2024 30th International Conference on Mechatronics and Machine Vision in Practice (M2VIP) | 979-8-3503-9191-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/M2VIP62491.2024.10746007

Using Physiological Signals for Pain Assessment: An Evaluation of Deep Learning Models

Jianan Zheng
Intelligent Human-Machine Systems Laboratory
Mechanical and Industrial Engineering
Northeastern University
Boston, USA
zheng.j@northeastern.edu

Yingzi Lin*
Intelligent Human-Machine Systems Laboratory
Mechanical and Industrial Engineering
Northeastern University
Boston, USA
yi.lin@northeastern.edu

***Abstract*—Pain assessment is of major significance in clinical environments. The current gold standard is self-reporting of pain based on the patient's subjective willingness. However, pain assessment based on physiological signals is developing rapidly due to the objectivity and real-time nature of physiological signals. This study aims to systematically compare the performance of convolutional neural networks (CNN) combined with long short-term memory networks (LSTM), bidirectional long short-term memory networks (BiLSTM), transformers, and gated recurrent units (GRU) models in pain classification tasks. We assessed these hybrid models' performance experimentally using a variety of metrics, such as accuracy, precision, recall, F1 score, training time, and inference time. The experimental results show that the CNN+Transformer model best performs in most evaluation metrics with an accuracy of 0.795, while the CNN+GRU model performs the worst with an accuracy of only 0.559. In addition, we also analyze the computational efficiency of each model in terms of training and inference time. Overall, this paper provides important direction for future research and real-world applications by thoroughly evaluating the effectiveness of various deep learning models in pain classification tasks.**

***Index Terms*—Physiological signals, pain assessment, and deep learning.**

I. Introduction

Chronic pain is persistent or recurring and has a serious impact on the patient's physical health and quality of life [1], [2]. It is crucial to develop a personalized treatment plan based on the patient's pain level. Pain assessment typically categorizes pain into three levels: no pain, mild pain, and severe pain [3]. Depending on the pain level, different pain management strategies are required. Accurate pain assessment can help clinicians develop more effective and personalized treatment plans [4]. Accurately assessing pain has become a major challenge in the clinical environment. Patients self-report using the verbal rating scale (VRS), visual analog scale (VAS), and numerical rating scale (NRS) as part of the traditional method for assessing pain. [5]–[7]. These pain assessment methods based on patient self-report have many limitations, mainly due to their strong subjectivity, inability to obtain real-time data, and difficulty in standardization [8]. Pain is not only a subjective feeling but also causes a variety of physiological reactions [9]. Because of the objectivity and real-time nature of physiological signals, it has become a new method for pain assessment [10]. Many researchers use physiological signals such as blood volume pulse (BVP), skin conductance (SC), electromyography (EMG), electrocardiography (ECG), electroencephalography (EEG), temperature, and respiration rate to assess pain [11]–[13]. Some researchers also use facial expressions and pupillary responses to assess pain [14], [15].

A significant challenge in pain level classification tasks is how to achieve high accuracy while maintaining computational efficiency. Convolutional neural networks (CNN) are frequently used for feature extraction because of their capacity to extract spatial hierarchies from data. [16]. But CNN can make the model work better when it is mixed with other architectures like long short-term memory networks (LSTM), bidirectional long short-term memory networks (BiLSTM), gated recurrent units (GRU), and transformers [17]–[21]. This is because they can use the best features of each architecture.

In this paper, we assessed the performance of CNN combined with LSTM, BiLSTM, Transformer, and GRU architectures in the pain level classification task based on multiple performance metrics. We show which model is the most effective and efficient by comparing these combined architectures in great detail. This helps us figure out which model does the best overall job of classifying pain levels. The rest of this paper is organized as follows: Section 2 shows the research methods, including the dataset (subject information and experimental procedure), data preprocessing, feature extraction, deep learning model architecture, and model performance evaluation. Section 3 shows the experimental results of the model. Section 4 shows the research findings and discussion. Section 5 shows the conclusion of this paper and proposes directions for future research.

II. Methodology

A. Dataset

1) Participants:

This experiment recruited 29 healthy subjects. Our inclusion criterion is no history of neurological, psychiatric, and cardiovascular problems, and no experiencing pain before the experiment. Before the experiment, the researcher provided the subjects with detailed experimental procedures in the written

consent form, and the researcher will explain and confirm that the subjects understand all of the experiment's details through verbal explanations. Northeastern University's Institutional Review Board (IRB#19-12-15) set forth the guidelines and regulations that the experiment followed.

