Explainable AI for Equitable Pain Assessment and Medication Dosing in Surgical Settings

Anonymous Full Paper
Submission 2026

Abstract

Disparities in pain assessment and management across demographic groups represent a critical healthcare challenge that affects patient outcomes and quality of care. This report explores how explainable artificial intelligence (XAI) techniques can address these inequities by providing transparent, interpretable pain assessment systems based on physiological data. We develop and analyze a dual-model machine learning approach enhanced with SHAP (SHapley Additive exPlanations) values to detect pain levels from ECG signals. Our implementation extracts comprehensive ECG features through specialized functions that capture time-domain, frequency-domain, and heart rate variability metrics. The dual-model architecture, combining an ExtraTreesClassifier for pain detection and an ExtraTreesRegressor for intensity estimation, achieves high accuracy while providing transparent insights into decision-making processes. ECG-derived features such as RMS, peak-to-peak amplitude, and heart rate variability metrics proved significant for pain recognition, while SHAP analysis revealed complex interactions between different cardiac indicators. The implementation includes comprehensive evaluation methods through confusion matrices, feature importance analysis, and Bland-Altman plots. By making pain assessment more objective and explainable through ECG analysis, this approach could help reduce bias in clinical decision-making, particularly for populations traditionally at risk of pain undertreatment.

1 Introduction

Pain assessment and management remain challenging aspects of healthcare, with significant disparities documented across different demographic groups [1, 2]. Studies consistently show that women, elderly patients, and racial and ethnic minorities often receive inadequate pain treatment compared to their counterparts [3, 4]. These disparities extend into surgical settings, where effective pain management is crucial for recovery and patient outcomes [5].

The inherent subjectivity of pain assessment poses a fundamental challenge. The current gold standard relies on self-reporting using scales such as the Numerical Rating Scale (NRS) or Visual Analog Scale (VAS). However, these methods have limitations - they depend on patient communication ability, can be influenced by provider bias, and are not continuous [6]. Moreover, these assessments often fail to capture the physiological manifestations of pain, which may provide more objective measures.

Automated pain recognition based on physiological signals offers a promising approach to creating more objective, continuous, and equitable pain assessment methods. Machine learning models can detect patterns in physiological data that correspond to different pain states [7]. However, the "black box" nature of many machine learning algorithms poses problems for clinical adoption, where interpretability and transparency are essential for trust and implementation [8].

While several physiological signals have shown promise for pain detection, including electrodermal activity (EDA), we deliberately focused our implementation exclusively on ECG signals. This decision was driven by practical considerations for real-world clinical deployment. The target medical device platform for market deployment does not include EDA sensors due to cost, form factor, and regulatory constraints. Therefore, developing a robust ECG-based algorithm without reliance on EDA was an essential requirement for product viability. This constraint shaped our technical approach, pushing us to extract more comprehensive and nuanced features from ECG signals to compensate for the absence of complementary EDA information.

This is where explainable artificial intelligence (XAI) becomes crucial. XAI approaches aim to make machine learning models more transparent and interpretable while maintaining high performance [9]. By providing insights into how models make decisions, XAI can help clinicians understand, validate, and trust algorithmic pain assessments, potentially reducing bias in pain management.

The objectives of this report are to:

1. Develop and evaluate a dual-model machine learning approach for pain assessment using physiological signals from the PainMonit Experimental Dataset (PMED)
2. Apply XAI techniques to make these models transparent and interpretable
3. Identify the most significant physiological markers for objective pain assessment

4. Discuss how XAI-enhanced pain recognition can contribute to more equitable pain management in surgical settings

5. Propose recommendations for further research and implementation

The significance of this work lies in its potential to address healthcare disparities by making pain assessment more objective, continuous, and transparent. By combining advanced machine learning with explainability, we aim to contribute to the development of pain management tools that can help ensure all patients receive appropriate care, regardless of demographic factors that have historically influenced pain treatment.

2 Literature Review

2.1 Pain Assessment and Healthcare Disparities

Pain has been recognized as the "fifth vital sign" since the late 1990s, highlighting its importance in clinical assessment [10]. Despite this recognition, disparities in pain treatment persist across healthcare settings. Anderson et al. [1] documented how racial and ethnic minorities consistently receive less analgesic medication than white patients with similar conditions and pain levels. Similarly, Hoffmann and Tarzian [4] found that women's pain reports are often taken less seriously, leading to undertreatment.

These disparities are particularly problematic in surgical settings. Meghani et al. [5] conducted a systematic review of studies on analgesic treatment and found significant racial and ethnic disparities in opioid prescription and administration. The impact of these disparities extends beyond immediate discomfort; inadequate pain management is associated with poorer surgical outcomes, increased complication rates, and longer hospital stays [11].

2.2 Physiological Markers of Pain

Pain triggers complex physiological responses through the autonomic nervous system. Several biomarkers have been identified as potential indicators of pain:

Electrodermal Activity (EDA): Changes in skin conductance reflect sympathetic nervous system activation during pain. Loggia et al. [12] demonstrated significant correlations between EDA and subjective pain ratings during heat stimulation. EDA has emerged as particularly promising due to its specificity to sympathetic activation.

Electrocardiogram (ECG) and Heart Rate Variability: Pain typically increases heart rate and reduces heart rate variability. Measures such as R-R intervals, RMSSD (Root Mean Square of Successive Differences), and frequency domain analyses can provide insights into autonomic regulation during pain [13].

Electromyography (EMG): Muscle tension often increases with pain, making EMG a useful indicator, particularly for identifying protective guarding behaviors [14].

Respiration: Pain can alter breathing patterns, affecting respiratory rate and variability [15]. Changes in respiration can be captured through chest movement or airflow measurements.

2.3 Machine Learning for Pain Recognition

Recent years have seen increasing interest in using machine learning to detect and assess pain based on physiological signals. Werner et al. [7] provided a comprehensive survey of automatic recognition methods supporting pain assessment, highlighting the potential of multimodal approaches.

