EJAIC

Eur J Anaesthesiol Intensive Care Med 2026; **5:**3(e0111)

OPEN

ORIGINAL ARTICLE

Refining multiple artificial intelligence strategies for automatic pain assessment investigations (RUGGI Study)

A study protocol

Marco Cascella, Alfonso Maria Ponsiglione, Vittorio Santoriello, Maria Romano, Francesco Amato, Francesco Sabbatino, Stefano Pepe and Ornella Piazza

BACKGROUND Chronic pain is a complex, multidimensional condition that severely impairs patients' quality of life. As conventional techniques for evaluating pain are based on subjective self-reporting, these approaches have crucial drawbacks, especially for individuals with communication difficulties. Artificial intelligence (AI) provides the opportunity to complement subjective self-reports through multimodal, data-informed analysis to enhance real-time pain assessment and care.

OBJECTIVE To develop, calibrate and validate AI models for automatic pain assessment (APA) in adult patients by merging physiological, behavioural and clinical data and, consequently, complement patient-reported information and support more personalised and effective pain management.

DESIGN Prospective, single-centre, noninterventional study.

SETTING University of Salerno Hospital, Italy.

PATIENTS AND PARTICIPANTS Adult patients (>18 years) with chronic primary or secondary pain (oncologic and non-oncologic), able to provide their informed consent. The main exclusion criteria are severe psychiatric or cognitive disorders and treatment with psychotropic medications.

PRIMARY OUTCOME MEASURES Predictive performance of AI models (sensitivity, specificity, area under the receiver operating characteristic curve, AUC-ROC) for automatic pain assessment based on collected multimodal data.

SECONDARY OUTCOMES Quality-of-life evaluation, analgesic treatment monitoring, development and analysis of a multidimensional dataset for APA and identification of correlations between clinical and physiological variables.

RESULTS N/A (study ongoing).

CONCLUSIONS This study will provide essential data for developing and validating integrated AI tools for objective, multidimensional pain assessment, with potential future clinical and therapeutic applications.

TRIAL REGISTRATION ClinicalTrials.gov Identifier: NCT07038434.

From the Anesthesia, Intensive Care and Pain Medicine, Department of Medicine, Surgery and Dentistry 'Scuola Medica Salernitana', University of Salerno, Baronissi (MC, OP), Department of Electrical Engineering and Information Technology, University of Naples 'Federico II', Naples (AMP, VS, MR, FA) and Oncology, Department of Medicine, Surgery and Dentistry, University of Salerno, Baronissi, Salerno, Italy (FS, SP)

Correspondence to Professor Ornella Piazza, Dipartimento di Medicina, Chirurgia Odontoiatria 'Scuola Medica Salernitana', Università degli Studi di Salerno, Via Allende, Baronissi (SA), 84081, Italy.
E-mail: opiazza@unisa.it

Received 31 July 2025 **Accepted** 6 March 2026

Published online 15 May 2026

DOI:10.1097/EA9.0000000000000111

KEY POINTS

- Chronic pain assessment based on patients' self-reporting has several limitations.
- Integration of physiological, behavioural and linguistic data via AI can improve diagnosis and management of chronic pain.
- This study will collect multimodal noninvasive physiological data and neurocognitive test results from patients with chronic pain to create validated AI models for precise, personalised pain management and serve future clinical research.

Introduction

Pain is among the most prevalent symptoms that call for medical intervention. According to estimates, the prevalence of chronic pain is 27.5% worldwide, with significant regional variation, ranging from 9.9 to 50.3%.[1] While acute pain, such as postoperative or trauma-related pain, acts as a physiological warning signal eliciting adaptive and homeostatic responses, chronic pain is a more complex phenomenon that encompasses biological, psychological and social components.[2] In temporal and pathophysiological terms, the distinction between acute and chronic pain adds more complexity.[3] Understanding the transition from acute to chronic pain and its clinical manifestations remains a critical research area, with many aspects still unclear. For instance, some forms of chronic pain may arise without a preceding acute pain phase.[4,5]

Recent advancements in computational techniques have made it easier to develop models that deepen our understanding of pain mechanisms.[6] Nociception, namely the process by which the body detects and responds to potentially harmful stimuli, involves the neural encoding and transmission of damaging or unpleasant signals to the central nervous system, where they are processed.[7,8] This processing integrates both cortical and subcortical systems; for example, the limbic system plays a key role in the emotional and affective aspects of pain. As a result, pain is a subjective experience that depends on how the brain interprets and perceives these signals across sensory, emotional and cognitive domains.[9]

Rationale for this study

Pain is influenced by many factors, including the type of stimulus, biological and psychological components and the social environment around us. These elements make precise assessment and measurement challenging. Therefore, reliable diagnostic standards are essential for developing effective treatments. While clinical examination and a detailed patient history should complement subjective reports, the complex biopsychosocial nature of pain and its emotional impact can introduce variability and uncertainty in self-reported information, highlighting the need for complementary assessment approaches for avoiding undertreatment and overtreatment risks.[10]

