

Personalized Ensemble Based Stress Detection Using Wearable Sensor Data

Kyungtaek Oh*, Jun Kyun Choi*, Hyunseo Park*, Seungjin Lee**

*School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

**School of AI·SW Convergence, Kyungnam University, Changwon, Republic of Korea

okt5380@kaist.ac.kr, jkchoi59@kaist.edu, tkf92001@kaist.ac.kr, sjinlee@kyungnam.ac.kr

Abstract— With the recent advancements in Internet of Things (IoT) technology, the digital healthcare industry has seen significant growth. Stress detection is a key area within digital healthcare, focusing on classifying human stress based on various physiological signals. The need for highly stable classification of human states based on wearable sensor data is becoming increasingly important. This study introduces an ensemble-based stress detection model that utilizes multimodal features and metadata to capture personalized patterns for each individual. Using the WESAD dataset, which contains physiological and motion signals, we compared the performance of our model to other deep learning and machine learning approaches. Our proposed model, comprising three CNN-based classifiers and an ensemble attention module, achieved the highest mean accuracy of 94.01%, demonstrating its robustness and reliability. This makes it a promising solution for stress management systems in workplace environments.

Keywords— Internet of things, Artificial intelligence, Metadata, Digital healthcare, Ensemble model, Stress detection

I. INTRODUCTION

Stress detection aims to identify a person's specific stress state using data from IoT sensors attached to the individual. With recent advancements in IoT technology, numerous IoT applications have emerged, offering innovative services [1], [2]. Similarly, the digital healthcare industry has expanded by utilizing sensor data to deliver medical services. The technologies such as smart watches have been developed for stress detection, utilizing various sensor data, including electroencephalography (EEG), three-axis acceleration (ACC), electrocardiogram (ECG), and heart rate (HR) [3].

Analyzing a stress in individual is crucial, as stress defined as the “nonspecific response of the body to any demand upon it” can lead to complex mental and physical health issues such as depression or heart disease [4], [5], [6]. Furthermore, there is a growing concern for the mental health of workers in the workplace, as their mental stress can lead to reduced productivity, high turnover rates, and increased sick leave, all of which contribute to higher social and economic costs [7], [8]. The American Institute of Stress claims that stress-related illnesses result in an annual expenditure of \$300 billion. Work-related stress cost EU businesses about 25 billion euros in 2013 [9], [10]. If an adequate intervention strategy is implemented by the organization for employees suffering from stress, the net

benefits could amount to €1000 per employee per year by reducing the stress cost [11]. To enhance work efficiency, it is necessary to develop methods to appropriately diagnose and respond to stress by analyzing the physiological data.

Stress detection is based on various forms of physiological data measured from sensors attached to the human, such as EEG, HR, ECG, and galvanic skin response (GSR). This is because the occurrence of stress induces a variety of emotional and physiological responses [12], [13]. These physiological data respond to changes in the human nervous system according to specific psychological states.

Numerous machine learning algorithms have been applied to distinguish stress states, but the sensor data must first undergo preprocessing before being used as input to these models [14], [15], [16]. Depending on the characteristics of the data, it is necessary to manually calculate many features in the frequency and time domains [17]. However, traditional machine learning models require extensive handcrafted feature engineering, which also demands specialized domain knowledge [18], [19]. Without sufficient analysis, the model is likely to perform poorly in classification.

Recently, with the rise of deep learning technology, which features automatic feature extraction, various deep learning-based models have been applied to the field of stress detection to address the issues inherent in traditional machine learning models [20]. In particular, some approaches have been proposed to handle the high-dimensional and multimodal nature of physiological data [21].

However, since data is collected from multiple individuals through wearable devices, the characteristics of the data can vary from person to person. For example, one person's heart rate during a stressful situation might be similar to another person's heart rate in a normal situation. Previous works have not fully addressed the issue of varying data distributions across the subjects. If additional information about the person (e.g., height, weight, age) is available, it could help mitigate this issue by aligning data features according to individual characteristics. People with similar physical attributes are likely to exhibit similar data distributions. Moreover, as the model learns the metadata for personal information, it can further capture personalized features, potentially leading to performance variations. To address this, an ensemble structure has been introduced. By complementing features that individual models

may have missed, this approach is expected to ultimately improve the model's generalization and stability.

