Physiological-based comparison of deep learning models for pain level estimation

Wenchao Zhu
Intelligent Human-Machine Systems Laboratory
Mechanical and Industrial Engineering Department
Northeastern University
Boston, MA, USA

Yingzi Lin*
Intelligent Human-Machine Systems Laboratory
Mechanical and Industrial Engineering Department
Northeastern University
Boston, MA, USA

2024 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI) | 979-8-3503-7916-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/SOLI63266.2024.10956064

***Abstract*— Chronic pain is an unpleasant and sensory experience that affects over 30% of the global population, posing a significant public health crisis. Traditional pain assessment relies heavily on self-report measures such as numerical rating scales, which are prone to biases and are not suitable for individuals unable to communicate their pain level effectively. This study aims to develop an automatic continuous pain level estimation model using physiological sensors and deep learning models. Twenty-nine participants performed a cold pain test while wearing physiological sensors including blood volume pulse (BVP), electromyography (EMG), electrodermal activity (EDA), and respiration (RR). We explored different sensor fusion architectures (data level and feature level fusion) and deep learning models (CNN-MLP, CNN-LSTM, and CNN-XGBoost) to assess their performance in pain level estimation. The result showed that feature level fusion outperforms data level fusion in pain level estimation, and among deep learning models, CNN-XGBoost and CNN-LSTM demonstrated better performance than CNN-MLP. The result also suggests that models using fewer, specific sensors like EDA can nearly match the performance of more complex multi-modality sensor systems. Future work should focus on evaluating the computational efficiency and cost-effectiveness of these models to enhance their applicability in real-time and mobile settings, thereby potentially stemming opioid misuse and improving overall patient outcomes and quality of life.**

Keywords—pain assessment, physiological sensor, deep learning, sensor fusion, pain level

I. INTRODUCTION

Chronic pain is prevalent in the world and disproportionately impacts adults living in underserved regions and developing countries. It is an unpleasant, sensory, and emotional experience that affects more than 30% of people globally [1]. In the United States, it affects 51.6 million adults (20.9% of the population) with 17.1 million (6.9%) suffering from high-impact chronic pain [2], [3]. In China, over 30% of the population suffers from chronic pain, and increases 10-20 million cases annually [4].

Chronic pain arises from a complex integration of sensory, emotional, cognitive, and behavioral components, influencing health outcomes, such as pain intensity level, emotional function, and physical activity [5], [6]. Currently, the gold standard for assessing pain relies primarily on self-reported measures, including the Numerical Rating Scale (NRS) [7], Visual Analogue Scale (VAS) [8], and Verbal rating scale (VRS). The NRS requires patients to rate their pain on a scale from 0 to 10, where 0 indicates 'no pain' and 10 signifies 'severe pain'. The VRS employs verbal adjectives to categorize pain into four levels, including 'no', 'mild', 'moderate', and 'severe'. However, self-reports are susceptible to biases from factors such as anxiety [9], [10], memories [11], current pain level, and physical function [12]. In addition, gold standards are inefficient for patients who are not able to self-report [13], such as stroke patients, patients with verbal impairments, and infants [14]. Therefore, developing validated instruments that objectively and continuously monitor pain, could significantly enhance chronic pain assessment and treatment, and potentially reduce the opioid crisis.