Automatic pain assessment (APA) is an emerging research field. It explores the objective components of pain using biosignal responses (e.g. galvanic skin response – GSR, photoplethysmography – PPG) and observable behaviour such as facial expressions and speech analysis. These methodologies are increasingly investigated through the application of artificial intelligence (AI) techniques.[11] AI systems can process large volumes of multimodal data, identifying patterns and correlations that may not be readily apparent to clinicians. For example, machine learning (ML) algorithms can be trained to recognise specific physiological responses associated with varying pain intensities, offering a more objective and consistent estimation of pain.[12] Moreover, computational language analysis has been used to detect verbal markers of pain[13] and emotional states.[14] Additionally, computer vision, especially deep learning, a subfield of AI that enables machines to interpret visual data through artificial neural architectures, has shown promise in analysing pain-related facial expressions.[15]

Despite the potential of APA methods to offer more objective insights into pain intensity, important limitations must be acknowledged. These include the scarcity of high-quality validation studies, uncertainty regarding which parameters are appropriate in different clinical settings and technical challenges such as timing and synchronisation of multimodal data acquisition. Therefore, comprehensive pain evaluation should ideally combine subjective self-assessment with objective measures to achieve a more holistic understanding of the pain experience and to support better-informed pain management strategies[16] (Fig. 1).

This protocol complies with the SPIRIT-AI 2020 extension for ensuring transparency, reproducibility and methodological rigour when evaluating AI interventions in healthcare.[17] The study is registered with ClinicalTrials.gov (Identifier: NCT07038434).

Aims of the study

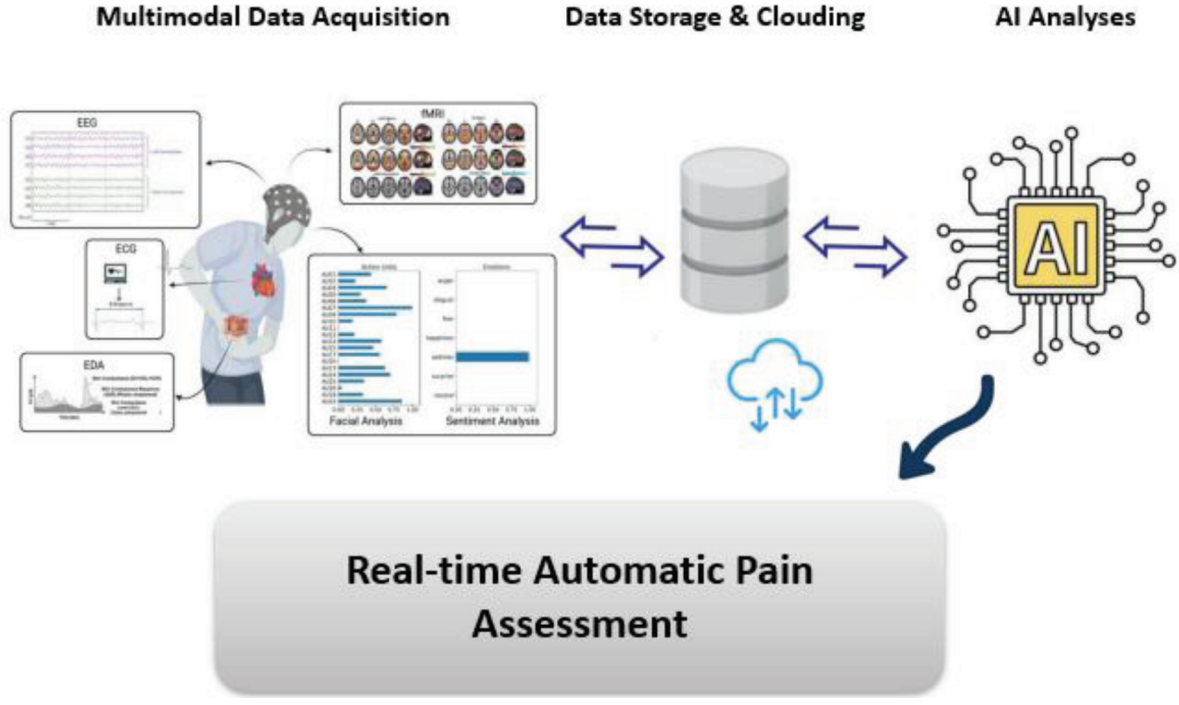
Primary objective

To develop and validate AI models that accurately predict chronic pain intensity and type (nociceptive, neuropathic, mixed) from multimodal clinical and physiological data in adult patients with chronic pain.

Secondary objectives

(1) To evaluate associations between AI-derived pain metrics and patient-reported quality of life and treatment responses.

Fig. 1 Study framework.



(2) To construct a comprehensive multimodal dataset for future AI research in pain assessment.
(3) To explore the potential of AI models in predicting pain flare events and analgesic treatment outcomes.

