Received: October 1, 2025. Revised: February 21, 2026. 784

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Hybrid Deep Model for Pain Intensity Classification Using Fused ECG, EMG, and GSR Signals

Wala'a N. Jasim[1,2]* **Adala Mahdi Chyad[2]**

[1]*Department of Computer Science, College of Computer Science and Information Technology, University of Basrah, Basrah, Iraq.*
[2]*Department of Pharmacognosy, College of Pharmacy, University of Basrah, Basrah, Iraq*
* Corresponding author's Email: walaa.jasim@uoabasrah.edu.iq

Abstract: Pain assessment is an active field of study with the accurate measurement of the extent of pain being a major problem particularly in individuals who are unable to communicate. Hence, the need to manage pain correctly triggered innovation. In this research, we propose a hybrid model that is based on bidirectional long-term memory (BiLSTM) with a multi-headed attention mechanism, integrated with convolutional neural networks (CNNs), to classify levels of pain based on physiological indicators. Experiments have been conducted using BioVid Heat Pain database (Part A), employing an efficient feature extraction method and simplified preprocessing to capture subtle patterns that might be missed by conventional models. The proposed design also leverages the integration of iterative and convolutional modules to extract self-referential time patterns. To analyze the impact of multimodal signal integration, three different arrangements of electrocardiogram (ECG), electromyography (EMG), and galvanic cutaneous response (GSR) signals were evaluated, and the proposed hybird approach was compared with two deep learning architectures: CNNs and BiLSTM. The results showed that the signal order significantly affected performance, with the best results achieved in the third order using the BiLSTM-Attention-CNN model, which once again proved that the system was effective in accurately identifying pain.

Keywords: Pain classification, Physiological signals, BiLSTM, CNN, Multi-head attention, Multi-head attention, BioVid heat pain, Hybrid deep model.

1. Introduction

The issue of pain assessment is a critical one in the contemporary medicine. The management and assessment of pain is essential in the work of clinicians as it aids in proper diagnosis, directs the process of treatment, and helps manage the pain properly. Based on the International Association to the Study of Pain (IASP), pain definition is: an unpleasant sensation and emotion that is related to, or similar to that, actual or possible tissue injury. Pain is an inherently multifaceted, which involves emotion and sensation. Neuropathic pain and psychogenic pain are conditions that show that one can experience pain in the absence of tissue damage [1-3].

The most common use of self-reporting in a clinical setting is regarded as the foundation of most of the pain assessment procedures. These procedures are very subjective and not very reliable in case the patient has communication problems including non-verbal disorders, dementia, or the neonate [4]. Due to the subjectivity of pain that is experienced by every individual and the inability of some individuals to express the pain, there is continuous development in automatic AI-based pain classification methods. This has been done in terms of physiological signals which include galvanic skin response (GSR), electrocardiogram (ECG), and electromyogram (EMG) [1].

Pain represents the fact that the body requires medical care. Majority of the conventional pain identification methods depend on the subjective judgment and observation of humans. In addition, the weakness of these approaches is the unavailability of universal and reliable guidelines. As such, automated

pain recognition is required. In such systems, behavioral and physiological data are used such as expressions, movements, sounds, or a combination of these is utilized in classification. The use of only behavior may be problematic since patients will repress or overdo displays of pain, and pain behavior is influenced by personality.

Therefore, physiological signal–based pain identification is crucial [5]. This technology is particularly helpful for individuals unable to communicate discomfort, including young children, neonates, and older adults with dementia.

Accurate pain assessment—subjective and multidimensional as it is—remains challenging. Although verbal reports, pain scales, and visual analog scales are common, they are often invalid for cognitively impaired individuals. Objective physiological parameters may provide a solution. Existing coding systems for facial expressions are expensive, inefficient, time-consuming, or insufficiently validated. Using lessons learned from prior work, the BioVid dataset was created with physiological as well as visual signals for developing automated pain recognition system; participants were exposed to controlled painful heat stimuli [6].

This work introduces a novel deep hybrid model that merges both Bidirectional LSTM, multi head attention and Convolutional Neural Network layers to classify pain intensity levels from electrocardiogram, galvanic skin response, and electromyography signals leveraging the publicly available BioVid part A dataset. The paper is organized as: we define the motivation and problem in Section 2; hypothesis to tackle such challenge is presented in Section 3; Section 4 provides an overview of related efforts in pain recognition and assessment using physiological indicators; Section 5 expounds on the materials, preprocessing steps, model architectures explored, and performance metrics employed; Section 6 discusses the outcomes of our experiments; in Section 7 offer conclusions and potential directions for future work; finally, Highlighting the main scientific contribution in Section 8.

2. Problem statement and research motivation

Health professionals still rely heavily on patients describing how much something hurts, yet such self-reports can be inconsistent and unreliable, especially for children, individuals with cognitive impairments, or non-native speakers. This gap is the cause of demand to have objective measurement of pain, and hope is that better data will lead to its proper treatment and minimize unnecessary pain. Accurate assessment of pain is not a clinical nicety, it can elevate mood, decrease anxiety, regain confidence, and assist patients in being connected to friends and family, rather than being confined within their own bodies [7]. In spite of the quick development of artificial intelligence (AI), automated pain evaluation continues to encounter basic issues that compromise the accuracy and performance of models in a real-world clinical environment. The first one is that available datasets are limited and biased, and in many cases, those do not represent diversity in terms of age, gender, ethnicity, and health status, which leads to poor generalization models. Moreover, pain manifestation is already a complicated matter and is diverse in individuals, both in terms of behavior and physiology. The combination of multimodal evidence, which includes facial expression, audio signals, and physiological signals, is an important technical issue that demands advanced algorithms and a large amount of computing resources.

Additionally, reliance on patient self-reports is unreliable because of over- or under-reporting due to psychological or social factors, which degrades data credibility. Pain assessment is further complicated in non-verbal populations such as dementia patients and individuals with communication disorders, necessitating non-verbal vital-sign analysis. Finally, most current studies lack unbiased, continuous pain-monitoring systems; real-time, automatic monitoring could greatly improve healthcare quality and support immediate, objective medical decisions [4, 5].

This work aims to create deep learning (DL) model which automates analyses traditionally requiring manual expertise, particularly precise feature extraction from physiological signals.

3. Research hypothesis

The problems that were explained in Section 2 are not going to go away in a hurry in this research, we argue, and that makes our hypothesis that the combined structures of such BiLSTM and CNN models will yield better pain-classification results compared to single-architecture all the more so. We base this hypothesis on the ability of mixed models to produce complex features from multimodal physiological data We also put forth a hypothesis that adding in temporal dynamics—following how the body's vital signs change over time—will reinforce the model's capability for tracking dynamic jumps between levels of pain. Static or instantaneous measurements only consider particular points at which there is a change; they ignore all of the important time frame around when pain starts and moves on.

4. Related work

Pain assessment and recognition have attracted considerable academic and clinical interest in recent years. Researchers have combined many databases—often obtained from hospitals or created speci fically for this purpose—with various AI approaches. Multimodal datasets containing physiological signals such as ECG, EMG, and GSR have proven especially valuable because of their ability to distinguish and evaluate pain of all types and levels. Consequently, a significant body of research has used biosignals to interpret pain. This section summarizes recent developments in automated pain assessment and recognition systems employing machine-learning and AI techniques; Table 1 provides an overview of key studies.

Philip Gouverneur et al. (2021) compared several methods for extracting features from physiological signals for pain recognition using a new database, PainMonit (PMDB), which contains both objective (e.g., temperature) and subjective (participant reports) assessments of thermal pain in 52 participants. Five methods—including manual feature engineering as well as deep feature learning—were tested on BioVid and PMDB.

The research established that even traditional feature-extraction techniques can compete with DL and that the complexity of the model is not always relevant to the improvement of the performance. The self-reports of participants were proven to be a good substitute of the objective data [8].

This is because the research is limited by the fact that it uses the EDA signal, which adds bias to it and limits its scope. Disregard of individual differences amongst the participants can decrease the accuracy and generalizability. In addition, the existing classification procedure is not very precise, which means that it can be enhanced with better multimodal data fusion and more analytical procedures, in particular [8].

Patrick Thiam et al. (2021) emphasized the benefits of deep neural networks, such as automatic feature learning and high adaptability, but two significant issues, namely, a large dataset is required, and overfitting may occur. They had suggested multimodal deep-learning methods with supervised and self-supervised learning, to enhance accuracy on the basis of physiological signals.

Self-supervised learning improved data efficiency by generating synthetic physiological data and updating models trained on small datasets [9].

The study is limited by scarce labeled data, the complexity of pain recognition requiring careful hyper-parameter tuning, and data collected in controlled settings, highlighting the need for validation in diverse environments [9].

A multimodal method of using electrodermal activity (EDA) signals as well as facial expressions for measuring post-operative pain in children was suggested by Busra T. Susam et al. (2021), as self-reports are inaccurate. The ability to combine EDA with facial expressions enhanced performance, with the accuracy of 90.91% (81.82% specificity, 100% sensitivity) to classify clinically significant pain. A weighted maximum-likelihood algorithm was successful in selecting features between the two sources [10].

Small pediatric dataset, noise in EDA and video signals, individual variation in pain responses, and high sensitivity of the models to close tuning, which can influence the generalizability and clinical applicability, are some of the limitations of the study. Even in spite of such shortcomings, their research gives a good base upon which non-invasive and objective methods of measuring pain among children can be developed, and more studies are required to broaden the range of application and enhance the accuracy of the prototypes [10].

Kelati et al. (2022) created an ML-based classifier to recognize the degree regarding pain depending on facial EMG signals (corrugator and zygomaticus), utilizing BioVid Heat Pain Data-base. It was found that the method had an accuracy of 99.4% in the detection of pain tolerance (P0 vs. P4) and did not involve subject bias [11].

The research mainly focuses on the possibility of facial EMG (fEMG) signals, restricting its application to the general settings. Also, a feature selection bias and additional optimization required in real world applications limits the applicability as well as strengths of the proposed approach [11].

Stefanos Gkikas et al. (2022) applied ECG signals to find heartbeat-interval features through the Pan-Tompkins algorithm. Three ML models have been evaluated for binary and multiclass pain classification. Existence of considerable gender differences developed: males were less sensitive to pain, particularly at high levels. Aging was associated with a reduction in pain perception, which puts the person at risk of injury. According to the authors, it was suggested to adapt pain-assessment instruments to the demographic factors [12].

This study has a number of shortcomings which are to be taken into consideration.

First, it has used only the electrocardiogram (ECG) signal without involving other physiological signals, which could be a constraint of the accuracy of the classification of pain levels.

Table 1. Summary of related work

Ref	Year	Methods	Dataset	Signals	Pain Type	Levels	Results
[8]	2021	CNN, MLP, HCF Combined	BioVid PMDB	EDA,ECG, EMG,IBI, HR, BVP	Heat	5 Levels & 6 Levels	CNN: 87.41% MLP: 86.50% HCF Combined 86.34%
[9]	2021	DDCAE, SSL	BioVid SenseEm-otion	EDA,ECG, EMG, RSP	Heat	5 Levels & 4 Levels	Accuracy 84.25%, 83.99%, 81.05% F1 Score 81.48%, 78.66%, 79.78%
[10]	2021	SVM	BioVid	EDA	Heat	5 Levels	Accuracy 90.91% Recall 81.82%
[11]	2022	KNN	BioVid	EMG	Heat	5 Levels	Accuracy 99.4%
[12]	2022	LDA, SVM LN, SVM RBF	BioVid	ECG	Heat	5 Levels	Accuracy 67.00% with SVM LN
[13]	2022	DT	BioVid	EDA,ECG, EMG	Heat	5 Levels	Accuracy exceeding 90%.
[5]	2023	CNN BiLSTM	BioVid Emopain	EDA,ECG, EMG, Sound, video	Heat Chronic Low Back Pain	3 Levels 0-10 levels	Accuracy 84.8%, 87.8%
[14]	2023	Video Masked AutoEncoder TimeSformer ViViT	UNBC BioVid	Video for FACS EDA,ECG, EMG	Chronic Shoulder Pain Heat	15 levels 5 levels	Accuracy 96% F1 Score 94%
[15]	2023	1D-CNN & BiLSTM	BioVid	EDA	Shoulder Pain Heat	5 Levels	Accuracy 91.5%
[16]	2023	MSCN SEResNet Transformer Encoder + TCN	BioVid	EDA	Heat	5 levels	Accuracy 85.56% F1 Score 85.49%
[17]	2024	Transformer	BioVid	ECG, EMG, GSR	Heat	5 levels	Accuracy Binary Class 82.74% multilevel 39.77%
[7]	2024	machine learning algorithms	BioVid	ECG, EMG, GSR	Heat	5 levels	Accuracy Binary Class 87% multilevel 52%
[18]	2024	MSWDCN+ ACFN+ TCCAN	BioVid	EDA	Heat	5 levels	81.25%
[19]	2024	CNN,LSTM BiGRU, DNN	BioVid Emopain	ECG, EMG, GSR Sound, video	Heat Chronic Low Back Pain	5 levels 0-10 levels	Accuracy 87.8%, 84.8%
[20]	2025	RF, KNN, SVM		ECG, EDA, RESP, PPG, PPT, FT	Emotionally Induced Pain	7 levels	F1 Score 0.83%
[21]	2025	multimodal fusion (RF/DNN/Auto-Encoder)	BioVid	ECG, EDA , Video	Heat	5 levels	Accuracy 83.05%, 82.69% For EDA 84.15%, 83.35% For (Video + EDA) F1 Score 81.66, 80.18 For EDA

							82.86, 82.36 For (Video + EDA)
[22]	2025	combines CNN, BI-LSTM, and BI-GRU	BioVid Emopain	ECG, EMG, GSR Sound, video	Heat Chronic Low Back Pain	5 levels 0-10 levels	Accuracy > 90

Moreover, the results regarding the models of classification have been poor, and the model could not distinguish among various pain levels with a maximum accuracy of 23.8% which means that it was difficult to distinguish similar levels of pain. Lastly, validation was done by the LOSO method that cannot grantee the generalizability regarding the findings to other populations [12].

According to Peter Bellmann et al. (2022), BioVid has been argued as one of the most complicated physiological-signal classification tasks to be automatically identified by its intensity of pain. They demonstrated with the help of chaos theory that the task is very complex and provided a numerical instrument to estimate the complexity and tested it on simpler tasks like handwritten-digit recognition. They suggested accounting nonlinear dynamics of signals and new measures of evaluation besides accuracy [13].

This study was limited to analyzing the complexity of pain classification using chaos theory, without focusing on improving the accuracy of the models. It relied on only one model (Decision Tree), which limited the results to its analytical framework. Furthermore, it used manual features taken from previous studies rather than features extracted using modern methods, which could lead to the limitation of the generalizability of results. The proposed "chaos check" test is a preliminary step, and further studies are needed to confirm its effectiveness on different databases and signal types[13].