2) Experiment Procedure:

The subjects signed a consent form about the experiment. Then the researcher placed biosensor devices on the subjects' hands, head, arms, and chest. These biosensor devices measured brain activity through electroencephalography (EEG), sweating through galvanic skin response (GSR), heart rate with electrocardiography (ECG), muscle tension with electromyography (EMG), skin temperature (ST), respiration rate, and eye movement. The maximum period for the experiment is 200 seconds. After recording the first period of physiological signal data for baseline purposes, the researcher will apply controlled stimuli. The subjects reported their level of pain every 30 seconds. The subject could terminate the experiment at any time when they could not tolerate the pain. After the experiment, the subject was reimbursed.

B. Data Preprocessing

The physiological signals selected as model inputs are blood volume pulse (BVP), galvanic skin response (GSR), electromyography (EMG), skin temperature (ST), and respiration rate (RR) [22]. We resampled these data to 50 Hz. The blood volume pulse (BVP) is filtered using a fifth-order Butterworth bandpass filter with a cutoff frequency of [0.5, 12] Hz. The galvanic skin response (GSR) is filtered out using a fifth-order 1 Hz low-pass Butterworth filter. The respiration rate (RR) is filtered by a fifth-order Butterworth bandpass filter with a cutoff frequency of [0.1, 1] Hz. Next, we split the data into groups of 2 seconds and removed all NaN values. Then, we classify the data based on the label value: we classify the group with a label value of 0 as no pain, the group with a label value between 0 and 5 as mild pain, and the group with a label value between 5 and 10 as severe pain. We pad the data to make sure that each group is the same size before splitting the dataset into 80% for training data and 20% for testing data.

C. Automatic Feature Extraction

Convolution layers and pooling layers work together to automatically extract features from a convolutional neural network (CNN). [23]. The convolution layer extracts local spatial features by sliding convolution kernels over the input data and doing dot product calculations. Downsampling is done by the pooling layer (max polling) in order to minimize the feature map's size and overfitting risk. CNN is able to learn complex and abstract feature representations by stacking multiple layers of convolution and pooling.

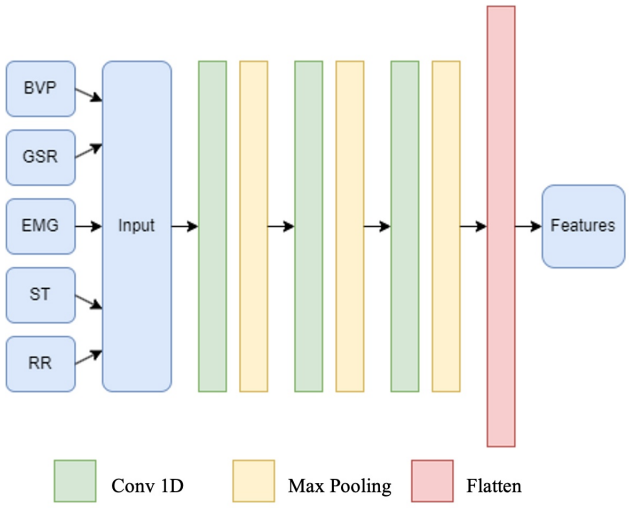

Fig. 1. CNN Structures for Automatic Feature Extraction

D. Deep Learning Models

1) Convolutional Neural Network (CNN) and Long Short-Term Memory Networks (LSTM) Model:

We use a hybrid of a convolutional neural network and long short-term memory networks. The LSTM is used for classification, and the CNN is used for feature extraction. The goal of this combined strategy is to extract temporal and spatial dependencies from the input data.

2) Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory Networks (BiLSTM) Model:

We combine the capabilities of a bidirectional long short-term memory and a convolutional neural network. The BiLSTM is used for classification, and the CNN is used for feature extraction. The combined method seeks to extract from the input data both bidirectional temporal dependencies and spatial dependencies.

3) Convolutional Neural Network (CNN) and Transformer Model:

We employ the CNN+Transformer model, which combines a transformer with a convolutional neural network. The transformer is used for classification, and the CNN is used for feature extraction. Through the transformer's self-attention mechanism, this combined approach seeks to capture both spatial features and intricate temporal dependencies in the input data.

4) Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) Model:

We employ the CNN+GRU model, which combines GRU with a convolutional neural network. The CNN is used for feature extraction, and the GRU is used for classification. This combined approach aims to leverage GRU's ability to handle temporal dependencies efficiently and enhance overall classification performance by integrating both spatial and temporal information.

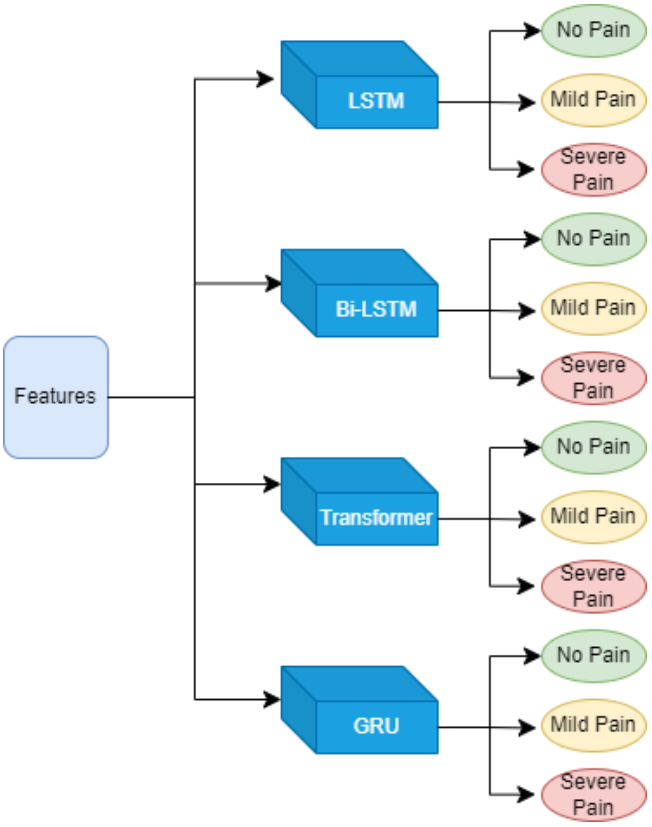

Fig. 2. Four Deep Learning Models Perform Classification Tasks on The Features Extracted by CNN: LSTM, BiLSTM, Transformer, and GRU Structures.

E. Performance Evaluation

The evaluation of deep learning models will start with multiple indicators. We will use the information obtained from the model's confusion matrix to calculate the true positives, true negatives, false positives, and false negatives. Next, we will use these values to determine the model's accuracy. To comprehensively evaluate the model's performance, we will measure precision, recall, and F1 score through macro-averaging, micro-averaging, and weighted-averaging. Finally, we will compute the training and inference times of the models.

1) Confusion Matrix:

The confusion matrix provides an intuitive way to visualize the performance of the deep learning model. We can use the confusion matrix to observe the model's prediction results in each category, including the number of correct and incorrect classifications. The confusion matrix of deep learning model can be used to compute accuracy, precision, recall, and F1 score, among other performance metrics. These indicators can evaluate the deep learning model more comprehensively, especially in the case of imbalanced categories. We can also analyze the confusion matrix to optimize the deep learning model.

TABLE I
CONFUSION MATRIX FOR DEEP LEARNING MODEL

	Predicted N	Predicted M	Predicted S
Actual N	E_{NN}	E_{NM}	E_{NS}
Actual M	E_{MN}	E_{MM}	E_{MS}
Actual S	E_{SN}	E_{SM}	E_{SS}

where:

- The category N denotes the level of no pain.
- The category M denotes the level of mild pain.
- The category S denotes the level of severe pain.
- E_{ij} denotes that the actual category is i and the predicted category is j.