Methods

Study design

This is a prospective, single-centre, interventional observational study conducted at the University of Salerno Hospital, Italy.

Ethics

The study is conducted in accordance with the Declaration of Helsinki (2013 amendment) and ICH-GCP guidelines. Ethical approval for this study (Comitato Etico Territoriale Campania 2, N° 2024/28590) was provided by the Ethical Committee Campania 2 on the 3rd of April 2025.

Written informed consent will be obtained from all participants before enrolment.

Study population

Inclusion criteria

(1) Adults, older than 18 years.
(2) Diagnosis of chronic pain (primary or secondary, oncologic and nononcologic) persisting for at least 3 months, as per IASP and ICD-11 definitions.
(3) Ability to provide informed consent.

Exclusion criteria

(1) Active psychiatric disorders or psychotropic medication interfering with study assessments.
(2) Cognitive impairments that preclude informed consent.
(3) Pregnancy or breastfeeding.

Data collection procedures

Participants undergo baseline clinical evaluation, including personal details, pain history, clinical scales (numeric rating scale, 0 to 10 NRS, DN-4, Brief Pain Inventory) and multimodal acquisitions for APA (T0 phase).

Multimodal physiological data are collected noninvasively using:

(1) BITalino platform for electrodermal activity (EDA), electromyography (EMG) and and heart rate variability (HRV).[18]
(2) BitBrain wearable devices for EEG via DIADEM and Versatile headsets, GSR and PPG.[19,20]
(3) Hybrid InfraRed Affective computing system (HIRA) for facial thermography and visual expression analysis. This technology allows dynamic analysis of temperature changes and facial expressions and will

enable real-time thermographic analysis with facial landmark tracking and processing.[21]

As chronic pain can cause attention deficit,[22] participants will also complete a Stroop cognitive interference test to assess pain-related cognitive effects. The test usually includes three conditions (blocks): a neutral condition, a congruent one and a final incongruent condition. In the neutral phase, participants are asked to quickly read a list of colour words printed in black on a white background. This part serves as a baseline to measure reading speed and accuracy. In the congruent condition, participants read colour words printed in the matching colour (e.g. the word 'red' printed in red). Here, the goal is to measure speed and accuracy when the ink colour matches the word's meaning. In the final incongruent condition, participants read colour words printed in a colour that does not match the word's meaning (e.g. the word 'red' printed in blue). This part is considered the most difficult because it requires inhibiting automatic responses and focusing on the ink colour rather than the word's meaning (Fig. 2). Finally, participants will undergo a brief videotaped semi-structured interview for linguistic and prosodic analysis and APA through facial action units (AUs) analyses.[11,23]

A follow-up evaluation is provided (T1 phase). This phase corresponds to the scheduled clinical evaluation according to the patient's therapeutic plan and is performed 15 to 30 days after the baseline visit. The follow-up assessment may be conducted in person or via telemedicine. It includes a routine clinical examination, administration of the Patient Global Impression of Change (PGIC) questionnaire and documentation of any adverse events related to ongoing therapies. Data collected at follow-up (T1) will be integrated into the longitudinal and predictive analyses to assess temporal trends in pain intensity, treatment response and model stability.

Outcome measures

Primary outcome

Performance metrics (AUC-ROC, sensitivity, specificity, precision and other parameters) of AI models in classifying pain intensity levels and types based on the integrated multimodal dataset.

Secondary outcomes

Quality-of-life scores, analgesic treatment responses and development of a comprehensive, labelled multimodal dataset for future research.

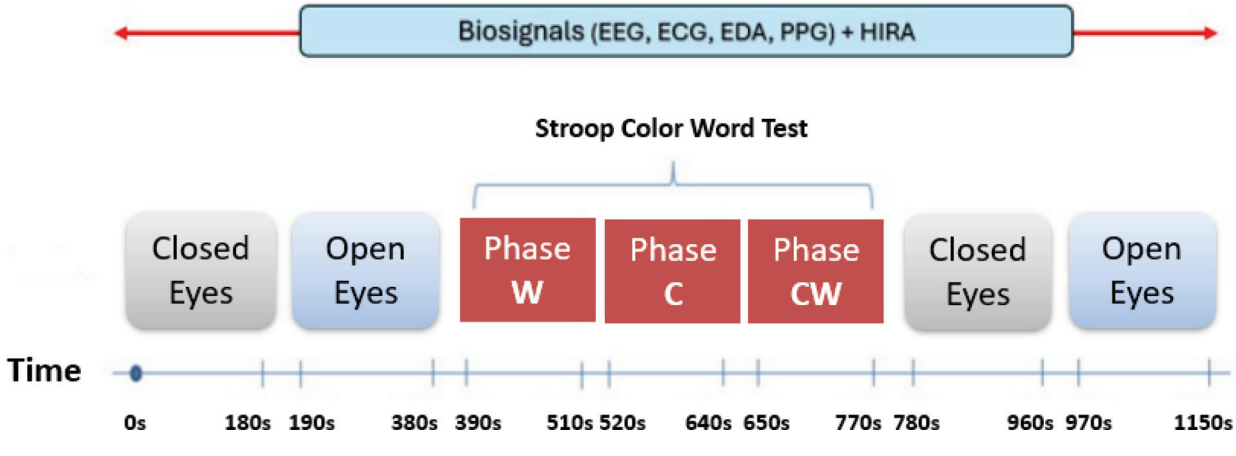
Sample size and statistical analysis

To estimate the sample size, we referred to an AUC value between 0.85 and 0.95 for the ROC curve. A recent study reported an ROC AUC of 0.977 by using facial expression to classify pain states.[24]

