A Deep Learning Approach Using WESAD Data for Multi-Class Classification with Wearable Sensors

Mohamed Abd Al Aleem
Biomedical Engineering Department
Faculty of Engineering, Helwan University
Helwan, Cairo
eng.alim_89@h-eng.helwan.edu.eg

Roaa Mubarak
Electronics and Communication Engineering Department
Faculty of Engineering, Helwan University
Helwan, Cairo
roaa_ibrahim@h-eng.helwan.edu.eg

Ibrahim Sadek
Biomedical Engineering Department
Faculty of Engineering, Helwan University
Helwan, Cairo
ibrahim_sadek@h-eng.helwan.edu.eg

Nancy M. Salem
Biomedical Engineering Department
Faculty of Engineering, Helwan University
Helwan, Cairo
nancy_salem@h-eng.helwan.edu.eg

2024 6th Novel Intelligent and Leading Emerging Sciences Conference (NILES) | 979-8-3503-7851-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/NILES63360.2024.10753228

***Abstract*— Stress is a common part of life, but chronic or intense stress can severely impact safety and disrupt daily activities. Early detection of mental stress can prevent many related health issues. Stress causes significant changes in bio-signals, which can be used to identify stress levels. This paper proposes various machine learning and deep learning techniques for detecting stress using multimodal datasets from wearable physiological and motion sensors, aiming to prevent stress-related health problems. Data from sensor modalities such as three-axis acceleration (ACC), electrocardiogram (ECG), blood volume pulse (BVP), body temperature (TEMP), respiration (RESP), electromyogram (EMG), and electrodermal activity (EDA) were taken from the WESAD dataset, covering amusement, neutral, and stress states. The best results for the Adam optimizer were achieved with a batch size of 16, yielding an accuracy of 90.26%, while the best results for the SGD optimizer were obtained with a batch size of 16, yielding an accuracy of 87.84%.**

Keywords—WESAD, Stress, DNN, Early Detection, Wearable Sensors, COVID-19

I. Introduction

Close to two billion individuals worldwide experience mental health challenges, characterized by impaired brain function and these issues can involve unusual perceptions, altered thoughts, irregular emotions, atypical behavior, and disrupted social interactions. The global cost of managing these issues is expected to reach $16 trillion by 2030 [1]. Despite their significance, Psychological health care has historically been undervalued compared to physical treatments, with many countries' services under-resourced, a situation worsened by the COVID-19 pandemic [2,3]. This strain has notably impacted Individuals affected by COVID-19, healthcare frontline workers, people with pre-existing mental health conditions, and the broader population, leading to a growing interest in developing accessible sensors for continuous mental status monitoring [4,5].

Stress, triggered by physical, cognitive, or emotional challenges, is managed by the interconnected autonomic nervous system (ANS) and hypothalamic-pituitary-adrenal (HPA) axis [6]. The ANS, with its sympathetic (fight-or-flight) and parasympathetic (rest-and-digest) branches, regulates body functions. During stress, the sympathetic system activates, increasing heart rate, respiration, and sweat production to meet the challenge. The HPA axis releases cortisol, further supporting the stress response through various physiological changes [6].

Acute stress is typically triggered by daily short-term stressors, which our bodies can usually handle. However, excessive or prolonged stress can cause physiological issues like elevated blood pressure and a heightened risk of mental health concerns such as anxiety [7]. To mitigate severe mental health challenges, it is essential to consistently or periodically track physiological and biochemical markers associated with the sympathetic nervous system (SNS) and hypothalamic-pituitary-adrenal (HPA) axis. Current stress level measurements rely on subjective methods like clinical interviews and self-report questionnaires, which are time-consuming and require trained clinicians, limiting their accessibility [8,9]. These methods may also yield unreliable results due to personal bias, potentially affecting clinical decisions. Consequently, there is growing interest in wearable sensors that can continuously collect real-time bio-signals related to mental status [10].

