Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Explainable Automated Pain Level Classification Using a Deep Neural Network with EEG Signals

Mohammed Nafizul Islam Chowdhury Nayan
University of Southern Queensland
SALVIN PRASAD
Fiji National University
RAVINESH DEO
ravinesh.deo@usq.edu.au

University of Southern Queensland
RAJENDRA ACHARYA
University of Southern Queensland
PRABAL BARUA
University of Southern Queensland
ARUNA DEVI
University of the Sunshine Coast
JEFFREY SOAR
University of Southern Queensland

Research Article

Keywords: EEG-based pain classification, Spatial-temporal CNN, Explainable AI (XAI), Bayesian optimisation, Subject-independent validation, Laser-evoked potentials, Pain intensity classification

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11448439/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Explainable Automated Pain Level Classification Using a Deep Neural Network with EEG Signals

Mohammed Nafizul Islam Chowdhury Nayan[1, *], Salvin Prasad[2], Ravinesh C Deo[1, *], Rajendra Acharya[1], Prabal Datta Barua[1], Aruna Devi[4] and Jeffrey Soar[5]

[1]Artificial Intelligence Applications Laboratory, School of Science, Engineering and Digital Technologies, University of Southern Queensland, QLD 4300, Australia

[2] School of Sciences, College of Engineering, and Technical Vocational Education and Training, Fiji National University, Suva, Fiji

[3]School of Education and Tertiary Access, University of the Sunshine Coast, Petrie, QLD, Australia

*Corresponding Author: Ravinesh C Deo (e-mail: ravinesh.deo@unisq.edu.au)

ABSTRACT

Background: Traditional and handcrafted approaches often neglect the spatial distribution and temporal dynamics of raw EEG signals, creating a gap in objective pain classification and clinical trust. As a result, existing pain classification systems overlook important neural patterns associated with pain perception, limiting both classification performance and clinical interpretability.

Objectives: In this paper, we introduce a novel electroencephalography (EEG)-based approach to develop a specialised Spatial-Temporal Convolutional Neural Network (PainNet-ST) that integrates an Explainable AI (XAI) method to establish a three-class metric model for pain severity classification (low, medium, and high).

Methods: Our method includes the conversion of the raw EEG data from the OpenNeuro ds005280 dataset (nociceptive-evoked potentials) into multi-spectral topography maps (2D images) to preserve spatial distribution and temporal dynamics. Our proposed PainNet-ST model integrates spatial-temporal decoupling, Bayesian optimisation (implemented via the Optuna approach), and dual-layer XAI. Our analysis focused on the neural feature representations across the somatosensory cortex and the N2-P2 temporal window. Proposed PainNet-ST model was benchmarked against Support Vector Machines (SVM), Random Forest (RF), 2D-CNN, and STAM-net. For explainability, SHAP-based visualisation was incorporated to address the gap in clinical trust.

Results: Our proposed PainNet-ST model was evaluated against the benchmarked models with 11 performance metrics using accuracy, precision, recall, specificity, F1-score, False Positive Rate (FPR), False Negative Rate (FNR), Negative Prediction Value (NPV), False Discovery Rate (FDR), Cohen's Kappa (k), and Area Under the Curve (AUC), with results showing that the proposed model PainNet-ST was better than the other benchmark models in every classification task. That is, in this three-class pain classification problem, an accuracy of 92.28, F1-score of 92.29%, precision of 92.35%, recall of 92.28%, and specificity of 96.14% were obtained.

Conclusion: In the multiclass raw EEG-based pain classification task (low, medium, high), the model reached 92.28% accuracy, demonstrating the efficacy of the proposed method for the classification of pain. The proposed PainNet-ST model successfully addresses multi-class raw EEG-based pain classification by integrating spatial-temporal decoupling, Bayesian optimisation, and dual-layer XAI.

KEYWORDS

EEG-based pain classification, Spatial-temporal CNN, Explainable AI (XAI), Bayesian optimisation, Subject-independent validation, Laser-evoked potentials, Pain intensity classification

1. INTRODUCTION

The International Association for the Study of Pain (IASP) defines pain as "an unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage" (Raja et al., 2020). From a neurocomputational perspective this definition is expanded to include distributed brain signals that can be objectively assessed through spatio-temporal brain activity (Fu et al., 2024; Hu & Iannetti, 2019; Huang et al., 2013). This broader definition highlights that pain is a complex physiological and psychological phenomenon. According to the revised IASP commentary, "Verbal description is only one of several behaviours used to express pain".

Globally, over 1.5 billion people, which is approximately 20% of the adult population, experience chronic pain, with 1 out of 10 adults receiving diagnosis (International Association for the Study of Pain (IASP) 2025). Despite its prevalence, 40.8% of patients in Australia receive suboptimal or "low value" care due to the lack of objective diagnostic tools (Australia, 2024; Mishriky et al., 2020). This diagnostic gap is significantly wider in vulnerable populations, including elderly patients with dementia and those in intensive care units, indicating a critical failure in equitable healthcare delivery (Pain, 2025).

In Australia, chronic pain remains a major public health challenge, affecting approximately 3.24 million individuals. It is estimated that 20% of people over the age of 45 live with persistent chronic pain (Australian Institute of Health and Welfare, 2024). The total financial cost of chronic pain in Australia was estimated at $139.3 billion annually, and is projected to rise to $215.6 billion by 2050, mainly contributed by the productivity losses and health system expenditures (Economics, 2019).

Bridging this gap between subjective self-reporting and objective AI- driven pain classification (at a fraction of estimated current productivity loss) is no longer only a clinical goal, but also an economic necessity.

Pain serves as a critical homeostatic signal; however, accurate pain assessment is clinically challenging for effective diagnosis, guiding clinical decisions, ensuring appropriate treatment planning, and maximizing patient's well-being and improving patient's quality of life. Currently, pain evaluation heavily relies on the self-reporting scale, such as the Visual Analog Scale (VAS) and Numerical Rating Scale (NRS) (Hawker et al., 2011). This reliance on subjective measures can be problematic for populations with impaired communication. Such dependence on an individual's perception, can lead to clinical inaccuracies resulting in either overtreatment or undertreatment (Mouraux & Iannetti, 2018). Consequently , there is an urgent need for reliable, non-invasive and objective systems capable of assessing pain using physiological or behavioural signals (Chen et al., 2025; Zis et al., 2022).

Electroencephalography (EEG) has emerged as a promising modality for objective pain classification. However, EEG signals are inherently high-dimensional, spanning multiple channels over time. They are also non-linear as it reflects the complex cortical interaction, and are notoriously noisy, due to contamination with physiological artefacts (Alshehri et al., 2025; Rodriguez-Bermudez & Garcia-Laencina, 2015). Furthermore, model generalisability is affected further by the variability of EEG signals across individual subjects (Rezzouk et al., 2025). Traditional signal processing algorithms depend heavily on manually hand-crafted features and domain-specific expertise. This signal processing involves extracting important information from EEG signals, which comes with the disadvantage of missing out complex temporal and spatial patterns associated with pain perception. As a result, researchers have moved mostly towards machine learning (ML) and deep learning (DL) to overcome the limitations of manual feature engineering.

Earlier studies often employed diverse validation tactics, such as partitioning data into train-test sets in altered proportions. Thiam et al. (2019) used a fixed train-test (e.g., 80:20) ratio. Cross-validation, like K-fold, is performed at random split up by the number of folds, and every time K=1 fold, data is trained once, and one-fold is put out for testing. If the split is done with a single person's EEG trials, the model learns subject-specific patterns rather than the pain itself. These approaches may introduce data leakage or train-test overlapping (Jui et al., 2026). Consequently, the evaluation metrics show improved performance for a certain dataset, but for completely unknown/unseen data, the model operates poorly (Rosenblatt et al., 2023).

Unlike traditional statistical methods, Deep Learning architectures have shown strong capability in learning hierarchical representations directly from raw EEG signals (Barnova et al., 2023). Deep Learning models are particularly effective in modelling non-linear EEG patterns. In particular, convolutional neural networks (CNN) and recurrent neural networks (RNN) have excelled in their strong capabilities in learning hierarchical representations from raw physiological signals. Recent studies have reported successful classification of pain states using DL-based EEG signals, including EEG-based brain-computer interface and clinical settings (Al-Nafjan et al., 2025; Hamwi et al., 2025). Automated pain classification systems have gradually started to focus on objectively identifying pain using different behavioural or physiological signals that include facial expressions, speech patterns, or neurophysiological measurements (Cascella et al., 2025; Hammal et al., 2025; Subramanian et al., 2025). Recent deep learning studies emphasize the need to capture spatio-temporal neural dynamics while ensuring model interpretability in EEG classification (Cao et al., 2024; Tveter et al., 2024). Extending this focus, Sun et al. (2026) demonstrated that temporal modelling with CNN-BiLSTM architectures effectively accounts for subject-level variability in automated pain assessment.

Current literature, however, remains restricted to several limitations. Existing studies have focused on binary pain detection rather than multi-level pain intensity classification. However, many approaches rely on manually designed architectures, which may fail to capture the nonlinear complexities of EEG signals. These approaches are also identified to have human bias (Abbasimehr & Paki, 2021). Although CNNs and attention-based models have shown better predictive results (Abbasimehr & Paki, 2021; Chen et al., 2022; Gunter et al., 2023; Liu et al., 2022; Nezam et al., 2018; Phan et al., 2022; Subakan et al., 2021; Yu et al., 2020), challenges related to parameter optimisation, model generalisability, and interpretability persist. Particularly, hyperparameter selection remains a complex and often inefficient process. This demands an automated optimisation strategy such as Bayesian optimisation (implemented via Optuna with the TPE sampler) (Imai et al., 2025; Kamhi et al., 2022; Sameen et al., 2020; Wu et al., 2022). While explainable artificial intelligence (XAI) methods have gained popularity in clinical applications, their integration into multi-class EEG-based pain classification still remains limited.

The need to publish this paper arises from a clear research gap that exists where current approaches do not provide an adequately unified framework. These research gaps include:

- lack of support for multi-level pain classification,
- limited subject-independent generalisation,
- absence of automated hyperparameter optimisation strategies, and
- insufficient model interpretability for clinical trust.

The proposed PainNet-ST model therefore addresses the above limitations by introducing a unified framework with the following contributions:

- A Spatial-temporal decoupled CNN architecture designed for learning EEG representations

- Integration of Bayesian optimisation-based hyperparameter tuning using Optuna (TPE sampler)
- Implementation of subject-independent validation using StratifiedGroupKFold
- Integration of Dual-layer Explainable AI (SHAP and LIME)
- Benchmarking against state-of-the-art Machine Learning and Deep Learning methods

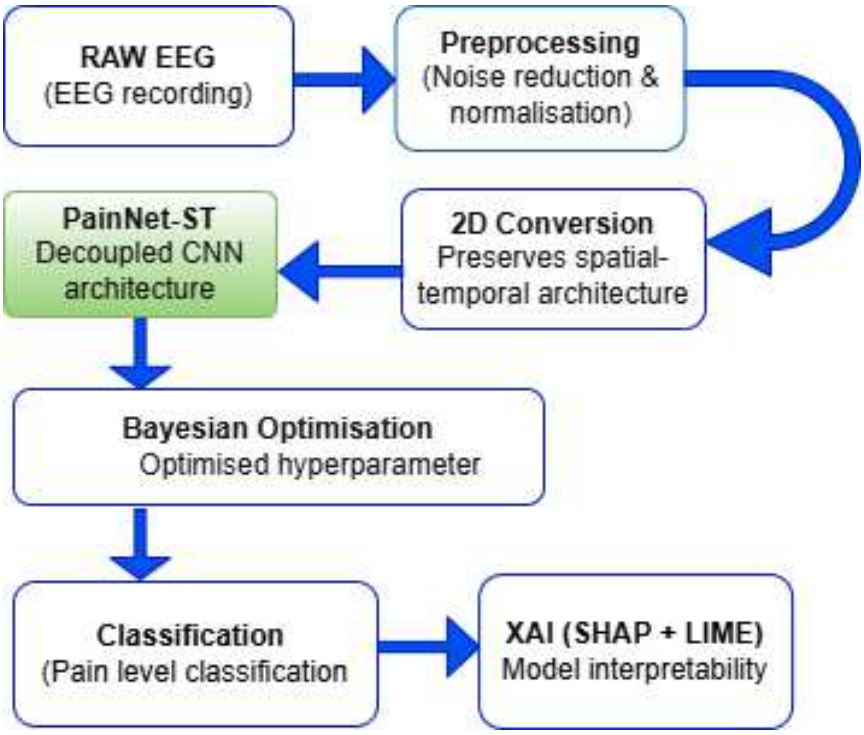


Fig. 1. Key contributions of the proposed PainNet-ST model.