The AUC reflects the model's ability to distinguish between two classes (i.e. low and high, given NRS >4). An AUC between 0.85 and 0.95 suggests a good level of discrimination. To calculate the minimum sample size needed for model fitting (the cohort required to define a model), we followed the approach below:

(1) Starting from a multidimensional dataset on pain, implemented in APA-related studies,[16] we performed data preparation, dataset balancing, data

Fig. 2 Experimental timeline. Biosignals (EEG, ECG, EDA, PPG) plus HIRA were recorded continuously. After an initial, eyes-closed rest followed by eyes open, participants performed the Stroop Colour Word Test in three blocks: Word (W), Colour (C) and Colour–Word (CW), and then completed a second eyes-closed and eyes-open rest (total duration: 1150 s ≈ 19 min).



augmentation and comparison of ML models using evaluation metrics.
(2) As the initial dataset was imbalanced, an oversampling procedure was applied to balance the dataset.
(3) The dataset was then randomly augmented through a resampling process and the addition of noise to introduce variability, until the desired target size was reached.
(4) Various dataset sizes were hypothesised to evaluate how model performance varied as a function of the number of available samples.
(5) The RFE (Recursive Feature Elimination) technique was then applied to select the best features of the dataset. In this case, the estimator used was a Support Vector Classifier (SVC) with a linear kernel, which helped identify the 15 most relevant features.
(6) Through a simulation-based training and testing process, we calculated the model's area under the ROC curve. The population size was tested until AUC values above 0.85 were achieved.

For model evaluation, we used logistic regression and random forest, as they are among the most commonly used models for binary classification in biomedical research. At the end of the analysis, we estimated that a sample size between 200 and 400 individuals allows achieving an area under the ROC curve greater than 0.90 for the random forest model and greater than 0.85 for the logistic regression model, using 15 predictive variables, as shown in Fig. 3.

Therefore, a sample size of at least 200 patients is considered an adequate and conservative estimate to reach the performance required.

Data analysis

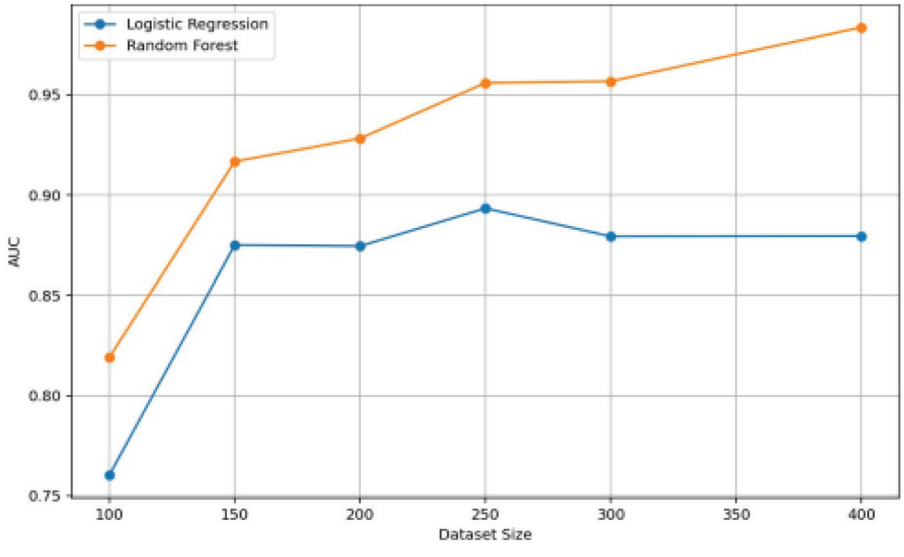
Statistics

Demographic and clinimetric variables (e.g., type, location and intensity of pain), unstructured data (imaging and text analysis, clinical notes and reports), questionnaires and multimodal data recorded during the study will be considered for descriptive analyses and to train AI models.

The variables will be described in tables, with categorical variables represented as frequencies and percentages, and numerical variables as mean (standard deviation) and median (interquartile range). Univariate analysis will consist of describing crude associations between the main variables and the dependent variable – primarily the score obtained from questionnaires on the NRS scale (0 to 10) and quality of life over time (screening T0, baseline T1) – using tests for comparing means/medians or contingency tables. Statistical tests will be applied at a 95% significance level ($P < 0.05$) to describe trends across the time-based assessments (T0 and T1), as well as the different steps of the Stroop test.