Kim Ngan Phan *etal*. (2023) have introduced a DL approach in order to classify pain vs. non-pain states using physiological signals without manual feature engineering. Multi-level context information was leveraged, outperforming single-context methods. On BioVid Part A and Emopain 2021, the model achieved 84.8 % ± 13.3 % accuracy (87 participants) and 87.8 % ± 11.4 % (67 participants, LOSO) [5].

Benavent-Lledo et al. (2023) achieved >96 % pain-estimation accuracy using single-frame facial-expression analysis and transformer-based architectures on BioVid and UNBC-McMaster datasets, establishing a new benchmark [14].

The study showed that incorporating temporal information did not improve pain detection accuracy, with difficulty distinguishing between categories due to overlapping data and poor performance regarding CNN model, along with the problem of overlearning and limited data, as well as the impact of cultural and psychological factors related to the subjective nature of pain [14].

Pinzon-Arenas et al. (2023) combined 1D-CNN and BiLSTM to detect continuous pain using phase-EDA features. With 36 participants, the model achieved 91.5 % accuracy on BioVid, confirming the effectiveness of neurocutaneous EDA signals [15].

Current pain assessment methods rely on subjective measures, and previous studies did not use sequence-to-sequence method for detecting persistent acute pain, resulting in limited accuracy with the highest F1-score being only 77.8% [15].

Zhenyuan Lu et al. (2023) suggested a transformer-encoder model of physiological-signal based pain classification named PainAttnNet. The model is a hybrid of MSCN, SEResNet and a transformer module with TCN and multi-head attention, which is more effective than earlier approaches on BioVid [16].

The model does not combine two or more physiological signals like ECG or EMG as opposed to single physiological signal (EDA) which can be restrictive to the completeness and accuracy of the assessment [16].

Stefanos Gkikas et al. (2024) suggested a multimodal framework combining facial video and ECG signals with AugmNet-based reinforcement learning, achieving 82.74 % binary and 39.77 % multilevel accuracy on BioVid [17].

The research has been carried out on small sample under controlled laboratory conditions, which could limit generalizability to real-world settings. Inter-subject variability, sensor noise, and challenges in multimodal signal fusion could affect model robustness. Additionally, the deep learning models provide limited interpretability of the underlying physiological mechanisms.generalizability to real-world settings. Inter-subject variability, sensor noise, and challenges in multimodal signal fusion could affect model robustness. Additionally, the deep learning models provide limited [17].

Da'ad Albahdal *etal*. (2024) have utilized ECG, GSR, and EMG signals from Bio-Vid in order to

develop an objective pain-classification tool. Random Forest achieved 87 % binary and 52 % multilevel accuracy, supporting field deployment via wearables [7].

The most limitation of this study pain assessment is the scarcity of datasets, preventing validation of models across diverse scenarios. The predefined heat pain window in the dataset (BioVid) is 5.5sec, which cannot be sufficiently long with regard to accurate automated pain intensity estimation. Moreover, emotions, anxiety, sleep, and other experiences can influence physiological signals responding to pain [7].

JiaHao Li et al. (2024) introduced EDAPainNet for pain assessment using EDA signals. Combining MSWDCN, ACFN, and TCCAN, the model achieved 81.25 % accuracy on BioVid, outperforming prior methods by >2 % [18].

The primary weaknesses of the study include: the model relies on few physiological signals that limits generalizability; the experiment was performed in controlled laboratory settings and it lowers the diversity of the participants; and the model does not account for psychological and individual differences factors that contribute to pain which limits the personalization related to the model and its adaptation to the individual variation [18].

Chandan Kumar Behera *etal*. (2024) have proposed a deep-learning method, which combines the classification of the multi-level information of pain by utilizing both BioVid and Emopain-2021 demonstrating the 87.8% accuracy (Pain-0 vs. Pain-4) in the absence of extensive medical knowledge [19].

Eun-Hye Jang et al. (2025) classified the perceived pain level in respect to the same stimulus by various physiological indicators (ECG, SC, respiration, PTT) and emotional ratings. Electrodermal responses (SC) were the most valuable biomarkers with 112 participants (F1 = 0.83 pain vs. baseline; F1 = 0.73 across pain levels), which proves the usefulness of the personalized pain-assessment models [20].

Limitations of the study are it consists of small sample sizes of some levels of pain, which could result in models' overfitting; a focus on static snapshots regarding physiological response with no dynamic time analysis; relying only or participant self-assessment regarding pain, limiting objectivity; and narrow range of emotional pain stimulus (anger); failing in fully representing diverse emotional experiences. Such constraints have implications on the generalizability and reliability of the model, indicating the necessity to conduct future research, which involves larger sample sizes, incorporating time dynamics, integrating multi-measurement and expand the emotional stimulus range [20].

A multimodal system, which Jaleh Farmani et al. (2025) suggested, combines facial expressions as well as physiological (ECG, EDA) signals from (BioVid). CNN-LSTM with video attention and LSTM on signals, EDA alone produced over 83% accuracy and video-EDA combination produced 84.15% accuracy (F1 = 82.86%) in binary no pain vs severe pain classification [21].

The model is limited by issues concerning the fact that it has been tested on a small experimental dataset that can lead to the generalizing the results to wide range of clinical or natural settings. In addition, the use of certain physiological signals and facial expression data in a certain situation is not exhaustive in the variety of pain patterns, types, and various backgrounds of the study participants [21].

Mattikurava Sowmya et al. (2025) introduced a hierarchical CNN-BiLSTM-BiGRU model using multi-level contextual information without complex preprocessing. High performance was achieved on BioVid and Emopain-2021, demonstrating scalability [22]. The limitations are associated with traditional methods that rely on manually extracted features informed by medical expertise

5. Materials and methods

5.1 Physiological interpretation of learned features

Physiological signals provide an objective window into the autonomic and neuromuscular responses of the body to pain.

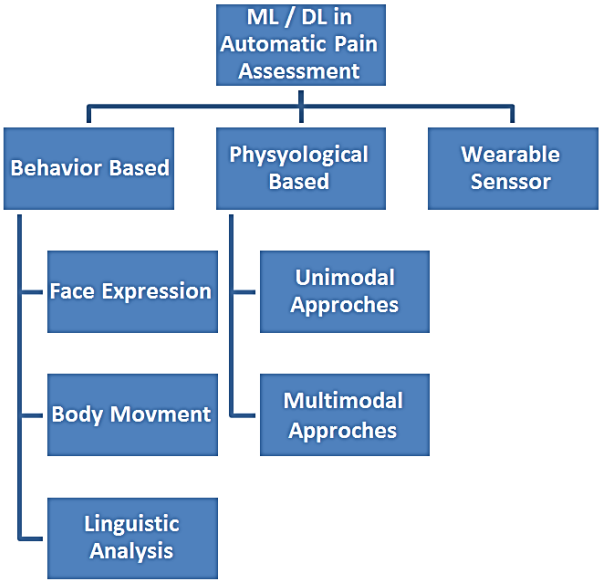


Figure. 1 A schematic representation of the structure of the related work [7]

Unlike self-reported scores, these biosignals capture real-time changes driven by sympathetic activation, cardiovascular modulation, and muscle-activity adaptations that accompany nociceptive processing.

Thus, ECG, EMG, and GSR analysis allows the model to identify the patterns of pain based on established biomedical processes. All signals represent a unique physiological pathway in which pain is involved and these signals present a multimodal representation of the bodily reaction toward noxious stimuli complementary to each other [23, 24].

5.1.1. Physiological interpretation of ECG in relation to pain

ECG activity especially metrics based on heart rate variability (HRV) have generally been utilized for indexing the response regarding autonomic nervous system to pain. In addition, the HRV represents the ratio of the sympathetic and parasympathetic systems, and it has been demonstrated by a series of studies that painful stimulation can change this balance considerably. As an example, a review of systematic literature discovered that experimentally induced pain correlates with higher sympathetic and lower parasympathetic (vagal) tone as indicated by measures of the HRV frequency domain [25, 26].

There is also meta-analytic evidence suggesting that chronic pain states are associated with a progressive decrease in high-frequency HRV, and it suggests a loss of parasympathetic control [25, 27]. The fact that these physiological alterations are consistent with the notion of pain causing a stress-like autonomic response makes HRV an excellent indicator of pain assessment [25, 28].

5.1.2. Physiological interpretation of EMG in relation to pain

EMG can be defined as a mode of muscle electrical activity, which represents recruitment of motor units, their rate of firing, and coordination. The effect of pain on voluntary EMG amplitude (e.g., by hypertonic saline injection) has always been reported to decrease, but without any peripheral muscle fiber properties - implying central (neural) inhibition of muscle activity [25, 29]. In addition, pain may change the co-ordination of agonist and antagonist muscles: in the case of dynamic movement in painful conditions, the agonist muscles may become less active, and antagonistic ones may become more active, which may be a protective mechanism [25, 30]. These patterns of EMG activity can be used to sustain a pain adaptation model where motor control is recognized under the influence of nociceptive input, and both protective and compensatory mechanisms are evident.

5.1.3. Physiological interpretation of GSR / electrodermal activity (EDA) in relation to pain

Electrodermal activity (EDA) or, commonly recorded through skin conductance (GSR) is a direct indication regarding sympathetic nervous system activity since the muscle of sweat glands is so carefully regulated by sympathetic fibers. It has been demonstrated that painful stimuli result in decisive changes in the phasic element of EDA (i.e., skin conductance responses, SCR), which is associated with the severity of the stimulus [25, 31].

More recently, spectral or derivative analyses of EDA have been proposed to better capture rapid fluctuations linked to pain; these methods have demonstrated high sensitivity in distinguishing different pain levels [25, 32]. Therefore, in pain contexts, EDA can reflect moment-to-moment sympathetic arousal, making it a powerful physiological marker of nociceptive processing.

5.2 Dataset

The way people feel pain is something that is beyond measure in an objective way and can only be judged subjectively and in many ways. At present, this problem has not been properly solved. The results you obtain from verbal methods like traditional pain scales and questionnaires, as well visual analogue scales (VAS), are widely used in clinical settings. When applied to people with cognitive impairments however they often lack reliability and validity. Another avenue may be to analyse the expressions of pain and examine biopotential signals related to them. There are already a few facial expression-based pain-coding systems in use, but they are frequently time-consuming, too costly, or not thoroughly assessed in light of psychometric theory.

BioVid Heat Pain Database is specialized database that combines physiological signals with facial video in order to improve automated pain recognition. The database enhances the performance regarding objective pain-assessment systems, assesses their psychometric quality, and helps their development. Participants were exposed to controlled heat stimuli under standardized laboratory conditions, enabling systematic collection of pain-related responses [33-35].

BioVid Heat Pain Database is a key experimental resource in automated pain recognition. Created

through a joint project led by Ulm University, Germany, it provides multimodal, standardized data for training and evaluating machine-learning and AI models to recognize human pain objectively and reliably. It combines biopotential recordings with high-resolution facial video, offering an ideal environment for multidimensional pain analysis [33-35].

Part A of BioVid Heat Pain Data-base includes experiments of heat stimulation without recording facial muscle signals. It contains frontal facial video along with biomarkers (ECG, trapezius EMG and GSR) in both raw and preprocessed forms. This part comprises 8,700 samples collected from 87 participants, each of whom underwent five levels of stimulation representing five different pain categories, with 20 trials per category per participant, each with a 5.5-second time window. This part was utilized in many studies [39,40] to estimate pain intensity depending on physiological signals.

In the presented work, Part A of the database has been utilized only for biomarkers without video, relying on three main signals: (ECG to measure cardiac activity, trapezius EMG to measure muscle activity, and GSR/EDA) as an indicator of pain-related physiological arousal. This data has been utilized for training and evaluating the proposed deep learning model for automatic pain level recognition.

Here is a brief overview of each signal [7]:

1) The Electrocardiogram (ECG) is a test used to examine heart problems and heart activity. changes. It captures the heart's electrical impulses. The ECG was the Empirical Mode Decomposition technique and has been filtered with the use of Butterworth band-pass filter (0.1-250Hz).
2) Galvanic skin response (GSR) is a segment regarding electrodermal activity (EDA) and changes in sweat gland activity. GSR changes in sweat activity are related to emotional arousal and indicate the strength of a particular emotion.
3) Electrical activity of the muscles is an arousal and is a reflection of psychophysiological general arousal of the body. Increased muscle tone and the activity related to sympathetic nervous system. Increased somatomotor activity is suggestive of predominant parasympathetic arousal. Dataset's age distribution is shown in Fig. 2.

The experimental procedure was Eighty-seven healthy adults (18–65 years) underwent a controlled protocol in which painful heat stimuli were delivered to the medial forearm via a Thermode TSA-II device. Four individualized pain intensities (T1–T4) were determined relative to each participant's pain threshold, plus a no-pain baseline (T0). This personalized temperature assignment enhances physiological-response validity and ecological realism [35].

5.2.1. Signal-modalities

The multimodal dataset integrates 32-channel electroencephalography (EEG), facial and shoulder electromyography (EMG), electrocardiography (ECG), electrical skin conductance (ESC, referred to as the galvanic skin response/GSR), and temporally aligned high-definition facial video optimized for computer-vision analysis.

5.2.2. Psychometric utility

BioVid supports the development of automatic pain-recognition systems and their evaluation against psychometric criteria (validity, reliability, sensitivity). It is widely used for multisignal fusion research, particularly with regard to non-verbal populations (e.g., ICU or cognitively impaired patients) [34, 35].

5.2.3. Availability

After its introduction at the 2013 IEEE International Conference on Cybernetics, the database was released to the research community and remains a primary reference for pain-recognition studies based on physiological and facial cues) [34, 35].

5.2.4. Data subset used in this study

Fig. 2 in the above page illustrates the ECG, GSR, and EMG partitions employed here. The 87 retained participants were roughly balanced across three age brackets—18–35, 36–50, and 51–65 years—with 15 males and 15 females per bracket before technical exclusions [17]. Fig. 3 depicts the experimental setup with the thermode-induced pain stimulus [36], and Fig. 4 presents exemplary ECG, GSR, and EMG traces used for pain assessment [14].