This study proposes a feed-forward DNN incorporating Dropout layers between the hidden layers for effective stress detection. Utilizing the publicly available WESAD dataset, this research explores and compares the performance of two different optimizers, namely Stochastic Gradient Descent (SGD) and Adam, across various patch sizes and a 20% Dropout layers. The analysis focuses on determining the optimal combination for stress detection, employing three-level classification (No-Stress/Medium/High).

II. Related Work

With significant advancements interest in wearable technology is growing, focusing on continuously tracking stress through various physiological signals [11,12]. [13] Studies have explored the links mmeasuring the distinction between suffering and stress the role of wearable sensors and diagnostic implants in this context. These sensors can detect physiological signals such as BP, EDA, HR, MA, BR, SC. The goal of researchers is to create wearable health service systems that assess stress and pain through the analysis of wearable sensors in healthcare.

The study [14] compares ML techniques for accuracy in 3-class and 2-class. Additionally, a feed-forward deep learning network is proposed for these classifications, achieving accuracies of up to 81.65% and 93.20% with ML, and up to 84.32% and 95.21% with DL, respectively.

Previous research explored stress assessment through social media sentiment analysis using ML and BERT for tweet classification [15]. Their model identified emotions (joy, sadness, neutrality, anger, fear) with 94% accuracy, highlighting the potential of NLP for stress detection.

In [16], two DNNs were developed: MLP and a 1D-CNN. These networks analyzed wrist and chest-worn physiological data for two tasks. For binary classification of stressed vs. non-stressed states, average accuracies were 99.65% (MLP) and 99.80% (1D CNN). For three-class classification (stressed, baseline, amused), accuracies were 98.38% (MLP) and 99.55% (1D CNN).

In [17], a convolutional neural network with multi-level deep learning capabilities was introduced. The model was evaluated using the WESAD dataset for mental health, demonstrating superior performance compared to state-of-the-art methods with an outstanding accuracy of 87.7%.

In [18], a 1D - CNN was developed to categorize arousal, valence, and liking based on peripheral physiological signals, such as EMG, BVP, hEOG, vEOG, SKT, and RSP. By leveraging time and frequency domain features, the CNN network obtained accuracies of 77.03% for arousal, 74.68% for liking, and 68.75% for valence.

In [19], researchers developed a novel framework called MFBPST-3D-DRIFT was developed for emotion recognition based on EEG signals, employing a parallel spatial-temporal 3D deep residual learning approach. This framework harnessed various EEG frequency bands (delta, theta, alpha, beta, gamma) to construct a 3D feature representation, which was subsequently trained using a 3D deep residual CNN model. It achieved an impressive accuracy of 96.67% on the SEED dataset, which includes positive, neutral, and negative emotions, and 88.21% on the SEED-IV dataset, which comprises happy, fear, sad, and neutral emotions.

In [20], research concentrated on recognizing emotions through EEG records introduced the Spatial-Temporal Information Learning Network (STILL), achieving 68.31% accuracy for arousal and 67.52% for valence classification. In [21], an attention-LRCN model was developed to mitigate motion artifacts in photo plethysmography (PPG) data. This model combined a baseline network with an attention mechanism and employed frequency domain features from PPG, resulting in an accuracy of 97.11%.

Bio-signals such as EEG, ECG, and PPG are widely used To evaluate physical and mental conditions such as stress, focus, and emotions. Open datasets like AMIGOS [22], DREAMER [23], SWELL [24], SEED [25], DEAP [26], and WESAD [27] are curated for research with experiments designed to induce stress and capture bio-signals. These datasets include eexperiments are meticulously crafted to elicit stress and collect bio-signals. AMIGOS and SEED used video clips to elicit emotions, gathering EEG and ECG data, while DEAP uutilized music videos to stimulate emotions over various arousal-valence quadrants while simultaneously peripheral signals and recording EEG.

The WESAD dataset gathered peripheral signals using video clips and the Trier Social Stress Test. SWELL exposed participants to office stressors like time pressure, and collecting ECG data during tasks including delivering presentations and writing a report. These datasets provide comprehensive physiological data, making them invaluable for understanding human stress and emotional responses. They have been adapted for a variety of studies due to their diverse stimuli and robust experimental designs. These resources are critical for advancing research in emotion recognition and stress detection.