Chronic nociceptive pain is caused by noxious stimuli that potentially damage the body, such as extreme temperatures, pressure, pinching, and chemical exposure [15], [16]. These stimuli activate sensory neurons, which transmit signals to the brain via the peripheral nervous system. The brain generates a pain sensation that prompts automatic responses, such as changes in heart rate, respiration, pupil dilation, and bioimpedance arousal [17], [18]. Physiological sensors have significantly advanced the objective assessment of human states and characteristics [19], [20]. Research has underscored the potential of these sensors for classifying pain intensity levels. For example, Guo et al. estimated three levels of cold pressor pain using facial expression by comparing three neural network models, and the personalized spatial-temporal framework using a convolutional long short-term memory model achieved the highest performance [21], [22]. Another study measured the pain level via features generated from the pupillometry data using a genetic algorithm with an artificial neural network classifier, and the best performance was obtained with an accuracy of 81% [23]. EEG studies have demonstrated statistical differences in central and occipital regions and were able to classify pain and no-pain states using multi-layer CNN frameworks [24], [25]. However, a single sensor may not provide complete information, necessitating the integration of data from multiple sensors to accurately determine pain levels. Multimodal physiological classification with decision-level fusion and feature-level fusion proved promising in pain level detection and classification [26].

This study conducted a comprehensive investigation of sensor fusion architectures to develop an automatic, continuous pain estimation framework. To achieve this aim, 29 subjects were recruited to perform a cold pain test while equipped with physiological sensors. Comparisons were made between deep

Corresponding author: Yingzi Lin (yi.lin@northeastern.edu), https://lin-yingzi.northeastern.edu/.

learning models (CNN-MLP, CNN-LSTM, and CNN-XGBoost) and fusion methods (data level and feature level fusion). Individual sensors and multiple modalities were also compared. Performance was evaluated based on a pain intensity level regression problem, with mean squared error and root mean square error as the measurement metrics.

II. Methods

A. Participants

The experiment recruited 29 healthy participants (11 males and 18 females), all from Northeastern University, aged between 19 and 22. Before the experiment, investigators explained the apparatus and procedures to ensure participants were fully informed. Written informed consent was obtained from all participants. The study was conducted in accordance with the guidelines and regulations of the Northeastern University Institutional Review Board (IRB # 17-01-25).

B. Apparatus

The study utilized various sensors to monitor physiological responses during a cold pain test. These physiological sensors (FlexComp Infiniti, Thought Technology, Canada) included a blood volume pulse (BVP) sensor for heart rate tracking through the middle finger of the non-dominant hand, a chest-mounted respiration rate sensor (RR), electromyography (EMG) sensor on the non-dominant forearm, and an electrodermal activity (EDA) sensor between the index and ring fingers of the non-dominant hand. Data collection and storage were facilitated using a Dell desktop computer.

C. Experiment Procedure

Upon arrival, participants were seated in a comfortable chair at a distance of one meter from a computer screen and instructed to follow on-screen prompts. The experiment commenced with a 5-minute baseline data recording, during which participants were asked to relax fully. The screen then signaled 'GO', prompting participants to immerse their right hand in iced water for 20 seconds, as it is shown in *Figure 1*. Subsequently, the screen displayed 'P', indicating the participants should report their pain intensity level on a scale from 0 (no pain) to 10 (severe pain). The investigator had three seconds to record the pain level. This 23-second procedure constituted one session, repeated until the participant opted to withdraw or reached 10 sessions. Each participant provided up to 11 pain level ratings before concluding the experiment, with the option to terminate at any time.

D. Data Preprocessing

Physiological data (BVP, EMG, EDA, RR) were synchronized by resampling at 50Hz. The BVP signal was filtered using a fifth-order Butterworth band-pass filter with cutoff frequencies between 0.5 and 12 Hz. EDA was processed with a fifth-order 1 Hz low-pass Butterworth filter, and RR with a fifth-order Butterworth band-pass with [0.1, 1] Hz as cutoff frequencies. Outliers were removed using the Interquartile (IQR) range method, which identifies extremes by comparing data points to the 25th and 75th percentiles. Data points outside 1.5 times of IQR from these percentiles were deemed outliers and were replaced via quantile-based flooring and capping.