In this paper, we propose an end-to-end wearable stress detection model that learns individual-specific cues through the fusion of single models based on metadata. The main contributions of this paper are summarized as follows:

- By utilizing metadata containing information about individuals as part of the training data, the model can learn cues related to each person, thereby reflecting individual characteristics in the model.
- To increase robustness against subject-based datasets with different data distributions, the ensemble model is proposed. The base models complemented each other's insufficient features, enabling the ensemble model to achieve a generalized representation with low variance.

II. RELATED WORKS

In the stress detection domain, numerous deep learning-based approaches have been proposed. Some research has focused on transforming time series data [22], [23]. For instance, the authors of [24] transformed ECG data into an image map using a Gram matrix and Markov transition matrix. By utilizing these matrix transformations, a CNN-based model can learn temporal local pattern representations. In [25], FFT was applied to ECG, GSR, and electromyogram (EMG) signals to classify human stress states. The authors assumed that each dominant component in different time windows in the frequency domain has inherent patterns that are significantly affected by a person's mental condition.

Meanwhile, researchers have attempted to detect stress by leveraging the multimodality of stress datasets [26], [27]. A common approach involves splitting the input features by modality so that a model learns multimodal representations by fusing the unimodal features. Using the DEAP dataset [28], each modality (EEG, EMG, GSR, ECG) was separately processed with transformer modules, and unimodal features were combined for classification in [29]. In [30], an end-to-end stress detection model was proposed, adapting a bidirectional cross and self-modal attention mechanism (BCSA). Using Blood Volume Pulse (BVP) and Electrodermal Activity (EDA) data, cross-modal attention and self-modal attention were used to capture the relationship between the two modalities.

Ensemble methods play a key role in reducing the generalization error of base models. The authors of [31] utilized an ensemble network for a synthesized dataset. Multiple datasets were combined under random sampling, and an ensemble network consisting of XGBoost [32] and a DNN model was used for classification. An ensemble network composed of CNN and LSTM was used for ECG signals in [33], where the input signal was transformed into a spectrogram for feature extraction in the frequency domain. An ensemble model combining random forest, LightGBM, and gradient boosting classifiers at level 0, with a random forest as the level 1 meta classifier, was proposed in [34]. The random forest served as the meta-classifier for predicting the final classification result. In [35], a hybrid deep learning-based ensemble approach was proposed. First, the CNN network and CNN-LSTM network of

the model were trained separately and then combined for stress state classification using the WESAD (wearable stress and affect detection) dataset [4].

Transforming time series data into other forms, such as an image using various signal processing algorithms, increases computational costs due to the enlarged dimensions of the data. Additionally, for modality-based approaches, calculating features for each modality and combining them requires high training costs. And previous approaches do not adequately consider the characteristics of human stress. The distribution of each subject data differs because the datasets typically collect wearable sensor data from multiple subjects. If we have some information (height, weight, etc.) about the subjects, it can help alleviate the aforementioned problem by using the information explicitly since subjects with similar personal characteristics are likely to have similar data distributions. A model can learn the metadata of subjects to capture customized features. An ensemble-based approach can be used to effectively enhance the model's learning capacity, as incorporating metadata introduces additional information that may increase the complexity of the learning process, potentially making convergence more challenging. The stability of the model can be enhanced by calibrating some features that the individual models fail to capture. Our approach ensures the learning of customized features and a stabilized classification result with lower variation.

III. PROPOSED METHODS

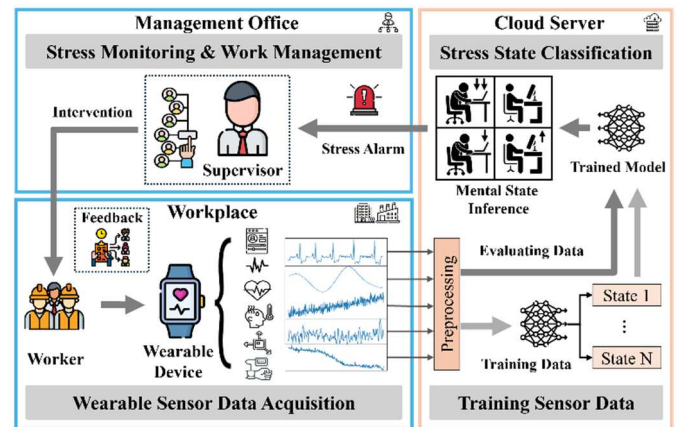


Figure 1. Human stress detection model in IoT system.