In this study, we utilized the WESAD dataset, a publicly available resource, to implement a feed-forward deep neural network (DNN) for distinguishing between different stress levels. We compared the performance of two optimizers, SGD and Adam, across various patch sizes. Additionally, we experimented with Dropout layers at 20% to enhance the model's generalization by preventing overfitting. The inclusion of Dropout layers helped mitigate the risk of overfitting by randomly deactivating a fraction of neurons during training, thereby improving the robustness of our model.

The next section details our methodology, providing an in-depth look at the techniques and configurations used in our analysis. This comprehensive approach allowed us to identify the optimal settings for effective stress detection.

III. Methodology

This study used the WESAD dataset, made publicly available in [27]. It includes movement and physiological data from fifteen participants, recorded with the Raspbian Advance chest apparatus and the Empatica E4 arm sensor during activities like baseline, preparation, amusement, stress, meditation, and recovery. Detailed sensor setup, positioning, methodology, and data collection processes are in [28]. The Raspbian measured heart rate, ACC, RESP, EMG, and TEMP at 700 Hz. The Empatica E4 recorded TEMP, RESP, ACC, and EDA at frequencies of 8 Hz, 16 Hz, 32 Hz, and 64 Hz.

The primary reason for employing a DNN in this study was its parameter-sharing capability, allowing a few filters to efficiently extract features from the entire input. This DNN was designed to process physiological signals from eight sensors, including ECG, EDA, EMG, RESP, TEMP, and a 3-axis ACC. Each ACC axis was treated as a separate input, resulting in eight signals for the DNN network. Data from each sensor were segmented into 5-second windows, providing simultaneous inputs for the DNN, which was configured to perform a three-class classification.

Dropout layers were incorporated between each hidden layer to prevent overfitting and enhance generalization. The DNN consists of five hidden layers with node capacities of 64, 128, 256, 128, and 64, respectively, all using the ReLU activation function, while the output layer, configured for three-level stress classification, uses the softmax function. This hierarchical design with Dropout layers enables nuanced feature extraction. The dataset was randomly scrambled and split into training and testing subsets with a 70:30 ratio, ensuring the model's robustness and generalizability by allowing it to learn from diverse samples and validating performance on unseen data.

Two optimization techniques, Adam and SGD, were evaluated for their impact on the model's classification performance. During training, the model underwent 40 iterations per epoch, and the influence of varying batch sizes (16 to 512) on performance was thoroughly investigated. This analysis provided valuable insights into optimizing the training process for DNN models in stress classification.

The primary metrics utilized to evaluate the deep neural network for each optimizer included accuracy, precision, recall, and F1-score, as shown in Equations 1 to 4, respectively.

$$Accracy = \frac{True\ positive + True\ negative}{Total\ population} \quad (1)$$

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \quad (2)$$

$$Recall = \frac{True\ positive}{True\ positive + False\ negative} \quad (3)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

IV. Results

The effectiveness of the feed-forward Neural Network was thoroughly evaluated in terms of its performance metrics for the three-class classification task. This assessment included two specific optimization methods: The Adam optimizer and the SGD optimizer, applied under different batch size settings. This comparative study aimed to uncover how optimization techniques and batch size variations influence the overall performance and efficiency of DNN models in stress classification. Additionally, Dropout layers were strategically incorporated between each hidden layer to enhance model robustness and prevent overfitting during training.

In our initial analysis, we began by evaluating the performance metrics of the feed-forward DNN model using the Adam optimizer and 20% Dropout layers across different numbers of epochs. Our assessment concentrated on key metrics such as F1-Score, precision, recall, and accuracy, as outlined in Table 1.