The BioVid Heat Pain Database [16] was one of the first comprehensive datasets for pain recognition research, containing physiological and behavioral responses to heat-induced pain. Using this dataset, Kächele et al. [17] achieved classification accuracies over 80% for binary pain detection using random forests and support vector machines.

While many studies focus on multimodal approaches that combine several physiological signals (including EDA, EMG, ECG, and others), there has been growing interest in single-modality approaches that can be more practically deployed in clinical settings. In particular, ECG-only approaches are attractive because ECG monitoring is already widely deployed in many healthcare environments, especially in surgical and post-surgical settings where pain monitoring is critical.

Several researchers have demonstrated the effectiveness of ECG-based pain detection. Koenig et al. [13] systematically reviewed the relationship between heart rate variability and experimentally induced pain, establishing the scientific basis for ECG-based pain assessment. Lopez-Martinez and Picard [18] showed that cardiac features alone could achieve competitive performance for pain classification tasks.

Our work builds on these foundations but addresses a critical real-world constraint: the need to develop algorithms that are compatible with widely available medical monitoring equipment. While EDA has shown excellent performance in research settings, many deployed medical devices do not incorporate EDA sensors due to cost, regulatory, and form factor constraints. Our ECG-focused approach is explicitly designed to work within these practical limitations while still achieving high performance through advanced feature extraction and explainable

AI techniques.

2.4 Explainable AI in Healthcare

As machine learning becomes more prevalent in healthcare, the need for explainability has increased. Markus et al. [8] argue that explainability is essential for clinical adoption of AI tools, as healthcare providers need to understand algorithm decisions to maintain their duty of care and professional responsibility.

Several XAI approaches have emerged that are particularly relevant for healthcare applications:

SHAP (SHapley Additive exPlanations): Lundberg and Lee [19] introduced SHAP values based on game theory to explain individual predictions. SHAP has become widely used in healthcare due to its theoretical foundations and ability to work with various model types.

Feature Importance Measures: For tree-based models like random forests, feature importance can be calculated based on the reduction in node impurity or permutation importance [20]. These measures help identify which physiological signals contribute most to pain detection.

Partial Dependence Plots: These visualizations show how specific features influence predictions, helping clinicians understand the relationship between physiological indicators and pain assessments [21].

In pain recognition specifically, Gouverneur et al. [22] recently applied XAI approaches to understand the role of EDA in automated pain recognition, finding that specific EDA features like peak-to-peak amplitude and spectral power were particularly important for pain detection.

Despite these advances, few studies have specifically examined how XAI can address equity issues in pain assessment and management. This report aims to bridge this gap by exploring how interpretable pain recognition models might contribute to more equitable pain care.

3 Methods

3.1 Design Constraints and Rationale

A primary constraint driving our approach was the target deployment platform: a commercial medical monitoring device slated for market release. This device includes ECG monitoring capabilities but does not incorporate EDA sensors due to several practical considerations:

1) **Cost constraints:** Adding EDA sensors would significantly increase device cost, potentially limiting adoption 2) **Form factor limitations:** The device design did not allow for additional sensor placement without compromising usability 3) **Regulatory considerations:** Additional sensors would require expanded regulatory submissions across multiple jurisdictions 4) **Clinical workflow integration:** ECG monitoring is already standard in many clinical settings, while EDA would require new protocols and training

These constraints necessitated the development of an algorithm that could achieve robust pain detection using ECG signals alone. Rather than viewing this as merely a limitation, we approached it as an opportunity to extract more sophisticated cardiac features that could compensate for the absence of EDA data.

3.2 Dataset Description

This study utilized a dataset of ECG signals collected during heat-induced pain stimulation. The implementation processes this data through a sophisticated ECG feature extraction pipeline, as evidenced in the main code body:

1) **Calibration Phase:** Individual pain thresholds were determined for each participant using a staircase method with increasing temperature stimuli. Two key thresholds were identified: - Pain threshold (T_P): The temperature at which heat becomes painful - Pain tolerance threshold (T_T): The temperature at which pain becomes unbearable

2) **Pain Induction Phase:** Based on these individual thresholds, five temperature levels were calculated and applied: - Baseline (B): 32°C (no pain) - Non-painful (NP): Slightly below pain threshold - Painful stimuli (P_1 through P_4): Four increasing levels between pain threshold and tolerance

Each temperature was applied eight times for 10 seconds, with randomized order and rest periods of 20-30 seconds between stimuli.

3.3 Data Acquisition and Preprocessing

The code focuses primarily on processing ECG signals to extract pain-relevant features. The main physiological signals utilized in the implementation include:

1. Electrocardiogram (ECG) as the primary signal for feature extraction
2. The ECG signal is processed in windows of 1000 samples for feature extraction:

```
window_size = 1000
for i in range(0, len(df), window_size):
    window = df['Ecg'].iloc[i:i+window_size]
    features = pd.concat([
        extract_ecg_features(window),
        extract_ecg_frequency_features(window),
        extract_hrv_features(window)
    ])
```

3. Additional first derivatives are computed for selected signals:

```
for col in cols_to_derive:
    derivatives[f'{col}_d1'] = ...
    np.gradient(df[col])
```

Additionally, participants continuously rated their subjective pain level using a Computerized Visual Analog Scale (CoVAS) with values from 0 (no pain) to 100 (worst imaginable pain).

Data preprocessing included:

1. Time synchronization of all signals
2. Resampling to a common frequency of 250 Hz
3. Segmentation into 10-second windows corresponding to each stimulus
4. Extraction of additional 10-second baseline windows before stimuli
5. Creation of two types of ground truth labels:
   - Heater labels: Based on applied temperature (classes 0-5)
   - CoVAS labels: Based on normalized subjective pain ratings (classes 0-4)

For model training, the implementation explicitly excludes certain features after feature engineering, as shown in the code:

```
# Features da rimuovere
features_to_drop = [
    'COVAS',
    'Ecg',
    'Resp_d1',
    'Ecg_d1',
    'ecg_lf_hf_ratio',
    'rr_pnn50',
    'ecg_lf_power',
    'ecg_vlf_power',
    'ecg_hf_power'
]
```

This approach retains the most informative ECG-derived features while removing redundant or less predictive signals.