Multivariate analysis will be performed using appropriate methods for repeated measures analysis where needed (GLMER – Generalized Linear Mixed Effects Models), to calculate risk factors (ORs) or linear associations

Fig. 3 Performance (area under the receiver operating characteristic curve) of the simulations using logistic regression and random forest models.



(betas) that characterise the statistical relationships (proportional or linear) between the main variables and the dependent variable of interest, with a 95% significance level ($P < 0.05$), while assessing the effect of time.

Regression analyses and hypothesis testing (parametric and nonparametric tests, depending on data normality) will be carried out to explore relationships between predictors (data extracted from biosignals) and clinical/sociodemographic characteristics, specifically focusing on dependent variables (NRS, quality of life, etc.) relevant to the study.

Structuring and cleaning of the dataset will include data standardisation, data integration (linking and integrating data from multiple sources into a single dataset) and accurate labelling. Missing data will be handled using Multiple Imputation by Chained Equations (MICE), assuming data are 'Missing at Random'. Quality control will be performed to avoid errors or anomalies and ensure regulatory compliance. Finally, we will implement measures to ensure patient data privacy and security (e.g. de-identified data).

Artificial Intelligence analysis

The 'RUGGI' study aims to apply various AI strategies for the APA. AI processes will be structured as follows:

(1) Multisource data and signal acquisition. Gathering data and signals from multiple heterogeneous sources (clinical and demographic data, questionnaires, interviews, biosignals, etc.).
(2) Exploratory data analysis. Various techniques will be developed, including nongraphical univariate analysis; graphical univariate analysis using histograms, boxplots, and so forth; multivariate analysis to highlight linear statistical relationships between covariates and dependent variables and graphical multivariate analysis
(3) Feature extraction and engineering. Relevant features will be extracted and selected from the data/signals acquired from multiple sources (such as facial expression features, text or prosodic elements, or electrodermal activity variations).
(4) Structured multimodal dataset. Organisation of a multimodal structured dataset gathering features from biosignals, questionnaires, interviews, as well as clinical/sociodemographic data from electronic health records. This step is necessary to establish a shared and uniform template for building the dataset with all the relevant features useful for the subsequent data analysis.
(5) Dataset preprocessing. It ensures that dataset features are interpretable by the algorithm, including data loading, normalisation and standardisation. If needed, dimensionality reduction techniques will be used to reduce model complexity and enhance computational performance.
(6) Train and test datasets. A random split of the whole multimodal dataset into a training set and an independent test set that will not be included in the subsequent cross-validation step will be made. This is necessary to set aside an independent set for testing and assessing model generalisation.
(7) Cross-validation for feature selection and model optimisation. Consensus Nested k-fold cross-validation (as reported in Parvandeh *et al.*[25]) with embedded feature selection will be adopted to avoid selection bias (to avoid leakage between training folds and the test fold: the training set will be split into k roughly equal subsets, and feature selection and classification that will occur on k-1 folds, with performance assessed accordingly on the remaining fold). For each ML algorithm, the most 'important' features will be identified (i.e. those strongly associated with the pain output classes), with association intensity evaluated using standard algorithms such as AUC-ROC analysis, ReliefF, LASSO and backward selection. Identification of optimal features and model hyperparameters will be made through consensus across the cross-validation procedure to ensure the selection of the most robust predictors and model parameters.
(8) Testing on an independent test set for model generalisation. Testing the best model on the independent test set and storing the obtained performance metrics.
(9) Repeating previous steps of cross-validation and testing. Cross-validation and testing on the independent test set will be repeated N times, and for different classification models, in order to assess the variability in the performance metrics across different models and different train/test splits. Average and standard deviation of the performance metrics will be calculated to have a measure of how stable the optimal model identified is, and how reliable the performance metrics are, when tested on an independent data set for generalisation.

Algorithms and models

ML and Deep Learning (DL) strategies will be adopted, with model comparison to select the best-performing one. These AI models will be compared in terms of accuracy and performance metrics in pain prediction.

(1) Supervised Models. Trained using labelled data, such as correlations between physiological signals and self-reported pain. Algorithms include Support Vector Machines (SVM), Naive Bayes (NB), k-Nearest Neighbors (kNN), Gradient Boosting Machines (GBM), tree-based models (e.g. Decision Tree,

Random Forest, Gradient Boosted Trees) and artificial neural networks.
(2) Unsupervised Models. Clustering and dimensionality reduction will be used to discover patterns in unlabelled data, identifying hidden structures and patient subgroups with similar characteristics.
(3) Deep Learning. Neural networks will model complex data relationships, for example, facial expression analysis (AUs) or physiological signal classification. Networks may be feedforward or recurrent, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). For facial expression analysis, a feedforward neural network and the CNN model YOLO ('You Only Look Once') will be implemented. YOLO is a real-time, open-source computer vision algorithm, already applied to pain datasets.[26]

Language analysis will include textual and phonetic-prosodic analysis. Specifically, textual analysis will build a 'pain corpus'. Texts will be preprocessed (normalisation, removal of punctuation, numbers, stop words), lemmatised/stemmed and tokenised. Semantic and lexical analysis will identify key terms and linguistic patterns associated with pain. Techniques like Word2Vec or BERT will explore semantic associations, and ML classifiers will categorise texts by pain type (e.g. nociceptive or neuropathic).