Table 1 Performance Metrics of the DNN Model Using the Adam Optimizer Across Different Batch Sizes

Optimizer	No. Class	No. Epochs	ACC.	F1-Score	Precision	Recall
Adam	3 Class	16	0.902	0.900	0.924	0.903
		32	0.871	0.867	0.907	0.872
		64	0.847	0.839	0.895	0.847
		128	0.871	0.866	0.906	0.871
		256	0.835	0.825	0.889	0.836
		512	0.862	0.857	0.902	0.863

From Table 1 it is clear that the Adam optimizer demonstrated strong performance across all batch sizes in the stress classification task, with accuracies ranging from 83.5% to 90.2%. The highest accuracy and F1-score 0.900 were achieved with a batch size of 16, highlighting the model's effective stress detection. Precision and recall were also highest with a batch size of 16, at 0.9245 and 0.9030, respectively. Smaller batch sizes yielded the best results, suggesting better generalization and faster convergence. Overall, the Adam optimizer proved robust and efficient, particularly with smaller batch sizes, for classifying stress using physiological signals.

To visually represent and clarify the best result of the DNN when using the Adam optimizer, Fig. 1 presents the confusion matrices. The optimal outcome was achieved with 16 epochs, demonstrating the model's performance under these conditions.

In the following analysis, we will examine the performance of the DNN with a 20% Dropout layer using the SGD optimizer. This evaluation involved a thorough investigation of different epoch settings, with performance metrics such as Accuracy, F1-Score, Precision, and Recall carefully recorded and detailed in Table 2 for in-depth comparison and analysis.

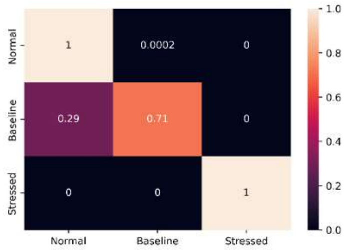

Fig. 1. Confusion matrices for DNN performance with Adam optimizer (16 epochs).

Table 2 Performance Metrics of the DNN Model Using the SGD Optimizer Across Different Batch Sizes

Optimizer	No. Class	No. Epochs	ACC.	F1-Score	Precision	Recall
SGD	3 Class	16	0.878	0.874	0.910	0.878
		32	0.851	0.844	0.896	0.851
		64	0.819	0.805	0.882	0.819
		128	0.774	0.746	0.864	0.774
		256	0.700	0.664	0.725	0.700
		512	0.772	0.754	0.816	0.772

The performance of the DNN model using the SGD optimizer showed noticeable variations across different batch sizes. The highest accuracy was achieved with a batch size of 16, reaching 87.84%. However, as the batch size increased, a decline in performance metrics was observed, with the lowest accuracy of 70.05% at a batch size of 256. The model with a batch size of 16 also demonstrated a high F1-Score of 0.875 and precision of 0.911. Overall, the results indicate that smaller batch sizes yielded better performance in stress classification tasks with the SGD optimizer.

To visually represent and clarify the best result of the DNN when using the SGD optimizer, Fig. 2 presents the confusion matrices. The optimal outcome was achieved with 16 epochs, demonstrating the model's performance under these conditions.

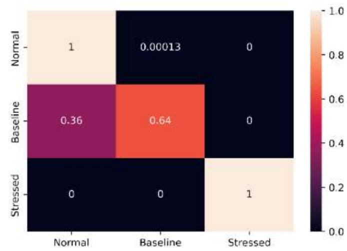

Fig.2. Confusion matrices for DNN performance with Adam optimizer (16 epochs).

Discussion

Our study explored stress classification using DNNs with Dropout layers, finding the Adam optimizer most effective for classifying three stress levels after 16 epochs. This research examines how the optimizer, number of epochs, and Dropout layers affect DL model efficiency, using chest patch signals for the first time.

A limitation of this study is that the model, trained on lab-collected data, may not generalize well to real-life conditions due to its inability to handle sensor noise in everyday environments. In Table 3, a review will be conducted based on the results from previous research.

Table 3Review of Results Based on Previous Research Findings

Ref.	ML/DL	Model	ACC.