The outcome of this study is therefore a PainNet-ST model that is expected to support clinicians in pain management through an accurate, interpretable clinical decision support. Our proposed methodology has been developed and evaluated using a large-scale electroencephalography (EEG) dataset of laser-evoked potential (LEPs) from 223 healthy subjects. Unlike previous studies that rely on subject-dependent validation or binary classification, our proposed framework is rigorously evaluated under a subject-independent multi-class setting. As a result, a clinically realistic assessment is produced.

Research Motivation

In this study, we aim to develop an explainable and trustworthy framework, developed as PainNet-ST model, to effectively extract and generalise neuro-physiological signatures of pain. EEG is considered highly sensitive to registering the complex functional interplay of the brain occurring in milliseconds. By leveraging high-density raw, non-invasive and cost-effective, potentially mobile EEG data, this model would objectively assess pain in real-time. This enables the transition from traditional self-reporting scale-based subjective pain assessment.

To the best of the authors' knowledge, this is among the first studies to engineer a decoupled Spatial-Temporal Convolutional Neural Network optimised using Optuna, which implements Bayesian optimisation through Tree-Structured Parzen Estimator (TPE) sampler, data augmentation, and directly using raw EEG signals. It is carefully validated using a StratifiedGroupKFold strategy to grade the pain into three levels: low, medium and high. To ensure clinical transparency and encourage trust, we have embedded a dual-layered explainable AI (XAI) architecture using SHAP and LIME.

This strategy allows us to identify both global neural biomarkers across populations and local trial-specific activations, e.g., N2-P2 ERP complex.

2. MATERIALS AND METHOD

In this study, the proposed model (*i.e.*, PainNet-ST) and comparative models (Deep Learning and Conventional approaches) were developed on a Windows 10 Pro system with an Intel Core i7- 9700 CPU at 3.00 GHz.

Python Software (Sanner, 1999) (version 3.10) was used to design the PainNet-ST model. The platform provides free access to open-source libraries like NumPy, pandas, scikit-learn (Pedregosa et al., 2011), matplotlib, seaborn, TensorFlow (Abadi et al., 2016), Keras (Ketkar, 2017) and Optuna, which support the advancement of most ML and DL models. Keras enables users to fully use its ease of setting up and operation, which includes quicker prototyping, user friendliness, a modular and extensive framework to integrate PainNet-ST as a real-time (i.e., web-based backend) running on both CPU and GPU (https://keras.io/)

To manage the computational demands, the modelling process was conducted in three stages. Initial Python development was performed using a web-based Jupyter Notebook (v7.0.8), followed by preliminary deep learning assessments on Google Colab to leverage cloud-based GPU acceleration. Finally, full-scale training and hyperparameter optimisation were transitioned to a High-Performance Computing (HPC) cluster. This provided the necessary resources including CPU, GPU, and memory to handle high-volume dataset and the computationally intensive Stratified Group K-Fold cross-validation protocol.

The prime objective of this study is to design and assess PainNet-ST model for three-class pain intensity classification (low, medium and high) and compare it against SVM, RF, STAM-Net, and 2D-CNN. Rigorous preprocessing of the data is performed to ensure data accessibility and integrity prior to training and testing sets.

2.1 Database Description

To build the PainNet-ST model, this study uses the OpenNeuro ds005280 dataset, one of the largest publicly available datasets of nociceptive-evoked EEG. It contains high-temporal EEG recordings from 223 healthy participants (Experiment 8). Further details on the public dataset are available at (Zhao et al., 2025). In principle, laser-evoked potentials (LEPs) were used to induce pain by delivering an infrared laser to the hand dorsum. The choice of this modality was to activate nociceptive-specific nerve endings. This ensures a clear N2-P2 complex for spatio-temporal analysis, which occurs between 150ms-400ms (0.15s-0.4s) after the laser hits the skin (Hu & Iannetti, 2019).

Table 1. Pain Class and Corresponding Laser Intensity Levels

Pain Class	Stimulus Method	Laser Intensity
Low	Laser-evoked (LEPs)	3.0 J
Medium	Laser-evoked (LEPs)	3.5 J
High	Laser-evoked (LEPs)	4.0 J

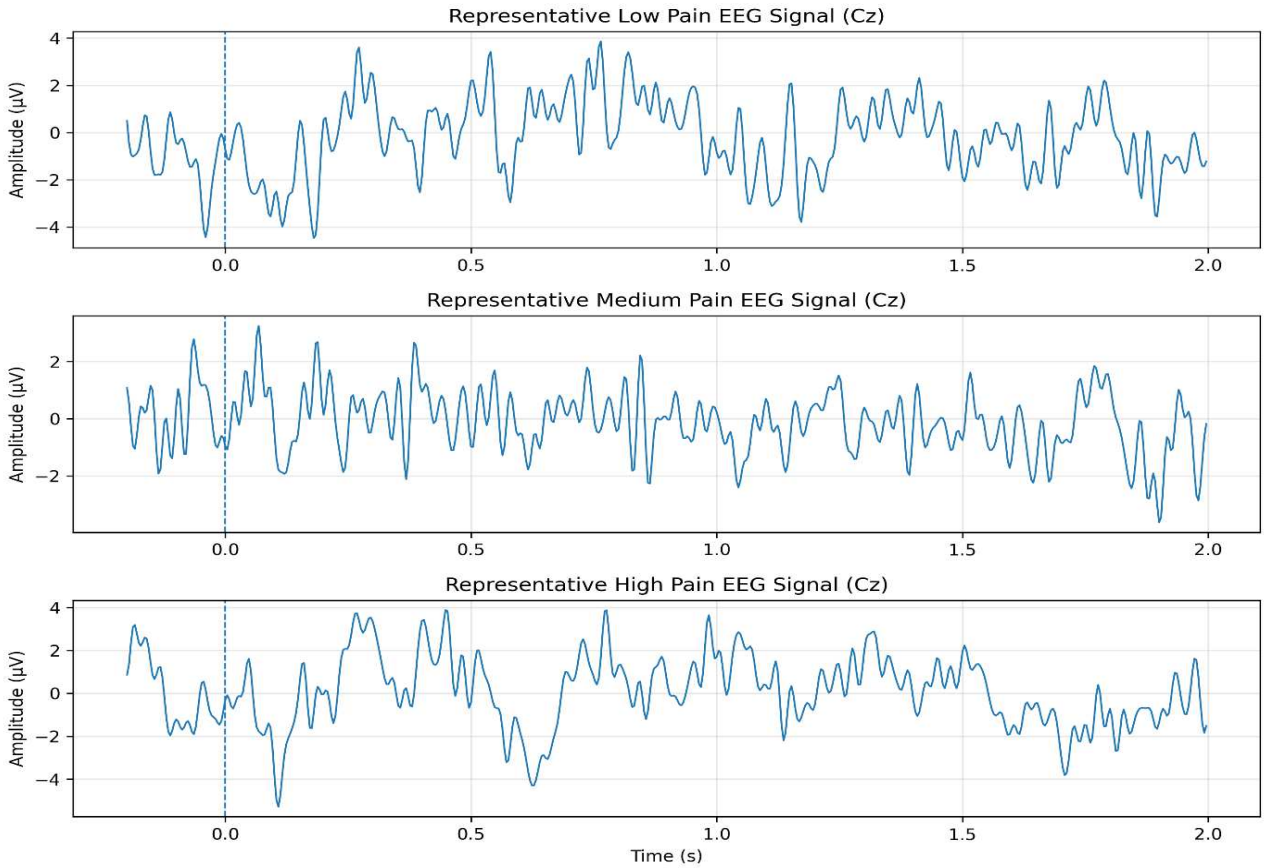


Fig. 2. Representative single-trial EEG signals at channel Cz for low, medium, and high pain classes.

Fig. 2 illustrates representative EEG signals selected from individual epochs that are closest to the class-wise mean waveform. Unlike averaged EEG signals, single-trial recordings exhibit high variability, making characteristic components such as the N2–P2 complex less distinguishable. In our proposed PainNet-ST model, subject-independent validation confirms that the model is evaluated on entirely unseen subjects, delivering a rigorous assessment of generalisation. In addition, the PainNet-ST modelling framework is compared against multiple machine learning and deep learning baselines to further support the robustness.

2.2 Data Preprocessing

Exploratory data analysis (EDA) was performed to better understand the BIDS-formatted dataset. Specifically, EDA involved verification of the availability and consistency of .vhdr, .eeg, .tsv files, inspection of event distributions across pain classes, assessment of signal length consistency, and identification of missing or corrupted recordings. Moreover, the presence of relevant central electrodes (e.g. Cz) for nociceptive processing was confirmed.

The dataset was pre-processed to standardize according to the Neurophysiological pipeline tailored for EEG-based deep learning. The preprocessing steps included channel selection to ensure only cortical activity was analysed, non-brain channels like EOG and ECG were filtered out first. To remove the DC offset and noise, a band-pass filter (1-40Hz) was applied to ensure that only the relevant EEG frequency bands were preserved.

For epoching, event markers from the .tsv files -were used to segment the continuous EEG into time-locked epochs ranging from -0.2 s to 2.0 s comparative to stimulus onset, which captures the nociceptive response window (N2-P2).

In addition, signals were downsampled using a decimation factor of 4 to reduce temporal dimensionality. To maintain inter-subject consistency and prevent data leakage, we applied Z-score normalisation within each fold of the subject-isolated standardization process. Following this normalisation, an Exponential Linear Unit (ELU) activation function was implemented to introduce non-linearity while allowing for negative outputs, which significantly improves gradient flow and overall training stability compared to standard ReLU.

Three augmentations were applied to the training data, including temporal shifting (rng. randint (-8,8), small amount of Gaussian Noise, and Magnitude Scaling of (+/-) 5%. Epoching was performed using a 2.2-second window, "tmin" = -0.2, "tmax" =2.0 and "baseline" = (-0.2 to 0). For more details, refer to (Zhao et al., 2025)

To ensure the clinical validity of this method, subject-independent Stratified Group 5-Fold Cross Validation was employed using StratifiedGroupKFold in this study. To prevent data leakage, we have used a subject-independent partitioning strategy, ensuring training and testing sets stay entirely distinct. No physiological data from one single individual was shared between training and testing. In each fold, where 80% of subjects were used for training and 20% for testing, while protecting the classification across subjects. This approach improves generalisation performance to unseen subjects.

2.3 Model Design

The proposed model PainNet-ST model is designed to capture both spatial and temporal characteristics of EEG signals for multi-class pain classification. The model architecture follows a decoupled spatial-temporal design, where independent extraction of channel-wise neural activity and temporal dynamics involved with nociceptive responses.

In principle, the proposed PainNet-ST model is based on a convolutional neural network (CNN) designed for high-dimensional, non-stationary EEG data. It employs a Spatial-Temporal decoupled convolutional strategy. The input of PainNet-ST consists of pre-processed EEG epochs represented as (channel x time), while preserving the electrode topology and temporal information within N2-P2 window. We have converted the 1D EEG signal into 2D matrix to resemble as a 2D image (Electrode x Time), where the rows are "Space" (electrode) and the columns are "Time".

The model architecture involves three main stages; Firstly, the spatial branch part of the model focuses on learning inter-channel relationships across the scalp model. The model architecture initially employs a (C,1) kernel to perform spatial transformation for channel-wise feature extraction. The first layer is (n_channels, 1), and this acts as a filter for the scalp. This spatial layer works by collapsing the 61 electrodes into 32 high-level feature maps, acting as a trainable spatial filter. This allows the model to focus on the signals from the midline vertex (Cz), where the N2-P2 complex is very prominent.