Phonetic-Prosodic Analysis will use Python's Librosa package to extract vocal features, including Mel-Frequency Cepstral Coefficients (MFCC), chromatic features, spectral contrast and zero-crossing rate. Following prior experience,[16,27] a 16 kHz sampling rate will be applied, collecting data every ~32 ms. All measurements will then be aggregated by computing the temporal mean. A Multilayer Perceptron (MLP) neural network will be built using Scikit-learn, processing 181 prosodic features to assign a score to each of the 6+1 emotion states (i.e. happiness, sadness, fear, anger, surprise, disgust, plus neutral) to be correlated with pain expressions.

Model evaluation

To conduct robust analysis, the dataset will be dynamically split into training and testing sets using k-fold cross-validation. In each iteration, different portions are used for training and testing.

Optimisation of model parameters will be performed through a systematic grid search process, iteratively evaluating predefined combinations of hyperparameters to identify the best-performing configuration. Model performance will be assessed using repeated k-fold cross-validation (100 repetitions). All algorithms will be trained and tested on the same prerandomised fold sets. Final results will be expressed as the average of all performance metrics across iterations.

Various metrics will assess model performance in pain prediction. To evaluate class imbalance, a normalised Shannon Entropy score will be calculated; given this strategy, the most reliable performance metrics for both feature selection as well as model optimisation and evaluation will be prioritised (e.g. F1-score, precision-recall curves and balanced accuracy).[28] Moreover, in the occurrence of high-class imbalance, the most resilient model to class imbalance issue will be preferred, with particular regard to tree-based algorithms such as Decision Trees[29] as well as Random Forest,[28] and methods based on Firth Logistic Regression.[30] Furthermore, other approaches to addressing imbalance could also be attempted, such as weighting the cost matrix, assigning higher misclassification costs to minority classes and, eventually, synthetic minority oversampling techniques (e.g. SMOTE – Synthetic Minority Over-sampling Technique and ROSE – Random Over-Sampling Examples) during the training phase. For regression problems (where the outcome is continuous or ordinal), we will calculate the mean absolute error (MAE), mean squared error (MSE), mean absolute percentage error (MAPE), root mean squared error (RMSE) and adjusted R-squared (adj-R^2). For classification problems, model performance will be evaluated using confusion matrix metrics. True positives, true negatives, false positives and false negatives will be implemented to obtain derived metrics such as accuracy, sensitivity (recall), specificity, precision, balanced accuracy (calculated as the arithmetic mean between sensitivity and specificity), positive-predictive value, negative-predictive value and F1 score (as the harmonic mean of precision and recall).

Feature selection and hyperparameter tuning within cross-validation

To optimise model performance, various combinations of model parameters and feature selection techniques will be explored and integrated within a Consensus Nested k-fold Cross-validation approach.[25] This approach involves combining outer and inner validation loops to ensure that feature selection and parameter tuning are validated against unseen data at multiple levels. Specifically, the outer loops are used for robust performance evaluation, while the inner loops handle feature selection and hyperparameter tuning for each outer fold's training data. The process begins by splitting the dataset into k outer folds for independent testing. For each outer fold, the remaining data are used for inner cross-validation loops, where feature selection algorithms are applied to the inner training folds, and hyperparameters are optimised through systematic iteration over a predefined grid of hyperparameter combinations to identify the best model settings. Features and hyperparameters that demonstrate consistency across inner folds are considered the most reliable. Finally, the model is trained using the selected features and the optimised hyperparameters on the inner training data and then tested on the held-out outer test

fold, with performance recorded to identify the optimal model. Among the feature selection algorithms embedded within the inner loop of the consensus nested cross-validation, AUC-ROC analysis (or precision-recall curves), ReliefF, LASSO and backward selection will be employed to evaluate the association strength of the most 'important' features (i.e. those highly related to the pain output classes). Consensus across the cross-validation process will determine the best features and model hyperparameters, ensuring the selection of the most reliable predictors and parameters.

Model explainability

To assess and visualise model performance and improve interpretability, various explainability techniques will be applied. They include feature importance, decision plots, heatmaps, *SHAP* (SHapley Additive exPlanations) values and *LIME* (Local Interpretable Model-agnostic Explanations) visualisations.

These strategies will help visualise the separation between predicted outcome groups and classes and illustrate the univariate and multivariate impact of key predictors, enabling researchers to understand model decision-making and identify the most important quantitative indicators of predicted outcomes. Furthermore, a qualitative clinical relevance score (e.g. based on a Likert scale) will be developed in collaboration with clinical experts to compare interpretability outputs with traditional model performance metrics. This strategy will offer insights for bridging the gap between algorithmic insights and clinical understanding.