[29]	ML	RF	0.856
[30]	ML	Fuzzy	0.765
[31]	DL	1D-CNN	0.705

Conclusion

This study demonstrated the effectiveness of using a DNN with Dropout layers and various optimization techniques for stress classification. By leveraging physiological signals from multiple sensors, the DNN was able to perform three-class stress classification with high accuracy. The analysis highlighted that the performance of the model significantly increased with a smaller number of epochs, especially when using the Adam optimizer. The use of publicly available datasets such as WESAD allowed for rigorous training and validation, ensuring the robustness and generalizability of the model. These findings underscore the potential of DNNs in enhancing stress detection and contribute to the broader field of physiological signal analysis for mental health assessment.

References

[1] Mental Health Wkly., "Report: Mental illness will cost the world $16 trillion (USD) by 2030," vol. 28, pp. 7-8, 2018.

[2] J. Alonso, M. Codony, V. Kovess, M.C. Angermeyer, S.J. Katz, J.M. Haro, G. De Girolamo, R. De Graaf, K. Demyttenaere, G. Vilagut, et al., "Population level of unmet need for mental healthcare in Europe," Br. J. Psychiatry, vol. 190, pp. 299-306, 2007.

[3] S. Rathod, N. Pinninti, M. Irfan, P. Gorczynski, P. Rathod, L. Gega, and F. Naeem, "Mental Health Service Provision in Low- and Middle-Income Countries," Health Serv. Insights, vol. 10, pp. 1178632917694350, 2017.

[4] M. Luo, L.X. Guo, M.Z. Yu, W.Y. Jiang, and H.Y. Wang, "The psychological and mental impact of coronavirus disease 2019 (COVID-19) on medical staff and general public—A systematic review and meta-analysis," Psychiatry Res., vol. 291, pp. 113190, 2020.

[5] K.Y. Pan, A.A.L. Kok, M. Eikelenboom, M. Horsfall, F. Jorg, R.A. Luteijn, D. Rhebergen, P. van Oppen, E.J. Giltay, and B.W.J.H. Penninx, "The mental health impact of the COVID-19 pandemic on people with and without depressive, anxiety, or obsessive-compulsive disorders: A longitudinal study of three Dutch case-control cohorts," Lancet Psychiatry, vol. 8, pp. 121-129, 2021.

[6] S. Rotenberg and J. J. McGrath, "Interrelation between autonomic and HPA axis activity in children and adolescents," Biol. Psychol., vol. 117, pp. 16-25, 2016.

[7] S. Gedam and S. Paul, "A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques," IEEE Access, vol. 9, pp. 84045-84066, 2021.

[8] T. Richter, B. Fishbain, G. Richter-Levin, and H. Okon-Singer, "Machine Learning-Based Behavioral Diagnostic Tools for Depression: Advances, Challenges, and Future Directions," J. Pers. Med., vol. 11, p. 957, 2021.

[9] W. M. L. Finlay and E. Lyons, "Methodological issues in interviewing and using self-report questionnaires with people with mental retardation," Psychol. Assess., vol. 13, pp. 319-335, 2001.

[10] M. Sheikh, M. Qassem, and P. A. Kyriacou, "Wearable, Environmental, and Smartphone-Based Passive Sensing for Mental Health Monitoring," Front. Digit. Health, vol. 3, p. 662811, 2021.

[11] M. Namavari, J. Lippoth, S. Knight, A. A. Jamali, M. Hedayati, R. J. Spiteri, and S. Syed-Abdul, "Photoplethysmography Enabled Wearable Devices and Stress Detection: A Scoping Review," J. Pers. Med., vol. 12, p. 1792, 2022.

[12] T. Iqbal, A. J. Simpkin, D. Roshan, N. Glynn, J. Killilea, J. Walsh, G. Molloy, S. Ganly, H. Ryman, E. Cohen, et al., "Stress Monitoring Using Wearable Sensors: A Pilot Study and Stress-Predict Dataset," Sensors, vol. 22, p. 8135, 2022.

[13] J. Chen, M. Abbod, and J. S. Shieh, "Pain and Stress Detection Using Wearable Sensors and Devices—A Review," Sensors, vol. 21, p. 1030, 2021

[14] Bobade, P.; Vani, M., "Stress detection with machine learning and deep learning using multimodal physiological data," in *Proceedings of the 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 15-17 July 2020, pp. 51-57.