Secondly, the dynamic changes in EEG signals over time are captured by the temporal branch of the model. The following two Temporal Layers use a (1,15) kernel that identifies the oscillations of pain in the time-domain, such as N2 (negative deflection) - P2 (positive peak) complex, and then pure temporal filtering is performed using a (1, T) kernel to capture time-dependent patterns of the N2-P2. Following the spatial reduction, two sequential temporal convolutional blocks are applied to capture pain dynamics.

Spatial and temporal convolutions are decoupled purposely to reduce the parameter space. This prevents the model from learning subject-specific noise. This design improves the generalisation of the PainNet-ST model to unseen subjects.

To stabilise the training process, we substituted standard ReLU with ELU Activation, which allows negative activations and improves gradient flow, without neurons dying or losing features. Batch normalisation is applied after every convolutional layer to reduce internal covariate shift and speed up the convergence. We applied a high dropout (p=0.5) to tackle the high inter-subject variability inherent in EEG data. Finally, the feature representations are fused and passed through fully connected layers. The last dense layer flattens feature maps into a 128-dimensional vector and further reduced to 64 units. In order to classify pain intensities into low, medium and high, a SoftMax output later with three neurons was used. Along with the Cross-Entropy Loss function, Label Smoothing (0.1) was applied to prevent model from becoming overconfident in classification and to encourage better feature clustering.

Although the proposed PainNet-ST system is a robust model, it is highly dependent on hyperparameters, requiring proper tuning. We applied hyperparameter optimisation using Optuna, which implements Bayesian optimisation via the TPE sampler and is integrated with subject-isolated cross-validation. To stop any subject appearing in both training and testing, we implemented a StratifiedGroupKFold cross-validation (cv) strategy in this research. Given the heavy inter-subject variability in EEG signals, this strict separation is not just a preference but a necessity for clinical validity. Without this, the model would simply learn to identify individuals instead of nociceptive responses.

Once the model has been strategized, the next major step in ML and DL modelling is hyperparameter tuning. When hyperparameters are tuned efficiently, they give the optimal result for the model. In this study, hyperparameter optimisation was performed using Optuna, which implements a Tree-structured Parzen Estimator (TPE) sampler, a Bayesian optimisation-based approach for efficient exploration of the search space.

In this study, 50 optimisation trials were conducted using 5-fold Stratified Group Cross Validation. The search for optimal configurations within our high-dimensional parameter search space, including learning rate, dropout rate, and weight decay, is shown in Table 2. Unlike previous studies that rely on grid search or manual training, this study provides a computationally efficient and automated hyperparameter tuning for optimal configuration (Alba et al., 2007; Srinivas & Katarya, 2022).

Fig. 3 shows the model architecture of the PainNet-ST model's spatial-temporal layers, Bayesian optimisation (implemented via Optuna with the TPE sampler) and subject-isolated cross-validation. For Bayesian optimisation (Fig. 3(b)) and subject-isolated cross-validation (Fig. 3(c)), a schematic for each is provided in the appendix (Appendix B).

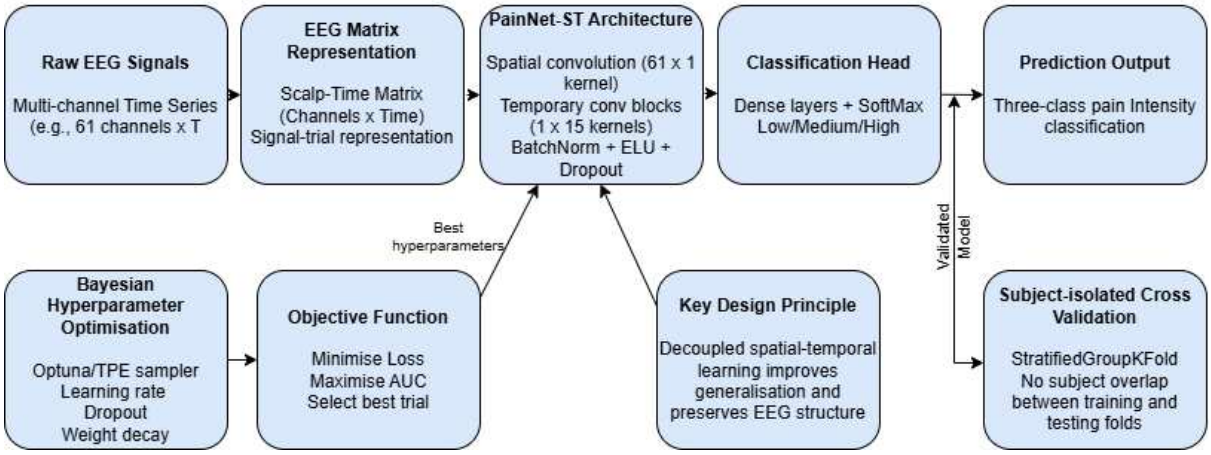


Fig. 3. The Model Architecture of the Proposed PainNet-ST model.

For this research work, 50 trials were conducted to hyperparameter-tune the proposed PainNet-ST model. The hyperparameter exploration space features are stated in Table 2. The table also shows the optimum hyperparameter for this study. Important hyperparameters for trial-based optimisation are shown in Fig. 4 and Fig. 5. The hyperparameter search space of the PainNet-ST model consisted of three continuous hyperparameters, which are related to optimisation and regularization. These hyperparameters are learning rate (1e-5 - 1e-2, log scale), dropout rate (0.1-0.7, uniform) and weight decay (1e-6 - 1e-3). 50 optimisation trials were conducted using 5-fold Stratified Group Cross Validation.

Table 2. Hyperparameter exploration space configured via Optuna and the optimal hyperparameter set for this research.

Hyperparameter	Search Space	Optimum Hyperparameter
Learning Rate	1e-5 - 1e-2 (log scale)	6.40e-4/ 0.00042
Dropout	0.1-0.7 (uniform)	0.5693
Weight Decay	1e-6 - 1e-3	2.21e-4
Batch Normalisation	Convolutional layer	True (Enabled)
Augmentation	Dynamic (70% probability)	Active
Trials	Training= 50	Trial #41 (Best)
Label Smoothing	F. cross_entropy= 0.1	0.1
Early stopping	15 Epochs	15
CV (StratifiedGroupKFold)	K=5 per trial	5-Fold
Epoch per trial	50	50
Batch Size	16, 32, 64, 128	32
Optimiser	Adam/AdamW	AdamW

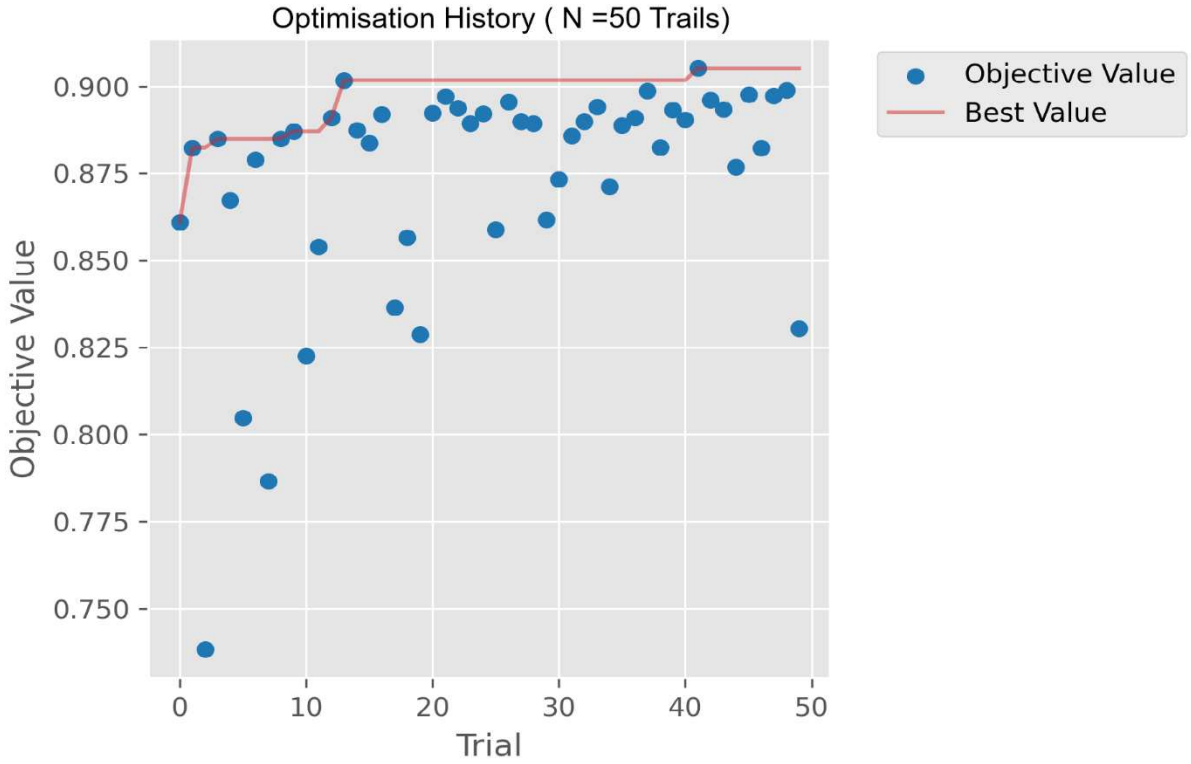


Fig. 4. Optimisation history plot of the 50 trials for the intended PainNet-ST model.

We first defined the optimal parameter set as the configuration that simultaneously minimised the loss function and maximised the Area Under the Curve (AUC), ensuring both high confidence and strong class discrimination. The chosen parameters are put in Table 2. To separate the performance

gains specifically attributable to Bayesian optimisation, we have maintained baseline configurations for the benchmark models (SVM, RF, STAM-Net, and 2D-CNN), while thoroughly tuning only the PainNet-ST framework. This helped to investigate the influence of Bayesian optimisation on PainNet-ST when experimented on the high-temporal ds005280 dataset.

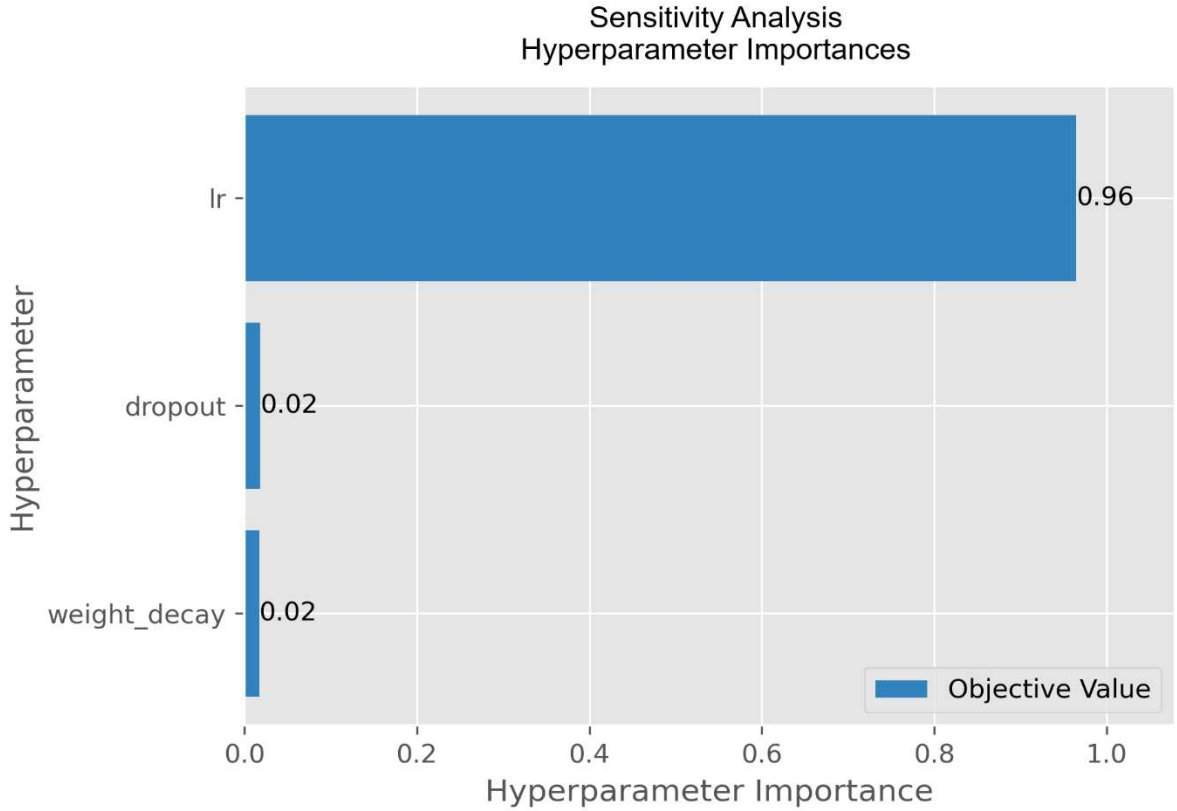


Fig. 5. Importance of hyperparameters to realize optimum performance for the proposed PainNet-ST model.