External validation

Where possible, the models will be validated on external datasets to confirm their robustness and generalisability.

Model building and analyses will be performed in Python using reproducible, version-controlled pipelines to ensure transparency and replicability. DL architectures will be implemented in *PyTorch* or *TensorFlow*, while classical ML algorithms will be developed using the *scikit-learn* library. For example, hyperparameter tuning will be conducted through the *GridSearchCV* function from *scikit-learn*. All analytic workflows will be documented and executed using standardised scripts and random seed control to guarantee full reproducibility of the results.

Discussion

Although pain remains an inherently subjective experience, AI-based multimodal systems may assist clinicians by identifying reproducible physiological and behaviour-based elements and patterns that can complement and improve self-reported measures, rather than replacing them. Therefore, this prospective interventional study aims to evaluate the performance and clinical utility of an AI-based diagnostic tool for APA. On these premises, this study actively implements different devices and approaches for multimodal data collection to develop predictive models that may transform pain evaluation into clinical practice.

Pain assessment traditionally relies on subjective patient reports, which can be limited by cognitive impairments, communication barriers and the fluctuating nature of pain. The integration of AI provides an opportunity to objectively quantify pain intensity and its multifaceted manifestations through real-time analysis of physiological, behavioural and linguistic data.[11] The biopsychosocial complexity of pain and its emotional weight can skew self-reported information, risking both undertreatment and overtreatment. Nevertheless, aiming for mechanical objectivity in pain management can generate the automation bias, which is the inclination to be unjustifiably self-reliant in algorithmic outputs. Many clinical and ethical concerns are consequent to the use of AI-based solutions to simplify pain management, which, instead, needs comprehensive multimodal management.[31]

We recently participated in the analysis of a survey on the application of APA for clinical use,[32] followed by expert consensus, which recommended using multimodal inputs (e.g. physiological signals, facial expressions, speech and clinical data) in these systems. Consequently, our approach addresses several gaps in current research by combining multiple biosignal modalities, validated cognitive testing and information obtained through patients' interviews, supported by advanced ML and DL algorithms.

The study design of this trial allows for controlled data acquisition under standardised conditions, optimising the training and validation of AI models. While prior studies have demonstrated promising results for individual AI techniques in pain detection, few have evaluated integrated systems within a structured clinical workflow. On this basis, this study will provide crucial evidence on the feasibility, accuracy and clinical relevance of AI-assisted pain diagnostics, potentially guiding personalised treatment strategies and improving patient outcomes.

Limitations

Limitations include its monocentric design and sample size constraints, which may affect generalisability. Future multicentre randomised controlled trials will be essential to validate these findings across diverse populations and clinical settings. Additionally, ethical considerations surrounding AI transparency, data privacy and clinician acceptance must be continuously addressed.

Conclusion

This clinical study represents a critical step forward in the development and clinical evaluation of AI-based diagnostic tools for chronic pain assessment. Multimodal biosignals and cognitive testing acquisitions, as well as computer vision investigations and natural language

analyses, will be performed to produce predictive models aimed at improving the objectivity and precision of pain measurement. The findings are expected to inform future clinical practice by enabling personalised pain management strategies and advancing patient-centred care.

Acknowledgements relating to this article

Assistance with the article: none.

Financial support and sponsorship: financial support (data manager grant) was provided by Campania Region (Linea Progettuale n.5 'Tecnologia Sanitaria').

Conflicts of interest: none.

Presentation: none.

This manuscript was handled by Esther M. Pogatzki-Zahn.