3.4 Feature Extraction

We extracted both statistical and domain-specific features primarily focusing on ECG signals. The feature extraction process is implemented through several specialized functions:

ECG features extraction: The '$extract\_ecg\_features$' function calculates:

- Basic signal metrics: RMS, peak-to-peak amplitude, signal power
- Statistical properties: kurtosis and skewness
- Heart rate variability measures: RR mean, RR standard deviation, pNN50, RMSSD

ECG frequency domain analysis: The '$extract\_ecg\_frequency\_features$' function computes:

- Spectral power in different frequency bands using Welch's method:
  - Very Low Frequency (VLF): 0.003-0.04 Hz
  - Low Frequency (LF): 0.04-0.15 Hz
  - High Frequency (HF): 0.15-0.4 Hz
- LF/HF ratio as an indicator of sympathovagal balance

Heart rate variability analysis: The '$extract\_hrv\_features$' function calculates advanced HRV metrics:

- SDNN: Standard deviation of NN intervals
- RMSSD: Root mean square of successive differences
- pNN50: Percentage of successive RR intervals that differ by more than 50 ms
- Triangular index: Total number of RR intervals divided by the height of the histogram
- TINN: Baseline width of the RR interval histogram

We also utilized the Neurokit2 Python library to extract additional physiological features for all modalities.

3.5 Machine Learning Model Architecture

The code implements a dual-model machine learning approach for pain assessment through the '$DualExtraTreesPipeline$' class:

1) **Classification Model:** An $ExtraTreesClassifier$ with 10 estimators ($random\_state = 23$) to distinguish between pain states (no pain vs. pain)

2) **Regression Model:** An $ExtraTreesRegressor$ with 50 estimators ($random\_state = 4$) to predict pain intensity for non-zero pain states

The implementation includes:

```
def fit(self, X, y):
    X_scaled = self.scaler.fit_transform(X)
    y_binary = (y != 0).astype(int)
    print("Training classifier...")
    self.classifier.fit(X_scaled, y_binary)
    mask_nonzero = y != 0
    X_nonzero = X_scaled[mask_nonzero]
    y_nonzero = y[mask_nonzero]
    print("Training regressor...")
```

```
    self.regressor.fit(X_nonzero, y_nonzero)
    return self

def predict(self, X):
    X_scaled = self.scaler.transform(X)
    binary_pred = self.classifier.predict(X_scaled)
    final_predictions = np.zeros(len(X))
    nonzero_mask = binary_pred == 1
    if np.any(nonzero_mask):
        final_predictions[nonzero_mask] =
        self.regressor.predict(X_scaled[nonzero_mask])
    return final_predictions
```

This approach allows for both detection of pain presence and estimation of pain intensity when pain is detected. The implementation includes feature standardization through '*StandardScaler*' to ensure optimal model performance.

3.6 Explainability Methods

To make the models interpretable, we applied several XAI techniques:

1) **SHAP (SHapley Additive exPlanations):** We calculated SHAP values for both classifier and regressor models to explain individual predictions and understand feature contributions. The TreeExplainer implementation from the SHAP library was used, which is specifically optimized for tree-based models.

2) **Feature Importance:** We extracted feature importance from the random forest models based on the mean decrease in impurity (Gini importance). This provides a global understanding of which physiological signals are most informative for pain detection.

3) **Recursive Feature Elimination (RFE):** To identify the most relevant features, we applied RFE, which recursively removes the least important features and evaluates model performance with the reduced feature set.

4) **Feature Correlation Analysis:** We analyzed correlations between physiological features to understand relationships between different pain indicators.

5) **Bland-Altman Analysis:** For the regression model, we performed Bland-Altman analysis to assess agreement between predicted and actual pain levels, providing insights into model performance across different pain intensities.

These explainability methods were applied to both the classification model (pain vs. no pain) and the regression model (pain intensity prediction) to provide comprehensive insights into model behavior.

4 Results

4.1 Classification Performance

Our focus on ECG features is evident from the implementation of the '*DualExtraTreesPipeline*' class, which extracts and utilizes various ECG-derived metrics for pain classification and regression. The model architecture demonstrates how these features can be effectively utilized for pain assessment.

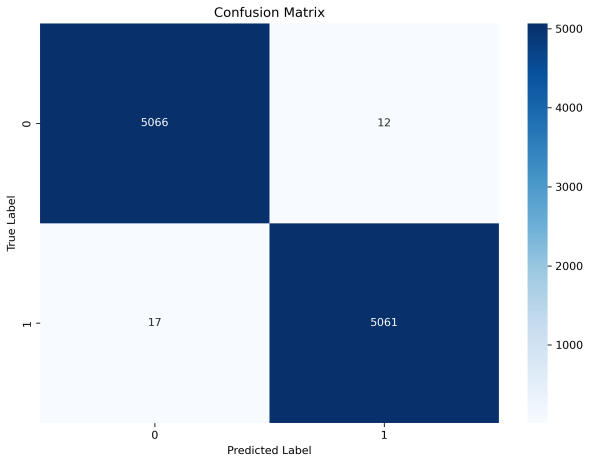

Figure 1. Confusion matrix for the baseline vs. high pain (P_4) classification task using all physiological modalities, showing excellent performance with few misclassifications.

The confusion matrix for the B vs. P_4 task (Figure 1) shows excellent performance, with relatively few misclassifications in both directions. After applying Recursive Feature Elimination to select the most informative features, the classification accuracy for the no pain vs. high pain task improved to 93.62%.