As illustrated in Figure 1, a workplace stress monitoring system consists of wearable devices that measure workers' physiological signals, a cloud server that acquires and analyzes the data, and a stress detection model that assesses the mental state of workers. All workers wear wearable devices that measure physiological signals. And the acquired data are transmitted to a central cloud server. The evaluation results are then sent to the supervisor, who adjusts the work schedule for individual workers and provides intervention for those experiencing stress. This system allows an organization to effectively manage workers' stress levels. The organization can implement several interventions, including offering counselling, or rescheduling tasks, leading to reduced costs associated with

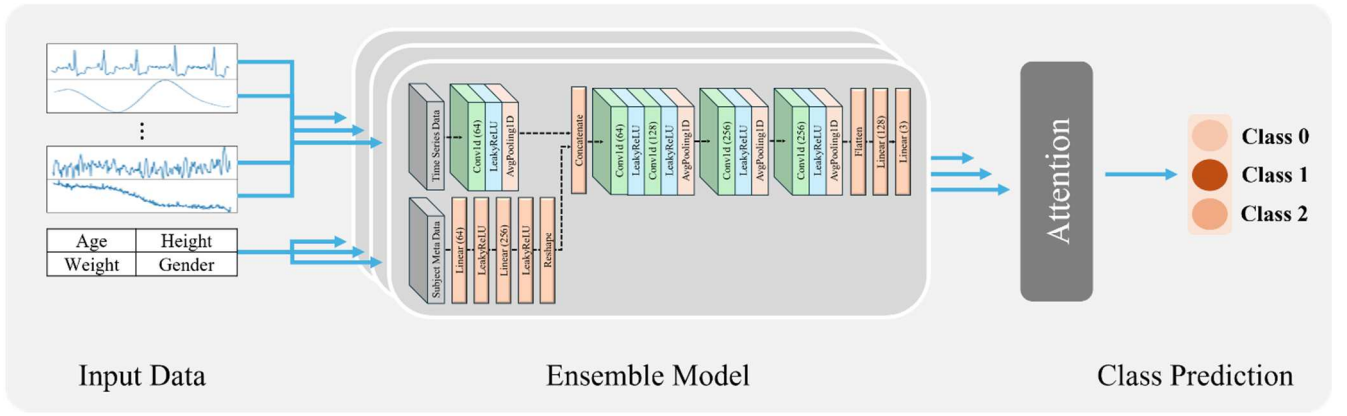


Figure 2. Proposed ensemble network architecture for stress detection

worker stress. The model also incorporates metadata as input to reflect the individual characteristics, enhancing the performance of the stress management system.

A. Overview of Models

Our ensemble model comprises three classifiers, each of which processes sensor signals and metadata for each subject, along with an attention module, as illustrated in Figure 2. The time-series input data and metadata are concatenated at an intermediate layer, where the temporal and meta features are combined spatially and processed through convolutional (Conv1D) blocks. The outputs from all classifiers are then integrated via the attention module for final classification.

B. Time Series Data Preprocessing

For this experiment, we used the WESAD dataset and focused solely on chest-worn sensor data. The dataset contains physiological and motion features, including ACC, ECG, EDA, EMG, temperature (TEMP), and respiration (RESP). The data was undersampled from 700Hz to 128Hz using a FFT method, followed by the application of min-max normalization. The dataset includes personal information for each subject, with metadata such as height, weight, age, and sex. Min-max normalization was also applied to the metadata, except for sex, which is encoded as binary values: 1 for male and 0 for female.

C. Ensemble Model Design

Our ensemble model consists of three classifiers, each of which contains multiple Conv1D blocks composed of CNN, LeakyReLU, and pooling layers. The ensemble model's architecture is presented in Figure 2. The time series input data passes through a Conv1D block and then is concatenated with reshaped metadata feature along the channel dimension. After passing through multiple blocks, the data is flattened, passes through an FC layer, and produces a prediction result. Each of the three models has a slightly different variation. The first model uses average pooling layers, the second uses max pooling layers, and the third model also uses average pooling layers for the Conv1D blocks but replaces the flattening and FC layer process with a global average pooling (GAP) layer. For all Conv1D modules, the kernel size and stride are set to 3 and 1, respectively.