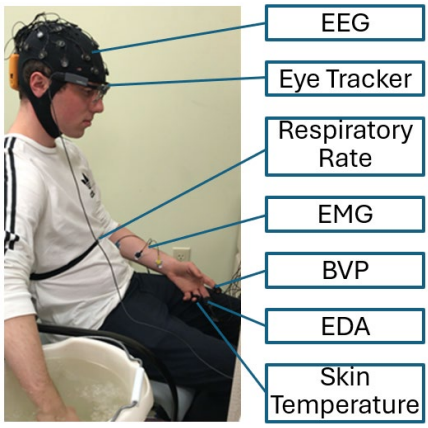

Fig. 1. Cold pressor pain experiment apparatus

To address the issue of class imbalance in our dataset, we employed the Synthetic Minority Oversampling Technique (SMOTE) [27]. This method involves oversampling the minority class by generating synthetic examples, as opposed to resampling with replacement. Initially, SMOTE randomly selects a minority class instance and identifies its k nearest neighbors within the same class. Subsequently, it randomly picks one of these neighbors and creates a line segment in the feature spacing connecting the two instances. Synthetic instances are then produced as a convex combination of these two selected instances. By using this data augmentation technique, we were able to generate a sufficient number of synthetic examples to balance the class distribution in our training dataset.

E. Regression task

Physiological sensor data were utilized as training inputs, while self-reported pain intensity levels served as training labels. We classified pain ratings into four levels: No pain (0), Mild pain (1-3), Moderate pain (4-6), and Severe pain (8-10). Data from each participant were segmented into 10-second time windows. For labeling, we assigned the mean pain rating of two consecutive windows (e.g., from 10s to 20s and 20s to 30s) to the training corresponding to the end of the second window (20s).

To evaluate our models, we employed 5-fold stratified cross-validation across all subjects. This method ensures each fold is representative of the overall dataset by maintaining a consistent proportion of each class, which is crucial in datasets with imbalanced class distributions.

F. Deep Learning Models

For automated pain level estimation from physiological signals, we implemented various deep learning algorithms alongside Extreme Gradient Boosting Regression (XGBoost) algorithms [28]. XGBoost was chosen for its computational efficiency and sensitivity in machine learning tasks involving human state recognition. For example, Pouromran et al. extracted features from ECG, EDA, and EMG signals and utilized XGBoost for predicting pain intensities, demonstrating superior performance compared to conventional methods such

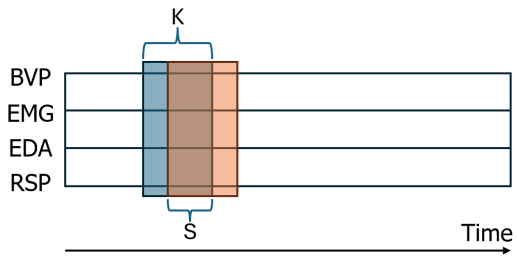

Fig. 2. 1D CNN layer in data level fusion. The Kernel size denoted in K moves along the x axis (Time) and extracts temporal features. After extracting features, the kernel moves to the right with a stride of length S.

as Naive Bayes, KNN and Random Forest [29]. Importantly, XGBoost does not require as extensive training data sets as many deep learning models, making it particularly suitable for our dataset.

We developed an end-to-end model capable of processing raw data to estimate pain levels. For feature extraction, we used a 1D conventional neural network (CNN) layer, which is optimal for time series data due to its one-dimensional operational nature [30], [31]. The 1D CNN layer diagram is illustrated in *Figure 2*. The kernel, designated as K, shifts to the right by a stride of length S after each extraction. This approach mimics the 2D CNN used in image classification but is adapted to the one-dimensional structure of time series data. Additionally, we developed two sensor fusion architectures: data level fusion, and feature level fusion. In both architectures, we employed 1D CNN for automatic feature extraction, followed by regressor models to estimate pain level. We choose multilayer perceptron (MLP), Long Short-Term Memory (LSTM), and XGBoost as the final regressor models.