TP_i (true positives) are the number of samples with the same actual and predicted categories.

$$TP_i = E_{ii} \tag{1}$$

FP_i (false positives) are the number of samples whose actual category is not i but are mistakenly predicted to be i.

$$FP_i = \sum_{j \neq i} E_{ji} \tag{2}$$

FN_i (false negatives) are the number of samples whose actual category is i but are mistakenly predicted to be other categories.

$$FN_i = \sum_{j \neq i} E_{ij} \tag{3}$$

TN_i (true negatives) are the number of samples whose actual category is not i and is not predicted to be i.

$$TN_i = \sum_{k \neq i} \sum_{j \neq i} E_{kj} \tag{4}$$

2) Accuracy:

Accuracy is defined as the percentage of true predictions (true positives and true negatives) among all predictions. It is a broad indicator of the frequency with which the deep learning model is accurate. When a model has a high accuracy rate, it typically means that most samples are classified correctly overall.

$$\text{Accuracy} = \frac{\sum_{i=1}^{3} TP_i + \sum_{i=1}^{3} TN_i}{\sum_{i=1}^{3} (TP_i + FP_i + FN_i + TN_i)} \tag{5}$$

3) Precision:

Precision is defined as the proportion of samples that are true positive among samples predicted as positive by the deep learning model, that is, how many of the positive samples predicted by the model are accurate. High precision usually means that the model is very accurate in predicting the positive class, which means that the number of false positives is small.

$$\text{Macro Precision} = \frac{1}{3} \sum_{i=1}^{3} \frac{TP_i}{TP_i + FP_i} \tag{6}$$

$$\text{Micro Precision} = \frac{\sum_{i=1}^{3} TP_i}{\sum_{i=1}^{3} (TP_i + FP_i)} \tag{7}$$

$$\text{Weighted Precision} = \frac{\sum_{i=1}^{3} (\text{Precision}_i \times (\text{TP}_i + \text{FN}_i))}{\sum_{i=1}^{3} (\text{TP}_i + \text{FN}_i)} \tag{8}$$

4) Recall:

Recall quantifies the percentage of correctly predicted positive samples. A high recall means that there are fewer false negatives, and the deep learning model can identify most of the samples that are actually positive. High recall is very important for applications where the cost of false negatives is high.

$$\text{Macro Recall} = \frac{1}{3} \sum_{i=1}^{3} \frac{TP_i}{TP_i + FN_i} \tag{9}$$

$$\text{Micro Recall} = \frac{\sum_{i=1}^{3} TP_i}{\sum_{i=1}^{3} (TP_i + FN_i)} \tag{10}$$

$$\text{Weighted Recall} = \frac{\sum_{i=1}^{3} (\text{Recall}_i \times (\text{TP}_i + \text{FN}_i))}{\sum_{i=1}^{3} (\text{TP}_i + \text{FN}_i)} \tag{11}$$

5) F1 Score:

The performance of precision and recall is combined to create the F1 score, which is the harmonic mean of precision and recall. It works especially well with the unequal distribution of classes. Recall and precision are satisfactorily balanced when the F1 score is high.

$$\text{Macro F1 Score} = \frac{1}{3} \sum_{i=1}^{3} 2 \cdot \frac{\text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{12}$$

$$\text{Micro F1 Score} = \frac{2 \cdot \text{Micro Precision} \cdot \text{Micro Recall}}{\text{Micro Precision} + \text{Micro Recall}} \tag{13}$$

$$\text{Weighted F1 Score} = \frac{\sum_{i=1}^{3} (\text{F1 Score}_i \times (\text{TP}_i + \text{FN}_i))}{\sum_{i=1}^{3} (\text{TP}_i + \text{FN}_i)} \tag{14}$$

6) Training Time:

Training time is the total time required for a model from the start of training to the end of training. It can measure the model training process's efficiency. For models that require frequent training, a shorter training time indicates increased development efficiency and reduced computing costs.

$$\text{Training Time} = \sum_{i=1}^{N} t_{\text{train}i} \tag{15}$$

where:

- t_{train_i} represents the time required for the i-th training iteration.
- N is the total number of training iterations.

7) Inference Time:

Inference time is the time it takes for a model to receive input data and output a prediction result. It is used to measure the response speed of the model in real-world applications. Short inference time is crucial for real-time systems and application scenarios with high response speed.

$$\text{Inference Time} = \frac{\sum_{j=1}^{M} t_{\text{infer}j}}{M} \tag{16}$$

where:

- t_{infer_j} represents the time required for the j-th inference.
- M is the total number of inferences.

III. Experiments and Results

A. Confusion Matrix

In this section, we present the confusion matrices for each of the deep learning models we have evaluated. These confusion matrices show how well each model separates into various classes and give a thorough analysis of each model's performance on the classification task. We can discover more about the benefits and drawbacks of each model and pinpoint particular classes that are commonly misclassified by looking at the confusion matrices. This analysis is crucial for refining our models and improving their accuracy and robustness in real-world applications.