D. Attention mechanism

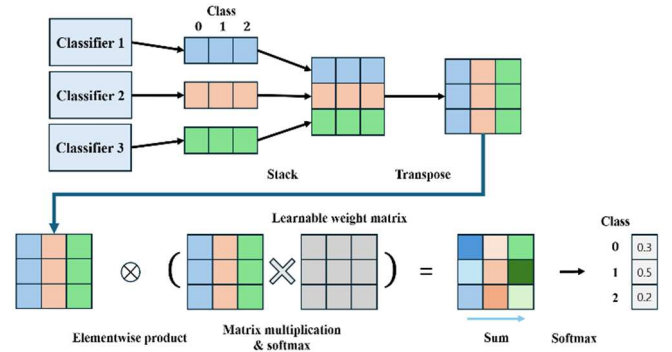


Figure 3. Proposed attention module for the ensemble model

There are total of N models, each producing a prediction vector of size C . The attention module takes the outputs from all classifiers. The initial prediction vectors are stacked into a matrix of size $N \times C$, which is then transposed. This transposed matrix is multiplied by a learnable weight matrix $W \in \mathbb{R}^{N \times N}$, followed by an element-wise product with the original transposed prediction matrix as shown in Figure 3. The resulting matrix is then summed along the dimension corresponding to the number of models. Finally, a softmax layer is applied to the output vector to produce the final prediction.

E. Loss Function

Let M be the number of classes, y_c the true class label, and p_{c_i} the predicted probability for the c^{th} class from i^{th} output of the ensemble model. Here, p_{c_0} comes from the attention module while the other p_{c_i} values come from each classifier model. In this case, $i \in \{0, 1, 2, 3\}$. The hyperparameter α controls the weighting between the output of the attention mechanism and the outputs of the classifiers. In this work, α is set to 0.4. Then entire cross-entropy loss is defined as:

$$L_{CE_i} = - \sum_{c=1}^M y_c \log(p_{c_i}), \quad (1)$$

$$L_{CE} = \alpha L_{CE_0} + (1 - \alpha)(L_{CE_1} + L_{CE_2} + L_{CE_3}) \quad (2)$$

TABLE 1. COMPARISON RESULTS OF DIFFERENT MODELS FOR HUMAN STRESS STATE CLASSIFICATION

Methods	Accuracy	F1-score	Precision	Recall
Ours w/ ensemble (attention) & metadata	0.9401 ± 0.0046	0.9268 ± 0.0048	0.9126 ± 0.0062	0.9470 ± 0.0025
Ours w/ metadata	0.9087 ± 0.0202	0.8902 ± 0.0224	0.8762 ± 0.0220	0.9152 ± 0.0177
Ours	0.8999 ± 0.0172	0.8857 ± 0.0193	0.8715 ± 0.0203	0.9116 ± 0.0126
Featurefusion	0.9177 ± 0.0155	0.8988 ± 0.0185	0.8850 ± 0.0150	0.9284 ± 0.0151
ECGTransformer	0.8842 ± 0.0200	0.8596 ± 0.0265	0.8504 ± 0.0273	0.8770 ± 0.0256
EnsembleCNN	0.8779 ± 0.0162	0.8612 ± 0.0161	0.8450 ± 0.0158	0.8972 ± 0.0122
DeepCNN	0.8655 ± 0.0215	0.8480 ± 0.0214	0.8383 ± 0.0154	0.8859 ± 0.0165
Linear Discriminant Analysis	0.6881	0.6367	0.6366	0.6489
AdaBoost DT	0.6951	0.6283	0.6249	0.6428
Random Forest	0.7798	0.7596	0.7457	0.7955

IV. EXPERIMENT AND RESULTS

A. Dataset Description

WESAD is a publicly available dataset for wearable stress and affect detection. Physiological (blood volume pulse, ECG, EDA, EMG, RESP, TEMP) and motion (ACC) signals were collected using wrist and chest-worn sensors from 15 participants. Each participant performed specific tasks under three affective conditions of normal, stress, and amusement. In the normal condition, the subjects sat or stood at a table and magazines were provided to read. In the amusement condition, they were requested to watch a set of eleven funny video clips. In the stress condition, they were asked to give a five-minute speech in front of a three-person panel about their strengths and weakness. And they were told to count down from 2023 to 0 by steps of 17 in reverse order.