2.4 Theoretical Overview

For completeness, we provide a brief overview of the proposed PainNet-ST model in this section. Detailed theoretical explanations of benchmark classifiers and supporting methods are outside the scope of this section, as they are established in the literature. Interested readers are referred to SVM (Boser et al., 1992), RF(Liu et al., 2012), STAMNet (Feng et al., 2022) ,CNN (Lawhern et al., 2018), Bayesian optimisation (implemented via Optuna with the TPE sampler) (Akiba et al., 2019), and Explainable Artificial Intelligence (XAI) methods, containing SHAP and LIME (Khare & Acharya, 2023).

In principle, our proposed PainNet-ST model is a hybrid structure designed to secure both the spatial distributions and sequential dependencies inherent in EEG signals. With the integration of convolutional layers for discriminative spatial feature extraction with temporal blocks that learn sequential patterns, the hybrid nature of the model enables it to effectively represent the complex neural pattern involved in pain intensity, outperforming conventional single-architecture approaches.

Deep learning architectures commonly rely on feed-forward learning with backpropagation, which remains a widely used optimisation strategy for training such models (LeCun et al., 2015). In this study, PainNet-ST is evaluated against the aforementioned benchmark models to demonstrate its effectiveness in handling noisy, nonstationary EEG data and improving classification performance across three pain-intensity levels.

2.5 XAI implementation

Explainable Artificial Intelligence (XAI) was integrated to improve the interpretability of the the proposed PainNet-ST model at both global and local levels (Jui et al., 2026). Thus, firstly, the SHAP (Shapley Additive Explanations) method was used for global interpretability. Based on the cooperative game theory, the SHAP method aims to compute the feature importance. This highlights the contributions of EEG channels (spatial domain) and temporal regions (time domain) that influence pain classification. In this way, dominant neurophysiological patterns related to different pain intensities were identified.

Secondly, the LIME (Local Interpretable Model-Agnostic Explanations) method was used for local interpretability. LIME approximates the model behaviour around an individual prediction by fitting a surrogate linear model. In this work, LIME was applied to individual EEG trials, which helped to interpret how specific temporal segments and channel activations contribute to a single prediction outcome. It takes the EEG sample and fits a sample surrogate model to interpret how variation in input features affects the output.

Both SHAP and LIME were applied to the trained PainNet-ST model using the pre-processed EEG epochs as input, which helps in the identification of both population-level patterns and trial-specific decision mechanisms.

2.6 Model performance criteria

To evaluate the performance of the proposed PainNet-ST model, a total of 11 performance metrics were used. The extensive choice of evaluation metrics makes a good judgment of the model, as reliance on a single metric is not sufficient to comprehend the model's trustworthiness and reliability. The metrics include accuracy, precision, recall, specificity, F1 score, False Positive Rate (FPR), False Negative Rate (FNR), Negative Prediction Value (NPV), False Discovery Rate (FDR), Cohen's Kappa (k), and area under the curve (AUC).

2.6.1 Reasoning for Metric Selection

Clinically, accuracy determines the overall correctness. However, relying solely on accuracy can be misleading, as EEG signals contain noise and may be subject to class imbalance. So, ROC-AUC is calculated, making it very useful in clinical diagnostics (Mandrekar, 2010). To further account for class imbalance, Cohen's Kappa (k) and F1 score metrics are calculated. Cohen's Kappa helps measure the agreement between predicted and actual labels (Viera & Garrett, 2005). The F1 score ensures that the model can perform equally on the low, medium and high classes of pain.

Recall (sensitivity), is critical for identifying high pain cases, while a low value False Negative Rate (FNR), justifies the clinical safety as it quantifies the risk of getting less treated or overlooked in pain classification. Negative Prediction Value (NPV) can effectively build confidence in clinicians where "Low pain" classification by the model means the patient may not require intervention immediately.

As EEG is vulnerable to artefacts (eye blinks, muscle movements), a high specificity underscores that the model can accurately distinguish the non-pain classes. False Positive Rate (FPR) is also considered to help reduce the unnecessary intervention and false alarms for the model in classifying pain. As all of the above evaluations come with both their benefits and flaws, a combined use reveals detailed information to the clinicians (Botchkarev, 2019; Naser & Alavi, 2023). This provides a more comprehensive clinical evaluation.

2.6.2 Statement of Model Reproducibility

A global random seed was settled (Seed=42) across all libraries, containing PyTorch, NumPy, and Optuna. We have also used a deterministic CuDNN backend to eliminate stochastic variance. We

used Stratified Group 5-Fold Cross-Validation. This is to ensure subject-independent performance. Final model weight has been preserved as a digital artefact and hyperparameter optimisation is also preserved to allow for independent verification. This would allow any other researcher to replicate the model and generate the same results.

3. RESULTS AND DISCUSSION

3.1 Necessity of a deep learning model

To create the robustness of the PainNet-ST model, an explanation of the experimental outcomes of the trials carried out in the experiment, and the assessment of the performance of all the machine learning and deep learning models to classify a 3-class metric for pain severity (low, medium and high) are described. Patterns for EEG-based pain intensity evaluation, comparison of proposed PainNet-ST model with baseline methods grounded on hand-crafted features, including SVM, RF, 2D-CNN, and STAMNET, was performed. Below are a few parameter backgrounds of these baseline methods:

Support Vector Machine (SVM): Before classification, EEG features are standardized and reduced using PCA (200 components). RBF kernel with C=1.0 and γ=scale, with balanced class weights. Evaluation is performed using 5-fold Stratified Group K-Fold cross validation.

Random Forest (RF): Implemented after standardization and PCA, using 400 decision trees with balanced class weights. Evaluation is performed using 5-fold Stratified Group K-Fold cross validation.

2D-CNN: The model learns spatial-temporal EEG patterns using spatial and temporal convolutional layers, supported by max-pooling and entirely linked layers with dropout. The learning rate is optimised using Optuna over the range 0.0001- 0.01. The final classification is performed using a SoftMax layer.

STAM-NET: It extends CNNs by incorporating spatial attention mechanisms to weigh EEG channels before convolutional feature extraction. Optimised using Optuna. Learning rate (lr)= 0.001, dropout=0.429, hidden units=46, epoch=30, trials=20. Evaluation is performed using 5-fold Stratified Group K-Fold cross validation.

Table 3. Classification evaluation comparison of the anticipated PainNet-ST model against the benchmarked methods in relation to accuracy, F1 score for the ds005280 dataset.

Model	Accuracy (%)	F1 Score (%)
SVM	71.36	71.31
RF	71.70	71.77
STAM-Net	53.63	48.11
2D-CNN	76.58	76.45
PainNet-ST	**92.28**	**92.29**

Table 4. Classification evaluation comparison of the anticipated PainNet-ST model against the benchmarked methods in relationship to accuracy, F1 score for the ds005280 dataset.

Model	Precision (%)	Recall (%)	Specificity (%)
SVM	71.41	71.36	85.68
RF	71.75	71.70	85.85
STAM-Net	67.72	53.63	76.81

2D-CNN	76.76	76.59	88.29
PainNet-ST	**92.35**	**92.28**	**96.14**

The proposed PainNet-ST model consistently outperforms all baseline models across the evaluation metrics. The most significant improvements are observed in accuracy (+15.7%) and specificity (+15.8%), indicating improved discrimination capability of the model. This advancement is attributed to the incorporation of spatial-temporal feature extraction and Bayesian optimisation.

To enable a rigorous evaluation of the proposed PainNet-ST model against its comparative counterpart models, multiple performance metrics are exploited. The outcomes recorded in Tables 2 and 3 were used for evaluation measures to compare with benchmarked models. These results present the evaluation metrics for the ds005280 dataset. Accuracy, precision, recall and specificity for PainNet-ST are 0.9228, 0.9235, 0.9228, 0.9229, respectively. Among all the models evaluated, these are the highest scoring. The PainNet-ST demonstrates a 15.7% upsurge in accuracy, a 15.7% upsurge in precision, a 15.69% increase in recall, a 15.84% upsurge in specificity, and a 7.85% upsurge in F1 over the competitive 2D CNN. This advancement can be attributed to the integration of Bayesian optimisation (implemented via Optuna with the TPE sampler) in improving pain intensity classification.

To help investigate the proposed model's robustness, Fig. 7 reveals records of four additional suitable metrics: FPR, FNR, NPV, and FDR. The FPR measure is employed to identify the erroneously classified positive class, whereas FNR is used to find out the erroneously predicted negative class variable. For an optimal model, lower FNR and FPR are preferred. Higher NPV is favoured for a best model, while a lower value of FDR is favoured during model contrast (Joseph et al., 2022).

3.2 Differentiation with deep learning models

Many studies have applied deep learning methods for EEG- based spatio-temporal learning to address the hand-crafted (manual extraction) of features. These approaches are superior and are a key strength of deep learning. This motivates the development of PainNet-ST for automated pain intensity classification. To validate the benefits of PainNet-ST, we have compared PainNet-ST with many SOTA (state-of-the-art) deep learning methods for EEG-based pain categorization.

Table 5 exhibits the classification performance of PainNet-ST and other DL models. Jiao et al. (2018) mentioned in a study that Convolutional Neural Network, by learning hierarchical features, maximizes diagnostic accuracy. By extracting the most informative features, the model remains robust on individual profiles after being trained on a multi-subject dataset. While previous study by Hefron et al. (2017) demonstrated that spectral and temporal dependencies could be captured using RNN with LSTM layers, our findings suggest that a hybrid CNN approach is superior for linking spatial and temporal information (Bashivan et al., 2015). In another study, Wu et al. (2022) employed Attention-Recurrent Neural Networks (ARCNN) to learn spatial-spectral-temporal representations from EEG signals. Within the PainNet-ST framework, the spatial block works on encoding the channels distribution, while the two temporal layers, refined via Bayesian optimisation, are capable of capturing the spectral-temporal dynamics of the EEG signals. Ultimately, this leads to superior classification performance. This can be attributed to the combined spatial and temporal feature learning with PainNet-ST.

Table 5 features the support of PainNet-ST to the discipline of EEG-based pain severity classification. High performance (accuracy, F1 score) attained in multi-class and subject-independent situations clearly showcases the high generalization power of PainNet-ST to completely unseen subjects or individuals, making it a reliable tool in clinical practice (Wu et al., 2022). The proposed framework

can work as a clinical device in pain treatment, supporting precise and timely identification of cases experiencing from different levels of pain.

Table 5. Classification evaluation contrast with deep learning methods

Model	Accuracy (%)	F1 Score (%)	Ref
CNN	87.80	87.25	(Jiao et al., 2018)
RNN	87.12	86.81	(Hefron et al., 2017)
CNN	86.52	86.52	(Al-Nafjan et al., 2025)
RNN	86.71	86.71	(Al-Nafjan et al., 2025)
CNN-RNN	90.47	89.73	(Bashivan et al., 2015)
ARCNN	92.14	92.11	(Wu et al., 2022)
PainNet-ST	**92.28**	**92.29**	**This study**

The fluctuation in the accuracy across models was primarily due to inter-subject differences, which is the initial encounter to realizing subject-independent classification. As shown in Table 5, PainNet-ST demonstrates improved generalisation performance, which can be attributed to its spatial-temporal decoupled structure. Hence, the proposed PainNet-ST is competent of learning the invariant EEG signal through different subjects and can generalise it to new subjects.