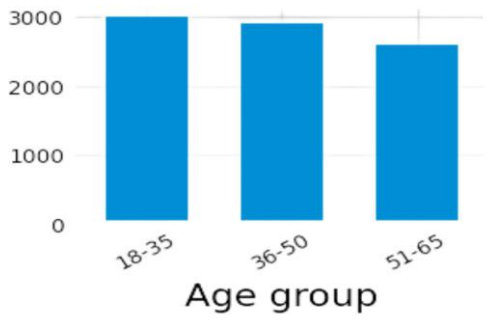


Figure. 2 Age Distribution in Dataset Used

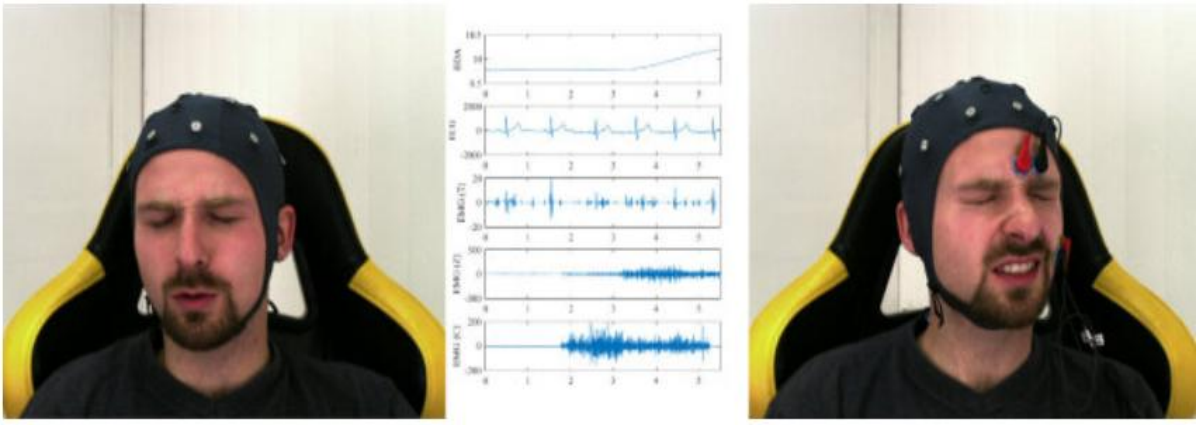
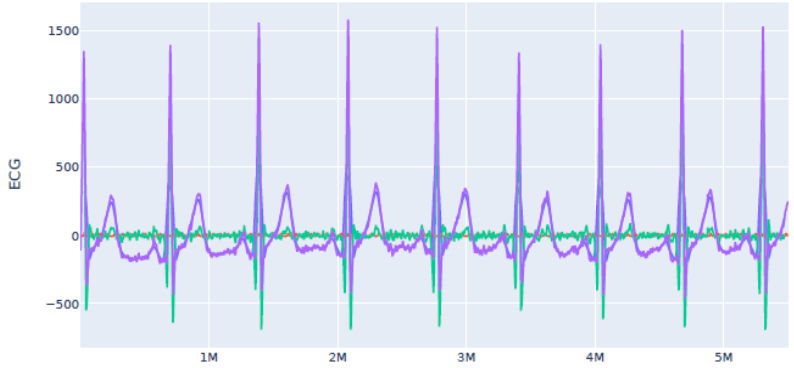
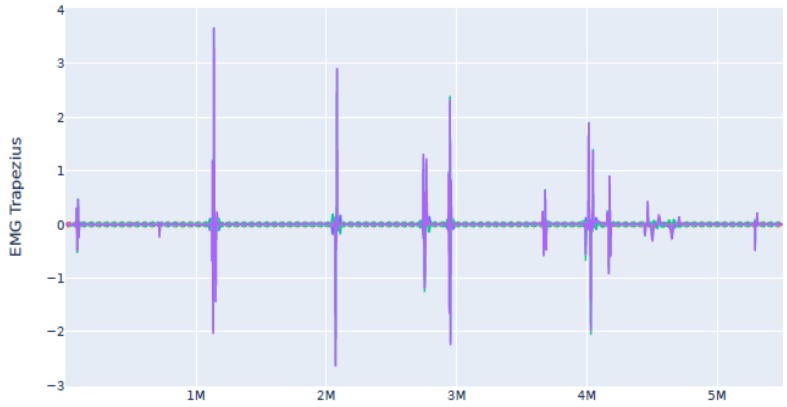


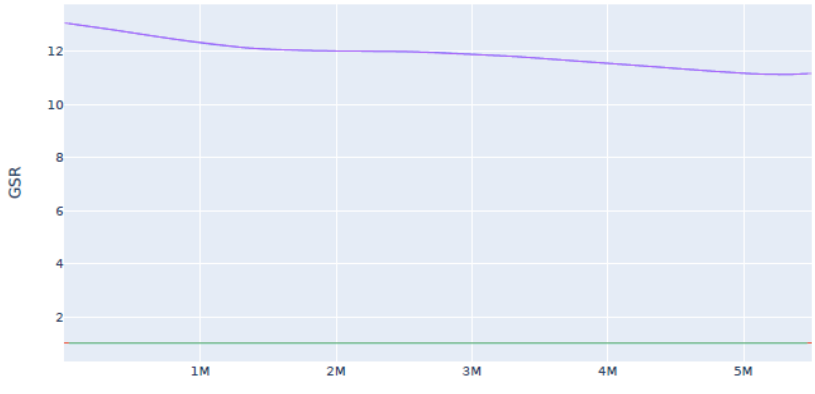
Figure. 3 Example of the measuring configuration with induced pain stimulus[36]



(a)



(b)



(c)

Figure. 4 Example of Physiological Signals [14]: (a) ECG, (b) EMG, and (c) GSR

5.3 Preprocessing

Preprocessing transforms raw physiological signals into a format suitable for deep learning. Our pipeline for the hybrid model comprises loading, cleaning, normalising, downsampling, format conversion, and augmentation. These steps yield uniform, low-noise data that improve both accuracy and generalisation [37, 38].

5.3.1. Class extraction

The pain label (e.g., BL1, PA1–PA4) is parsed from each file name and mapped to an integer 0–4.

5.3.2. File aggregation

All CSV files are read from a dedicated folder and grouped into five lists according to their numeric label.

5.3.3. Augmentation

To enlarge the training set we apply:

- Additive Gaussian noise;
- Random scaling (multiplicative factors < ±10 %);
- Time shifting (circular left/right translation).

5.3.4. Missing-value detection

Any NaNs are identified and linearly interpolated to prevent gradient errors and normalisation bias.

5.3.5. Normalisation

Each channel (ECG, EMG, GSR) is z-scored ($\mu = 0$, $\sigma = 1$) to ensure uniform dynamic range and faster convergence.

5.3.6. Formatting-split

Signals are reshaped into 2-D tensors (time steps × channels). Labels are one-hot encoded. With regard to training and validation, the proposed model was trained using K-Fold Cross-Validation to ensure reliable evaluation and minimize data splitting bias. The data was divided into several folds, some utilized for the training and others for validation. Within every fold, the model was re-initiated and re-trained from scratch to guarantee evaluation independence. Subsequently, a final model was trained on the entire training data-set before the final evaluation has been carried out on the independent testing dataset.

Three experiments were conducted for each model. Each experiment contained three signals combined in different arrangements: (Experiment1 [GSR,ECG,EMG],Experiment2:[ECG,GSR,EMG], and Experiment3:[EMG,GSR,ECG].

5.3.7. Techniques for controlling overfitting and underfitting

To mitigate overfitting and underfitting during model training, we employed several techniques: random dropout layers (Dropout and SpatialDropout1D) to reduce the model's dependence on specific units; batch normalization techniques to enhance training stability; EarlyStopping to halt training when the performance of the model on validation set didn't improve; and ReduceLROnPlateau for dynamically adjusting the rate of learning when the model's loss was constant. Furthermore, improved data augmentation had led to the enhancement of the model's generalization ability and improved performance on the test data [3,5].

5.4 Architectures

5.4.1. Bidirectional long short-term memory (BiLSTM)

BiLSTM augments a standard LSTM with a second layer that processes the sequence in reverse order. Concatenating forward as well as backward hidden states improves late-sequence information retention and provides more context. In pain-detection tasks, BiLSTMs automatically extract temporal features from physiological streams, reducing reliance on hand-crafted features and boosting classification accuracy [39, 40].

In LSTM networks, out-of-date information is gradually attenuated via forget gates, while memory blocks are updated with both prior and current inputs. Bidirectional LSTM (BiLSTM) extends LSTM by learning the temporal dependencies in backward as well as forward directions.

As shown in Fig.4, a BiLSTM layer receives the input x_i simultaneously into two LSTM paths. Forward weights (W_f) and backward weights (W_b) are updated according to the forget and input gate activations, which depend on previous hidden state h_{i-1}, current input x_i, and cell state c_{i-1} [40]. Although BiLSTM networks have demonstrated strong performance in pain detection, challenges remain in generalising these models across diverse populations and varying physiological conditions; additional research is required to enhance their robustness and clinical applicability.

The BiLSTM can be defined an extension regarding traditional LSTM model, enabling the processing of data sequences in both directions: forward and backward. This allows it to simultaneously capture preceding and subsequent

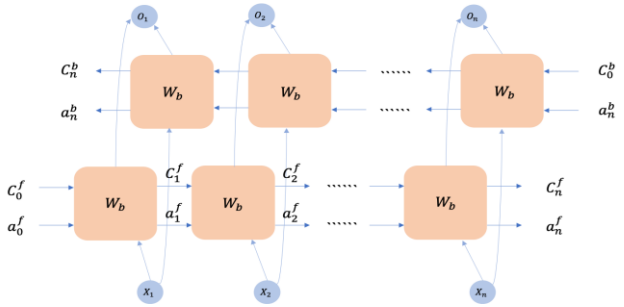


Figure. 5 Architecture of BiLSTM [41]

temporal relationships. This property is particularly important in biosignal analysis, as it contributes to a more accurate representation of temporal patterns associated with pain. The mathematical formulation of the BiLSTM network can be derived from the basic LSTM equations, applied in both directions.

The operation of LSTM layers can be described mathematically in the following way:

$$\Gamma i = \sigma(Wi[a(t-1),x(t)]+bi) \quad (1)$$

$$\Gamma f = \sigma(Wf[a(t-1),x(t)]+bf) \quad (2)$$

$$\Gamma o = \sigma(Wo[a(t-1),x(t)]+bo) \quad (3)$$

$$C\sim(t) = \tanh(Wc[a(t-1),x(t)]+bc) \quad (4)$$

$$C(t) = \Gamma i.\ C\sim(t) + \Gamma f.\ C(t-1) \quad (5)$$

$$a(t) = \Gamma o\ .\ \tanh(C(t)) \quad (6)$$

where input gate G_i controls input information that is added to cell state.

The past information that will be retained in long-term memory is determined by forget gate G_f. How regulated information is made available as output is determined by output gate [40]. An addition to LSTM cells, BiLSTM has been demonstrated to enhance performance in numerous applications. BiLSTM processes sequential data in both reverse and forward order by stacking two layers regarding LSTM cells. The representation related to sequential data in the architecture is created by concatenating the hidden states regarding backward and forward layers. It has been demonstrated that processing the data in reverse order improves the ability of the model to retain information close to the conclusion of sequence [40, 42].

5.4.2. Convolutional neural network (CNN)

CNN architectures for pain recognition have evolved rapidly, incorporating standard backbones, off-the-shelf pre-trained models, and lightweight designs tailored for specific cohorts, including neonates [41]. Every CNN contains multiple hidden layers, input layer, and output layer. With regard to feed-forward networks, the intermediate layers are termed "hidden" because their internal representations remain opaque until the final convolution and activation. CNN's hidden layers consist of three core components:

- Convolutional layers
- Pooling layers
- Fully connected (FC) layers

Collectively, these sub-structures form a Convolutional Neural Network [42].

Convolution operations are carried out using filters in convolution layer (CONV) [43]:

$$\boldsymbol{zi}=\boldsymbol{bi}+\Sigma\boldsymbol{xi}\in\boldsymbol{Wij}*\boldsymbol{xi} \quad (7)$$

To add non-linearity following convolution, an activation function like ReLU can be utilized[43]:

$$\boldsymbol{gi}=\boldsymbol{ReLU}(\boldsymbol{zi})=\mathbf{max}(\mathbf{0},\boldsymbol{zi}) \quad (8)$$

Pooling Layer: In DL, pooling is a crucial operation. The number of parameters, feature dimension, temporal complexity, and computation difficulty could all be decreased via pooling operations [42, 43]:

$$Pi=\max(zi:i+s)$$

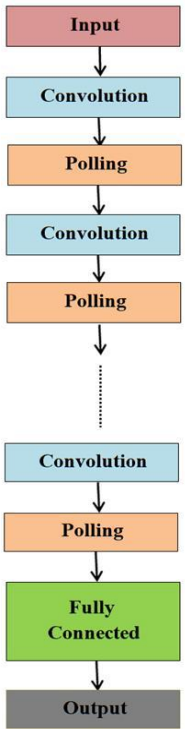


Figure. 6 Architecture of CNN [44]

FC Layer: Every input is connected to each one of the neurons in fully connected layer (FC), which functions on flattened input [43]:

$$oj=\sigma(\Sigma wij.aiN-1i=0+bj) \quad (9)$$

5.4.3. Attention mechanism (AM)

In many remote sensing (RS) image processing, computer vision applications, and signal processing as physiological signals, ML, especially DL, has emerged as a key as well as cutting-edge technique. Through creating novel network architectures and/or approaches, researchers are always attempting to enhance the performance regarding DL systems. AMs are among the most crucial regarding such processes. In spite of its type, the AM was extensively employed in numerous applications for enhancing the performance related to the current DL techniques since it was suggested [45]. The capacity of DL models to recognize the most significant components in data and give them higher weights throughout processing—as opposed to treating all input equally—is known as AM.

This concept has helped modern models understand complex relationships within sequences and improve the quality of representation and context, making attention a fundamental component in many advanced architectures such as Transformer. A summary of the types of AMs follows [45].

1. Attention softness: 'Soft attention' refers to AM, also called deterministic attention, that was first suggested through Bahdanau et al. [46]. The final context vector in the network is determined by taking into account all input components, in which average is computed at each one of the weights. Stochastic attention also known as hard attention, however, takes into account only the randomly sampled constituent in determining the final context vector [45, 46]. Hence, it saves on computational time. Furthermore, there is the idea regarding local and global attentions in relation to the classification that is typical in image processing as well as computer vision tasks [45, 47]. Since global attention takes into account all input components, it is most similar to soft attention overall. Whereas local attention represents a combination of hard and soft attentions, global attention can be defined as a simplification regarding soft attention because it utilizes the output from current time step instead of the prior time step. This strategy alleviates the problem of hard attention being non-differentiable, while at the same time being less heavy on the resources. Hence, also it is more efficient than the remainder on computational resources as it looks at a limited subset of input elements at a time.

2. Forms of input features: AMs could be categorized both location-wise and item-wise according to their input needs. Types of input features: location-wise and item-wise attention methods could be categorized according to their input needs. Inputs that are either explicitly known to the model or generated through a pre-process are necessary for item-wise attention. Nevertheless, location-wise attention does not always necessitate that the model handle input items which are hard to discern [45, 48–50].