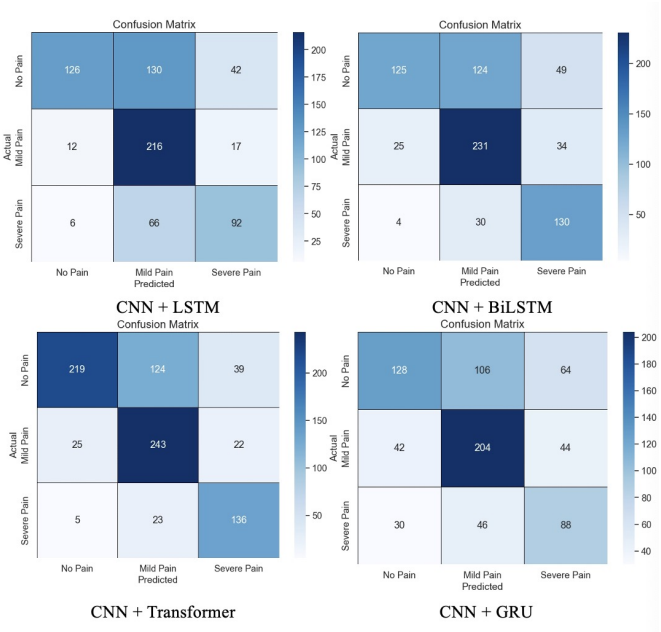

Fig. 3. Confusion Matrix for Four Deep Learning Models: CNN+LSTM, CNN+BiLSTM, CNN+Transformer, and CNN+GRU.

B. Model Performance

In this section, we summarize the performance metrics of each deep learning model, including accuracy, macro-average precision, micro-average precision, weighted precision, macro-average recall, micro-average recall, weighted

F1 score, macro-average F1 score, micro-average F1 score, weighted F1 score, training time, and inference time.

To compute the macro-average, the indicators of each category are computed independently, and the average of these indicators is subsequently determined. The performance of each category is taken into account by the macro-average, which does not take into account the quantity of samples within each category. The micro-average is used to calculate all samples with the same weight, which can reflect the performance of the overall model. In order to fairly represent the performance of each category, the weighted average computes the indicators for each category based on the quantity of samples in each category.

TABLE II
DEEP LEARNING MODELS PERFORMANCE

	Model 1	Model 2	Model 3	Model 4
Accuracy	0.637	0.646	0.795	0.559
Macro precision	0.685	0.674	0.788	0.554
Micro precision	0.637	0.646	0.795	0.559
Weighted precision	0.700	0.686	0.805	0.573
Macro recall	0.628	0.670	0.801	0.557
Micro recall	0.637	0.646	0.795	0.559
Weighted recall	0.637	0.646	0.795	0.559
Macro F1 score	0.618	0.642	0.790	0.545
Micro F1 score	0.637	0.646	0.795	0.559
Weighted F1 score	0.623	0.634	0.796	0.554
Training time	40.01s	48.96s	29.13s	49.06s
Inference time	0.43s	0.66s	0.29s	0.46s

where:

- Model 1 is CNN+LSTM.
- Model 2 is CNN+BiLSTM.
- Model 3 is CNN+Transformer.
- Model 4 is CNN+GRU.

IV. DISCUSSION

The outcomes demonstrate that the four models' performances differ significantly from each other. The CNN+LSTM model's accuracy is 0.637, and its performance is comparable to, but slightly inferior to, the CNN+BiLSTM model. Its confusion matrix shows particular difficulty in distinguishing the category of severe pain. The CNN+BiLSTM model's accuracy is 0.646, which is better than the CNN+LSTM model, but not as good as the CNN+Transformer model. Despite the demonstrated improvements, distinguishing between the categories of no pain and mild pain still poses challenges. With an accuracy of 0.795, the CNN+Transformer model outperforms other models, and the macro-average and micro-average metrics are also the highest. Its confusion matrix shows that the model performs well in correctly identifying all three categories, demonstrating the effectiveness of the transformer architecture. Finally, the CNN+GRU model's accuracy is 0.559, and both macro-average and micro-average indicators are lower than other models. The confusion matrix shows that the model has considerable difficulty distinguishing the three categories, particularly between the categories of no pain and mild pain. It is important to note that the confusion matrices reveal an underlying issue of dataset imbalance within the dataset, which may exacerbate the difficulty in classifying certain categories. Although this paper's main objective is to compare the performance of various deep learning models, future research could place greater emphasis on addressing the dataset imbalance issue. Overall, the CNN+Transformer model performs best across all metrics and effectively reduces misclassification, indicating that the Transformer architecture and its attention mechanism are very effective for this classification task.