B. Experiment Settings

We followed the inclusive split procedure outlined in [36], ensuring the class ratio was maintained. The first 50% of intervals from each subject's affective state were combined into the training set, while the remaining 50% were combined into the test set. A batch size of 128, window size of 512, and the Adam optimizer with a learning rate of 1e-3 were used for training. 30 epochs were used for training the ensemble model and other reference models. The models were implemented in PyTorch version 2.1.0 and GeForce RTX 3060 GPU.

C. Evaluation Metrics

The performance of the model is evaluated using common metrics: accuracy, F1-score, precision, and recall. By using the metrics together, comprehensive evaluation of the model's performance is provided, ensuring that it not only classifies instances correctly but also handles false positives and false negatives effectively in class imbalanced dataset. The metrics are computed using TP_c (true positives), TN_c (true negatives), FP_c (false positives), FN_c (false negatives) for each class c , respectively. The metrics for each class are computed independently, and the mean value is taken across all classes.

$$Accuracy = \frac{\sum_{c=1}^M TP_c}{\sum_{c=1}^M TP_c + \sum_{c=1}^M FP_c} \quad (3)$$

$$Precision = \frac{1}{M} \sum_{c=1}^M \frac{TP_c}{TP_c + FP_c} \quad (4)$$

$$Recall = \frac{1}{M} \sum_{c=1}^M \frac{TP_c}{TP_c + FN_c} \quad (5)$$

$$F1 - score = \frac{1}{M} \sum_{c=1}^M \frac{2TP_c}{2TP_c + FP_c + FN_c} \quad (6)$$

D. Experiment Results

In Table I, the overall performance of our model compared to other baseline models is presented. Each model was run for 5 iterations, and the mean and standard deviation of the evaluation metrics are reported in the table. Our baseline model is a single classifier without a metadata processing module. Four deep learning-based models and three machine learning-based models were compared to our proposed models. Only mean values are reported for the machine learning-based models. Below are brief explanations of the deep learning-based baseline models.

- Featurefusion [37]: This approach employs two modules, each containing Conv1D and LSTM layers, to capture spatio-temporal features. The extracted features are then fused using a tanh gate.
- ECGTransformer [38]: Three class human states were classified by Transformer-based network. The model consists of CNN layers and transformer modules.
- EnsembleCNN [35]: An ensemble mode with hybrid learning was proposed in [35]. A CNN model and CNN-LSTM based model was trained separately. The ensemble model predicts a class by soft voting.
- DeepCNN [39]: This method uses Conv1D layers and a multilayer perceptron neural network to extract features from each individual sensor, which are then combined for analysis. Chest data was used for classification of three different human states (normal, stress, amusement).

Overall, our method, which adopts an ensemble network, showed the best classification results compared to other models.