1) Data level fusion

In the data level fusion approach, we integrated signals from four different physiological sensors – BVP, EMG, EDA, and RR – into a single input tensor for our deep learning model. The input tensor was structured with dimensions of 4x2560, where '4' represents the number of sensor modalities and '2560' corresponds to the data points collected over a 10-second window at a sampling rate of 256 Hz per second. This fusion approach is illustrated in *Figure 3(a)*. Feature extraction from this integrated data was performed using a 1D CNN layer that processes all sensor data simultaneously, ensuring that temporal dynamics across different modalities are captured effectively. Our CNN configuration included a kernel size of 512 and a stride of 64, which was followed by a max pooling layer to reduce dimensionality and enhance the detection of dominant features. Subsequently, a dropout layer with a rate of 0.2 was implemented to prevent overfitting. The architecture was completed with a batch normalization layer to standardize inputs to the next layer, followed by a flatten layer to convert the multi-dimensional output into a 1D array suitable for regression.

2) Feature level fusion

In the feature level fusion approach, features were first extracted individually from each of the four physiological sensors—BVP, EMG, EDA, and RR—using separate 1D CNN layers. This method ensures that the unique characteristics of each sensor's data are captured independently. Each sensor's data was processed as a 1x2560 input tensor, reflecting 2560 data points over a 10-second window, similar to the data level fusion setup. Following the CNN layers, each sensor's output was subjected to a max pooling layer, followed by a dropout layer set at 0.2 to mitigate overfitting and batch normalization. After processing, each stream of data was transformed into a flat format using a flatten layer. The outputs from the individual flatten layers were then concatenated into a unified feature vector. The structure and process of feature level fusion are illustrated in *Figure 3(b)*.

G. Hyperparameter Tuning

We used the Adaptive Moment estimation algorithm (Adam) as the optimizer for our model. Model parameters are updated