V. CONCLUSION

The primary research objective of this paper is to compare the overall performance of four different deep learning models in the pain level classification task and comprehensively explore the effectiveness and differences of these models in handling classification tasks. Experimental results show that the CNN+Transformer model performs best among all models, with an accuracy of 0.795, and outperforms other models in both macro-average and micro-average indicators. This shows that the Transformer architecture and its attention mechanism have significant advantages in handling classification tasks.

The results of this paper not only show the performance differences between different deep learning models in classification tasks but also emphasize the importance of choosing a suitable model architecture. In particular, the Transformer model's excellent performance provides strong support for applying this architecture to similar tasks in the future.

In summary, this paper offers insightful information about the application of deep learning models to classification tasks and provides guidance for future research and practice. Future research can further optimize the model architecture and hyperparameter settings, as well as explore more innovative deep learning methods to improve classification performance.

ACKNOWLEDGMENT

This work has been financially supported by a U.S. National Science Foundation project entitled "Novel Computational Methods for Continuous Objective Multimodal Pain Assessment Sensing System (COMPASS)" under the award #1838796.

REFERENCES

[1] A. K. Szewczyk, A. Jamroz-Wiśniewska, N. Haratym, and K. Rejdak, "Neuropathic Pain and Chronic Pain as An Underestimated Interdisciplinary Problem," Int J Occup Med Environ Health, vol. 35, no. 3, pp. 249–264, doi: 10.13075/ijomeh.1896.01676.

[2] P. R. Tutelman et al., "Epidemiology of chronic pain in children and adolescents: a protocol for a systematic review update," BMJ Open, vol. 11, no. 2, p. e043675, Feb. 2021, doi: 10.1136/bmjopen-2020-043675.

[3] W. Zhu, Y. Xiao, and Y. Lin, "A novel labeling method of physiological-based pressure pain assessment among patients with and without chronic low back pain," Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 68, Sep. 2024, doi: 10.1177/10711813241284506.

[4] D. Hui and E. Bruera, "A Personalized Approach to Assessing and Managing Pain in Patients With Cancer," J Clin Oncol, vol. 32, no. 16, pp. 1640–1646, Jun. 2014, doi: 10.1200/JCO.2013.52.2508.

[5] M. J. Hjermstad et al., "Studies Comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for Assessment of Pain Intensity in Adults: A Systematic Literature Review," Journal of Pain and Symptom Management, vol. 41, no. 6, pp. 1073–1093, Jun. 2011, doi: 10.1016/j.jpainsymman.2010.08.016.

[6] C. T. Hartrick, J. P. Kovan, and S. Shapiro, "The numeric rating scale for clinical pain measurement: a ratio measure?," Pain Practice, vol. 3, no. 4, pp. 310–316, 2003.

[7] M. P. Jensen, C. Chen, and A. M. Brugger, "Interpretation of visual analog scale ratings and change scores: a reanalysis of two clinical trials of postoperative pain," J Pain, vol. 4, no. 7, pp. 407–414, 2003.

[8] Y. Kang and G. Demiris, "Self-report pain assessment tools for cognitively intact older adults: Integrative review," International Journal of Older People Nursing, vol. 13, no. 2, p. e12170, 2018, doi: 10.1111/opn.12170.

[9] Y. Lin, L. Wang, Y. Xiao, R. D. Urman, R. Dutton, and M. Ramsay, "Objective Pain Measurement based on Physiological Signals," Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care, vol. 7, no. 1, pp. 240–247, Jun. 2018, doi: 10.1177/2327857918071056.

[10] W. Zhu, C. Liu, H. Yu, Y. Guo, Y. Xiao, and Y. Lin, "COMPASS App: A Patient-centered Physiological based Pain Assessment System," Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 67, no. 1, pp. 1361–1367, Sep. 2023, doi: 10.1177/21695067231192200.