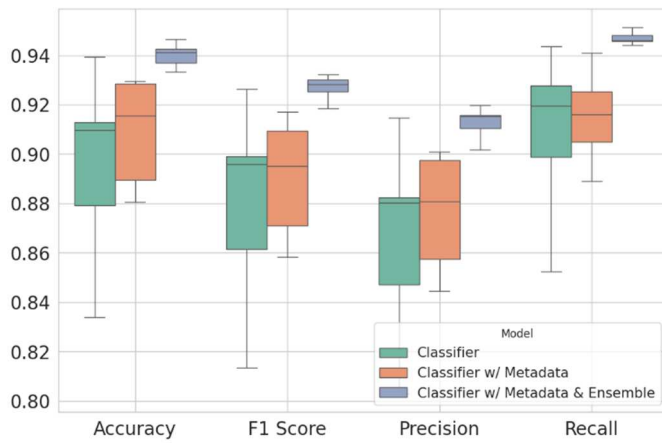


Figure 4. Box plot of classification metrics for the proposed models

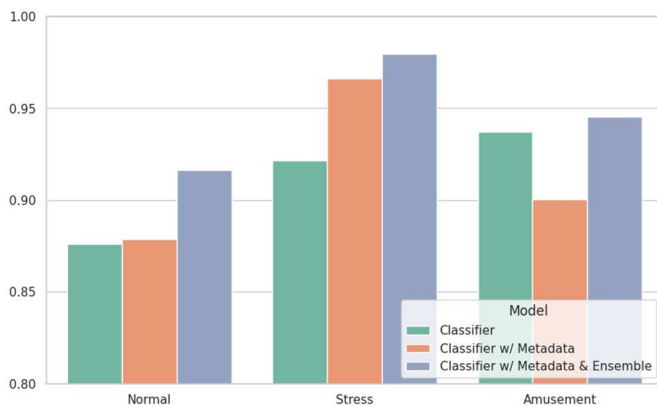


Figure 5. Comparison of the accuracy of the proposed models for each class

The proposed ensemble model successfully reduced the variance in classification performance from 2.02% to 0.46% in terms of accuracy, compared to the classifier without the ensemble. The model outperformed the second-best model by 2.24% and other ensemble-based model by 6.22%. The proposed attention module successfully assigns appropriate weights to the prediction values, enhancing the performance.

As shown in Figure 4, the box plot for mean metric values over 5 iterations are provided. It is evident that the single classifier model with metadata achieved higher accuracy than the model without metadata. By using the ensemble network, we achieved both higher accuracy and lower variance compared to the single classifier without metadata. This suggests that the ensemble model is better able to leverage the features provided by the metadata, thereby enhancing the model's ability to classify human states more accurately.

In Figure 5, It appears that injecting metadata into the classifier results in lower accuracy for the amusement class while improving the performance for the other classes consistently. This is likely because the additional information made the network unstable, leading to performance degradation. However, by adopting an ensemble architecture, the overall accuracy across all classes improved with mean accuracy improvements of 3.73% for normal class, 1.33% for stress class and 4.49% for amusement class comparing to the metadata-only model for 5 iterations.

V. CONCLUSIONS

In this study, we proposed an ensemble stress detection model that captures inherent multimodal features tailored for each individual by incorporating metadata. Using the WESAD dataset, which contains physiological and motion signals, the performance of the proposed model was compared to other deep learning and machine learning-based models. Our model achieved the highest mean accuracy of 94.01% and the lowest classification variance among the comparative models, as well as the highest F1-score, precision, and recall. This indicates that the proposed model is well-suited for integration into stress management systems in workplace environments, supporting IoT applications.

ACKNOWLEDGMENT

This work was supported by the Korea Industrial Complex Corporation (KICOX) with funding from the Ministry of Trade, Industry and Energy (MOTIE) of the Republic of Korea. (Project No. SG20240201, Development and Demonstration of AI Services for Manufacturing Industry Specialization).

REFERENCES

- [1] G. H. Lee, H. Park, J. W. Jang, J. Han, and J. K. Choi, "PPO-based autonomous transmission period control system in IoT edge computing," *IEEE Internet of Things Journal*, vol. 10, no. 24, pp. 21 705–21 720, 2023.
- [2] H. Park, G. H. Lee, J. Han, and J. K. Choi, "Multiclass autoencoder-based active learning for sensor-based human activity recognition," *Future Generation Computer Systems*, vol. 151, pp. 71–84, 2024.
- [3] Y. S. Alshehry, K. A. Sidek, M. G. M. Johar et al., "Stress recognition using electroencephalogram (eeg) signal," in *Journal of Physics: Conference Series*, vol. 1502, no. 1. IOP Publishing, 2020, p. 012052.
- [4] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, pp. 400–408.
- [5] P. Bobade and M. Vani, "Stress detection with machine learning and deep learning using multimodal physiological data," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2020, pp. 51–57.
- [6] L. Huynh, T. Nguyen, T. Nguyen, S. Pirttikangas, and P. Siirtola, "Stressnas: Affect state and stress detection using neural architecture search," in *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 2021, pp. 121–125.
- [7] E. K. Kelloway, J. K. Dimoff, and S. Gilbert, "Mental health in the workplace," *Annual Review of Organizational Psychology and Organizational Behavior*, vol. 10, no. 1, pp. 363–387, 2023.
- [8] K. Plarre, A. Raij, S. M. Hossain, A. A. Ali, M. Nakajima, M. Al'Absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott et al., "Continuous inference of psychological stress from sensory measurements collected in the natural environment," in *Proceedings of the 10th ACM/IEEE international conference on information processing in sensor networks*. IEEE, 2011, pp. 97–108.
- [9] Y. S. Can, B. Arnrich, and C. Ersoy, "Stress detection in daily life scenarios using smart phones and wearable sensors: A survey," *Journal of biomedical informatics*, vol. 92, p. 103139, 2019.
- [10] J. Hassard, K. Teoh, T. Cox, M. Cosmar, R. Grundler, D. Flemming, B. Cosmans, and K. Van den Broek, "Calculating the cost of workrelated stress and psychosocial risks," 2014.
- [11] B. F. Wijnen, J. Lokkerbol, C. Boot, B. M. Havermans, A. J. van der Beek, and F. Smit, "Implementing interventions to reduce work-related stress among health-care workers: An investment appraisal from the employer's perspective," *International archives of occupational and environmental health*, vol. 93, no. 1, pp. 123–132, 2020.