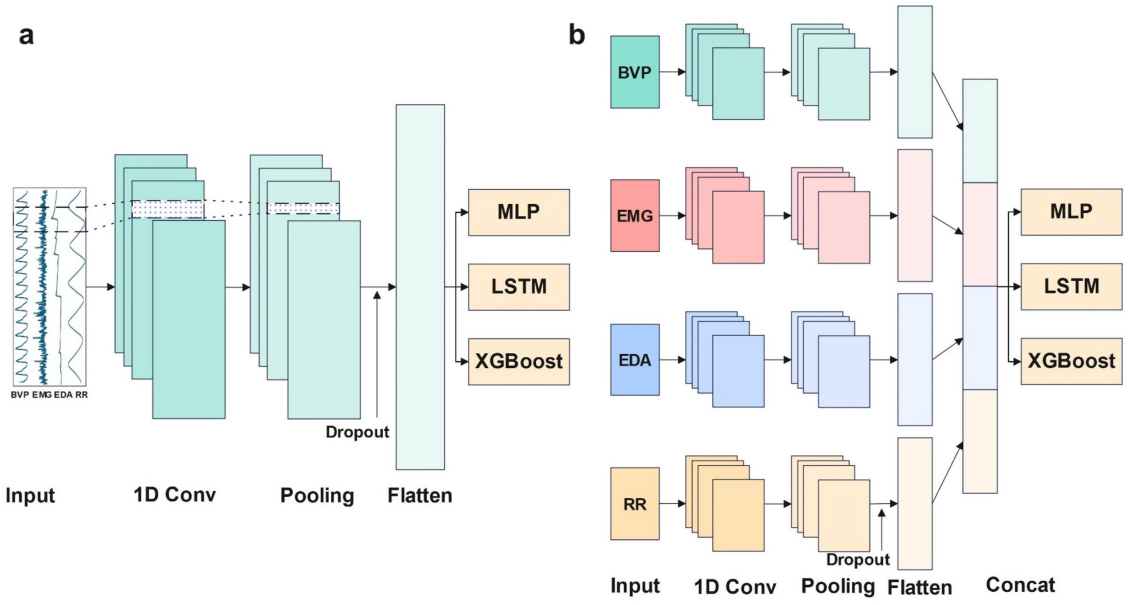

Fig. 3 Diagram of (a) data level fusion and (b) feature level fusion

Algorithm	Parameter	Values/Range
MLP	Hidden layers	[1,2,3]
	Units	[32,64,128]
	epochs	[10,20,30]
	Learning rate	[0.001,0.01,0.1]
LSTM	Layers	[32,64,128]
	Epochs	[50,100,200]
	Learning rate	[0.001,0.01,0.1]
XGBoost	Min child weight	[1,5,10]
	Gamma	[0.5,1,1.5,2,5]
	Subsample	[0.6,0.8,1]
	Max depth	[3,4,5]

Table. 1 Hyperparameter tuning space for each algorithm

Fusion Method	Model	MAE	RMSE
Data level fusion	CNN-MLP	0.964 ± 0.04	1.224 ± 0.06
	CNN-LSTM	0.856 ± 0.02	1.043 ± 0.04
	CNN-XGBoost	0.864 ± 0.05	1.085 ± 0.06
Feature level fusion	CNN-MLP	0.915 ± 0.03	1.135 ± 0.03
	CNN-LSTM	**0.843 ± 0.04**	**1.035 ± 0.07**
	CNN-XGBoost	**0.857 ± 0.02**	**1.026 ± 0.03**

Table. 2 Performance of different models under data level and feature level fusion methods. Values are given in Mean ± STD.

Sensor	Model	MAE	RMSE
BVP	CNN-MLP	1.145 ± 0.17	1.384 ± 0.21
	CNN-LSTM	1.059 ± 0.02	1.278 ± 0.01
	CNN-XGBoost	1.001 ± 0.04	1.226 ± 0.06
EMG	CNN-MLP	1.190 ± 0.17	1.363 ± 0.21
	CNN-LSTM	1.009 ± 0.02	1.263 ± 0.03
	CNN-XGBoost	0.993 ± 0.02	1.155 ± 0.04
EDA	CNN-MLP	0.930 ± 0.03	1.155 ± 0.06
	CNN-LSTM	**0.858 ± 0.04**	**1.013 ± 0.04**
	CNN-XGBoost	**0.883 ± 0.02**	**1.080 ± 0.03**
RR	CNN-MLP	1.283 ± 0.10	1.493 ± 0.09
	CNN-LSTM	0.949 ± 0.03	1.171 ± 0.04
	CNN-XGBoost	0.944 ± 0.03	1.169 ± 0.05

Table. 3 Performance of different models for individual sensor modalities. Values are given in Mean ± STD.

using the backpropagation algorithm and gradient descent method, where the error between the desired output and the actual output is quantified using a loss function. The hyperparameters were explored using grid search for each regression algorithm are presented in *Table 1*.

To evaluate the performance of regressor models, we used Mean Absolute Errors (MSE) and Root Mean Square Errors (RMSE) as performance metrics,

$$MAE = \frac{\sum_{k=1}^{n} |y - \hat{y}|}{n}$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^{n} (y - \hat{y})^2}{n}}$$

Where y is a true value, $\hat{y}$ is the predicted value, and n is the number of samples.

III. Results

Given that pain levels are ordinal, it is more appropriate to assess pain intensity levels as a regression task rather than a classification task. This is because the consequences of misclassifying pain levels vary significantly; misclassifying mild pain as severe pain is more problematic than treating mild pain as moderate pain. The regression approach allows for more accurate pain level estimation, aligning better with effective pain assessment and management.

A. Comparison between data level and feature level fusion

Table 2 shows the comparison between data level and feature level fusion. We found that feature level fusion performs better than data level fusion in general. The best performance was achieved in CNN-LSTM and CNN-XGBoost in feature level fusion. The CNN-LSTM model achieved the lowest MAE of 0.843 and an average RMSE of 1.035 from 5-fold. The CNN-XGBoost model achieved an average MAE of 0.857 and the lowest RMSE of 1.026. There is no significant difference between the two models.