[11] M. Yu, M. Dong, J. Han, Y. Lin, L. Zhu, X. Tang, G. Sun, Y. He, Y. Guo, "EEG-based tonic cold pain assessment using extreme learning machine," Intelligent Data Analysis, vol. 24, no. 1, pp. 163–182, 2020, doi: 10.3233/IDA-184388.

[12] Y. Lin, Y. Xiao, L. Wang, Y. Guo, W. Zhu, B. Dalip, S. Karmarthi, K. L. Schreiber, R. R. Edwards, R. D. Urman, "Experimental Exploration of Objective Human Pain Assessment Using Multimodal Sensing Signals," Front Neurosci, vol. 16, Feb. 2022, doi: 10.3389/fnins.2022.831627.

[13] L. Wang, Y. Xiao, R. D. Urman, and Y. Lin, "Cold pressor pain assessment based on EEG power spectrum," SN Appl Sci, vol. 2, no. 12, Dec. 2020, doi: 10.1007/s42452-020-03822-8.

[14] J. Zheng and Y. Lin, "An Objective Pain Measurement Machine Learning Model through Facial Expressions and Physiological Signals," in 2022 28th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), IEEE, Nov. 2022, pp. 1–4. doi: 10.1109/M2VIP55626.2022.10041105.

[15] L. Wang, Y. Guo, B. Dalip, Y. Xiao, R. D. Urman, and Y. Lin, "An experimental study of objective pain measurement using pupillary response based on genetic algorithm and artificial neural network," Applied Intelligence, vol. 52, no. 2, pp. 1145–1156, Jan. 2022, doi: 10.1007/s10489-021-02458-4.

[16] M. Jogin, Mohana, M. S. Madhulika, G. D. Divya, R. K. Meghana, and S. Apoorva, "Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning," in 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), May 2018, pp. 2319–2323. doi: 10.1109/RTEICT42901.2018.9012507.

[17] Y. Guo, L. Wang, Y. Xiao, and Y. Lin, "A Personalized Spatial-Temporal Cold Pain Intensity Estimation Model Based on Facial Expression," IEEE Journal of Translational Engineering in Health and Medicine, vol. 9, pp. 1–8, 2021, doi: 10.1109/JTEHM.2021.3116867.

[18] F. Pouromran, Y. Lin, and S. Kamarthi, "Personalized Deep Bi-LSTM RNN Based Model for Pain Intensity Classification Using EDA Signal," Sensors, vol. 22, no. 21, Art. no. 21, Jan. 2022, doi: 10.3390/s22218087.

[19] J. Cheng, Q. Zou, and Y. Zhao, "ECG signal classification based on deep CNN and BiLSTM," BMC Med Inform Decis Mak, vol. 21, no. 1, p. 365, Dec. 2021, doi: 10.1186/s12911-021-01736-y.

[20] H. Liu, S. Cui, X. Zhao, and F. Cong, "Detection of obstructive sleep apnea from single-channel ECG signals using a CNN-transformer architecture," Biomedical Signal Processing and Control, vol. 82, p. 104581, Apr. 2023, doi: 10.1016/j.bspc.2023.104581.

[21] L. Lu, C. Zhang, K. Cao, T. Deng, and Q. Yang, "A Multichannel CNN-GRU Model for Human Activity Recognition," IEEE Access, vol. 10, pp. 66797–66810, 2022, doi: 10.1109/ACCESS.2022.3185112.

[22] S. Moscato, W. Zhu, Y. Guo, S. Kamarthi, C. A. Colebaugh, K. L. Schreiber, R. R. Edwards, R. D. Urman, Y. Xiao, L. Chiari, and Y. Lin, "Comparison of autonomic signals between healthy subjects and chronic low back pain patients at rest and during noxious stimulation," in National Congress of Bioengineering Proceedings, 2023.

[23] M. Yu, Y. Sun, B. Zhu, L. Zhu, Y. Lin, X. Tang, Y. Guo, G. Sun, M. Dong, "Diverse frequency band-based convolutional neural networks for tonic cold pain assessment using EEG," Neurocomputing, vol. 378, pp. 270–282, Feb. 2020, doi: 10.1016/j.neucom.2019.10.023.