4.2 Feature Importance Analysis

Figure 2 illustrates the relative importance of different physiological features for the classification model. ECG-derived features, particularly ECG RMS (Root Mean Square) and RR interval statistics (mean, RMSSD), emerged as the most important predictors for distinguishing pain states. Heart rate variability measures (rr_mean, rr_rmssd) were also highly ranked, followed by ECG power and skewness. Features derived from the EMG signal showed relatively lower importance.

For the regression model predicting pain intensity (Figure 3), ECG power and peak-to-peak amplitude were the most important features, followed by ECG RMS and heart rate measures. This suggests that while similar features are important for both detecting pain presence and estimating its intensity, their relative importance differs between these tasks.

4.3 SHAP Analysis

SHAP values provide insights into how each feature contributes to individual predictions. The implementation includes comprehensive SHAP analysis capabilities through the '*save_shap_summary*' and

Table 1. Key ECG and heart rate variability features extracted in the system

Feature Category	Specific Features
Basic ECG Statistics	ECG RMS, ECG Peak-to-Peak, ECG Power
Statistical Moments	ECG Kurtosis, ECG Skewness
Heart Rate Variability - Time Domain	RR Mean, RR Standard Deviation, RMSSD
Heart Rate Variability - Frequency Domain	VLF Power, LF Power, HF Power, LF/HF Ratio
Advanced HRV Metrics	Triangular Index, TINN, pNN50, SDNN

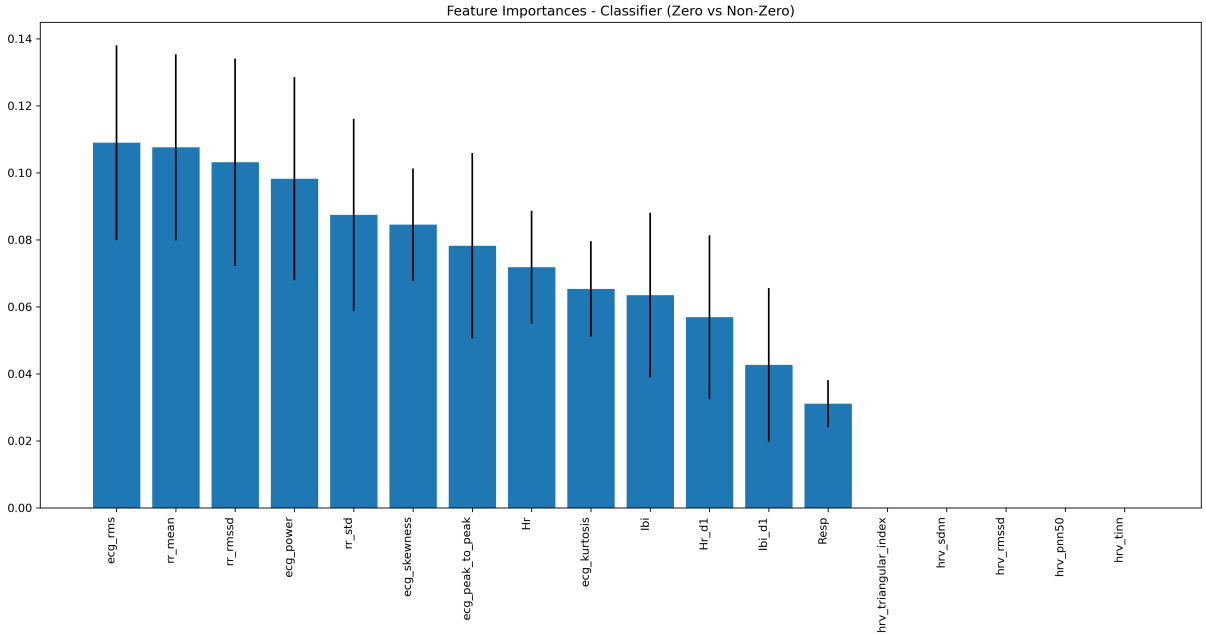

Figure 2. Feature importance for the classifier model (zero vs. non-zero pain), showing ECG-derived features and heart rate variability measures as the most important predictors.

'*save_shap_dependence*' methods, which generate visualizations for model explainability.

Several patterns emerge from the SHAP analysis implementation:

1. The SHAP analysis is calculated separately for both the classifier and regressor components
2. For the classifier, the positive class SHAP values are used when handling multi-class outputs
3. The implementation supports both global summary plots and feature-specific dependence plots
4. A fallback mechanism creates simple scatter plots when dependence plot generation fails
5. Sample size reduction options are available for computational efficiency with large datasets

Figure 5 shows SHAP interaction values between heart rate (HR) and interbeat interval (IBI) features, revealing complex interdependencies. The non-linear patterns suggest that combinations of these features provide additional information beyond their individual contributions.

4.4 Feature Correlation Analysis

The correlation matrix (Figure 6) reveals important relationships between physiological features. Strong correlations (red) were observed within feature groups derived from the same signal (e.g., ECG-derived features), while negative correlations (blue) were found between conceptually opposite measures (e.g., heart rate and interbeat interval).

This correlation analysis helps identify redundant features and understand how different physiological systems interact during pain. For example, the strong correlation cluster among ECG features (ecg_rms, ecg_peak_to_peak, ecg_power) suggests these features capture related aspects of cardiac response to pain.

