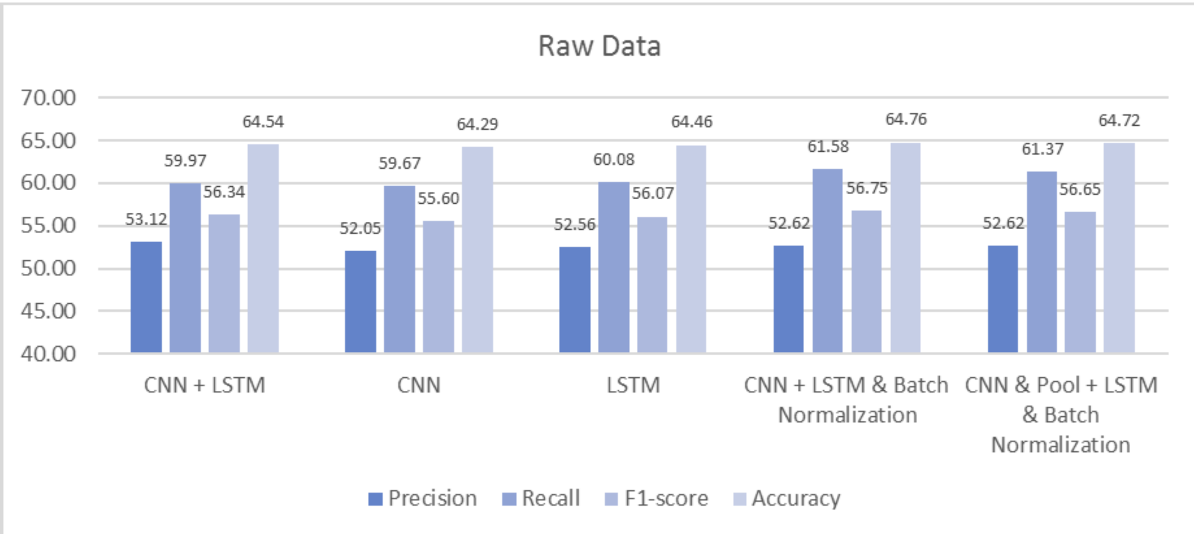

FIGURE 5. Graphic result for ECG raw data

unskew training data to make 1020 3-minute overlapping segments with each segment having 6.6 seconds overlap with the following segment. While this study uses 5-minute HRV data from the dataset without an overlapping. The comparison between this study and other studies using WESAD dataset is displayed in Table 5.

Although the time domain and frequency domain had a good result, raw ECG data did not get a close result for the comparison. WESAD dataset had a much different ratio

TABLE 5. Comparison with other studies using WESAD dataset

Model	Signal	Model	Accuracy
Chandrasekaran et al. [12]	3-minute HRV with 6.6-second overlap	Reservoir computing	93%
Hwang et al. [6]	ECG	CNN + LSTM	64.72%
Proposed model	5-minute HRV	CNN + LSTM	97.07%

between baseline and stress data. Since raw ECG data was unprocessed WESAD data, raw ECG data will get the same amount of baseline class and stress class. The unbalanced data made lower result than other types of data. The customized Deep ECGNet model only got 64.72% with the pooling layer and 64.76% without the pooling layer. Since the WESAD dataset only had one ECG channel, most of the previous deep learning studies with ECG signals [7,8,10,11] cannot be compared to detect stress with the WESAD dataset. The other deep learning study, DeepER Net [9] used ECG and RESP signal so it will not be compared since this study only focused on using ECG signal. In the HRV study from Giannakakis et al. [13], they compared machine learning techniques and did not build their optimal machine learning as Chandrasekaran et al. [12] did. From this result, this study proved that feature extraction is one of the most important steps to detect stress with machine learning. Although deep learning did not focus on feature extraction, feature extraction can become one of the alternatives to detecting stress with ECG signals.

5. **Conclusion and Future Works.** This study proposed a simpler CNN + LSTM model for stress detection using ECG signals. The proposed model was evaluated on ECG data in the WESAD dataset. The proposed model achieved the best result with 97.07% accuracy using HRV data. The result with raw ECG signal achieved lower result from HRV signal because WESAD dataset did not have balance data for baseline and stress class. From that result, this study concludes that feature extraction is one of the most important steps to detect stress with machine learning. The study also suggested doing feature extraction when using a dataset with one ECG channel such as the WESAD dataset.