B. Comparison between individual sensor modalities

Table 3 shows the comparison results between individual sensor modality. Each sensor, including BVP, EMG, EDA, and RR, went through the same deep learning architecture. We found that the best performance was achieved in EDA, while the worst performance was achieved in RR. In addition, there is no significant difference between the performance of EDA-based CNN-LSTM and EDA-based CNN-XGBoost models. Prior research has achieved similar results to the current study, where authors proposed EDA is the most information-rich sensor for continuous pain level prediction [29].

Comparing the performance between multi-modality models and individual modality models, we found that the EDA-based deep learning model is quite close to the multi-modality deep learning model. These remarkable results underscore the potential to estimate pain levels using fewer physiological sensors, achieving nearly the same accuracy as more complex multi-modality systems.

IV. Conclusion

In this work, we developed an automatic pain level estimation model using physiological sensors that can substitute patients' self-report pain information. We investigated different fusion methods, sensor modalities, and deep learning models. The best performance was achieved in feature level fusion using all physiological signals. EDA is the best signal compared with BVP, EMG, and RR for continuous pain estimation. Future research should assess the computational efficiency and cost-effectiveness of deep learning models, as well as their adaptability and effectiveness across diverse population groups. These directions have the potential to make pain assessment products more applicable in scenarios requiring real-time data processing and immediate feedback, thus enhancing personalized pain management and improving patient outcomes and quality of life.

ACKNOWLEDGMENT

This work has been financially supported by a collaborative 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] S. P. Cohen, L. Vase, and W. M. Hooten, "Chronic pain: an update on burden, best practices, and new advances," *The Lancet*, vol. 397, no. 10289. Elsevier B.V., pp. 2082–2097, May 29, 2021. doi: 10.1016/S0140-6736(21)00393-7.

[2] S. M. Rikard, A. E. Strahan, K. M. Schmit, and G. P. Guy, "Chronic Pain Among Adults — United States, 2019–2021," *MMWR Morb Mortal Wkly Rep*, vol. 72, no. 15, pp. 379–385, Apr. 2023, doi: 10.15585/mmwr.mm7215a1.

[3] S. N. Raja *et al.*, "The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises," *Pain*, vol. 161, no. 9. Lippincott Williams and Wilkins, pp. 1976–1982, Sep. 01, 2020. doi: 10.1097/j.pain.0000000000001939.

[4] Y. Jiang *et al.*, "The prevalence and management of chronic pain in the Chinese population: findings from the China Pain Health Index (2020)," *Popul Health Metr*, vol. 20, no. 1, Dec. 2022, doi: 10.1186/s12963-022-00297-0.

[5] R. R. Edwards *et al.*, "Patient phenotyping in clinical trials of chronic pain treatments: IMMPACT recommendations," *Pain Reports*, vol. 6, no. 1. Lippincott Williams and Wilkins, p. E896, Jan. 21, 2021. doi: 10.1097/PR9.0000000000000896.

[6] H. Flor and D. C. Turk, *Chronic pain: an integrated biobehavioral approach*. Lippincott Williams & Wilkins, 2015.

[7] 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.

[8] 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.

[9] O. Athnaiel, S. Cantillo, S. Paredes, and N. N. Knezevic, "The Role of Sex Hormones in Pain-Related Conditions," *International Journal of Molecular Sciences*, vol. 24, no. 3. MDPI, Feb. 01, 2023. doi: 10.3390/ijms24031866.

[10] T. Miettinen, J. Sverloff, O. P. Lappalainen, S. J. Linton, K. Sipilä, and E. Kalso, "Sleep problems in pain patients entering tertiary pain care: The role of pain-related anxiety, medication use, self-reported diseases, and sleep disorders," *Pain*, vol. 163, no. 7, pp. E812–E820, Jul. 2022, doi: 10.1097/j.pain.0000000000002497.