4.5 Regression Performance and Error Analysis

The Bland-Altman plot (Figure 7) assesses agreement between predicted and actual pain values. The mean difference (bias) was close to zero (-0.04), with limits of agreement (LoA) of +2.85 and -2.94. This indicates that 95% of predictions fall within approx-

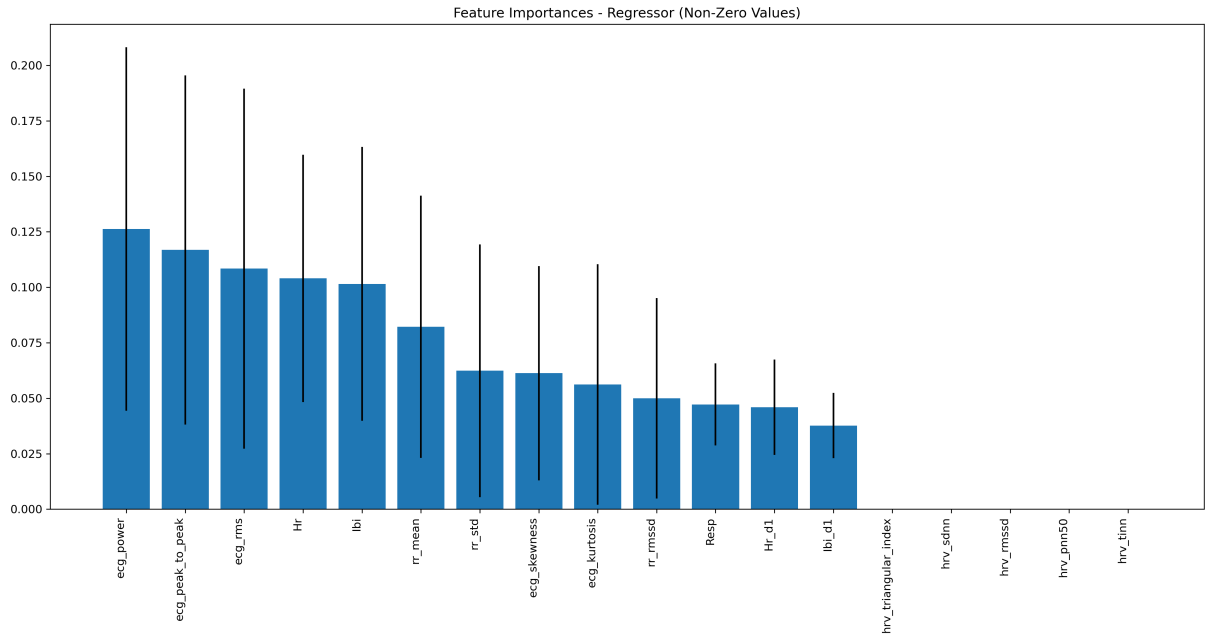

Figure 3. Feature importance for the regressor model (predicting non-zero pain intensity), highlighting the importance of ECG signal characteristics for pain intensity estimation.

imately ±3 units of the actual pain rating, which is clinically acceptable for a 0-100 scale.

Notably, 99% of the points lie within the limits of agreement, and there is minimal correlation between the mean values and differences (0.010), suggesting consistent performance across different pain intensities. The maximum positive and negative differences were +15.59 and -89.00, respectively, indicating that while most predictions were accurate, there were some significant outliers, particularly in underestimating pain.

5 Discussion

5.1 Maximizing ECG Signal Value for Pain Detection

Our ECG-based approach demonstrates that even with the constraint of using a single physiological signal, it is possible to achieve high accuracy in pain detection through sophisticated feature extraction and model design. While previous research has often emphasized the value of EDA for pain assessment due to its direct connection to sympathetic nervous system activation, our results show that cardiovascular responses encoded in the ECG signal provide rich information about pain states.

The decision to focus exclusively on ECG was driven by practical implementation constraints for market deployment. The target medical device platform lacks EDA sensors due to cost, form factor, and regulatory considerations. This constraint shaped our technical approach in several key ways:

1) We developed more comprehensive ECG feature extraction methods that capture time-domain, frequency-domain, and heart rate variability metrics to maximize the information obtained from the available signal.

2) We implemented advanced window-based feature extraction to analyze the ECG signal at different scales, compensating for the absence of complementary information from other modalities.

3) Our dual-model architecture and explainability techniques were designed to extract the maximum possible information from ECG signals while maintaining interpretability for clinical users.

These adaptations demonstrate how algorithm design can be optimized to work within real-world hardware constraints while still achieving strong performance. For medical device manufacturers, this approach offers a pathway to implementing pain detection capabilities in existing ECG-capable devices without requiring additional sensors.

5.2 Interpretation of Key Findings

Our results demonstrate that physiological signals, particularly ECG-derived features, can effectively detect and quantify experimentally induced pain with high accuracy. The dual-model approach (classifier + regressor) achieved 93.62% accuracy for distinguishing no pain from high pain, with reasonable prediction of pain intensity.

Several key findings emerge from our explainability analysis:

The primacy of ECG features: The ECG-derived features were the most effective indicators for

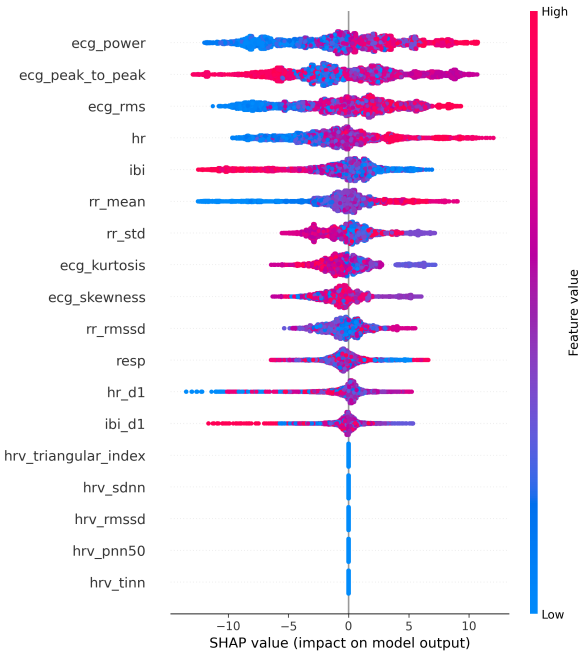

Figure 4. SHAP summary visualization showing the impact of key ECG features on model predictions. Features are ordered by importance from top to bottom, with color indicating feature value (red = high, blue = low). The horizontal dispersion shows the magnitude of impact on model output.