References

1 Zimmer Z, Fraser K, Grol-Prokopczyk H, *et al.* A global study of pain prevalence across 52 countries: examining the role of country-level contextual factors. *Pain* 2022; **163**:1740–1750.
2 Rahman S, Kidwai A, Rakhamimova E, *et al.* Clinical diagnosis and treatment of chronic pain. *Diagnostics* 2023; **13**:3689.
3 Cohen SP, Vase L, Hooten WM. Chronic pain: an update on burden, best practices, and new advances. *Lancet* 2021; **397**:2082–2097.
4 Clauw DJ. From fibrositis to fibromyalgia to nociplastic pain: how rheumatology helped get us here and where do we go from here? *Ann Rheum Dis* 2024; **83**:1421–1427.
5 Kaplan CM, Kelleher E, Irani A, *et al.* Deciphering nociplastic pain: clinical features, risk factors and potential mechanisms. *Nat Rev Neurol* 2024; **20**: 347–363.
6 Lang VA, Lundh T, Ortiz-Catalan M. Mathematical and computational models for pain: a systematic review. *Pain Med* 2021; **22**:2806–2817.
7 Kutafina E, Becker S, Namer B. Measuring pain and nociception: through the glasses of a computational scientist. Transdisciplinary overview of methods. *Front Netw Physiol* 2023; **3**:1099282.
8 L'Haridon L, Cañamero L. The effects of stress and predation on pain perception in robots. In: 2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII), Cambridge, MA, USA, 2023, pp. 1–8.
9 Treede R-D, Rief W, Barke A, *et al.* Chronic pain as a symptom or a disease: the IASP Classification of Chronic Pain for the International Classification of Diseases (ICD-11). *Pain* 2019; **160**:19–27.
10 Nicholas MK. The biopsychosocial model of pain 40 years on: time for a reappraisal? *Pain* 2022; **163**:S3–S14.
11 Cascella M, Schiavo D, Cuomo A, *et al.* Artificial intelligence for automatic pain assessment: research methods and perspectives. *Pain Res Manag* 2023; **2023**:6018736.
12 Lötsch J, Ultsch A, Mayer B, *et al.* Artificial intelligence and machine learning in pain research: a data scientometric analysis. *Pain Rep* 2022; **7**: e1044.
13 Albashayreh A, Bandyopadhyay A, Zeinali N, *et al.* Natural language processing accurately differentiates cancer symptom information in Electronic Health Record Narratives. *JCO Clin Cancer Inform* 2024; **8**: e2300235.
14 Machová K, Szabóova M, Paralič J, *et al.* Detection of emotion by text analysis using machine learning. *Front Psychol* 2023; **14**:1190326.
15 D'Antoni F, Russo F, Ambrosio L, *et al.* Artificial intelligence and computer vision in low back pain: a systematic review. *Int J Environ Res Public Health* 2021; **18**:10909.
16 Cascella M, Di Gennaro P, Crispo A, *et al.* Advancing the integration of biosignal-based automated pain assessment methods into a comprehensive model for addressing cancer pain. *BMC Palliat Care* 2024; **23**:198.
17 Cruz Rivera S, Liu X, Chan AW, *et al.* on behalf of the SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. *BMJ* 2020; **370**:m3210.
18 Cascella M, Vitale VN, D'Antò M, *et al.* Exploring biosignals for quantitative pain assessment in cancer patients: a proof of concept. *Electronics* 2023; **12**:3716.
19 Bitbain. Didem and Ring. Available at: https://www.bitbrain.com/neurotechnology-products/dry-eeg/diadem and https://www.bitbrain.com/neurotechnology-products/biosignals/ring. [Accessed 16 April 2026].
20 Bitbrain. Versatile. Available at: https://www.bitbrain.com/neurotechnology-products/water-based-eeg/versatile-eeg and https://www.bitbrain.com/neurotechnology-products/biosignals/versatile-bio. [Accessed 16 April 2026].
21 Next2u. HIRA. Available at: https://www.next2u-solutions.com/. [Accessed 16 April 2026].
22 Andersson G, Haldrup D. Personalized pain words and Stroop interference in chronic pain patients. *Eur J Pain* 2003; **7**:431–438.
23 Cascella M, Vitale VN, Mariani F, *et al.* Development of a binary classifier model from extended facial codes toward video-based pain recognition in cancer patients. *Scand J Pain* 2023; **23**:638–645.
24 Tan CW, Du T, Teo JC, *et al.* Automated pain detection using facial expression in adult patients with a customized spatial temporal attention long short-term memory (STA-LSTM) network. *Sci Rep* 2025; **15**:13429.
25 Parvandeh S, Yeh HW, Paulus MP, *et al.* Consensus features nested cross-validation. *Bioinformatics* 2020; **36**:3093–3098.
26 Cascella M, Shariff MN, Lo Bianco G, *et al.* Employing the Artificial Intelligence Object Detection Tool YOLOv8 for real-time pain detection: a feasibility study. *J Pain Res* 2024; **17**:3681–3696.
27 Cascella M, Cutugno F, Mariani F, *et al.* AI-based cancer pain assessment through speech emotion recognition and video facial expressions classification. *Signa Vitae* 2024; **20**:28–38.
28 Thölke P, Mantilla-Ramos YJ, Abdelhedi H, *et al.* Class imbalance should not throw you off balance: choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *Neuroimage* 2023; **277**: 120253.
29 Rodríguez JJ, Díez-Pastor JF, García-Osorio C. Sansone C, Kittler J, Roli F. Ensembles of decision trees for imbalanced data. *Multiple Classifier Systems. MCS 2011. Lecture Notes in Computer Science, vol 6713* Berlin, Heidelberg: Springer; 2011.
30 Puhr R, Heinze G, Nold M, *et al.* Firth's logistic regression with rare events: accurate effect estimates and predictions? *Stat Med* 2017; **36**: 2302–2317.
31 Buchman DZ. AI and the ethics of techno-solutionism in pain management. *Pain* 2025; **166**:469–470.
32 Cascella M, Ponsiglione AM, Santoriello V, *et al.* Expert consensus on feasibility and application of automatic pain assessment in routine clinical use. *J Anesth Analg Crit Care* 2025; **5**:29.