- [12] C. L. Stewart, A. Folarin, and R. Dobson, "Personalized acute stress classification from physiological signals with neural processes," *arXiv preprint arXiv:2002.04176*, 2020.
- [13] B. Chu, K. Marwaha, T. Sanvictores, A. O. Awosika, and D. Ayers, "Physiology, stress reaction," in StatPearls [Internet]. StatPearls Publishing, 2024.
- [14] S. Aziz, S. Ahmed, and M.-S. Alouini, "Ecg-based machine-learning algorithms for heartbeat classification," *Scientific reports*, vol. 11, no. 1, p. 18738, 2021.
- [15] S. Celin and K. Vasanth, "Ecg signal classification using various machine learning techniques," *Journal of medical systems*, vol. 42, no. 12, p. 241, 2018.
- [16] M. Wasimuddin, K. Elleithy, A.-S. Abuzneid, M. Faezipour, and O. Abuzaghlh, "Stages-based ecg signal analysis from traditional signal processing to machine learning approaches: A survey," *IEEE Access*, vol. 8, pp. 177 782–177 803, 2020.
- [17] S. Gedam and S. Paul, "A review on mental stress detection using wearable sensors and machine learning techniques," *IEEE Access*, vol. 9, pp. 84 045–84 066, 2021.
- [18] H. Park, N. Kim, G. H. Lee, and J. K. Choi, "Multicnn-filterlstm: Resource-efficient sensor-based human activity recognition in IoT applications," *Future Generation Computer Systems*, vol. 139, pp. 196–209, 2023.
- [19] G. H. Lee, J. Han, and J. K. Choi, "MPdist-based missing data imputation for supporting big data analyses in IoT-based applications," *Future Generation Computer Systems*, vol. 125, pp. 421–432, 2021.
- [20] Y. Dev, M. Namdev, R. Shrivastava, and R. Srivastava, "Lstm based mental stress level detection using wearable sensor devices," *Current Trends in Technology & Science*, vol. 11, no. 1, pp. 1–4, 2022.
- [21] R. Katmah, F. Al-Shargie, U. Tariq, F. Babiloni, F. Al-Mughairbi, and H. Al-Nashash, "A review on mental stress assessment methods using eeg signals," *Sensors*, vol. 21, no. 15, p. 5043, 2021.
- [22] T. H. Priya, P. Mahalakshmi, V. Naidu, and M. Srinivas, "Stress detection from eeg using power ratio," in *2020 International conference on emerging trends in information technology and engineering (ic-ETITE)*. IEEE, 2020, pp. 1–6.
- [23] A. Islam, A. K. Sarkar, and T. Ghosh, "Eeg signal classification for mental stress during arithmetic task using wavelet transformation and statistical features," in *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)*. IEEE, 2021, pp. 1–6.
- [24] M. Quadrini, A. Capuccio, D. Falcone, S. Daberdaku, A. Blanda, L. Bellanova, and G. Gerard, "Stress detection with encoding physiological signals and convolutional neural network," *Machine Learning*, pp. 1–29, 2024.
- [25] D. Chatterjee, S. Dutta, R. Shaikh, and S. K. Saha, "A lightweight deep neural network for detection of mental states from physiological signals," *Innovations in Systems and Software Engineering*, pp. 1–8, 2022.
- [26] S. Li, W. Li, Z. Xing, W. Yuan, X. Wei, X. Zhang, and B. Hu, "A personality-guided affective brain-computer interface for implementation of emotional intelligence in machines," *Frontiers of Information Technology & Electronic Engineering*, vol. 23, no. 8, pp. 1158–1173, 2022.
- [27] A. Nagrani, S. Yang, A. Arnab, A. Jansen, C. Schmid, and C. Sun, "Attention bottlenecks for multimodal fusion," *Advances in neural information processing systems*, vol. 34, pp. 14 200–14 213, 2021.
- [28] 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 transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2011.
- [29] R. Wang, W. Jo, D. Zhao, W. Wang, A. Gupte, B. Yang, G. Chen, and B.-C. Min, "Husformer: A multi-modal transformer for multi-modal human state recognition," *IEEE Transactions on Cognitive and Developmental Systems*, 2024.
- [30] X. Zhang, X. Wei, Z. Zhou, Q. Zhao, S. Zhang, Y. Yang, R. Li, and B. Hu, "Dynamic alignment and fusion of multimodal physiological patterns for stress recognition," *IEEE Transactions on Affective Computing*, 2023.
- [31] G. Vos, K. Trinh, Z. Samyai, and M. R. Azghadi, "Ensemble machine learning model trained on a new synthesized dataset generalizes well for stress prediction using wearable devices," *Journal of Biomedical Informatics*, vol. 148, p. 104556, 2023.
- [32] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- [33] M. Kang, S. Shin, J. Jung, and Y. T. Kim, "Classification of mental stress using cnn-lstm algorithms with electrocardiogram signals," *Journal of Healthcare Engineering*, vol. 2021, no. 1, p. 9951905, 2021.
- [34] S. Chatterjee and Y.-C. Byun, "Eeg-based emotion classification using stacking ensemble approach," *Sensors*, vol. 22, no. 21, p. 8550, 2022.
- [35] G. Singh, O. C. Phukan, and R. Kumar, "Stress recognition with multi-modal sensing using bootstrapped ensemble deep learning model," *Expert Systems*, vol. 40, no. 6, p. e13239, 2023.
- [36] J. Li, P. Washington et al., "A comparison of personalized and generalized approaches to emotion recognition using consumer wearable devices: machine learning study," *JMIR AI*, vol. 3, no. 1, p. e52171, 2024.
- [37] Feng, Mingxu, et al. "Affect and stress detection based on feature fusion of LSTM and 1DCNN." *Computer Methods in Biomechanics and Biomedical Engineering* 27.4 (2024): 512-520.
- [38] B. Behinaein, A. Bhatti, D. Rodenburg, P. Hungler, and A. Etemad, "A transformer architecture for stress detection from ecg," in *Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 2021, pp. 132–134.
- [39] Li, Russell, and Zhandong Liu. "Stress detection using deep neural networks." *BMC Medical Informatics and Decision Making* 20 (2020): 1-10.