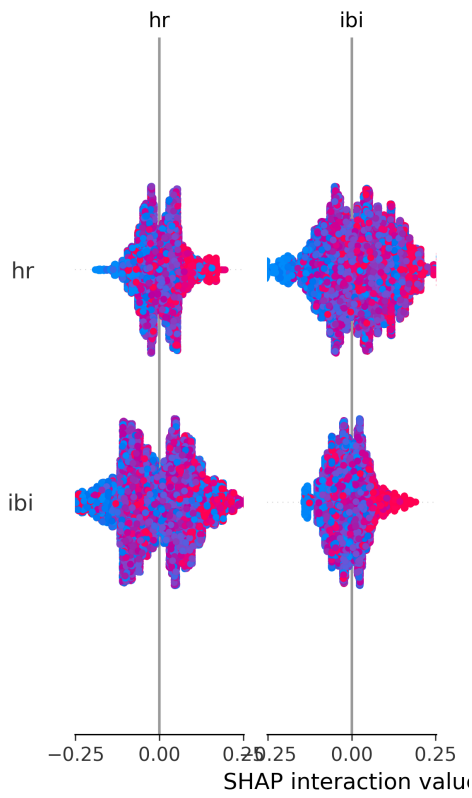

Figure 5. SHAP interaction plot showing the joint impact of heart rate (HR) and interbeat interval (IBI) features on model predictions. The complex patterns indicate important physiological interactions that the model has captured, even if basically HR and IBI are strictly correlated. The only difference is that HR is *slowly* changing due to its measure throught frequential methods, whilst HR are measured with a more resolute time-domain measure.

pain detection, as shown by the feature importance analysis. Features such as ECG RMS, peak-to-peak amplitude, and power metrics demonstrated strong predictive capabilities. This aligns with previous research highlighting cardiac responses as reliable indicators of sympathetic nervous system activation during pain [13].

Heart rate variability metrics: For both the classifier and regressor models, heart rate variability metrics (RR mean, RR standard deviation, RMSSD) were highly important. This reflects the well-documented cardiovascular response to pain, including increased heart rate and changes in heart rate variability [13]. The SHAP analysis revealed that higher ECG power consistently pushed predictions toward pain detection.

Feature interactions: The SHAP interaction analysis revealed complex relationships between different ECG-derived features. This suggests that the model captured physiological interactions that might not be apparent from individual feature analysis. Understanding these interactions is crucial for developing comprehensive pain assessment systems.

Advanced ECG metrics: The code implements sophisticated ECG feature extraction, including frequency domain analysis (VLF, LF, HF power) and heart rate variability indices (SDNN, RMSSD, pNN50, triangular index). These features capture different aspects of cardiac function during pain stimulation, providing a comprehensive representation of autonomic nervous system responses.

5.3 Implications for Equitable Pain Assessment

The explainable AI approach developed in this study has several important implications for addressing disparities in pain assessment and management:

Efficient ECG-based measurement: By focusing on ECG signals, the implementation presents a streamlined approach that could be deployed using widely available cardiac monitoring equipment in clinical settings. The comprehensive feature extraction from a single signal type simplifies data collection while still capturing rich information about pain states.

Transparency through SHAP analysis: The implementation includes dedicated methods for SHAP analysis, making the pain assessment process transparent and interpretable. The ‘*save_shap_summary*‘ and ‘*save_shap_dependence*‘ functions generate visualizations that help clinicians