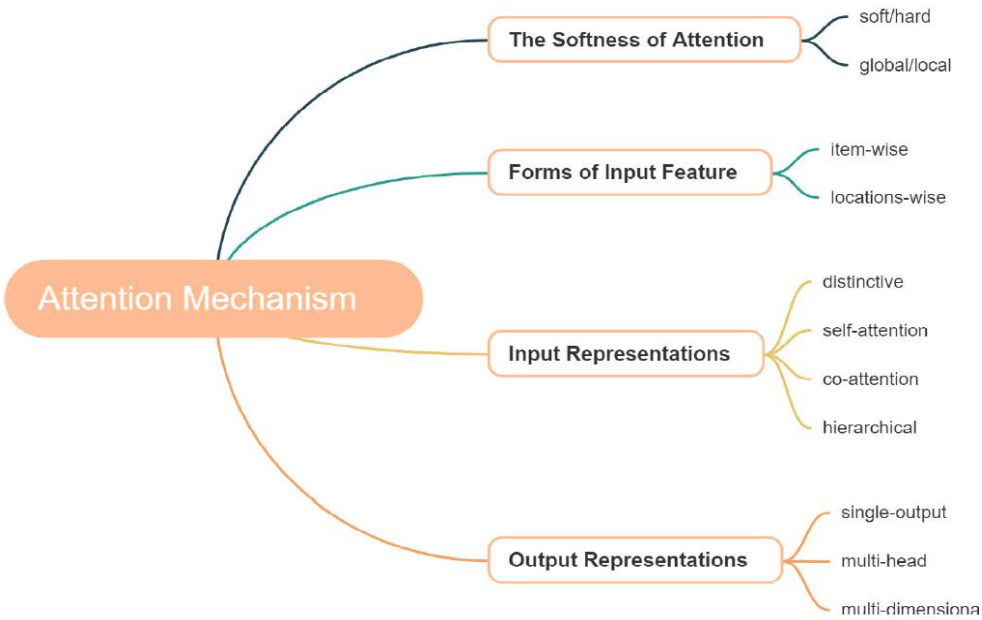


Figure. 7 An overview of typical AM approaches [45]

3. Input Representations: On the one hand, there exist single-input models, and on the other hand, there exist multi-input models, as with the attention networks described in the references, [45, 51]. During the development stages of the models, there were some variations in how the models processed the inputs. Single-input attentional networks, which currently dominate the attention networks, process their inputs in a sequence of independent steps, i.e., in a distinctive fashion. Co-attention model, at the same time, is multi-input attention network that applies the AM in parallel to two distinct inputs before combining the outcomes. Self-attention networks generate their attention maps solely from the model inputs. This mechanism is advantageous for models operating on background complex images / complex signals, as the model can direct its attention towards the areas corresponding to the complex background. [52]. The Hierarchical AM assigns weights to the inputs on different levels, including the raw input. This mechanism, also referred to as fine-grained attention, can improve the performance of image classification systems [45].

4- Output representations: The most common type regarding attention output representation in AMs is single-output attention, which generates weight scores by processing one feature at a time. Additionally, there are various types of multi-head as well as multidimensional attention [45]. Multi-head attention processes the inputs in linear fashion in several smaller groups and, at the end, merges the attention to form the final attention weights and is particularly beneficial in the case when using AM together with CNN approaches [45, 52, 53], which type was employed in this study. Multidimensional attention, mostly used in NLP, calculates weights using a matrix representation of the features rather than a vector-based representation [45, 54, 55].

In our research, we used a multi-headed AM (MHAT), and the following paragraphs outline how this type of AM works.

As observed in the presented study, such type is distinguished by its excellent capability to concentrate on the most significant features emerging from BiLSTM network. Although MHAT mechanism relies heavily on attention, this model differs essentially in the fact that it is capable of handling complicated information through several distributed computations [56].

There are four steps to compute attention [56]:

- Similarity is taken into account while calculating each key as well as query weight. For calculating similarity, the suggested model is utilized as dot product.
- The operation of scaling is the following step for calculating attention, in which factor $\sqrt{d_k}$ is utilized as moderator so that the dot-product isn't too big.
- Resulting weights are normalized using Softmax method.
- The corresponding principal value V and similarity add up to weighted sum.

Using previously described procedures, the next formula has been obtained:

$$\text{Attention(Q, K,V)} = \text{softmax}(\frac{QK^T}{\sqrt{d_k}}) \quad (10)$$

Conventional AM was improved with MHAT, which performs exceptionally well. MHAT mechanism's architecture is depicted in Fig. 8. K, Q, and V are first transformed linearly to become scaled dot-product attention's input. As a result, the process computes one head at a time. As a result, h, or multi-head, must be performed. Every linear transformation regarding K, Q, and V has a separate set of parameters, W. The value generated by linear transformation is used as MHAT output after the concatenation of every scaled dot-product attention output of m time [56, 57]. The following is an expression for the formula:

$$\text{headi} = \text{attention}\ (QW_i^Q,\ KW_i^K,\ VW_i^Q) \quad (11)$$

$$\text{Multi-head(Q, K,V)} = \text{concat(headi,..., headm)}W^o \quad (12)$$

5.4.4. Proposed model architecture

CNNs are traditionally used for image classification; however, by integrating a hybrid block of BiLSTM + Multi-Head Attention + 1-D CNN, we

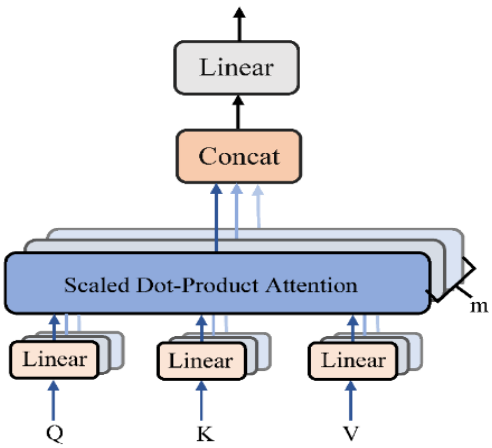


Figure. 8 MHAT Mechanism Architecture [56]

constructed a novel architecture specifically tailored for multi-channel physiological pain recognition [44].

Overal-lPipeline

Raw ECG, ,EMG and GSR streams are early-fused into a single 3-channel tensor (time-steps × channels). This tensor is then processed sequentially by:

5.4.4.1. BiLSTM-encoder

A bidirectional LSTM layer of 128 units—returning full sequences—feeds a second bidirectional LSTM of 64 units to capture temporal dependencies spanning several seconds of data, while dropout of 0.3 applied after each layer mitigates overfitting.

5.4.4.2. Multi-head-attention

Eight attention heads re-weight the BiLSTM outputs, letting the model focus on the most pain-relevant temporal segments while preserving diversity across heads.

5.4.4.3. 1-D CNN feature extractor

- 64 filters, kernel size 3, ReLU activation, same padding
- Captures local cross-channel interactions revealed by the attention weights.
- Batch Normalization stabilises training; MaxPooling1D (pool_size=2) halves the sequence length and retains salient features.

5.4.4.4. Global-average-pooling

- Compresses the feature map into a fixed-length vector, drastically reducing parameters compared with flattening.

5.4.4.5. Classification-head

- Dense 128, ReLU, Dropout 0.3
- Dense(num_classes), Softmax yields five-level pain probability.

5.4.4.6. Design rationale

The architecture combines a BiLSTM that encodes slow physiological modulations to provide long range context with an AM that emphasises moment to moment discriminative epochs and a CNN that distils fine grained shared patterns across modalities while early fusion forces the network to learn synergistic ECG EMG GSR relationships that mimic the human body's integrated pain response as illustrated in Fig.7.

While the overall structure combines BiLSTM, Attention, and 1D-CNN layers, the novelty regarding the presented study lies not in the individual components but in the way they are integrated. The following subsection highlights the methodological innovation introduced in this study.

5.4.5. Proposed methodological innovation

Strategically positioned between BiLSTM and 1D-CNN modules. Unlike conventional pain recognition architectures that either apply attention after convolutional layers or use static attention weights, the proposed design enables the AM to act as an information filter—selectively amplifying physiologically relevant temporal features before spatial feature extraction occurs. The BiLSTM component in this architecture initially captures the bidirectional temporal dependencies, encoding the involving dynamic of pain-related physiological variables. These temporal embeddings are then analyzed by the adaptive attention layer to determine the most informative parts that reflect meaningful physiological variations like sudden amplitude changes or oscillatory variations that are associated with pain events. The model is trained through a dynamic weighting method to attach importance to important signal intervals and constrain redundant or noisy data. Such refined attention-weighted features are then fed into the 1D-CNN layer, which does localized convolution, to obtain high-level discriminative patterns, depending on the attention-based temporal focus.

This mechanism ensures that the CNN receives only the most relevant temporal information, enhancing both interpretability and accuracy.

By embedding the AM between temporal and spatial encoding stages, the proposed framework introduces a novel interaction between sequence modelling and spatial abstraction, resulting in a model that is not only empirically superior but also conceptually distinct from existing pain recognition architectures. Tabel 2. shows the sequential of hybrid (BiLSTM+MHAT+CNN) model with all parameters.

5.5 Evaluation metrics

With the use of common bio-classification benchmarks, suggested BiLSTM–CNN hybrid model has been assessed on BioVid Heat Pain Database (ECG,EMG and GSR): Accuracy can be defined as percentage of samples that have been correctly classified; precision is defined as the reliability related to positive predictions; sensitivity or recall is defined as the capacity for identifying all the true positive cases; and specificity is defined as

the capacity for preventing false positives, which is essential for differentiating between pain and no pain. The confusion matrix reveals patterns regarding interclass misclassification, while F1 score offers harmonic mean of recall and precision which is robust to class imbalance. As shown in Table 3, tenfold cross validation guarantees the generalizability and stability of results, whereas ROC AUC measures discriminating performance across probability thresholds, particularly for binary pain versus no pain tasks.

6. Results and discussion

6.1 CNN model

This section presents experimental results of applying a 1-D convolutional neural network (1D-CNN) model to three physiological signals (ECG, EMG and GSR) using BioVid heat pain data-base (Part A). To analyze the effect of the order of multimodal signals on early integration, three experiments were designed with different signal arrangements, keeping the model structure and training settings constant in all cases. The model relied on stacked Conv1D layers to extract temporal features, followed by clustering layers and full-connections to classify pain levels. The results showed that the order of the physiological signals significantly affects classification performance, highlighting the importance of carefully selecting an early integration strategy in multimodal pain recognition systems.

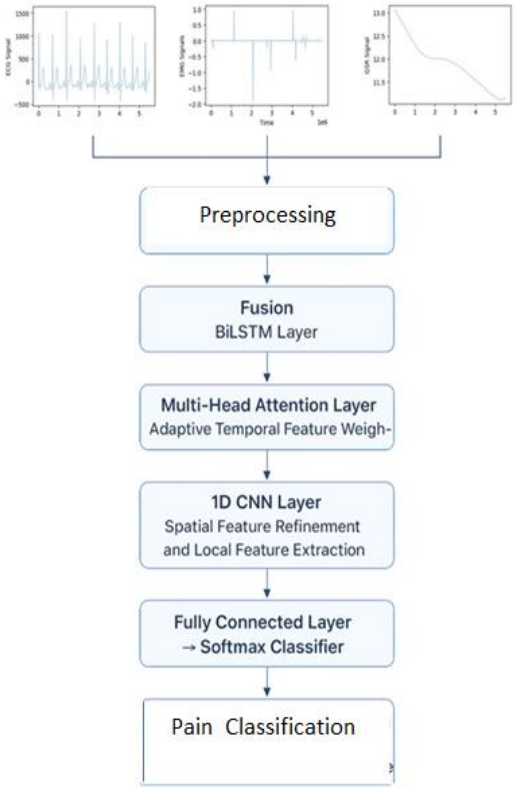


Figure. 9 General Architecture of Proposed Multi-Signals Deep Learning

Table 3. Classification performance measures[58]

Measure	Formula
Accuracy	$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$
Precision	$\text{Precision} = \frac{TP}{TP + FP}$
Recall	$\text{Recall} = \frac{TP}{TP + FN}$
F1 score	$\text{F1 score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$

Table 2. illustrates the sequential of hybrid (BiLSTM+MHAT+CNN) model with all parameters

Type of Layer	Output Shape	No. Parameter
input_layer (InputLayer)	(None, 400, 5)	0
Bidirectional (Bidirectional)	(None, 400, 256)	137,216
Dropout1 (Dropout)	(None, 400, 256)	0
Bidirectional (Bidirectional)	(None, 400, 128)	164,352
Dropout2 (Dropout)	(None, 400, 128)	0
multi_head_attention(MultiHeadAttention)	(None, 400, 128)	66,048
add(Add)	(None, 400, 128)	0
layer_normalization(LayerNormalization)	(None, 400, 128)	256
conv1d (Conv1D)	(None, 400, 64)	24,640
max_pooling1d (MaxPooling1D)	(None, 200, 64)	0
batch_normalization (BatchNormalization)	(None, 200, 64)	256
Dropout3(Dropout)	(None, 200, 64)	0
flatten(Flatten)	(None, 12800)	0
Dense (Dense)	(None, 128)	1,638,528
Dropout4(Dropout)	(None, 128)	0
Dense (Dense)	(None, 5)	645

Table 4 (A). Fold-wise Classification Performance of the CNN Model Using 5-Fold Cross-Validation