[11] S. E. Berger, É. Vachon-Presseau, T. B. Abdullah, A. T. Baria, T. J. Schnitzer, and A. V. Apkarian, "Hippocampal morphology mediates biased memories of chronic pain," *Neuroimage*, vol. 166, pp. 86–98, Feb. 2018, doi: 10.1016/j.neuroimage.2017.10.030.

[12] K. M. Naugle, T. Ohlman, K. E. Naugle, Z. A. Riley, and N. R. Keith, "Physical activity behavior predicts endogenous pain modulation in older adults," *Pain*, vol. 158, no. 3, pp. 383–390, Mar. 2017, doi: 10.1097/j.pain.0000000000000769.

[13] K. Herr, P. J. Coyne, E. Ely, C. Gélinas, and R. C. B. Manworren, "Pain Assessment in the Patient Unable to Self-Report: Clinical Practice Recommendations in Support of the ASPMN 2019 Position Statement," *Pain Management Nursing*, vol. 20, no. 5, pp. 404–417, Oct. 2019, doi: 10.1016/j.pmn.2019.07.005.

[14] R. C. B. Manworren, "Development of the Pediatric Healthcare Provider's Knowledge and Attitudes Survey Regarding Pain.," *Pediatr Nurs*, vol. 47, no. 6, 2021.

[15] C. Hartley *et al.*, "Nociceptive brain activity as a measure of analgesic efficacy in infants," *Sci Transl Med*, [Online]. Available: https://www.science.org

[16] M. M. Cobo *et al.*, "Early life inflammation is associated with spinal cord excitability and nociceptive sensitivity in human infants," *Nat Commun*, vol. 13, no. 1, Dec. 2022, doi: 10.1038/s41467-022-31505-y.

[17] D. Naranjo-Hernández, J. Reina-Tosina, and L. M. Roa, "Sensor technologies to manage the physiological traits of chronic pain: A review," *Sensors (Switzerland)*, vol. 20, no. 2. MDPI AG, Jan. 02, 2020. doi: 10.3390/s20020365.

[18] S. Moscato *et al.*, "Comparison of autonomic signals between healthy subjects and chronic low back pain patients at rest and during noxious stimulation," in *8th National Congress of Bioengineering, GNB 2023*, 2023.

[19] 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.

[20] W. Zhu, A. Kucyi, A. F. Kramer, and Y. Lin, "Multimodal Physiological Assessment of the Task-related Attentional States in a VR Driving Environment," in *2022 28th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, 2022, pp. 1–5.

[21] Y. Guo, L. Wang, Y. Xiao, and Y. Lin, "A Personalized Spatial-Temporal Cold Pain Intensity Estimation Model Based on Facial Expression," *IEEE J Transl Eng Health Med*, vol. 9, 2021, doi: 10.1109/JTEHM.2021.3116867.

[22] 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.

[23] 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.

[24] M. Yu *et al.*, "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.

[25] 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.

[26] Y. Lin *et al.*, "Experimental Exploration of Objective Human Pain Assessment Using Multimodal Sensing Signals," *Front Neurosci*, vol. 16, Feb. 2022, doi: 10.3389/fnins.2022.831627.

[27] A. Fernández, S. García, F. Herrera, and N. V Chawla, "SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary," 2018.

[28] 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.

[29] F. Pouromran, S. Radhakrishnan, and S. Kamarthi, "Exploration of physiological sensors, features, and machine learning models for pain intensity estimation," *PLoS One*, vol. 16, no. 7 July, Jul. 2021, doi: 10.1371/journal.pone.0254108.

[30] M. Yu *et al.*, "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.

[31] M. Yu *et al.*, "An eye detection method based on convolutional neural networks and support vector machines," *Intelligent Data Analysis*, vol. 22, no. 2, pp. 345–362, 2018, doi: 10.3233/IDA-173361.