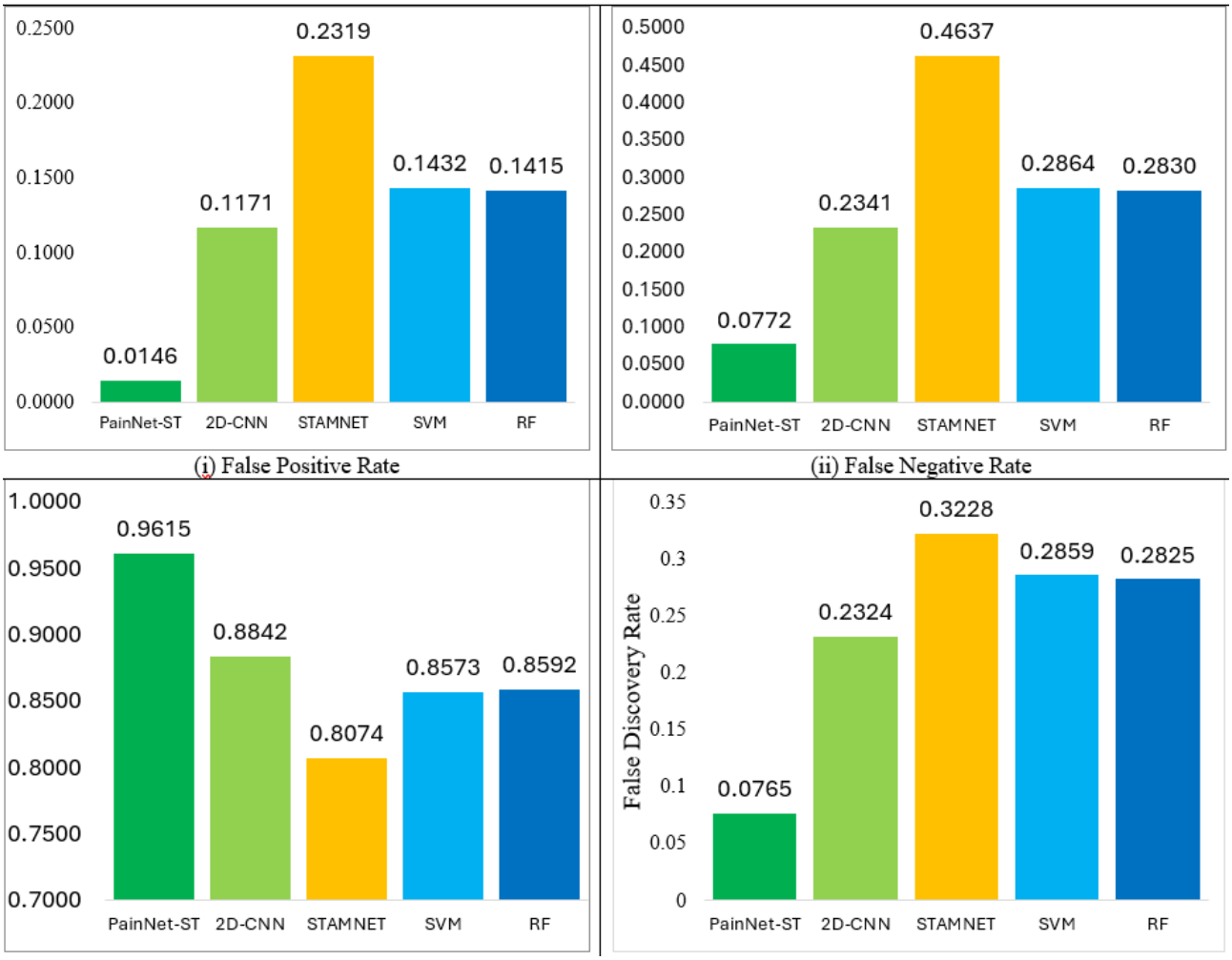


Fig. 6. Summary of bar plots for (i) False Positive Rate (FPR), (ii) False Negative Rate (FNR), (iii) Negative Prediction Value (NPV), (iv) False Discovery Rate (FDR) for dataset ds005280 (Exp. 8).

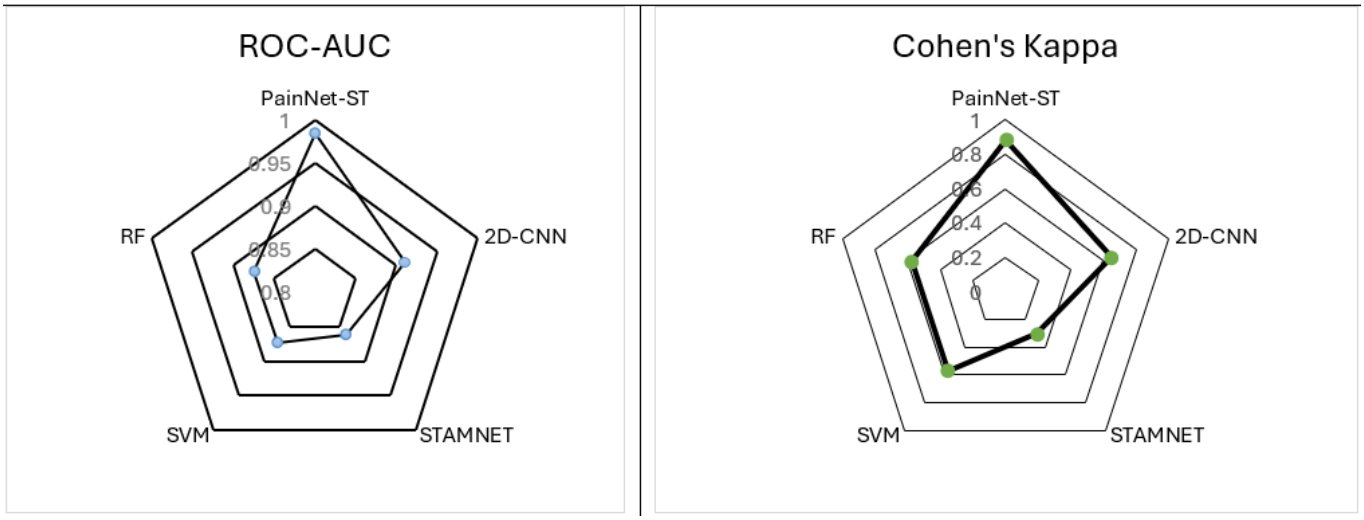


Fig. 7(i) Radar plots of area under the receiver operating characteristic curve (ROC-AUC), and 7(ii) Radar plots of Cohen's Kappa evaluation for ds005280 dataset.

3.3 Multi class classification limitations

Pain is subjective and varies from person to person on the tolerance level. These differences in the perception of pain create a divergence when pain is subjectively determined during LEPs. Discrepancy in pain increases as the pain class increases. One could subjectively categorise a pain as "low", while others may classify it as "high". This deviation may potentially cloud the classifier, resulting in an error rate. Therefore, Fig. 8 shows the Confusion matrix for three-class pain classification (low, medium, high) on completely held-out test data in this work. It achieved high accuracy across all the categories, reflecting the model's strength in capturing objective neural signatures of pain. The diagonal elements show high true positive (1226 for Low, 1218 for Medium, and 1246 for High). It can be observed that the decoupled spatial-temporal architecture of PainNet-ST has captured the oscillatory patterns.

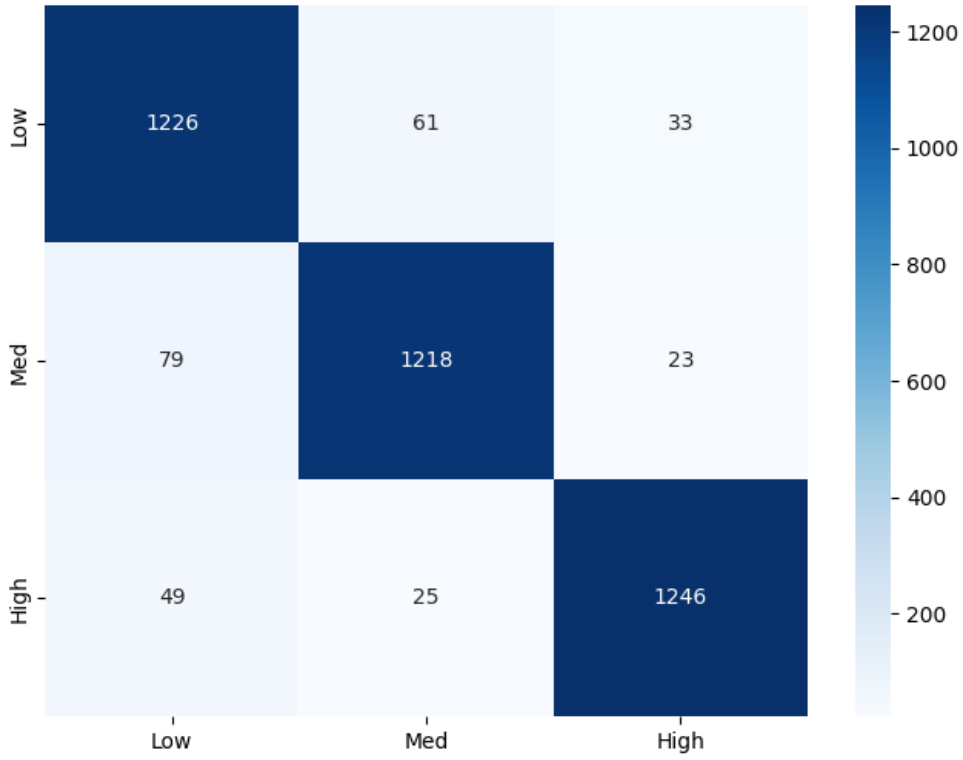


Fig. 8. Confusion matrix for three-class pain classification.

Misclassifications are primarily confined to adjacent intensity levels (e.g., Low to High). This behaviour is expected due to the overlap in neural representation between adjacent pain intensity levels (e.g. Low to High). This can be attributed to the gradual transition in pain perception across intensity levels. The high sensitivity for the High class (1278 out of 1320 total actual High samples), reflects high precision in identifying High pain class. This reduces the risk of undetected patients in an objective assessment scenario. And low confusion between the Low and High (33 out of 49). It has been observed that PainNet-ST has achieved strong performance in brain regions related to pain. Our PainNet-ST has recognized an effective sequence of brain regions, channel labels, and channel-time pairs that yielded the highest performance.

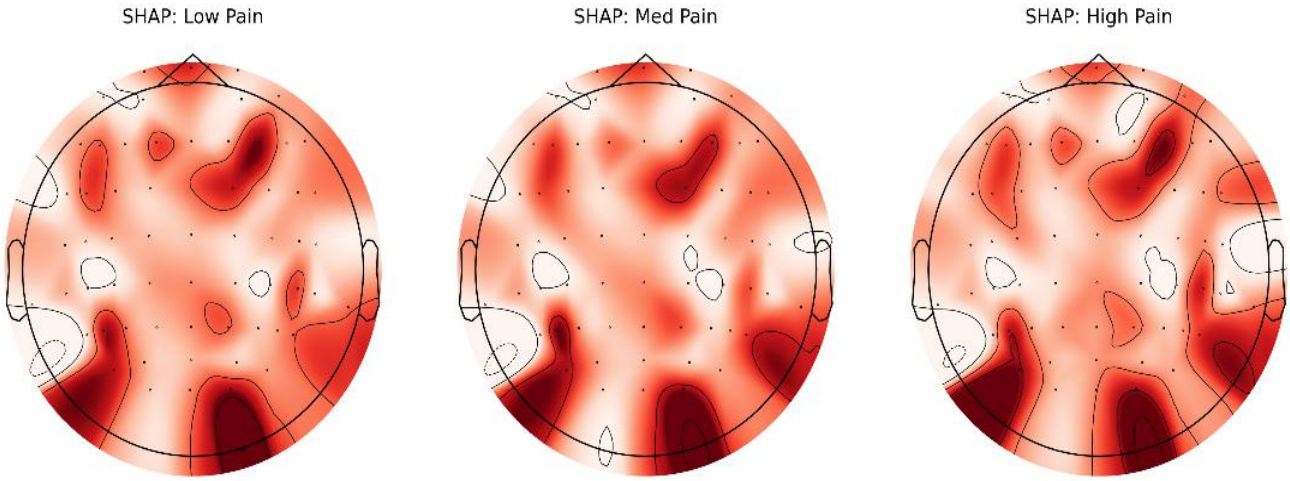


Fig. 9. The SHAP topographic map showing the global feature importance.