Experience 1 (GSR–ECG–EMG)				
Fold No.	**Levels of Pain**	**Precision**	**Recall**	**F1 Score**
1	BL1	0.9395	0.9408	0.9402
	PA1	0.9418	0.9349	0.9383
	PA2	0.9462	0.9362	0.9412
	PA3	0.9516	0.9454	0.9485
	PA4	0.9360	0.9573	0.9465
	Accuracy	-	-	0.9430
	Macro Avg	0.9430	0.9429	0.9429
	Weight Avg	0.9430	0.9430	0.9430
2	BL1	0.9426	0.9527	0.9476
	PA1	0.9604	0.9519	0.9561
	PA2	0.9535	0.9410	0.9472
	PA3	0.9581	0.9440	0.9510
	PA4	0.9414	0.9660	0.9535
	Accuracy	-	-	0.9511
	Macro Avg	0.9512	0.9511	0.9511
	Weight Avg	0.9513	0.9511	0.9511
3	BL1	.9469	0.9524	0.9496
	PA1	0.9552	0.9377	0.9464
	PA2	0.9358	0.9379	0.9368
	PA3	0.9480	0.9249	0.9363
	PA4	0.9375	0.9698	0.9534
	Accuracy	-	-	0.9445
	Macro Avg	0.9447	0.9445	0.9445
	Weight Avg	0.9446	0.9445	0.9445
4	BL1	0.9228	0.9433	0.9329
	PA1	0.9359	0.9353	0.9356
	PA2	0.9501	0.9206	0.9351
	PA3	0.9442	0.9442	0.9442
	PA4	0.9447	0.9537	0.9492
	Accuracy	-	-	0.9394
	Macro Avg	0.9395	0.9394	0.9394
	Weight Avg	0.9395	0.9394	0.9394
5	BL1	0.9153	0.9408	0.9279
	PA1	0.9323	0.9153	0.9238
	PA2	0.9438	0.9189	0.9312
	PA3	0.9401	0.9246	0.9323
	PA4	0.9157	0.9460	0.9306
	Accuracy	-	-	0.9292
	Macro Avg	0.9295	0.9291	0.9291
	Weight Avg	0.9295	0.9292	0.9292
All Fold	BL1	9332	0.9460	0.9396
	PA1	0.9453	0.9352	0.9402
	PA2	0.9458	0.9309	0.9383
	PA3	0.9483	0.9365	0.9424
	PA4	0.9350	0.9586	0.9467
	Accuracy	-	-	0.9414
	Macro Avg	0.9415	0.9414	0.9414
	Weight Avg	0.9415	0.9414	0.9414
Experience 2 (ECG–GSR–EMG)				
Fold No.	**Levels of Pain**	**Precision**	**Recall**	**F1 Score**
1	BL1	0.9517	0.9408	0.9462
	PA1	0.9466	0.9452	0.9459
	PA2	0.9479	0.9439	0.9459
	PA3	0.9444	0.9389	0.9416
	PA4	0.9374	0.9587	0.9479
	Accuracy	-	-	0.9455
	Macro Avg	0.9456	0.9455	0.9455
	Weight Avg	0.9456	0.9455	0.9455
2	BL1	0.9335	0.9549	0.9441
	PA1	0.9540	0.9409	0.9474
	PA2	0.9501	0.9425	0.9463
	PA3	0.9529	0.9410	0.9469
	PA4	0.9470	0.9580	0.9525
	Accuracy	-	-	0.9474
	Macro Avg	0.9475	0.9474	0.9474
	Weight Avg	0.9475	0.9474	0.9474
3	BL1	0.9456	0.9546	0.9501
	PA1	0.9445	0.9362	0.9404
	PA2	0.9435	0.9408	0.9422
	PA3	0.9416	0.9194	0.9304
	PA4	0.9466	0.9712	0.9587
	Accuracy	-	-	0.9444
	Macro Avg	0.9444	0.9444	0.9443
	Weight Avg	0.9444	0.9444	0.9443
4	BL1	0.9257	0.9296	0.9277
	PA1	0.9387	0.9268	0.9327
	PA2	0.9261	0.9315	0.9288
	PA3	0.9429	0.9328	0.9378
	PA4	0.9327	0.9457	0.9391
	Accuracy	-	-	0.9332
	Macro Avg	0.9332	0.9333	0.9332
	Weight Avg	0.9332	0.9332	0.9332
5	BL1	0.9302	0.9501	0.9400
	PA1	0.9472	0.9299	0.9385
	PA2	0.9394	0.9374	0.9384
	PA3	0.9440	0.9325	0.9382
	PA4	0.9416	0.9518	0.9467
	Accuracy	-	-	0.9404
	Macro Avg	0.9405	0.9403	0.9404
	Weight Avg	0.9404	0.9404	0.9404
All Fold	BL1	0.9372	0.9460	0.9416
	PA1	0.9462	0.9358	0.9410
	PA2	0.9414	0.9392	0.9403
	PA3	0.9451	0.9328	0.9389
	PA4	0.9411	0.9572	0.9491
	Accuracy	-	-	0.9422
	Macro Avg	0.9422	0.9422	0.9422
	Weight Avg	0.9422	0.9422	0.9422
Experience3 (EMG–GSR–ECG)				
Fold No.	**Levels of Pain**	**Precision**	**Recall**	**F1 Score**
1	BL1	0.9374	0.9394	0.9384
	PA1	0.9430	0.9312	0.9371
	PA2	0.9439	0.9552	0.9495
	PA3	0.9440	0.9454	0.9447
	PA4	0.9557	0.9523	0.9540
	Accuracy	-	-	0.9448
	Macro Avg	0.9448	0.9447	0.9447
	Weight Avg	0.9448	0.9448	0.9448

2	BL1	0.9516	0.9441	0.9479
	PA1	0.9415	0.9409	0.9412
	PA2	0.9453	0.9308	0.9380
	PA3	0.9397	0.9425	0.9411
	PA4	0.9383	0.9580	0.9480
	Accuracy	-	-	0.9432
	Macro Avg	0.9433	0.9433	0.9432
	Weight Avg	0.9433	0.9432	0.9432
3	BL1	0.9446	0.9488	0.9467
	PA1	0.9432	0.9501	0.9467
	PA2	0.9437	0.9313	0.9375
	PA3	0.9458	0.9333	0.9395
	PA4	0.9481	0.9621	0.9550
	Accuracy	-	-	0.9451
	Macro Avg	0.9451	0.9451	0.9451
	Weight Avg	0.9451	0.9451	0.9451
4	BL1	0.9283	0.9476	0.9378
	PA1	0.9379	0.9254	0.9316
	PA2	0.9391	0.9228	0.9309
	PA3	0.9470	0.9335	0.9402
	PA4	0.9319	0.9552	0.9434
	Accuracy	-	-	0.9368
	Macro Avg	0.9368	0.9369	0.9368
	Weight Avg	0.9369	0.9368	0.9368
5	BL1	0.9198	0.9494	0.9344
	PA1	0.9404	0.9219	0.9311
	PA2	0.9450	0.9296	0.9373
	PA3	0.9423	0.9261	0.9341
	PA4	0.9365	0.9561	0.9462
	Accuracy	-	-	0.9366
	Macro Avg	0.9368	0.9366	0.9366
	Weight Avg	0.9368	0.9366	0.9366
All Fold	BL1	0.9361	0.9458	0.9410
	PA1	0.9412	0.9339	0.9375
	PA2	0.9434	0.9341	0.9387
	PA3	0.9438	0.9361	0.9399
	PA4	0.9421	0.9568	0.9494
	Accuracy	-	-	0.9413
	Macro Avg	0.9413	0.9413	0.9413
	Weight Avg	0.9413	0.9413	0.9413

Table 4 presents the detailed results of all experiments using 5-fold cross-validation, divided into three main sections. Table 4(A) shows the classification performance of each individual fold, including accuracy, fine-tuning, recall, and F1 metrics, allowing for analysis of model performance variability across different folds. Table 4(B) focuses on the model's computational efficiency and training and validation performance, including indicators related to training time and learning stability.

The final pooled results, obtained through five-fold cross-validation Table 5. , show that the sequential 1D-CNN model performed consistently and highly across all three physiological signal

Table 4 (B). Computational Efficiency and Training–Validation Performance of the CNN Model

Experience 1 (GSR,ECG,EMG)					
Fold No.	**Train Acc.**	**Valid. Acc.**	**Train loss**	**Valid. Loss**	**Train Time (**Sec.)
1	0.9233	0.9430	0.1995	0.1898	247.12
2	0.9385	0.9511	0.1651	0.1680	243.67
3	0.9353	0.9445	0.1677	0.1224	244 c
4	0.9272	0.9394	0.1868	0.1872	242.90 .
5	0.9213	0.9292	0.2091	0.2412	244.94
Total Execution Time for Experience 1225.78 Sec					
Experience 2 (ECG–GSR–EMG)					
1	0.9327	0.9455	0.1766	0.1798	256.89
2	0.9286	0.9474	0.1913	0.1616	255.42
3	0.9370	0.9444	0.1689	0.1884	256.32
4	0.9366	0.9332	0.1741	0.2126	252.34
5	0.9289	0.9404	0.1838	0.1838	252.66
Total Execution Time for Experience 1276.74 Sec					
Experience 3 (EMG–GSR–ECG)					
1	0.9397	0.9448	0.1625	0.1864	248.29
2	0.9260	0.9432	0.1960	0.1855	242.21
3	0.9303	0.9451	0.1802	0.1884	242.71
4	0.9305	0.9368	0.1839	0.1839	242.32
5	0.9356	0.9366	0.1711	0.2107	242.03
Total Execution Time for Experience 1220.70 Sec.					

arrangements. Accuracy, recall, and F1 values ranged between 0.9402 and 0.9422, with low standard deviations (SD ≤ 0.0081), which had reflected strong statistical stability and good generalizability. Furthermore, the clear overlap of the 95% confidence intervals between trials indicates that the observed differences in performance are subtle but systematic, primarily due to the effect of the multimodal signal arrangement during early integration, rather than random variation in the training process.

In terms of comparative performance, the first trials achieved the highest accuracy and F1 values (0.9422), along with high Matthews correlation coefficients (MCC ≈ 0.9277), demonstrating an excellent balance between sensitivity and specificity and enhancing the model's reliability even in cases of class imbalance. Furthermore, all trials exhibited very high AUC values (≥ 0.9951) with extremely small standard deviations, confirming the model's high discriminatory power in separating different pain levels regardless of signal order.

As for computational efficiency, all three trials achieved relatively similar training times (≈ 1220–1276 seconds), with no undue computational burden associated with any particular signal order. This temporal convergence indicates that differences in classification performance are not due to differences in training time, but rather to how physiological

Table 5. Final Comparative Results of CNN and Hybrid (BiLSTM-MHAT-CNN) Models Using 5-Fold Cross-Validation (Mean ± SD and 95% Confidence Interval)

Model	CNN Model			Hybrid (BiLSTM_MHAT_CNN) Model		
Experience No.	**1**	**2**	**3**	**1**	**2**	**3**
Acc_mean	0.9414	0.9422	0.9413	0.805	0.800	0.789
Acc_sd	0.0081	0.0057	0.0043	0.009	0.006	0.023
Acc_ci_low	0.9314	0.9352	0.936	0.793	0.792	0.761
Acc_ci_high	0.9515	0.9492	0.9466	0.816	0.808	0.818
F1_mean	0.9414	0.9422	0.9413	0.804	0.800	0.789
F1_sd	0.0081	0.0056	0.0043	0.010	0.006	0.024
F1_ci_low	0.9314	0.9352	0.936	0.793	0.792	0.760
F1_ci_high	0.9514	0.9492	0.9466	0.816	0.808	0.818
Precision_mean	0.9416	0.9422	0.9414	0.805	0.801	0.790
Precision_sd	0.0081	0.0057	0.0042	0.009	0.006	0.023
Precision_ ci_low	0.9314	0.9257	0.9198	0.794	0.793	0.762
Precision_ci_high	0.9514	0.9540	0.9557	0.817	0.809	0.818
Recall_mean	0.9414	0.9422	0.9413	0.805	0.800	0.789
Recall_sd	0.0081	0.0057	0.0043	0.009	0.006	0.023
Recall_ ci_low	0.9314	0.9352	0.936	0.793	0.792	0.761
Recall_ ci_high	0.9515	0.9492	0.9466	0.816	0.808	0.818
Mcc_mean	0.9268	0.9277	0.9267	0.756	0.750	0.737
Mcc_sd	0.0101	0.0071	0.0053	0.012	0.008	0.029
Mcc_ ci_low	0.9143	0.919	0.9201	0.742	0.740	0.701
Mcc_ ci_high	0.9393	0.9365	0.9333	0.771	0.760	0.773
Auc_mean	0.9951	0.9956	0.9952	0.953	0.952	0.948
Auc_sd	0.0013	0.0007	0.0007	0.005	0.002	0.009
Auc_ ci_low	0.9935	0.9948	0.9943	0.947	0.949	0.937
Auc_ ci_high	0.9967	0.9964	0.9961	0.958	0.955	0.959

signals are organized within the model. This highlights the significance of the selection of optimal signal sequence for achieving the optimal level of balance between high performance and time efficiency, a crucial factor in the practical applications of automated pain recognition systems.

Overall, these results confirm that the design of the early integration strategy and the sequence of physiological signals are pivotal elements in multimodal deep learning models, where even a small improvement in signal sequence can lead to significant gains in performance and stability without increasing computational costs.

Fig. 10 show The final confusion matrix for each trial is presented to illustrate the model's performance in the differentiation between different categories, with the values reflecting the number of correct and incorrect predictions for every category. The final classification report is also presented as a bar graph that shows metrics of precision, recall, and F1-score for each category by Fig. 11 . This visual representation allows for easy comparison of the model's performance across all of the categories. The three trials are ranked by performance: Trial 1, 2, , and 3, demonstrating the evolution of performance between the different trials and providing a basis for determining the optimal model setup.

To ensure the fairness of the comparison, the hybrid model was applied to the same experiments, training settings, and cross-validation strategy previously used, allowing for a direct assessment of the hybrid architecture's impact on final performance.

The CNN model is adopted as the baseline in this study; therefore, the proposed hybrid BiLSTM–Multi-Head Attention–CNN model is primarily compared against the CNN architecture to emphasize the performance gains achieved by incorporating temporal modeling and attention mechanisms.

As a baseline comparison, Table 5 presents a comprehensive statistical evaluation of the suggested BiLSTM–Multi-Head Attention–CNN hybrid model using five-fold cross-validation. The table reports the mean performance, standard deviation, and 95% confidence intervals for multiple evaluation metrics, including accuracy (Acc), recall, MCC, F1-score, precision, and area under the ROC curve (AUC), across all experiments. This comparison with the baseline configuration enables a robust and reliable assessment of the model’s stability, discriminative capability, and generalization performance under different experimental conditions and signal-ordering strategies.