Kyungtaek Oh received the B.Sc. degree in electrical and computer engineering from University of Seoul, Seoul, Republic of Korea, in 2023. He is currently pursuing the M.S. degree in electrical engineering in Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea. His main research interests include machine learning and Internet of Things (IoT).



Jun Kyun Choi received the B.Sc. (Eng.) from Seoul National University in electronics engineering, Seoul, Korea in 1982, and M.Sc. (Eng.) and Ph.D. degree in 1985 and 1988, respectively, in electronics engineering from Korea Advanced Institute of Science and Technology (KAIST). From June 1986 until December 1997, he was with the Electronics and Telecommunication Research Institute (ETRI). In January 1998, he joined the Information and Communications University (ICU), Daejeon, Korea as Professor. At the year of 2009, he moved to Korea Advanced Institute of Science and Technology (KAIST) as Professor. He is a Senior Member of IEEE.



Hyunseo Park received the B.Sc. degree in electrical and computer engineering from University of Seoul, Seoul, Republic of Korea, in 2017, completed the M.Sc. in electrical engineering from Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea in 2019, and the Ph.D. degree in 2024 in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea. He is currently working as a post-doctoral researcher in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea. His current research interests include active learning and other machine learning tasks.



Seungjin Lee received the B.S. degree in Information and Communication Engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2005, and the M.S. and Ph.D. degrees in Information and Communication Engineering from KAIST, Daejeon, South Korea, in 2008 and 2023, respectively. From March 2023 to February 2024, Dr. Lee served as a post-doctoral researcher at the Institute for IT Convergence, KAIST. In 2024, he joined Kyungnam University, Changwon, South Korea, as an Assistant Professor in the School of AI-SW Convergence. His research interests include artificial intelligence, autonomous manufacturing, autonomous driving, and machine learning for networking.