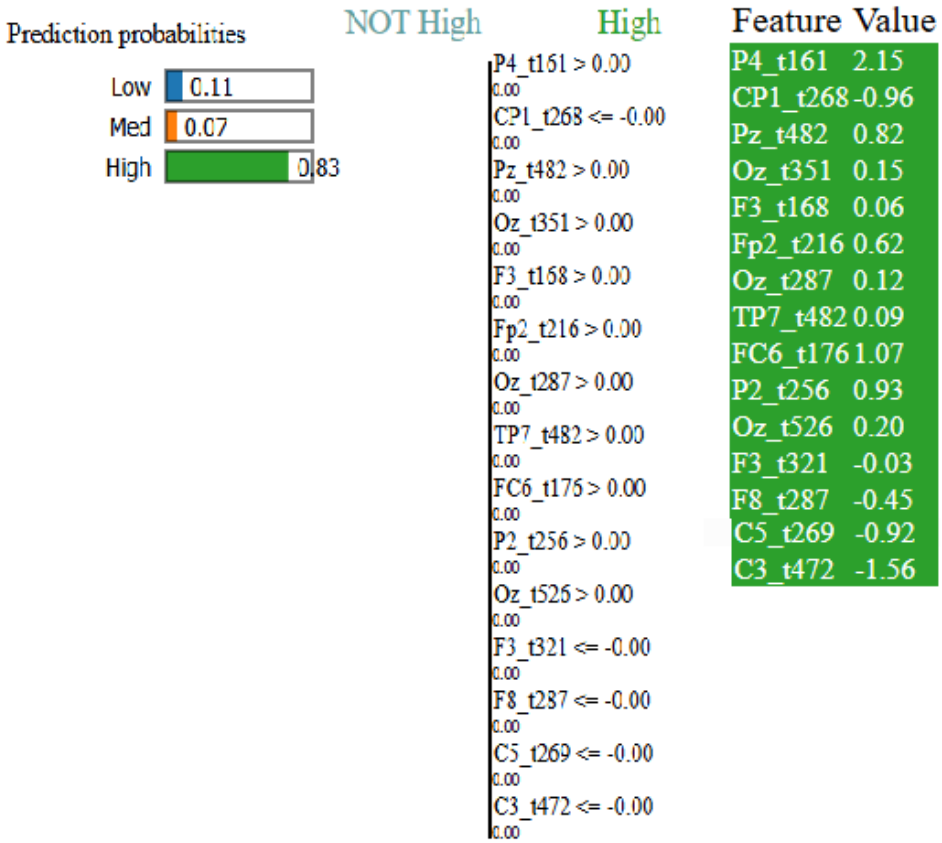


Fig. 10. The LIME explanation approach applied for high pain prediction.

LIME report identifies posterior parietal (P4, P2, Pz), central (CP1, FC6, CP5), and occipital (Oz) regions, which is an indication of somatosensory and higher-order sensory integration. The dense clustering of features across distinct time points (e.g., t161, t168, t216, t287) suggests that PainNet-ST captured complex spatio-temporal dynamics rather than relying only on a single late-latency peak.

These feature clusters correspond to electrode positions situated directly over the Posterior Parietal Cortex and Primary Somatosensory Cortex. By picking features at specific time indices (like t161 and t216), our PainNet-ST is effectively tuning in to the N2-P2 ERP complex. The strength of the model PainNet-ST is evident by its accuracy of 92.28% and its selective focus on spatio-temporal features in the centro-parietal regions (P4, CP1, Pz). These regions correspond to the primary somatosensory cortex and parietal integration zones. Consistent with known pain-related brain regions.

3.4 Contrast with earlier research for pain intensity evaluation

In medical environments, a model trained under subject-independent conditions should generalise to completely unseen subjects. This generalization is crucial in practical applications. However, as the extent of pain classes increases in a multi-class scenario, the precision of the pain state recognition diminishes. This introduces a challenge, but it is necessary for precision.

Table 6 relates PainNet-ST to the state-of-the-art methods for pain severity classification. Lovelace et al. (2024) have used Elastic Net as primary ML to extract Spectral power features (alpha, beta bands) of EEG to find out chronic pain, Binary classification (Pain vs No Pain), implementing Train/Test (70% of data) and Hold Out (30%), achieving an accuracy of 79.6%.

It is imperative to mention that Al-Nafjan et al. (2025) have employed convolutional neural networks and recurrent neural networks to categorise wavelet features obtained via time-frequency domain exploration of EEG using a BCI framework. They have classified pain intensities in low, moderate and high, achieving an accuracy of 86.71%. (Mari et al., 2025) have trained a Random Forest model with cortical event-related potentials (ERPs) used cross-validation and achieved an accuracy of 78% on binary classification of pain (Pain vs non-Pain). In another study, Imai et al. (2025) have employed random forest to classify pain intensity using frequency power spectrum into 2-, 3-, and 5-level categories.

A subject-wise 5-fold cross-validation was employed to assess the model. (Chen et al., 2022) have employed Ensemble ML for voting and have obtained time-domain, frequency-domain, and non-linear features to train several classifiers for a 3-level chronic pain classification. Subject-wise, nested CVs. The cross-validated model performed on a held-out test and has achieved an accuracy of 78%.

Table 6: Comparison of EEG-based pain classification methods.

Study	Dataset/Subjects	Task Type	Features	Method	Validation Strategy	Limitations	Performance
(Lovelace et al., 2024)	Chronic pain EEG	Binary (Pain vs No Pain	Spectral Power	Elastic Net (ML)	Train (70%), Test (30%)	Limited dataset size and variability	Acc:79.6%, Sens:82.2%, Spec: 66.7%
(Al-Nafjan et al., 2025)	EEG (BCI framework)	3-Level	Wavelet (time-frequency)	CNN and RNN	Train (80%), Test (20%)	Need hand-crafted feature extraction.	Acc:86.71%
(Imai et al., 2025)	30-channel EEG	Multi-Level	PSD features	XGBoost	Subject-wise 5-fold	Performance diminishes with higher classes	82% (2-level), 60% (3-level), 40% (5-level)

(Chen et al., 2025)	Dual-channel EEG	3-level	Band power (θ, α)	Ensemble ML	Subject-independent nested cross-validation with hold-out test	Limited spatial resolution	Acc: 78%
(Witjes et al., 2025)	EEG brain signals	Binary	PSD features	ML classifier	LOSO	Poor correlation with subjective pain	Acc:76%, Sens:80%, Spec:72%
(Mathra wala et al., 2025)	EEG (healthy subjects	Binary	Time + spectral features	SVM	Training 80%, Testing 20%	Limited to binary classification	Acc: 94.2%
(Mari et al., 2025)	EEG dataset	Binary	Band power	RF	Cross-validation (stratified k-fold)	Poor generalisability	Acc: 76%
Present Study	**ds005280 (LEP EEG)**	**3-level**	**Spatial-temporal EEG features**	**PainNet-ST**	**Subject-independent (SGKF)**	----	**Acc: 92.28%, Prec:0.9235, F1:0.9229**

Note: *Direct comparison across studies have been carefully interpreted as there are differences in datasets, feature extraction methods, and validation strategies. Many studies have used subject-dependent validation or focused on binary classification tasks, and this may have led to optimistic performance estimates. In comparison, our proposed PainNet-ST model is evaluated using subject-independent StratifiedGroupKFold cross-validation, a rigorous and clinically realistic assessment of generalisation performance was observed.*

The comparison in Table 6 shows that several studies have reported competitive accuracy, and many are limited to binary classification or rely on subject-dependent validation. A key reason for this is the difficulty in modelling inter-subject variability in EEG signals. In contrast, the proposed PainNet-ST model achieves strong performance (92.28% accuracy) under a more challenging subject-independent evaluation setting. This improved execution can be attributed to the spatial-temporal feature learning and Bayesian optimisation, which improve generalisation across unseen subjects. Unlike machine learning approaches, which depend on hand-crafted aspects, the proposed PainNet-ST framework automatically learns discriminative representations directly from raw EEG signals. This learning from the raw EEG signals makes the PainNet-ST model appropriate for real-world clinical applications.

3.5 Application of the proposed framework

Implementation of the proposed model alongside explainability remains a main challenge for commercial adoption. Commercialization of these models must go through rigorous testing on the broader capabilities of model's explainability. The capacity of the model should be in correctly explaining which characteristics are most important and any biases, if they will deter the model's performance. This makes the model decision fair and accurate (Escalante et al., 2018). In this study, we have described the global features, which brain regions (Spatial) and frequency (Temporal) are most important using SHapley Additive explanations (SHAP), and local (which single trial for a patient) features using local interpretable model-agnostic explanation (LIME) XAI models. Both global and local importance are vital to establishing trustworthiness in models (Jui et al., 2026).

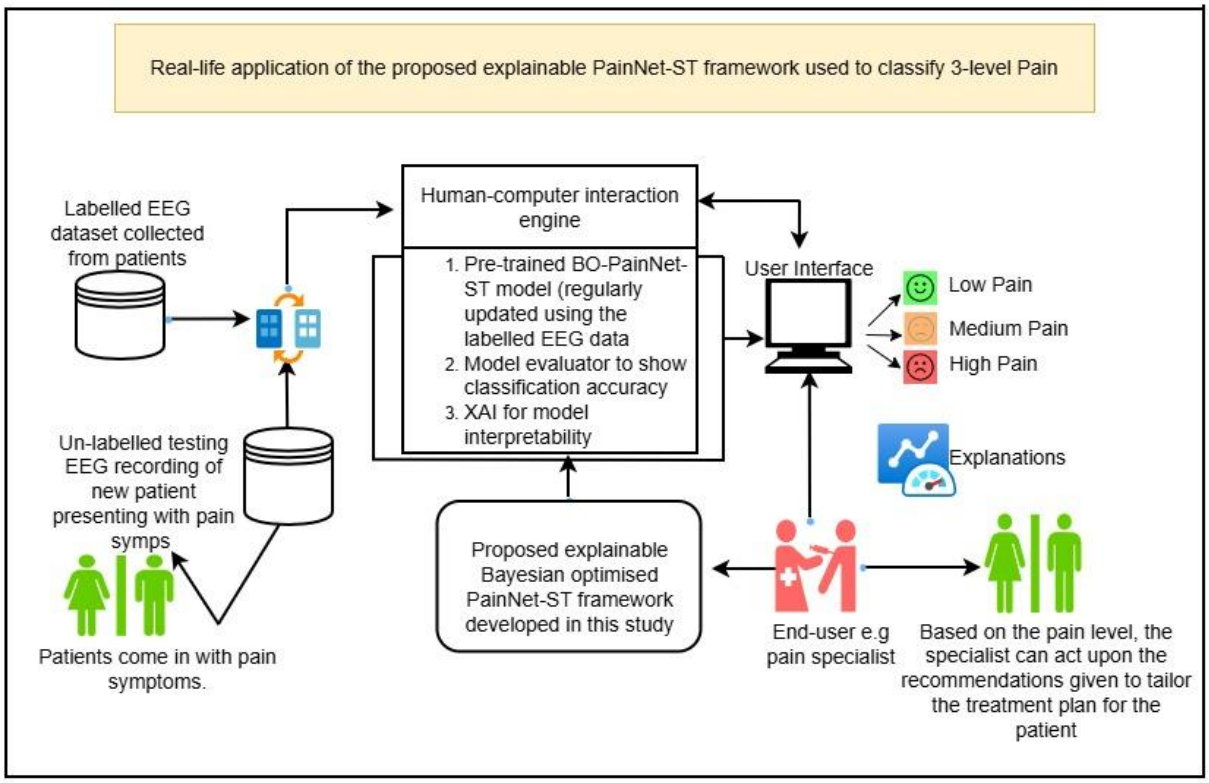


Fig. 12. Real-life implementation of the intended explainable PainNet-ST framework used to classify 3-level pain.

A real-life implementation of the intended explainable framework is illustrated in Fig. 12. If appropriately implemented, it would help quickly identify patients with pain, save time and cost, and thus aid in making the required treatment plan. Such implementation could be achieved by collecting EEG data with wearable devices. The collected data can then be fed into the model for precise pain classification. Real-time data is collected from patients, and clinicians can analyse and see the level of pain immediately. The implementation of such low-cost and portable diagnostics tools offers significant potential for resource-constrained environments, providing a scalable solution for objective pain assessment in developing nations.