Future work for this study would be detecting stress with other types of physiological signal such as EMG or RESP to detect stress; stress detection with the multimodal class such as amusement, baseline, and stress; and detecting stress with their level such as high stress, normal stress, or barely stress.

REFERENCES

[1] American Psychological Association, *Stress in America 2020: A National Mental Health Crisis*, 2020.

[2] G. Fink, Stress: Concepts, definition and history, *Reference Module in Neuroscience and Biobehavioral Psychology*, 2017.

[3] G. Giannakakis, D. Grigoriadis, K. Giannakak, O. Simantiraki, A. Roniotis and M. Tsiknakis, Review on psychological stress detection using biosignals, *IEEE Trans. Affect. Comput.*, DOI: 10.1109/TAFFC.2019.2927337, 2019.

[4] R. Lampert, ECG signatures of psychological stress, *J. Electrocardiogr.*, vol.48, no.6, pp.1000-1005, DOI: 10.1016/j.jelectrocard.2015.08.005, 2015.

[5] H.-G. Kim, E.-J. Cheon, D.-S. Bai, Y. H. Lee and B.-H. Koo, Stress and heart rate variability: A meta-analysis and review of the literature, *Psychiatry Investig.*, vol.15, no.3, pp.235-245, DOI: 10.30773/pi.2017.08.17, 2018.

[6] B. Hwang, J. You, T. Vaessen, I. Myin-Germeys, C. Park and B.-T. Zhang, Deep ECGNet: An optimal deep learning framework for monitoring mental stress using ultra short-term ECG signals, *Telemed. E-Health*, vol.24, no.10, pp.753-772, DOI: 10.1089/tmj.2017.0250, 2018.

[7] H.-M. Cho, H. Park, S.-Y. Dong and I. Youn, Ambulatory and laboratory stress detection based on raw electrocardiogram signals using a convolutional neural network, *Sensors*, vol.19, no.20, DOI: 10.3390/s19204408, 2019.

[8] G. Giannakakis, E. Trivizakis, M. Tsiknakis and K. Marias, A novel multi-kernel 1D convolutional neural network for stress recognition from ECG, *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, Cambridge, DOI: 10.1109/ACIIW.2019.8925020, 2019.

[9] W. Seo, N. Kim, S. Kim, C. Lee and S.-M. Park, Deep ECG-Respiration Network (DeepER Net) for recognizing mental stress, *Sensors (Basel)*, vol.19, no.13, DOI: 10.3390/s19133021, 2019.

[10] M. N. Rastgoo, B. Nakisa, F. Maire, A. Rakotonirainy and V. Chandran, Automatic driver stress level classification using multimodal deep learning, *Expert Syst. Appl.*, vol.138, 112793, DOI: 10.1016/j.eswa.2019.07.010, 2019.

[11] M. Kang, S. Shin, J. Jung and Y. T. Kim, Classification of mental stress using CNN-LSTM algorithms with electrocardiogram signals, *J. Healthc. Eng.*, vol.2021, 9951905, DOI: 10.1155/2021/9951905, 2021.

[12] S. T. Chandrasekaran, S. P. Bhanushali, I. Banerjee and A. Sanyal, A bio-inspired reservoir-computer for real-time stress detection from ECG signal, *IEEE Solid-State Circuits Lett.*, vol.3, pp.290-293, DOI: 10.1109/LSSC.2020.3016924, 2020.

[13] G. Giannakakis, K. Marias and M. Tsiknakis, A stress recognition system using HRV parameters and machine learning techniques, *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, Cambridge, DOI: 10.1109/ACIIW.2019.8925142, 2019.

[14] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger and K. V. Laerhoven, Introducing WESAD, a multimodal dataset for wearable stress and affect detection, *Proc. of the 20th ACM International Conference on Multimodal Interaction (ICMI'18)*, New York, pp.400-408, DOI: 10.1145/3242969.3242985, 2018.

[15] K. Nkurikiyeyezu, A. Yokokubu and G. Lopez, The effect of person-specific biometrics in improving generic stress predictive models, *J. Sens. Mater.*, vol.32, no.2, pp.703-722, DOI: 10.18494/SAM.2020.2650, 2020.

[16] A. Camm, M. Malik, J. Bigger et al., Heart rate variability: Standards of measurement, physiological interpretation and clinical use. Task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, *Circulation*, vol.93, no.5, pp.1043-1065, DOI: 10.1161/01.CIR.93.5.1043, 1996.