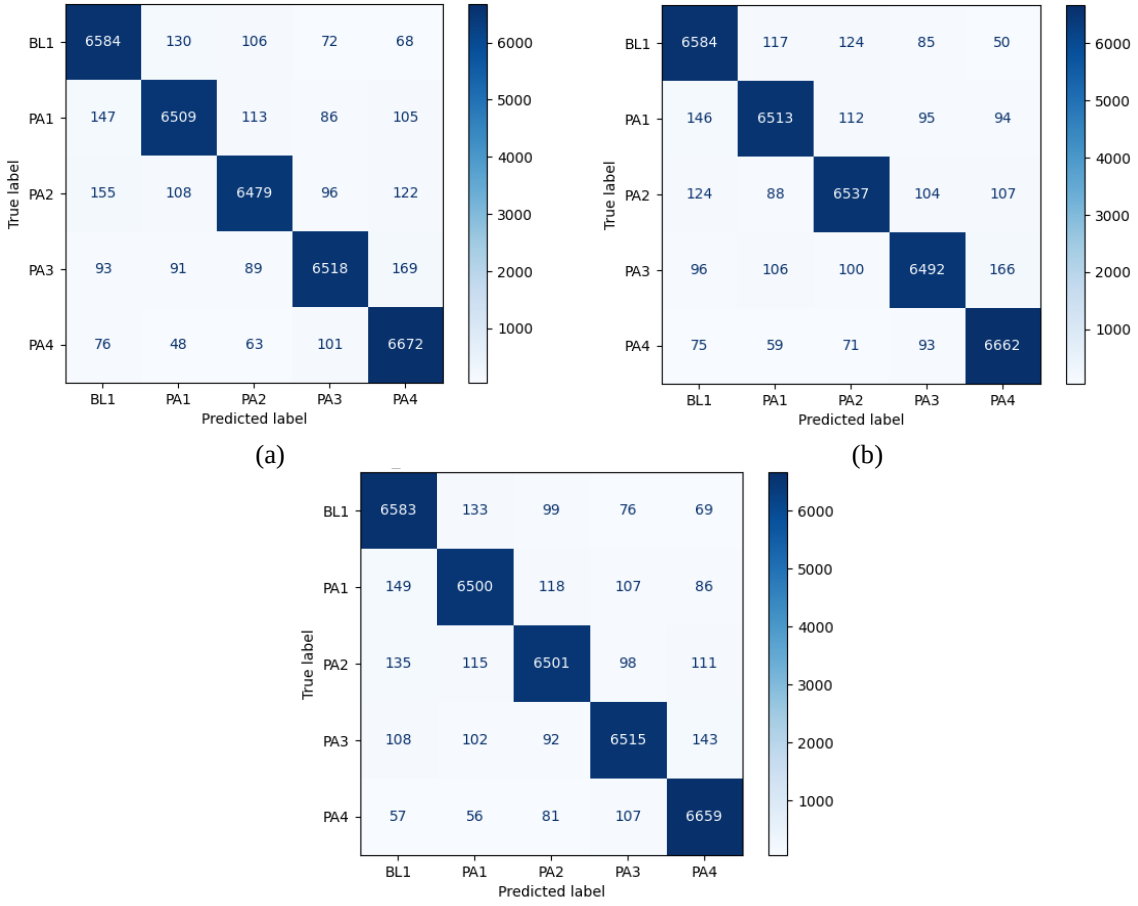


Figure. 10 Final Confusion Matrix for All Experiments for CNN Model: (a) Final Confusion Matrix for Experience 1, (b) Final Confusion Matrix for Experience 2, and (c) Final Confusion Matrix for Experience 3

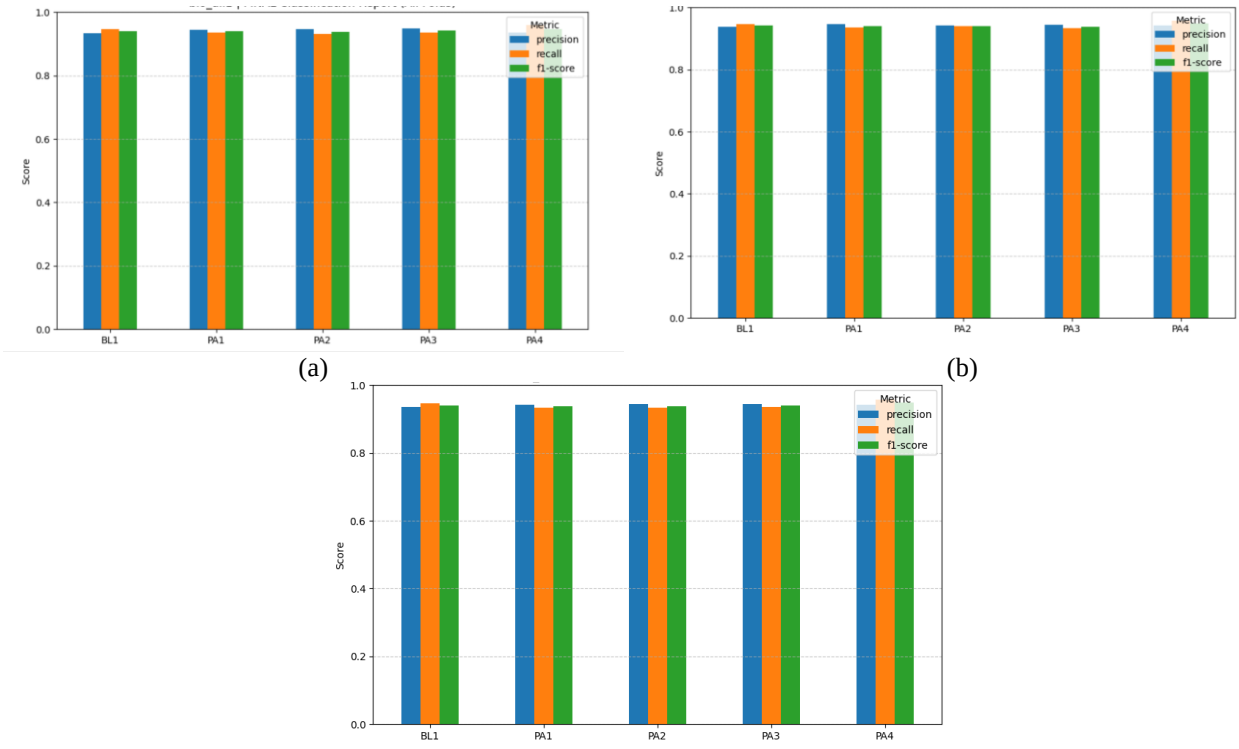


Figure. 11 Final Classification Report for All Experiments for CNN Model; (a) Final Classification Report for Experience 1, (b) Final Classification Report for Experience 2, and (c) Final Classification Report for Experience 3

Table 6 (A). Fold-wise Classification Performance of the Hybrid (BiLSTM_MHAT_CNN) Model Using 5-Fold Cross-Validation

Experience 1 (GSR–ECG–EMG)				
Fold No.	**Levels of Pain**	**Precision**	**Recall**	**F1 Score**
1	BL1	0.78	0.86	0.82
	PA1	0.81	0.79	0.80
	PA2	0.79	0.78	0.78
	PA3	0.83	0.77	0.80
	PA4	0.85	0.85	0.85
	Accuracy	-	-	0.81
	Macro Avg	0.81	0.81	0.81
	Weight Avg	0.81	0.81	0.81
2	BL1	0.75	0.83	0.79
	PA1	0.80	0.77	0.78
	PA2	0.79	0.75	0.77
	PA3	0.80	0.77	0.79
	PA4	0.82	0.84	0.83
	Accuracy	-	-	0.79
	Macro Avg	0.79	0.79	0.79
	Weight Avg	0.79	0.79	0.79
3	BL1	0.77	0.85	0.81
	PA1	0.79	0.77	0.78
	PA2	0.79	0.76	0.77
	PA3	0.82	0.76	0.79
	PA4	0.83	0.86	0.85
	Accuracy	-	-	0.80
	Macro Avg	0.80	0.80	0.80
	Weight Avg	0.80	0.80	0.80
4	BL1	0.78	0.83	0.81
	PA1	0.80	0.80	0.80
	PA2	0.82	0.77	0.79
	PA3	0.83	0.78	0.81
	PA4	0.81	0.86	0.83
	Accuracy	-	-	0.81
	Macro Avg	0.81	0.81	0.81
	Weight Avg	0.81	0.81	0.81
5	BL1	0.81	0.83	0.82
	PA1	0.79	0.79	0.79
	PA2	0.81	0.78	0.80
	PA3	0.82	0.80	0.81
	PA4	0.84	0.86	0.85
	Accuracy	-	-	0.81
	Macro Avg	0.81	0.81	0.81
	Weight Avg	0.81	0.81	0.81
All Fold	BL1	0.779	0.841	0.809
	PA1	0.779	0.783	0.791
	PA2	0.779	0.769	0.783
	PA3	0.819	0.777	0.797
	PA4	0.830	0.853	0.842
	Accuracy	-	-	0.805
	Macro Avg	0.805	0.805	0.804
	Weight Avg	0.805	0.805	0.804

Experience2 (ECG–GSR–EMG)				
Fold No.	**Levels of Pain**	**Precision**	**Recall**	**F1 Score**
1	BL1	0.76	0.85	0.80
	PA1	0.82	0.77	0.80
	PA2	0.80	0.78	0.79
	PA3	0.83	0.78	0.80
	PA4	0.84	0.87	0.85
	Accuracy	-	-	0.81
	Macro Avg	0.81	0.81	0.81
	Weight Avg	0.81	0.81	0.81
2	BL1	0.76	0.84	0.80
	PA1	0.81	0.78	0.79
	PA2	0.79	0.76	0.78
	PA3	0.81	0.78	0.79
	PA4	0.84	0.84	0.84
	Accuracy	-	-	0.80
	Macro Avg	0.80	0.80	0.80
	Weight Avg	0.80	0.80	0.80
3	BL1	0.79	0.83	0.81
	PA1	0.80	0.77	0.79
	PA2	0.81	0.78	0.79
	PA3	0.81	0.77	0.79
	PA4	0.81	0.86	0.84
	Accuracy	-	-	0.80
	Macro Avg	0.80	0.80	0.80
	Weight Avg	0.80	0.80	0.80
4	BL1	0.75	0.83	0.79
	PA1	0.77	0.78	0.78
	PA2	0.80	0.74	0.77
	PA3	0.82	0.75	0.78
	PA4	0.82	0.86	0.84
	Accuracy	-	-	0.79
	Macro Avg	0.79	0.79	0.79
	Weight Avg	0.79	0.79	0.79
5	BL1	0.76	0.82	0.79
	PA1	0.79	0.77	0.78
	PA2	0.80	0.76	0.78
	PA3	0.81	0.78	0.79
	PA4	0.82	0.85	0.84
	Accuracy	-	-	0.80
	Macro Avg	0.80	0.80	0.80
	Weight Avg	0.80	0.80	0.80
All Fold	BL1	0.763	0.834	0.797
	PA1	0.799	0.774	0.786
	PA2	0.800	0.764	0.782
	PA3	0.814	0.772	0.792
	PA4	0.827	0.856	0.841
	Accuracy	-	-	0.800
	Macro Avg	0.801	0.800	0.800

	Weight Avg	0.801	0.800	0.800
Experience3 (EMG–GSR–ECG)				
Fold No.	**Levels of Pain**	**Precision**	**Recall**	**F1 Score**
1	BL1	0.79	0.86	0.82
	PA1	0.78	0.79	0.78
	PA2	0.81	0.77	0.79
	PA3	0.82	0.76	0.79
	PA4	0.83	0.84	0.83
	Accuracy	-	-	0.80
	Macro Avg	0.80	0.80	0.80
	Weight Avg	0.80	0.80	0.80
2	BL1	0.74	0.79	0.76
	PA1	0.74	0.72	0.73
	PA2	0.75	0.70	0.72
	PA3	0.78	0.69	0.74
	PA4	0.74	0.84	0.78
	Accuracy	-	-	0.75
	Macro Avg	0.75	0.75	0.75
	Weight Avg	0.75	0.75	0.75
3	BL1	0.78	0.83	0.81
	PA1	0.80	0.78	0.79
	PA2	0.79	0.76	0.77
	PA3	0.81	0.77	0.79
	PA4	0.83	0.86	0.85
	Accuracy	-	-	0.80
	Macro Avg	0.80	0.80	0.80
	Weight Avg	0.80	0.80	0.80
4	BL1	0.75	0.84	0.79
	PA1	0.80	0.76	0.78
	PA2	0.80	0.77	0.78
	PA3	0.81	0.78	0.79
	PA4	0.83	0.83	0.83
	Accuracy	-	-	0.79
	Macro Avg	0.80	0.79	0.79
	Weight Avg	0.80	0.79	0.79
5	BL1	0.77	0.81	0.79
	PA1	0.79	0.77	0.78
	PA2	0.80	0.77	0.79
	PA3	0.80	0.79	0.80
	PA4	0.82	0.86	0.84
	Accuracy	-	-	0.80
	Macro Avg	0.80	0.80	0.80
	Weight Avg	0.80	0.80	0.80
All Fold	BL1	0.765	0.825	0.794
	PA1	0.783	0.764	0.773
	PA2	0.789	0.753	0.771
	PA3	0.803	0.759	0.781
	PA4	0.809	0.846	0.827
	Accuracy	-	-	0.789
	Macro Avg	0.790	0.789	0.789
	Weight Avg	0.790	0.789	0.789

Table 6 (B). Computational Efficiency and Training Validation Performance of the Hybrid (BiLSTM_MHAT_CNN) Model

Experience 1 (GSR,ECG,EMG)					
Fold No.	**Train Acc.**	**Valid. Acc.**	**Train loss**	**Valid. Loss**	**Train Time (**Sec.)
1	0.8073	0.8107	0.4847	0.6289	4725
2	0.8110	0.7918	0.4782	0.7204	4203.5
3	0.7948	0.7990	0.5024	0.6784	4680.2
4	0.8110	0.8051	0.4724	0.6484	4616.1
5	0.8166	0.8116	0.4610	0.6413	4700.0
Total Execution Time for Experience Sec					
Experience 2 (ECG–GSR–EMG)					
1	0.8149	0.8083	0.4704	0.6511	4738.4
2	0.8134	0.8044	0.4707	0.6392	4269.4
3	0.8202	0.8076	0.4648	0.6505	4733.6
4	0.7715	0.7937	0.5533	0.6838	4729.4
5	0.7987	0.7985	0.5140	0.6599	4264.7
Total Execution Time for Experience 22737.98 Sec					
Experience 3 (EMG–GSR–ECG)					
1	0.8099	0.8147	0.4826	0.6037	4499.7
2	0.7226	0.7482	0.7096	0.7411	4623.1
3	0.8133	0.8013	0.4705	0.6941	4657.6
4	0.8100	0.7971	0.4837	0.7049	4651.9
5	0.8016	0.8007	0.5064	0.6722	4728.2
Total Execution Time for Experience 23163.09 Sec.					

6.2 Hybrid (BiLSTM_MHAT_CNN) model

This section presents results of the proposed BiLSTM–Multi-Head Attention–CNN hybrid model in classifying pain levels based on multimodal physiological signals. This model combines BiLSTM's ability to model long-term time dependencies, the Multi-Head Attention mechanism to enhance the representation of the most time-sensitive features, and CNN layers to extract distinct spatial features. The results showed that integrating these components into a single architecture leads to more stable improvement in classification performance compared to individual models and enhances the model's ability to generalize across different signal arrangements.

Fig. 12 illustrates the final confusion matrix of the proposed hybrid model, demonstrating its ability to differentiate between various categories based on the number of correct and incorrect predictions for each category. Fig. 13 presents the final classification report as a bar chart of ACC, recall, and F1 scores for each category, facilitating visual comparison of the model's performance across all categories.