When the patient's EEG data are ready via wearable devices, the collected EEG signals are fed to the pre-trained model as unlabelled data. Then the model would classify the patient's pain intensity, showing accuracy along with other evaluation metrics. Both the local and global descriptions would be projected on the user interface to increase the model interpretability, reliance, and trust. Based on the information, a clinician can devise a likely medication and suggestions for the patient.

4. Limitations of the study

This study has several limitations that should be taken into account when interpreting the results.

- Dependency on a single dataset may limit the generalisability of the results to other EEG-based datasets due to varied acquisition conditions and subject populations. Future work will assess the proposed framework across different independent EEG datasets for robustness and generalisability.
- Real-world utility of the proposed framework for patients with chronic pain or complex neurological disorders still remains an essential area for future exploration.
- Despite extensive preprocessing, residual artifacts from muscle activity and eye movement can persist within high-dimensional EEG data. Achieving universal generalisability for the

proposed framework remains challenging within this noise, coupled with inter-subject variability. In addition, the PainNet-ST incorporates multiple components that increase computational complexity and limit real-time deployment. These components include spatial-temporal feature extraction, Bayesian optimisation and explainable AI.

- Pain a multi-modal experience. This study primarily focuses on EEG signals. Future research can also consider incorporating additional physiological and behavioural signals (speech signal and facial image). The explainable AI techniques used, including SHAP and LIME, provide approximate explanations. These may not fully correspond to the model's underlying decision-making process.
- Distinguishing between low, medium and high pain intensities remains problematic because neural patterns often overlap, and pain thresholds are subjective among individuals. To improve classification accuracy around this challenging task, future research may explore personalised or adaptive models.
- An ablation study was not led, as evaluating overall model performance and generalisability under subject-independent conditions was the primary objective of this work. The comparative evaluation across multiple baseline models provides indirect evidence of the competence of the proposed framework. Future work will include organized ablation experiments to quantify the involvement of individual components.

5. Conclusions

This study contributes to ongoing progress in implementing artificial intelligence methods for pain assessment (Al-Nafjan et al., 2025; Chen et al., 2025; Imai et al., 2025; Mari et al., 2025; Witjes et al., 2025), has proposed an explainable model, denoted as PainNet-ST, which assimilates a deep learning Convolutional Neural Network (CNN) tuned using Bayesian optimisation implemented via Optuna, and is trained and evaluated on a single dataset. The PainNet-ST has been evaluated using subject-independent predictions of three-class pain intensity levels (low, medium, and high) derived from laser-evoked EEG signals. For the model's input, it is the pre-processed, time-locked epochs capturing nociceptive responses within the N2-P2 temporal window.

The proposed Pain-ST framework uses a spatial-temporal decoupled convolutional architecture to enhance the accuracy and robustness. These spatial filters capture channel-wise neural distributions and temporal filters model the dynamic evolution of pain-related brain activity. To improve model generalisation across unseen subjects, Bayesian optimisation is integrated to find optimal hyperparameter configurations. The investigation reveal that the proposed PainNet-ST model outperforms conventional machine learning approaches (e.g., SVM, RF) and baseline deep learning models (e.g. 2D-CNN, STAM-Net) across multiple evaluation metrics, including accuracy and precision. To provide meaningful insights into the neurophysiological patterns underlying model predictions, explainable artificial intelligence has also been integrated.

From a clinical perspective, a low False Negative Rate, which contributes to the high sensitivity and reliability of PainNet-ST suggests significant potential for automated pain classification and monitoring in environments where patients cannot express their pain, such as Intensive Care Units (ICUs). By a subject-independent evaluation using StratifiedGroupKFold cross-validation, PainNet-ST maintains consistent performance and confirms that no subject data is shared between training and testing sets. The strong performance across all the evaluation metrics shows the models' ability to generalise effectively in unseen subjects

A high Negative Predictive Value (96.15%) indicates the strong predictive power of PainNet-ST compared with all baseline models. The False Discovery Rate (7.65%) remains significantly lower

than competing approaches. This confirms that the proposed model PainNet-ST produces fewer incorrect positive predictions.

Integration of Explainable Artificial Intelligence (XAI) methods, such as SHAP and LIME, enhances model interpretability by identifying the EEG channels and temporal regions that contribute to classification. These explanations support the neurophysiological markers, including the N2-P2 complex.

For future research, EEG-based pain assessment will focus on validating the architecture against spontaneous and chronic pain signatures to expand its diagnostic scope beyond Laser-Evoked Potentials (LEPs). Work is currently underway to extend this methodology into a multi-modal framework that fuses EEG signals with behavioural modalities such as speech and facial expression, enhancing the robustness of automated pain classification in real clinical settings.

Credit authorship contribution statement

Nafizul Islam Chowdhury: Writing- review & editing, Resources, Project administration, Methodology, Conceptualisation. **Ravinesh C. Deo:** Writing- review & editing, Supervision, Resources, Project administration. **Rajendra Acharya:** Writing- review & editing, Supervision, Resources, Project administration. **Prabal Datta Barua:** Writing- review & editing, Resources, Project administration, Funding acquisition. **Aruna Devi:** Writing- review & editing. **Salvin Prasad:** Writing- review & editing. **Jeffrey Soar:** Writing- review & editing, Resources, Project administration.

Declaration of competing interest

The authors declare that there are no financial or personal relationships that could have influenced the work presented in this article.

Funding Declaration

This research is supported by the University of Southern Queensland International PhD Fee Scholarship (2024-2027) and an International PhD Stipend Scholarship from Cogninet Australia Pty Ltd, under the Industry Research Scholarship Agreement. The authors also thank the reviewers and editors for their valuable feedback on the initial manuscript.

Data availability

Supplementary data for manuscript is available at (Zhao et al. (2024). Laser-evoked potentials in 223 healthy subjects. OpenNeuro. doi:10.18112/openneuro.ds005280.v1.0.0)

Ethical statement

The authors confirm that only public datasets were acquired and used in accordance with relevant laws and institutional guidelines and have been approved by institutional committee under Ethics Approval ETH2024–1025. No primary data were collected from the human subjects, and no personal identifiers were available to the project.

Appendix A.

List of acronyms (Table A.1)

List of acronyms

Topic	Author Supplied keywords	Topic	Author Supplied keywords
AI	Artificial Intelligence	LEPs	Laser-Evoked Potentials
ARCNN	Attention-Recurrent Convolutional Network	LIME	Local Interpretable Model-Agnostic Explanations
AUC	Area Under the Curve	ML	Machine Learning
BO	Bayesian optimisation	NRS	Numerical Number Scale
BIDS	Brain Imaging Data Structure	NPV	Negative Predictive Value
CNN	Convolutional Neural Network	Optuna	Hyperparameter Optimisation Framework
CPU	Central Processing Unit	PainNet-ST	Convolutional Neural Network-Spatial Temporal
CV	Cross Validation	PCA	Principal Component Analysis
DL	Deep Learning	RF	Random Forest
EDA	Exploratory Data Analysis	RNN	Recurrent Neural Network
EEG	Electroencephalography	ROC	Receiver Operating Characteristic
ELU	Exponential Linear Unit	SHAP	SHapley Additive explanations
ERP	Event-Related Potential	STAM-Net	Spatial-Temporal Attention Module Network
FDR	False Discovery Rate	SVM	Support Vector Machine
FNR	False Negative Rate	TPR	True Positive Rate
FPR	False Positive Rate	TPE	Tree-structured Parzen Estimator
GPU	Graphics Processing Unit	VAS	Visual Analog Scale
HPC	High Performance Computing	XAI	Explainable Artificial Intelligence
IASP	International Association for the Study of Pain		

Appendix B.

Schematics of Bayesian optimisation (implemented via Optuna with the TPE sampler) and subject-isolated cross-validation (Table B.1)

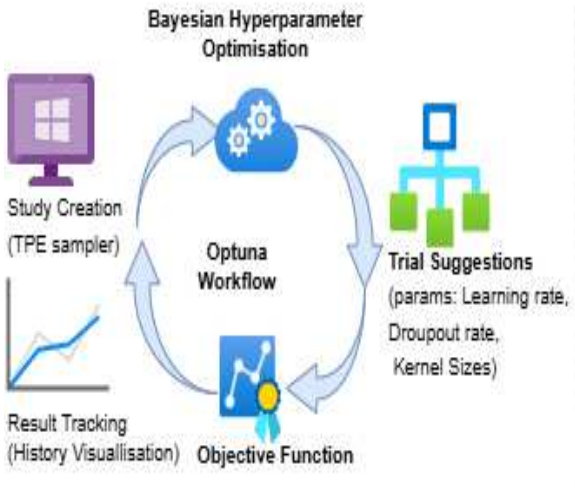


Fig. 3(b) Bayesian Hyperparameter Optimisation of PainNet-ST shows the trial suggestions for the objective function.

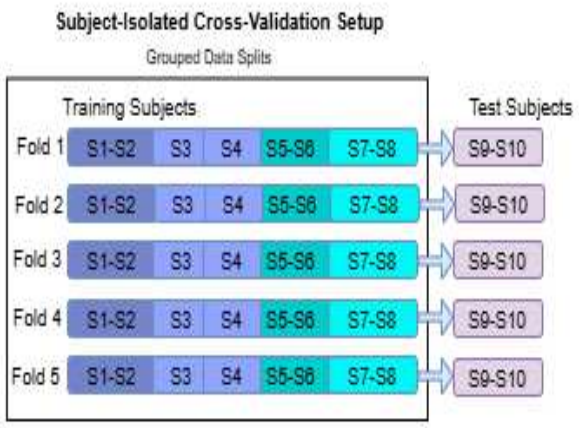


Fig. 3(c) Subject-isolated Cross-Validation PainNet-ST shows no subject is appearing in both training and testing.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., & Isard, M. (2016). {TensorFlow}: a system for {Large-Scale} machine learning. 12th USENIX symposium on operating systems design and implementation (OSDI 16),

Abbasimehr, H., & Paki, R. (2021). Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization. *Chaos, Solitons & Fractals*, *142*, 110511.

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining,

Al-Nafjan, A., Alshehri, H., & Aldayel, M. (2025). Objective pain assessment using deep learning through EEG-based brain–computer interfaces. *Biology*, *14*(2), 210.

Alba, E., Luque, G., Garcia-Nieto, J., Ordonez, G., & Leguizamon, G. (2007). MALLBA: a software library to design efficient optimisation algorithms. *International Journal of Innovative Computing and Applications*, *1*(1), 74-85.

Alshehri, H., Al-Nafjan, A., & Aldayel, M. (2025). Decoding pain: A comprehensive review of computational intelligence methods in electroencephalography-based brain–Computer interfaces. *Diagnostics*, *15*(3), 300.

Australia, C. P. (2024). *2024 National Pain Report: The social and economic cost of chronic pain*.

Australian Institute of Health and Welfare. (2024). *Chronic pain in Australia*. AIHW. https://www.aihw.gov.au/reports/chronic-disease/chronic-pain-in-australia

Barnova, K., Mikolasova, M., Kahankova, R. V., Jaros, R., Kawala-Sterniuk, A., Snasel, V., Mirjalili, S., Pelc, M., & Martinek, R. (2023). Implementation of artificial intelligence and machine learning-based methods in brain–computer interaction. *Computers in Biology and Medicine*, *163*, 107135.

Bashivan, P., Rish, I., Yeasin, M., & Codella, N. (2015). Learning representations from EEG with deep recurrent-convolutional neural networks. *arXiv preprint arXiv:1511.06448*.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory,

Botchkarev, A. (2019). A new typology design of performance metrics to measure errors in machine learning regression algorithms. *Interdisciplinary Journal of Information, Knowledge, and Management*, *14*, 045-076.

Cao, J., Li, G., Shen, J., & Dai, C. (2024). IFBCLNet: Spatio-temporal frequency feature extraction-based MI-EEG classification convolutional network. *Biomedical Signal Processing and Control*, *92*, 106092.