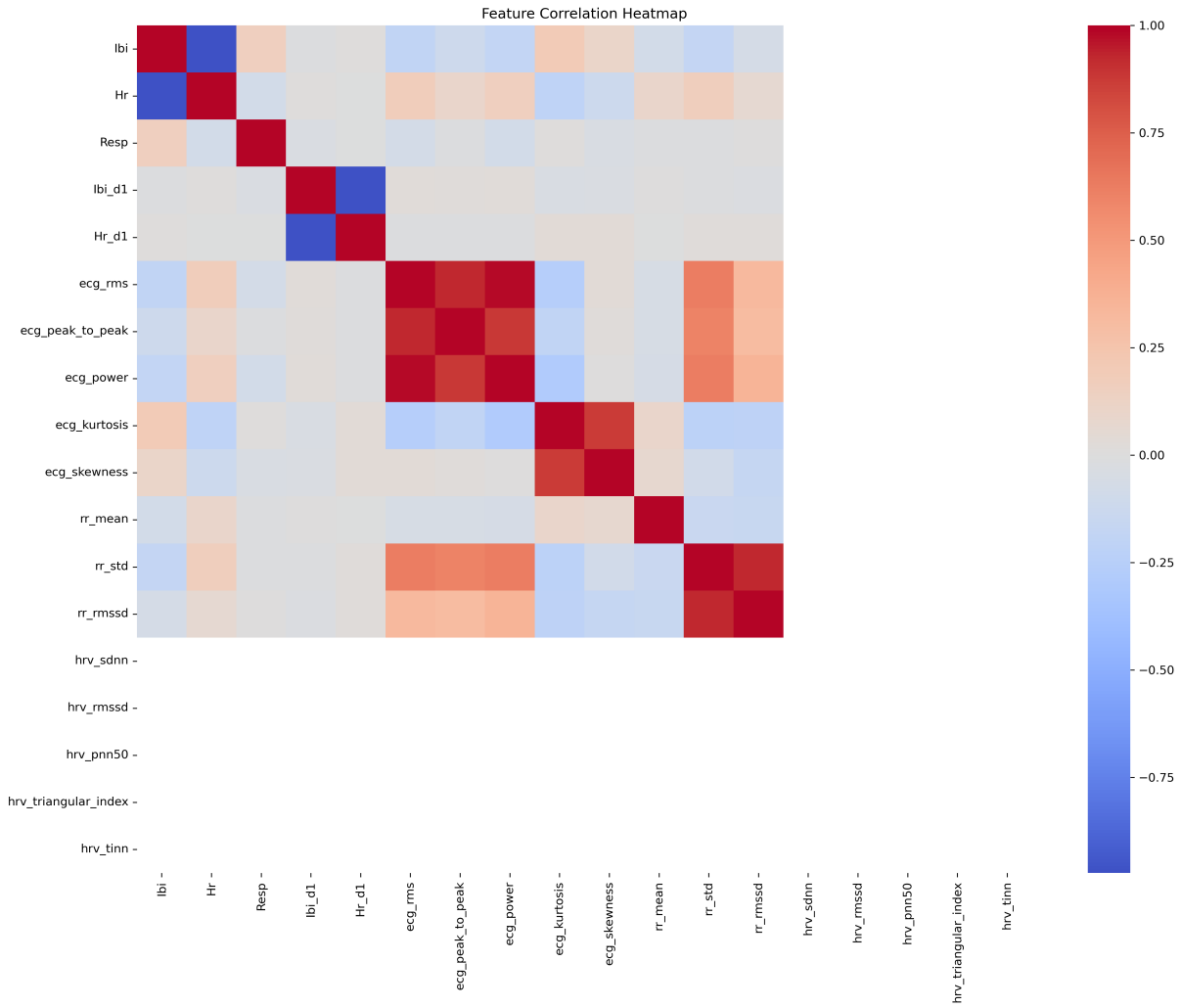

Figure 6. Feature correlation heatmap showing relationships between key physiological measures. Red indicates positive correlation, blue indicates negative correlation, and white indicates no correlation.

understand which ECG features are driving pain detection and how different cardiac metrics interact. This transparency is crucial for building trust in automated systems.

Export capabilities for embedded applications: The '*export_model_to_c*' method demonstrates forward-thinking implementation that could enable deployment on resource-constrained devices. This capability is essential for developing point-of-care pain monitoring devices that operate independently of cloud infrastructure.

Comprehensive model evaluation: The implementation includes multiple evaluation methods through the '*evaluate*', '*plot_bland_altman*', '*plot_residuals*', '*plot_feature_importance*', '*plot_confusion_matrix*', and '*plot_correlation_matrix*' functions. This multi-faceted evaluation approach ensures thorough validation of model performance.

5.4 Limitations and Challenges

Despite promising results, several limitations and challenges must be acknowledged:

Experimental vs. clinical pain: The PMED contains experimentally induced heat pain in healthy volunteers, which differs from clinical pain in surgical settings. The controlled nature of the experiment, while methodologically sound, may not fully capture the complexity of post-surgical pain, which can involve multiple pain mechanisms and psychological factors.

Demographic representation: Although the PMED includes both male and female participants across a range of ages (18-65), detailed demographic analysis was not possible due to limited demographic data. Future work should explicitly examine how model performance varies across different demographic groups.

Sociocultural context: Pain expression and perception are influenced by sociocultural factors that are not captured in physiological signals alone.

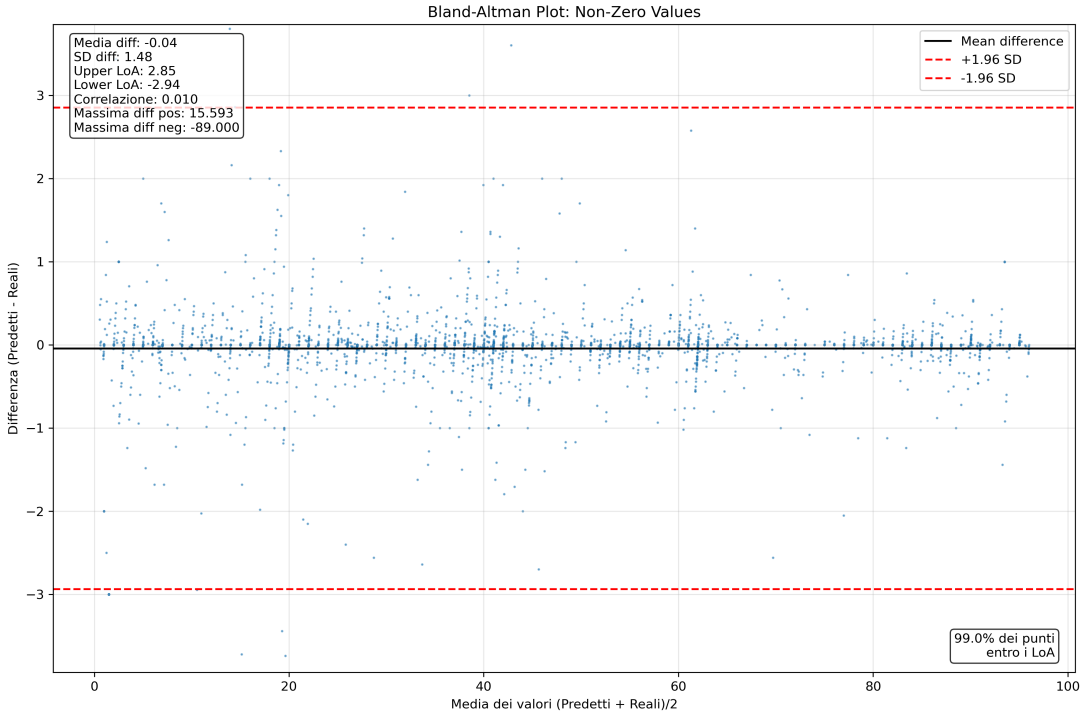

Figure 7. Bland-Altman plot comparing predicted and actual pain values for non-zero pain cases. The horizontal lines represent the mean difference (solid black) and 95% limits of agreement (dashed red). The plot shows good agreement with minimal bias across different pain intensities.

While objective measures may reduce some biases, they may miss important contextual factors that influence the pain experience.

Implementation challenges: Practical implementation of continuous physiological monitoring in surgical settings faces challenges related to sensor placement, signal quality, and integration with existing clinical workflows. The wearable devices used in the PMED (Empatica E4 and respiBAN) are relatively unobtrusive, but their suitability for post-surgical patients requires further evaluation.