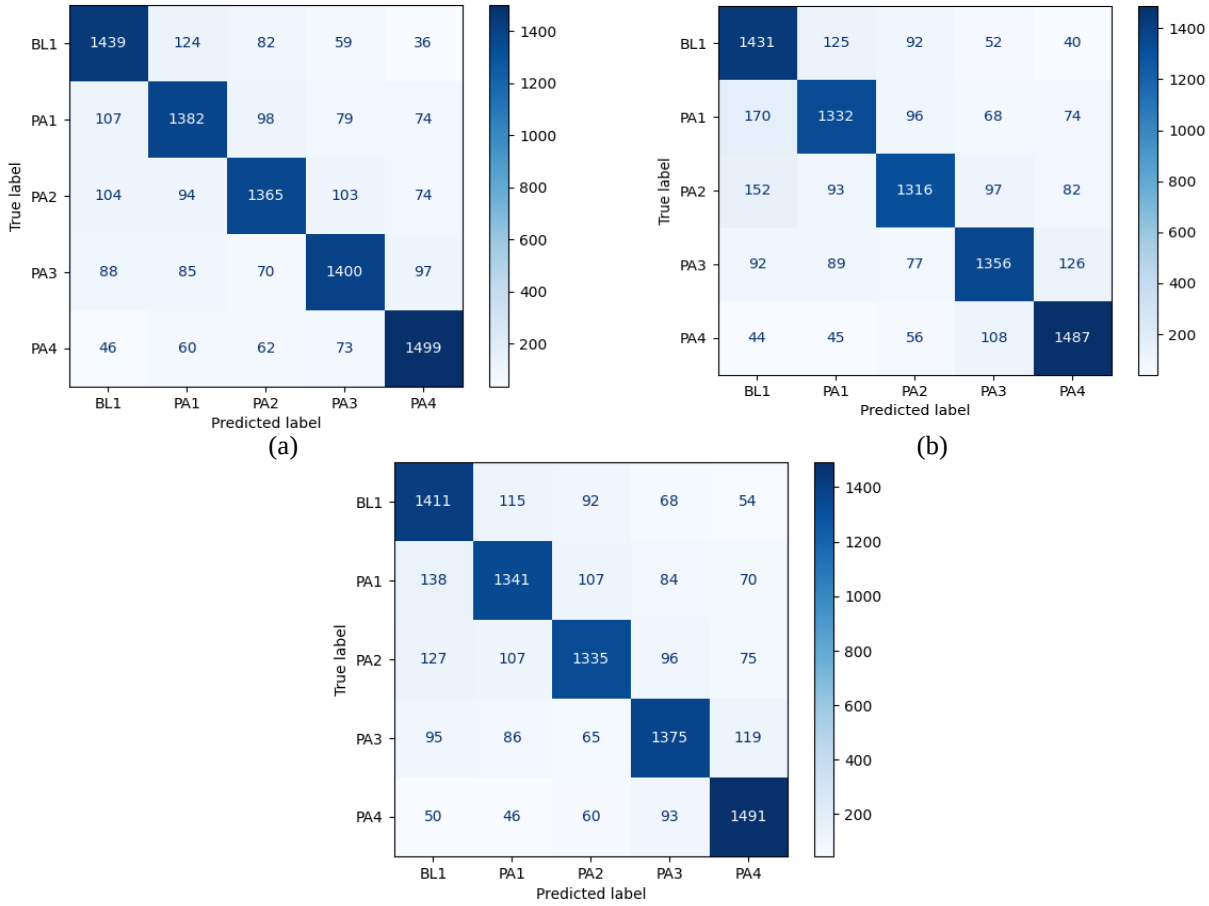


Figure. 12 Final Confusion Matrix for All Experiments for Hybrid (BiLSTM_MHAT_CNN) Model: (a) Final Confusion Matrix for Experience 1, (b) Final Confusion Matrix for Experience 2, and (c) Final Confusion Matrix for Experience 3

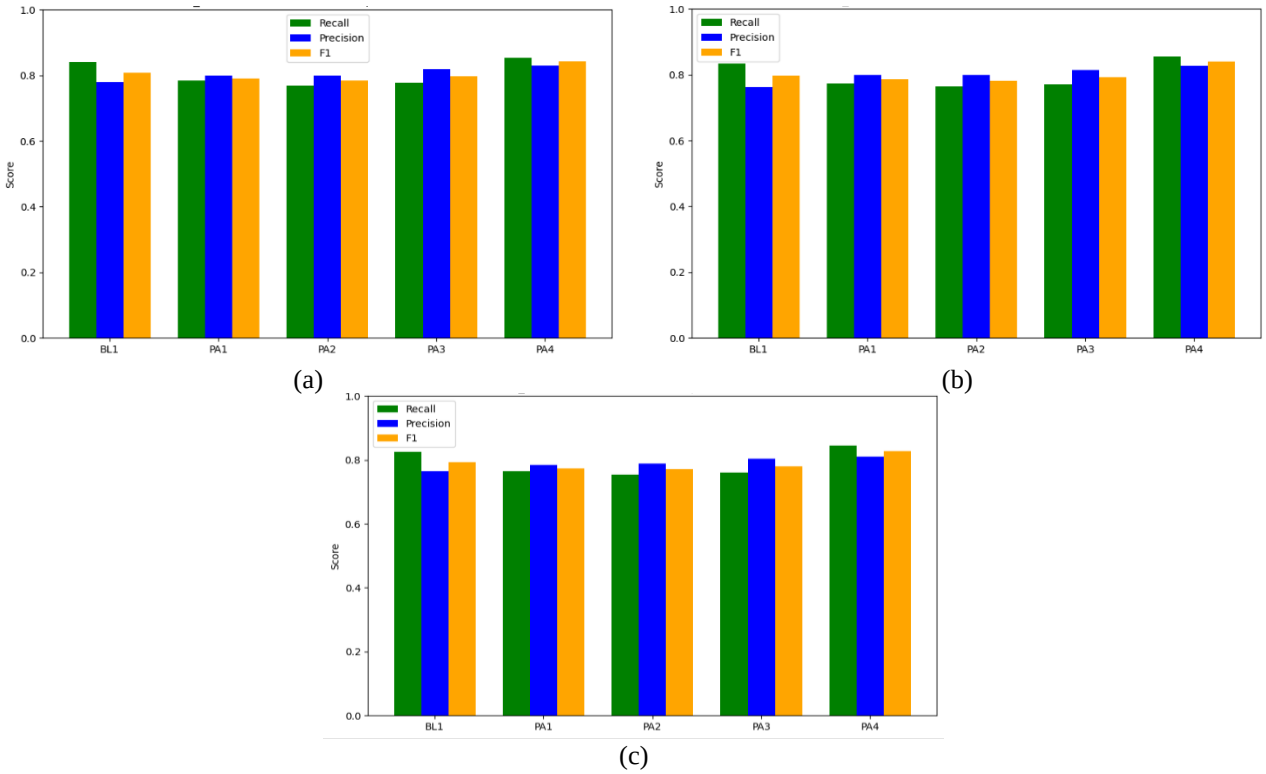


Figure. 13 Final Classification Report for All Experiments for Hybrid (BiLSTM_MHAT_CNN) Model: (a) Final Classification Report for Experience 1, (b) Final Classification Report for Experience 2, and (c) Final Classification Report for Experience 3

Experiments 1, 2, and 3 are ranked according to signals , highlighting incremental improvement and identifying the optimal model setup , which will be explained in comparative analysis section.

6.3 Comparative analysis

The performance of the CNN model as a baseline was compared with the proposed BiLSTM-Attention-CNN hybrid model using the same dataset and experimental settings to ensure a fair comparison. The experiments included three different signal set settings: EXP1 containing ['GSR', 'ECG', 'EMG_trapezius'], EXP2 containing ['ECG', 'GSR', 'EMG_trapezius'], and EXP3 containing ['EMG_trapezius', 'GSR', 'ECG']. The original dataset was augmented from 8,700 samples to 34,800 samples using augmentation techniques. Batch Size = 128 and Number of Epochs = 100 were used for all experiments and models.

The results showed that CNN model achieved the best numerical performance in all of the experiments when compared to the hybrid model. The highest absolute accuracy of the CNN model was achieved in EXP2 (Acc_mean = 0.9422), followed by EXP3 (0.9413) and EXP1 (0.9414), indicating that the arrangement of the combined signals can affect model performance. Data arrangement may facilitate the convolutional model's ability to capture the most influential local patterns. In contrast, the BiLSTM-Attention-CNN hybrid model exhibited lower numerical performance, with its highest accuracy in EXP1 (0.805) and lower performance in EXP3 (0.789). This reflects the sensitivity of complex models to data arrangement but also demonstrates greater stability when verified across K-folds.

Despite the lower numerical performance of the hybrid model compared to CNN, the standard deviation and 95% CI showed relatively stable results across different folds, suggesting greater generalizability and reliability. This is particularly important for critical applications, as the hybrid model provides more realistic performance estimates when using new data or unfamiliar distributions. The hybrid model also leverages the temporal structure of the data and attention to capture long-term dependencies between signals, something the CNN model cannot do as efficiently, even with data scaling via augmentation.

Furthermore, the results indicate that the order of the signals combined in the experiments can significantly impact performance. EXP2 and EXP1 outperformed EXP3 for both the CNN and the hybrid model, highlighting the importance of considering signal order when designing hybrid models that rely on chronological sequences and relationships between different channels. Although the training time for the hybrid model was considerably higher due to the complexity of the layers and attention, this computational investment enhances the model's long-term stability and reliability.

Based on this analysis, it has been concluded that the CNN model achieved the best numerical performance quickly when using Augmentation data, while the BiLSTM-Attention-CNN model provides higher reliability and stability with the ability to generalize results, especially when applying K-Fold Cross-Validation, and that the arrangement of data combined in experiments can significantly improve performance when choosing the optimal channel arrangement.

7. Conclusion and future work

The issue of automatic pain recognition is a difficult process because the experience and expression of pain is variable and depends on several elements including personality, past experiences, and pain origin. In this work, the researchers tried to increase the accuracy and reliability of pain recognition methods through combining the various sources of signals, such as physiological signals (ECG,EMG and GSR) and applyed three experiments in a different order at early fusion signals.

The evaluation of pain remains to be a serious but difficult part of medical treatment, especially among patients who are not capable of self-reporting. In order to meet this objective, non-invasive pain assessment requirement, we suggested a hybrid DL model combining BiLSTM, CNNs network, and multi-head attention to classify the level of pain based on multimodal physiological measures (ECG, EMG, and GSR) using Biovid dataset.

A systematic comparison was conducted between the proposed hybrid model based on BiLSTM–Multi-Head Attention–CNN and a baseline model based solely on CNN, to evaluate the impact of integrating two-way time modeling and a multi-head attention mechanism on improving classification performance. While the baseline model also achieved high and acceptable results across various evaluation indicators, reflecting its efficiency in extracting spatial features, the proposed model demonstrated more consistent and stable superiority when relying on five-fold cross-validation. This superiority was reflected in improved accuracy, F1 score, positive accuracy, recall, MCC, and AUC, along with a lower standard deviation and closer 95% confidence intervals. This confirms that the strong performance of baseline model does not detract from the

robustness of the suggested model, but rather highlights the added value of integrating time modeling and attention mechanisms in enhancing generalizability and stability, despite its lower results compared with baseline model.

In conclusion, findings of this research highlight the promising potential of hybrid deep modeling, which integrates temporal and spatial modeling with advanced attentional mechanisms, in intelligent healthcare systems, particularly for objective pain assessment tasks based on vital signs. The ability to achieve stable and generalizable performance, even when compared to robust baselines, suggests the viability of such systems as clinical decision support tools, contributing to reduced reliance on patient self-assessment and improved accuracy in continuous monitoring. Furthermore, this approach opens up future possibilities for developing non-invasive and scalable systems that can be integrated into clinical settings or remote monitoring systems, supporting early personalized care and enhancing the quality of healthcare services provided.

Future research should focus on several key directions:

1- Real-world validation in clinical environments with heterogeneous pain etiologies.
2- Integration of additional biosignals (e.g., EEG, facial expressions) to increase diagnostic granularity.
3- Edge-computing adaptations to enable point-of-care deployment and real-time analysis.

By bridging advances in artificial intelligence with clinical needs, this research contributes to the evolving paradigm of precision pain management, advocating for solutions that are both accurate and patient-centric.

8. Main scientific contribution

The major scientific contribution related to the presented work could be summarized as:

1- Building a robust multi-class classifier capable of distinguishing pain levels using multiple physiological signals with an MHAT that highlights the most informative temporal features, enabling accurate distinction between different pain intensity levels.
2- The MHAT enables the model to assign greater weight to time segments or features that carry more information relevant to pain discrimination, while ignoring irrelevant or noisy segments.
3- In pain signals, such as ECG, EMG, and GSR, not all time segments are equally important; the physiological changes associated with pain (such as increased heart rate or cutaneous conduction) occur only at specific moments.
4- The conventional model (such as LSTM or CNN alone) treats every part of the signal with equal importance, whereas AM allows the model to:
   - Adaptive focus on segments with characteristic pain activity;
   - Improved interpretability, since the weights generated by the AM point to the time regions most influential in the final decision;
   - Improved accuracy by enhancing important contextual representation.
5- Introducing a dynamic AM to highlight the most important time periods in the signal.
6- Proposing novel architecture combining Attention, BiLSTM, and 1D-CNN to enhance feature extraction.
7- Providing conceptual advancements by linking predictions to important physiological signal segments, increasing interpretability and methodological innovation.
8- Demonstrating that the order of vital signals directly affects performance, as the optimal order differs between the CNN model and the BiLSTM-MHAT-CNN algorithm, reflecting the sensitivity of sequence-dependent models and attention to the organization of multi-signal channels.

Conflicts of Interest

The authors declare that there aren't any conflicts of interest regarding this paper's publication.

Author Contributions

Conceptualization and methodology: Wala'a N. Jasim and Dr. Adala Mahdi Chyad; Software: Wala'a N. Jasim; Validation: Wala'a N. Jasim and Dr. Adala Mahdi Chyad; Formal analysis: Wala'a N. Jasim; Resources: Wala'a N. Jasim; Writing—Original Draft Preparation: Wala'a N. Jasim; Writing—Review and Editing: Wala'a N. Jasim and Dr. Adala Mahdi Chyad; Supervision: Dr. Adala Mahdi Chyad; Project administration: Dr. Adala Mahdi Chyad .

Acknowledgments

The authors would like to extend their gratitude to editors and reviewers for their valuable and constructive feedback. In addition to that, we wish to express our thanks to the university of Basrah and our College "College of Computer Science and Information Technology " for their continuous prop of our team.

References

[1] B. Alves, S. Brás, and R. Sebastião, "Explaining Pain: On the Impact of Physiological Signals in Pain Prediction", In: *Proc. of International Symposium: From Data to Models and Back*, pp. 70–87, 2024.

[2] S. N. Raja, D. B. Carr, M. Cohen, N. B. Finnerup, H. Flor, S. Gibson, F. J. Keefe, *et al.*, "The Revised International Association for the Study of Pain Definition of Pain: Concepts, Challenges, and Compromises", *Pain*, Vol. 161, No. 9, pp. 1976–1982, 2020.

[3] R. R. Bhatt, "Decoding the Neurological and Genetic Underpinnings of Chronic Pain", *Ph.D. dissertation*, Univ. of Southern California, 2024.