Cascella, M., Esposito, D., Muzio, M. R., Cascella, V., & Cerrone, V. (2025). Artificial intelligence for pain assessment via facial expression recognition (2015–2025): a systematic review. *Exploration of Medicine*, *6*, 1001370.

Chen, D., Zhang, H., Kavitha, P. T., Loy, F. L., Ng, S. H., Wang, C., Phua, K. S., Tjan, S. Y., Yang, S.-Y., & Guan, C. (2022). Scalp EEG-based pain detection using convolutional neural network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *30*, 274-285.

Chen, Y., Tian, Y., Zhou, H., Xu, J., Sun, Y., Zhao, C., & Liu, H. (2025). Assessment and biomarker exploration of chronic pain using wearable EEG device. *Discover Sensors*, *1*(1), 28.

Economics, D. A. (2019). The cost of pain in Australia.

Escalante, H. J., Escalera, S., Guyon, I., Baró, X., Güçlütürk, Y., Güçlü, U., van Gerven, M., & van Lier, R. (2018). *Explainable and interpretable models in computer vision and machine learning*. Springer.

Feng, L., Cheng, C., Zhao, M., Deng, H., & Zhang, Y. (2022). EEG-based emotion recognition using spatial-temporal graph convolutional LSTM with attention mechanism. *IEEE Journal of Biomedical and Health Informatics*, *26*(11), 5406-5417.

Fu, Z., Zhu, H., Zhang, Y., Huan, R., Chen, S., & Pan, Y. (2024). A spatiotemporal deep learning framework for scalp eeg-based automated pain assessment in children. *IEEE Transactions on Biomedical Engineering*, *71*(6), 1889-1900.

Gunter, K. M., Brink-Kjær, A., Mignot, E., Sørensen, H. B., During, E., & Jennum, P. (2023). SViT: A spectral vision transformer for the detection of REM sleep behavior disorder. *IEEE Journal of Biomedical and Health Informatics*, *27*(9), 4285-4292.

Hammal, Z., Walter, S., & Bianchi-Berthouze, N. (2025). The Fifth Edition of the Automated Assessment of Pain (AAP 2025). Proceedings of the 27th International Conference on Multimodal Interaction,

Hamwi, L., Du, H., Jasim, S., Wang, X., Shah, V., Cheng, C., Fabrizi, L., Fitzgerald, M., Meek, J., & Racine, N. (2025). Machine learning classification of EEG responses to pain-related vs non-pain-related stimulus in preterm infants. *Pain reports*, *10*(5), e1332.

Hawker, G. A., Mian, S., Kendzerska, T., & French, M. (2011). Measures of adult pain: Visual analog scale for pain (vas pain), numeric rating scale for pain (nrs pain), mcgill pain questionnaire (mpq), short-form mcgill pain questionnaire (sf-mpq), chronic pain grade scale (cpgs), short form-36 bodily pain scale (sf-36 bps), and measure of intermittent and constant osteoarthritis pain (icoap). *Arthritis care & research*, *63*(S11), S240-S252.

Hefron, R. G., Borghetti, B. J., Christensen, J. C., & Kabban, C. M. S. (2017). Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation. *Pattern Recognition Letters*, *94*, 96-104.

Hu, L., & Iannetti, G. (2019). Neural indicators of perceptual variability of pain across species. *Proceedings of the National Academy of Sciences*, *116*(5), 1782-1791.

Huang, G., Xiao, P., Hung, Y., Iannetti, G. D., Zhang, Z., & Hu, L. (2013). A novel approach to predict subjective pain perception from single-trial laser-evoked potentials. *Neuroimage*, *81*, 283-293.

Imai, H., Kanie, Y., Yoshimoto, S., Yamamoto, N., Furuya, M., Fujimori, T., & Okada, S. (2025). Classification accuracy of pain intensity induced by leg blood flow restriction during walking using machine learning based on electroencephalography. *Scientific Reports*, *15*(1), 27955.

International Association for the Study of Pain (IASP) (2025). *Global Year 2025: Pain Management in Low- and Middle-Income Settings*. IASP. https://www.iasp-pain.org/globalyear

Jiao, Z., Gao, X., Wang, Y., Li, J., & Xu, H. (2018). Deep convolutional neural networks for mental load classification based on EEG data. *Pattern Recognition*, *76*, 582-595.

Joseph, L. P., Joseph, E. A., & Prasad, R. (2022). Explainable diabetes classification using hybrid Bayesian-optimized TabNet architecture. *Computers in Biology and Medicine*, *151*, 106178.

Jui, S. J. J., Deo, R. C., Acharya, R., Barua, P. D., Soar, J., & Devi, A. (2026). Automated stress detection with explainable-AI transformer TabNet technique and physiological signals. *Biomedical Signal Processing and Control*, *113*, 109208.

Kamhi, S., Zhang, S., Ait Amou, M., Mouhafid, M., Javaid, I., Ahmad, I. S., Abd El Kader, I., & Kulsum, U. (2022). Multi-classification of motor imagery EEG signals using Bayesian optimization-based average ensemble approach. *Applied Sciences*, *12*(12), 5807.

Ketkar, N. (2017). Introduction to keras. In *Deep learning with python: a hands-on introduction* (pp. 97-111). Springer.

Khare, S. K., & Acharya, U. R. (2023). An explainable and interpretable model for attention deficit hyperactivity disorder in children using EEG signals. *Computers in Biology and Medicine*, *155*, 106676.

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of neural engineering*, *15*(5), 056013.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

Liu, K., Yang, M., Yu, Z., Wang, G., & Wu, W. (2022). FBMSNet: A filter-bank multi-scale convolutional neural network for EEG-based motor imagery decoding. *IEEE Transactions on Biomedical Engineering*, *70*(2), 436-445.

Liu, Y., Wang, Y., & Zhang, J. (2012). New machine learning algorithm: Random forest. International conference on information computing and applications,

Lovelace, J. A., Miller, J., Jacobs, S., Koppes, W., Minella, F., Porta, F., White, F. A., Hutchinson, M. R., Miller, J., & Jacobs, S. (2024). Development of machine learning algorithms using EEG data to detect the presence of chronic pain. *medRxiv*, 2024.2009. 2018.24313903.

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of thoracic oncology*, *5*(9), 1315-1316.

Mari, T., Ali, S. H., Pacinotti, L., Powsey, S., & Fallon, N. (2025). Machine learning classification of active viewing of pain and non-pain images using EEG does not exceed chance in external validation samples. *Cognitive, Affective, & Behavioral Neuroscience*, *25*(3), 814-831.

Mathrawala, A., Kurup, D., & Lau, J. (2025). EEG-Based Acute Pain Classification: Machine Learning Model Comparison and Real-Time Clinical Feasibility. *arXiv preprint arXiv:2510.05511*.

Mishriky, J., Stupans, I., & Chan, V. (2020). An investigation of the practices of Australian adults experiencing pain and their views of Australian community pharmacy pain management services. *Pharmacy*, *8*(4), 187.

Mouraux, A., & Iannetti, G. D. (2018). The search for pain biomarkers in the human brain. *Brain*, *141*(12), 3290-3307.

Naser, M., & Alavi, A. H. (2023). Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences. *Architecture, Structures and Construction*, *3*(4), 499-517.

Nezam, T., Boostani, R., Abootalebi, V., & Rastegar, K. (2018). A novel classification strategy to distinguish five levels of pain using the EEG signal features. *IEEE Transactions on Affective Computing*, *12*(1), 131-140.

Pain, I. A. f. t. S. o. (2025). *Fact Sheet: Challenges and opportunities in pain management, research, education and in low-and-middle-income settings*. International Association for the Study of Pain. Retrieved 17 March 2026 from https://www.iasp-pain.org/resources/fact-sheets/introduction-to-the-iasp-2025-global-year-pain-management-research-and-education-in-low-and-middle-income-settings/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

Phan, H., Mikkelsen, K., Chén, O. Y., Koch, P., Mertins, A., & De Vos, M. (2022). Sleeptransformer: Automatic sleep staging with interpretability and uncertainty quantification. *IEEE Transactions on Biomedical Engineering*, *69*(8), 2456-2467.

Raja, S. N., Carr, D. B., Cohen, M., Finnerup, N. B., Flor, H., Gibson, S., Keefe, F. J., Mogil, J. S., Ringkamp, M., & Sluka, K. A. (2020). The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises. *Pain*, *161*(9), 1976-1982.

Rezzouk, M., Gagnon, F., Champagne, A., Roy, M., Albouy, P., Coll, M.-P., & Subakan, C. (2025). Towards Generalizable Learning Models for EEG-Based Identification of Pain Perception. 2025 IEEE 35th International Workshop on Machine Learning for Signal Processing (MLSP),

Rodriguez-Bermudez, G., & Garcia-Laencina, P. J. (2015). Analysis of EEG signals using nonlinear dynamics and chaos: a review. *Applied mathematics & information sciences*, *9*(5), 2309.

Rosenblatt, M., Tejavibulya, L., Jiang, R., Noble, S., & Scheinost, D. (2023). The effects of data leakage on connectome-based machine learning models. *bioRxiv*.

Sameen, M. I., Pradhan, B., & Lee, S. (2020). Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. *Catena*, *186*, 104249.

Sanner, M. F. (1999). Python: a programming language for software integration and development. *J Mol Graph Model*, *17*(1), 57-61.

Srinivas, P., & Katarya, R. (2022). hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost. *Biomedical Signal Processing and Control*, *73*, 103456.

Subakan, C., Ravanelli, M., Cornell, S., Bronzi, M., & Zhong, J. (2021). Attention is all you need in speech separation. ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),

Subramanian, A., Cao, R., Naeini, E. K., Aqajari, S. A. H., Hughes, T. D., Calderon, M.-D., Zheng, K., Dutt, N., Liljeberg, P., & Salanterä, S. (2025). Multimodal pain recognition in postoperative patients: Machine learning approach. *JMIR Formative Research*, *9*, e67969.

Sun, M., Liu, Y., Geng, D., Wu, X., & Yang, D. (2026). Pain intensity classification and evaluation of individual differences in subjects based on hybrid CNN–BiLSTM approach. *Biomedical Signal Processing and Control*, *119*, 109815.

Thiam, P., Kessler, V., Amirian, M., Bellmann, P., Layher, G., Zhang, Y., Velana, M., Gruss, S., Walter, S., & Traue, H. C. (2019). Multi-modal pain intensity recognition based on the senseemotion database. *IEEE transactions on affective computing*, *12*(3), 743-760.

Tveter, M., Tveitstøl, T., Hatlestad-Hall, C., Pérez T, A. S., Taubøll, E., Yazidi, A., Hammer, H. L., & Haraldsen, I. R. H. (2024). Advancing EEG prediction with deep learning and uncertainty estimation. *Brain Informatics*, *11*(1), 27.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam med*, *37*(5), 360-363.

Witjes, B., Starmans, M. P., Huygen, F. J., & de Vos, C. C. (2025). Classification of chronic pain and spinal cord stimulation response using machine learning in magnetoencephalography data. *PloS one*, *20*(12), e0337726.

Wu, F., Mai, W., Tang, Y., Liu, Q., Chen, J., & Guo, Z. (2022). Learning spatial-spectral-temporal EEG representations with deep attentive-recurrent-convolutional neural networks for pain intensity assessment. *Neuroscience*, *481*, 144-155.

Yu, M., Sun, Y., Zhu, B., Zhu, L., Lin, Y., Tang, X., Guo, Y., Sun, G., & Dong, M. (2020). Diverse frequency band-based convolutional neural networks for tonic cold pain assessment using EEG. *Neurocomputing*, *378*, 270-282.

Zhao, X., Zhou, J., Zhang, L., Zhuang, Y., Duan, H., Wei, S., Yao, S., Lu, X., Bi, Y., & Hu, L. (2025). A comprehensive EEG dataset of laser-evoked potentials for pain research. *Scientific Data*, *12*(1), 1536.

Zis, P., Liampas, A., Artemiadis, A., Tsalamandris, G., Neophytou, P., Unwin, Z., Kimiskidis, V. K., Hadjigeorgiou, G. M., Varrassi, G., & Zhao, Y. (2022). EEG recordings as biomarkers of pain perception: where do we stand and where to go? *Pain and therapy*, *11*(2), 369-380.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FundingDeclaration.docx
- Ethicalstatement.docx