[15] T. Nijhawan, G. Attigeri, and T. Ananthakrishna, "Stress detection using natural language processing and machine learning over social interactions," J. Big Data, vol. 9, p. 33, 2022.

[16] R. Li and Z. Liu, "Stress detection using deep neural networks," BMC Med. Inform. Decis. Mak., vol. 20, p. 285, 2020.

[17] A. Kumar, K. Sharma, and A. Sharma, "Hierarchical deep neural network for mental stress state detection using IoT based biomarkers," Pattern Recognition Letters, vol. 145, pp. 81-87, 2021.

[18] S. Vijayakumar, R. Flynn, P. Corcoran, and N. Murray, "CNN-based Emotion Recognition from Multimodal Peripheral Physiological Signals," in Proceedings of the IMX'22: ACM International Conference on Interactive Media Experiences, Aveiro, Portugal, 22-24 June 2022.

[19] M. Miao, L. Zheng, B. Xu, Z. Yang, and W. Hu, "A multiple frequency bands parallel spatial-temporal 3D deep residual learning framework for EEG-based emotion recognition," *Biomed. Signal Process. Control*, vol. 79, p. 104141, 2023.

[20] Y. Tang, Y. Wang, X. Zhang, and Z. Wang, "STILN: A novel spatial-temporal information learning network for EEG-based emotion recognition," Biomedical Signal Processing and Control, vol. 85, p. 104999, 2023.

[21] J. Choi, J. S. Lee, M. Ryu, G. Hwang, G. Hwang, and S. J. Lee, "Attention-LRCN: Long-term Recurrent Convolutional Network for Stress Detection from Photoplethysmography," in Proceedings of the 2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Messina, Italy, 22-24 June 2022, pp. 1-6. New York, NY, USA: IEEE, 2022.

[22] J.A. Miranda-Correa, M.K. Abadi, N. Sebe, and I. Patras, "Amigos: A dataset for affect, personality and mood research on individuals and groups," IEEE Trans. Affect. Comput., vol. 12, pp. 479-493, 2018.

[23] S. Katsigiannis and N. Ramzan, "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices," IEEE J. Biomed. Health Inform., vol. 22, pp. 98-107, 2017.

[24] S. Koldijk, M. Sappelli, S. Verberne, M.A. Neerincx, and W. Kraaij, "The swell knowledge work dataset for stress and user modeling research," in Proc. 16th Int. Conf. Multimodal Interact., Istanbul, Turkey, 2014, pp. 291-298.

[25] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," IEEE Trans. Auton. Ment. Dev., vol. 7, pp. 162-175, 2015.

[26] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis using physiological signals," IEEE Trans. Affect. Comput., vol. 3, pp. 18-31, 2011.

[27] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in Proc. 20th ACM Int. Conf. Multimodal Interact., Boulder, CO, USA, 16–20 Oct. 2018, pp. 400-408.

[28] X. Long, B. Yin, and R.M. Aarts, "Single-accelerometer-based daily physical activity classification," in Proc. Int. Conf. IEEE Eng. Med. Biol. Soc., Minneapolis, MN, USA, 2009, pp. 6107-6110. doi: 10.1109/IEMBS.2009.5334925.

[29] Y. S. Can, N. Chalabianloo, D. Ekiz, J. Fernandez-Alvarez, G. Riva, and C. Ersoy, "Personal Stress-Level Clustering and Decision-Level Smoothing to Enhance the Performance of Ambulatory Stress Detection With Smartwatches," IEEE Access, vol. 8, pp. 38146–38163, 2020.

[30] S. Pourmohammadi and A. Maleki, "Continuous mental stress level assessment using electrocardiogram and electromyogram signals," *Biomed. Signal Process. Control*, vol. 68, p. 102694, 2021.

[31] A. Lisowska, S. Wilk, and M. Peleg, "Catching patient's attention at the right time to help them undergo behavioural change: Stress classification experiment from blood volume pulse," in *Artificial Intelligence in Medicine*, ser. Lecture Notes in Computer Science, Cham, Switzerland: Springer International Publishing, 2021, pp. 72–82.