[4] S. N. El-Tallawy, J. V. Pergolizzi, I. Vasiliu-Feltes, R. S. Ahmed, J. K. LeQuang, H. N. El-Tallawy, G. Varrassi, and M. S. Nagiub, "Incorporation of "Artificial Intelligence" for Objective Pain Assessment: A Comprehensive Review", *Pain and Therapy*, Vol. 13, No. 3, pp. 293–317, 2024.

[5] K. N. Phan, N. K. Iyortsuun, S. Pant, H.-J. Yang, and S.-H. Kim, "Pain Recognition with Physiological Signals Using Multi-Level Context Information", *IEEE Access*, Vol. 11, pp. 20114–20127, 2023.

[6] S. Walter, S. Gruss, H. Ehleiter, J. Tan, H. C. Traue, P. Werner, *et al.*, "The BioVid Heat Pain Database Data for the Advancement and Systematic Validation of an Automated Pain Recognition System", In: *Proc. of 2013 IEEE International Conference on Cybernetics (CYBCO)*, pp. 128–131, 2013.

[7] D. Albahdal, W. Aljebreen, and D. M. Ibrahim, "PainMeter: Automatic Assessment of Pain Intensity Levels from Multiple Physiological Signals Using Machine Learning", *IEEE Access*, Vol. 12, pp. 48349–48365, 2024.

[8] P. Gouverneur, F. Li, W. M. Adamczyk, T. M. Szikszay, K. Luedtke, and M. Grzegorzek, "Comparison of Feature Extraction Methods for Physiological Signals for Heat-Based Pain Recognition", *Sensors*, Vol. 21, No. 14, 4838, 2021.

[9] P. Thiam, H. Hihn, D. A. Braun, H. A. Kestler, and F. Schwenker, "Multi-Modal Pain Intensity Assessment Based on Physiological Signals: A Deep Learning Perspective", *Frontiers in Physiology*, Vol. 12, 720464, 2021.

[10] B. T. Susam, N. T. Riek, M. Akcakaya, X. Xu, V. R. De Sa, H. Nezamfar, *et al.*, "Automated Pain Assessment in Children Using Electrodermal Activity and Video Data Fusion via Machine Learning", *IEEE Transactions on Biomedical Engineering*, Vol. 69, No. 1, pp. 422–431, 2021.

[11] A. Kelati, E. Nigussie, I. B. Dhaou, J. Plosila, and H. Tenhunen, "Real-Time Classification of Pain Level Using Zygomaticus and Corrugator EMG Features", *Electronics*, Vol. 11, No. 11, 1671, 2022.

[12] S. Gkikas, C. Chatzaki, E. Pavlidou, F. Verigou, K. Kalkanis, and M. Tsiknakis, "Automatic Pain Intensity Estimation Based on Electrocardiogram and Demographic Factors", In: *Proc. of ICT4AWE*, pp. 155–162, 2022.

[13] P. Bellmann, P. Thiam, H. A. Kestler, and F. Schwenker, "Machine Learning-Based Pain Intensity Estimation: Where Pattern Recognition Meets Chaos Theory—An Example Based on the Biovid Heat Pain Database", *IEEE Access*, Vol. 10, pp. 102770–102777, 2022.

[14] M. Benavent-Lledo, D. Mulero-Perez, D. Ortiz-Perez, J. Rodriguez-Juan, A. Berenguer-Agullo, A. Psarrou, and J. Garcia-Rodriguez, "A Comprehensive Study on Pain Assessment from Multimodal Sensor Data", *Sensors*, Vol. 23, No. 24, 9675, 2023.

[15] J. O. Pinzon-Arenas, Y. Kong, K. H. Chon, and H. F. Posada-Quintero, "Design and Evaluation of Deep Learning Models for Continuous Acute Pain Detection Based on Phasic Electrodermal Activity", *IEEE Journal of Biomedical and Health Informatics*, Vol. 27, No. 9, pp. 4250–4260, 2023.

[16] Z. Lu, B. Ozek, and S. Kamarthi, "Transformer Encoder with Multiscale Deep Learning for Pain Classification Using Physiological Signals", *Frontiers in Physiology*, Vol. 14, 1294577, 2023.

[17] S. Gkikas, N. S. Tachos, S. Andreadis, V. C. Pezoulas, D. Zaridis, G. Gkois, *et al.*, "Multimodal Automatic Assessment of Acute Pain through Facial Videos and Heart Rate Signals Utilizing Transformer-Based Architectures", *Frontiers in Pain Research*, Vol. 5, 1372814, 2024.

[18] J. Li, J. Luo, Y. Wang, Y. Jiang, X. Chen, and Y. Quan, "Automatic Pain Assessment Based on Physiological Signals: Application of Multi-Scale Networks and Cross-Attention", In: *Proc. of 2024 13th International Conference on Bioinformatics and Biomedical Science*, pp. 113–122, 2024.

[19] C. K. Behera and D. Padhy, "Multi-Level Analysis for Pain Recognition from Physiological Signals", In: *Proc. of 2024 2nd International Conference on Computer*, Communication and Control (IC4), pp. 1–6, 2024.

[20] E.-H. Jang, Y.-J. Eum, D. Yoon, and S. Byun, "Classifying Emotionally Induced Pain Intensity Using Multimodal Physiological Signals and Subjective Ratings: A Pilot Study", *Applied Sciences*, Vol. 15, No. 13, 7149, 2025.

[21] J. Farmani, A. Giuseppi, G. Bargshady, and R. F. Rojas, "Multimodal Automatic Acute Pain Recognition Using Facial Expressions and Physiological Signals", In: *Proc. of International Conference on Neural Information Processing*, pp. 49–62, 2024.

[22] M. Sowmya, N. Dinakar, and A. J. Heiner, "Pain Recognition via Multi-Context Integration of Physiological Signals", In: *Proc. of 2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI)*, pp. 1559–1564, 2025.

[23] S. Moscato, P. Cortelli, and L. Chiari, "Physiological Responses to Pain in Cancer Patients: A Systematic Review", *Computer Methods and Programs in Biomedicine*, Vol. 217, 106682, 2022.

[24] A. Khalaf, M. Nabian, M. Fan, Y. Yin, J. Wormwood, E. Siegel, K. S. Quigley, *et al.*, "Analysis of Multimodal Physiological Signals within and between Individuals to Predict Psychological Challenge vs. Threat", *Expert Systems with Applications*, Vol. 140, 112890, 2020.

[25] R. Sebastião, A. Bento, and S. Brás, "Analysis of Physiological Responses during Pain Induction", *Sensors*, Vol. 22, No. 23, 9276, 2022.

[26] J. Koenig, M. N. Jarczok, R. J. Ellis, T. K. Hillecke, and J. F. Thayer, "Heart Rate Variability and Experimentally Induced Pain in Healthy Adults: A Systematic Review", *European Journal of Pain*, Vol. 18, No. 3, pp. 301–314, 2014.

[27] L. M. Tracy, L. Ioannou, K. S. Baker, S. J. Gibson, N. Georgiou-Karistianis, and M. J. Giummarra, "Meta-Analytic Evidence for Decreased Heart Rate Variability in Chronic Pain Implicating Parasympathetic Nervous System Dysregulation", *Pain*, Vol. 157, No. 1, pp. 7–29, 2016.

[28] G. Forte, G. Troisi, M. Pazzaglia, V. De Pascalis, and M. Casagrande, "Heart Rate Variability and Pain: A Systematic Review", *Brain Sciences*, Vol. 12, No. 2, 153, 2022.

[29] D. Farina, L. Arendt-Nielsen, and T. Graven-Nielsen, "Experimental Muscle Pain Decreases Voluntary EMG Activity but Does Not Affect the Muscle Potential Evoked by Transcutaneous Electrical Stimulation", *Clinical Neurophysiology*, Vol. 116, No. 7, pp. 1558–1565, 2005.

[30] T. Graven-Nielsen, P. Svensson, and L. Arendt-Nielsen, "Effects of Experimental Muscle Pain on Muscle Activity and Co-ordination during Static and Dynamic Motor Function", *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control*, Vol. 105, No. 2, pp. 156–164, 1997.

[31] D. S. Bari, H. Y. Yacoob Aldosky, C. Tronstad, H. Kalvøy, and Ø. G. Martinsen, "Electrodermal Activity Responses for Quantitative Assessment of Felt Pain", *Journal of Electrical Bioimpedance*, Vol. 9, No. 1, p. 52, 2018.

[32] Y. Kong, H. F. Posada-Quintero, and K. H. Chon, "Sensitive Physiological Indices of Pain Based on Differential Characteristics of Electrodermal Activity", *IEEE Transactions on Biomedical Engineering*, Vol. 68, No. 10, pp. 3122–3130, 2021.

[33] L. Zhang, S. Walter, X. Ma, P. Werner, A. Al-Hamadi, H. C. Traue, and S. Gruss, "BioVid Emo DB: A Multimodal Database for Emotion Analyses Validated by Subjective Ratings", In: *Proc. of 2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–6, 2016.

[34] M. F. H. Siddiqui, P. Dhakal, X. Yang, and A. Y. Javaid, "A Survey on Databases for Multimodal Emotion Recognition and an Introduction to the VIRI (Visible and Infrared Image) Database", *Multimodal Technologies and Interaction*, Vol. 6, No. 6, 47, 2022.

[35] S. Walter, S. Gruss, H. Ehleiter, J. Tan, H. C. Traue, P. Werner, *et al.*, "The BioVid Heat Pain Database Data for the Advancement and Systematic Validation of an Automated Pain Recognition System", In: *Proc. of 2013 IEEE International Conference on Cybernetics (CYBCO)*, pp. 128–131, 2013.

[36] [Online]. Available: https://github.com/philippwerner/pain-

database-list?tab=readme-ov-file#unbc-mcmaster-shoulder-pain-expression-archive-database.

[37] P. Silva and R. Sebastião, "Using the Electrocardiogram for Pain Classification under Emotional Contexts", *Sensors*, Vol. 23, No. 3, 1443, 2023.

[38] U. Naseem, I. Razzak, S. K. Khan, and M. Prasad, "A Comprehensive Survey on Word Representation Models: From Classical to State-of-the-Art Word Representation Language Models", *Transactions on Asian and Low-Resource Language Information Processing*, Vol. 20, No. 5, pp. 1–35, 2021.

[39] 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, 8087, 2022.

[40] Z. Yu, V. Ramanarayanan, D. Suendermann-Oeft, X. Wang, K. Zechner, L. Chen, J. Tao, A. Ivanou, and Y. Qian, "Using Bidirectional LSTM Recurrent Neural Networks to Learn High-Level Abstractions of Sequential Features for Automated Scoring of Non-Native Spontaneous Speech", In: *Proc. of 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 338–345, 2015.

[41] R. Wang, K. Xu, H. Feng, and W. Chen, "Hybrid RNN-ANN Based Deep Physiological Network for Pain Recognition", In: *Proc. of 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 5584–5587, 2020.

[42] Z. N. Nemer, S. A. W. Saddam, R. J. Al-Sukeinee, and E. A. J. Harfash, "A Comparative Performance Analysis of CNN, LSTM, and CNNLSTM Models for Classifying Sleep Stages Using Electroencephalography Signals", *Informatica*, Vol. 49, No. 28, 2025.

[43] Z. Tao, X. Chang, H. Lu, X. Ye, Y. Liu, and X. Zheng, "Pooling Operations in Deep Learning: From 'Invariable' to 'Variable'", *BioMed Research International*, Vol. 2022, No. 1, 4067581, 2022.

[44] W. N. Jasim, Z. N. Nemer, and E. A. J. Harfash, "Implementation of Multiple CNN Architectures to Classify the Sea Coral Images", *Informatica*, Vol. 47, No. 1, 2023.

[45] S. Ghaffarian, J. Valente, M. Van Der Voort, and B. Tekinerdogan, "Effect of Attention Mechanism in Deep Learning-Based Remote Sensing Image Processing: A Systematic Literature Review", *Remote Sensing*, Vol. 13, No. 15, 2965, 2021.

[46] D. Bahdanau, "Neural Machine Translation by Jointly Learning to Align and Translate", *arXiv Preprint*, arXiv:1409.0473, 2014.

[47] J. Ma, Q. Ma, X. Tang, X. Zhang, C. Zhu, Q. Peng, and L. Jiao, "Remote Sensing Scene Classification Based on Global and Local Consistent Network", In: *Proc. of 2020 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 537–540, 2020.

[48] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks", In: *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, 2018.

[49] L. Wang, J. Peng, and W. Sun, "Spatial–Spectral Squeeze-and-Excitation Residual Network for Hyperspectral Image Classification", *Remote Sensing*, Vol. 11, No. 7, 884, 2019.

[50] A. S. Alswayed, H. S. Alhichri, and Y. Bazi, "SqueezeNet with Attention for Remote Sensing Scene Classification", In: *Proc. of 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS)*, pp. 1–4, 2020.

[51] J. Lu, J. Yang, D. Batra, and D. Parikh, "Hierarchical Question-Image Co-Attention for Visual Question Answering", *Advances in Neural Information Processing Systems*, Vol. 29, 2016.

[52] H. Wu, S. Zhao, L. Li, C. Lu, and W. Chen, "Self-Attention Network with Joint Loss for Remote Sensing Image Scene Classification", *IEEE Access*, Vol. 8, pp. 210347–210359, 2020.

[53] Z. Niu, G. Zhong, and H. Yu, "A Review on the Attention Mechanism of Deep Learning", *Neurocomputing*, Vol. 452, pp. 48–62, 2021.

[54] S. Zhang, C. Li, S. Qiu, C. Gao, F. Zhang, Z. Du, and R. Liu, "EMMCNN: An ETPS-Based Multi-Scale and Multi-Feature Method Using CNN for High Spatial Resolution Image Land-Cover Classification", *Remote Sensing*, Vol. 12, No. 1, 66, 2019.

[55] T. Shen, T. Zhou, G. Long, J. Jiang, S. Pan, and C. Zhang, "DiSAN: Directional Self-Attention Network for RNN/CNN-Free Language Understanding", In: *Proc. of AAAI Conference on Artificial Intelligence*, Vol. 32, No. 1, 2018.

[56] H. S. S. Al-deen, Z. Zeng, R. Al-sabri, and A. Hekmat, "An Improved Model for Analyzing Textual Sentiment Based on a Deep Neural Network Using Multi-Head Attention Mechanism", *Applied System Innovation*, Vol. 4, No. 4, 85, 2021.

[57] X. Wan, "Co-Training for Cross-Lingual Sentiment Classification", In: *Proc. of Joint*

Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, pp. 235–243, 2009.

[58] S. L. Mirtaheri and R. Shahbazian, *Machine Learning: Theory to Applications*, 1st ed. Boca Raton, FL, USA: CRC Press, 2022.