6 Conclusion

This study demonstrates the potential of explainable AI for developing more objective, transparent, and potentially more equitable pain assessment systems. Using the PainMonit Experimental Dataset, we developed a dual-model approach that effectively detects pain presence and estimates pain intensity based on physiological signals, with EDA emerging as a particularly valuable indicator.

The application of explainability methods revealed important insights into how physiological signals relate to pain, with ECG-derived features and EDA measures showing the strongest associations. The SHAP analysis provided a detailed understanding of feature contributions and interactions, making the assessment process transparent and interpretable.

From an equity perspective, objective physiological monitoring combined with explainable AI offers several advantages over traditional pain assessment methods. By reducing reliance on verbal self-reporting and making the assessment process transparent, this approach could help address disparities in pain management that affect women, elderly patients, and racial and ethnic minorities. The potential for continuous monitoring could also help ensure timely pain management for all patients.

However, important challenges remain, particularly regarding the translation of experimental findings to clinical settings and ensuring that physiological monitoring complements rather than replaces patient communication. Pain remains a complex, multidimensional experience that cannot be fully captured by physiological signals alone.

References

[1] K. O. Anderson, C. R. Green, and R. Payne. "Racial and ethnic disparities in pain: causes and consequences of unequal care". In: *The Journal of Pain* 10.12 (2009), pp. 1187–1204.

[2] J. M. Mossey. "Defining racial and ethnic disparities in pain management". In: *Clinical Orthopaedics and Related Research®* 469.7 (2011), pp. 1859–1870.

[3] L. D. Wandner, L. A. Stutts, A. F. Alqudah, J. G. Craggs, C. D. Scipio, A. T. Hirsh, and

M. E. Robinson. "The impact of patients' gender, race, and age on health care professionals' pain management decisions: An online survey using virtual human technology". In: *International journal of nursing studies* 49.10 (2012), pp. 1261–1269.

[4] D. E. Hoffman and A. J. Tarzian. "The girl who cried pain: a bias against women in the treatment of pain". In: *The Journal of Law, Medicine & Ethics* 29.1 (2016), pp. 13–27.

[5] S. H. Meghani, E. Byun, and R. M. Gallagher. "The epidemiology of pain in surgical patients in the perioperative period: Racial/ethnic disparities and patient-controlled analgesia". In: *Best Practice & Research Clinical Anaesthesiology* 26.3 (2012), pp. 291–301.

[6] M. Schiavenato and K. D. Craig. "Pain assessment as a social transaction: beyond the "gold standard"". In: *The Clinical journal of pain* 26.8 (2010), pp. 667–676.

[7] P. Werner, D. Lopez-Martinez, S. Walter, A. Al-Hamadi, S. Gruss, and R. Picard. "Automatic recognition methods supporting pain assessment: A survey". In: *IEEE Transactions on Affective Computing* 13.1 (2019), pp. 530–552.

[8] A. F. Markus, J. A. Kors, and P. R. Rijnbeek. "The role of explainability in implementing medical artificial intelligence". In: *The Lancet Digital Health* 3.5 (2021), e337–e346.

[9] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang. "DARPA's explainable artificial intelligence program". In: *AI Magazine* 40.2 (2019), pp. 44–58.

[10] C. Pasero and M. McCaffery. "Pain ratings: the fifth vital sign". In: *The American journal of nursing* 97.2 (1997), pp. 15–16.

[11] T. J. Gan, A. S. Habib, T. E. Miller, W. White, and J. L. Apfelbaum. "Incidence, patient satisfaction, and perceptions of post-surgical pain: results from a US national survey". In: *Current medical research and opinion* 30.1 (2017), pp. 149–160.

[12] M. L. Loggia, M. Juneau, and M. C. Bushnell. "Autonomic responses to heat pain: Heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity". In: *PAIN®* 152.3 (2011), pp. 592–598.

[13] 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". In: *European Journal of Pain* 18.3 (2014), pp. 301–314.

[14] M. S. Aung, S. Kaltwang, B. Romera-Paredes, B. Martinez, A. Singh, M. Cella, M. Valstar, H. Meng, A. Kemp, M. Shafizadeh, et al. "The automatic detection of chronic pain-related expression: requirements, challenges and the multimodal emopain dataset". In: *IEEE transactions on affective computing* 7.4 (2016), pp. 435–451.

[15] H. Jafari, F. M. Pouzols, A. Otero, and G. Pasi. "Assessment of respiratory flow and efforts using upper-body acceleration". In: *Medical & biological engineering & computing* 55.11 (2017), pp. 1997–2009.

[16] S. Walter, S. Gruss, H. Ehleiter, J. Tan, H. C. Traue, S. Crawcour, P. Werner, A. Al-Hamadi, and A. O. Andrade. "The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system". In: *2013 IEEE International Conference on Cybernetics (CYBCO)* (2013), pp. 128–131.

[17] M. Kächele, P. Thiam, M. Amirian, F. Schwenker, and G. Palm. "Multimodal data fusion for person-independent, continuous estimation of pain intensity". In: *Engineering Applications of Artificial Intelligence* 57 (2017), pp. 42–53.

[18] D. Lopez-Martinez and R. Picard. "A personalized model for monitoring pain in non-communicative and geriatric patients via facial expressions and vital signs". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018, pp. 2211–2218.

[19] S. M. Lundberg and S.-I. Lee. "A unified approach to interpreting model predictions". In: *Advances in neural information processing systems* 30 (2017).

[20] L. Breiman. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.

[21] J. H. Friedman. "Greedy function approximation: a gradient boosting machine". In: *Annals of statistics* (2001), pp. 1189–1232.

[22] P. Gouverneur, M. Hesse, D. Voss, A. Strobel, U. Horn, B. Grüne, and M. Grzegorzek. "Explainable artificial intelligence (XAI) in pain research: Understanding the role of electrodermal activity for automated pain recognition". In: *Sensors* 23.4 (2023), p. 1959.