Enhancing Pain Level Assessment in Post-surgery Patients Using Artificial Intelligence Algorithms

G. Ben Othman¹, E. Yumuk¹,², D. Copot¹,³, A. R. Ynineb¹, H. Farbakhsh¹, I. R. Birs¹,³
C. I. Muresan², R. De Keyser¹, I. Chihí⁴, C. M. Ionescu¹,³, and M. Neckebroek⁵

Abstract—Control performance decreases significantly in the presence of uncertainty in variable availability, measurement noise, or instrumentation failure. In cluttered environments such as the Post Anesthesia Care Unit (PACU), clinical measures are often influenced by noise and artifacts. An important component in post-operative treatment is the assessment and management of pain levels. However, reliable information is critical for clinically relevant results and improved patient outcomes. From a control engineering point of view, variables are often estimated and interpolated to allow a suitable flow of feedback information to control loops for the optimization of drug dosages. In contrast, Artificial Intelligence (AI) stands as a promising tool to augment pain level assessment. This study introduces and compares two AI-based approaches for predicting continuous Numerical Rating Scales (NRS) based on heart rate (HR) data. The first approach uses polynomial regression, lasso regression, and ridge regression, while the second employs an LSTM model. Notably, the AI prediction model, independent of traditional interpolation techniques, outperforms the approach relying on interpolation. The proposed AI-based method promises continuous estimation and can serve as an estimator for model-based control. This proof of concept study underscores the potential of AI tools to enhance pain level assessment.

Index Terms—AI regression model; pain level; PACU; closed loop control; Long Short-Term Memory (LSTM).

I. INTRODUCTION

The success of machine learning tools in medicine is exponential for labeled data and apriori-defined classification features, mostly used in image-based diagnostics [1], [2]. This, paired with the pandemic situation of non-contact advisory general practitioner and specialist support emphasized the role of machine learning as a great opportunity window for breakthrough in clinical practice. However, moving away from diagnostics, things are rather grim when it comes to unlabelled data and large inter- and intra-patient variability. A typical example can be found in nociception monitoring for computer-based control of general anesthesia, as recently proposed in a benchmark simulator platform [3]. Models for anesthesia are typically Wiener - Hammerstein type, whereas the lack of persistent excitation in the input signal (i.e. analgesic drug such as Remifentanil) makes identification very difficult or limited in its applicability [4], [5]. Instead, piece-wise linear models are used which reduce the situation to a Wiener process, i.e. a linear dynamic model followed by a nonlinear gain but piece-wise linear on anesthesia intervals [6]. Some recent attempts to use artificial intelligence tools in anesthesia are reported in [7]–[9].

When it comes to nociception monitoring devices to evaluate the degree of analgesia present in an anesthetized patient, they are scarce and subject to a set of limitations of surgical applicability [10]. For instance, certain drugs used in cardiac surgery may not be applicable, as they inherently affect the evaluation formalism of nociception due to their impact on hemodynamic variables [11]. Alternatively, those devices based on bioimpedance monitoring may be biased by the use of electrosurgery instrumentation [12]. As computer-based control of hypnosis becomes a mature technology in clinical practice, recent advances report computerized optimal dosing control schemes for multi-drug infusion-based sedation [13]–[15]. Clinical studies are still scarce and the need for large data sets is crucial when machine learning is envisaged. NRS is a subjective measure used in post-surgery awake patients who can communicate to the anesthesiologist the level of pain they experience from 0 (no pain) to 10 (extreme pain). This is commonly used in clinical practice during PACU. However, the measurement intervals of subjective data such as the NRS often fail to provide the continuous and precise data needed for computerized forms of optimal closed-loop control. In response to this challenge, this study explores the use of AI to predict the NRS index with higher accuracy than the use of computerized forms of optimal closed-loop control and uses the manual assessment solely as a calibration of the AI algorithm. In this research, two different AI-based approaches are presented and juxtaposed to predict continuous NRS using heart rate (HR) data. The initial approach involves employing polynomial regression,
lasso regression, and ridge regression, whereas the second approach uses a Long Short-Term Memory (LSTM) classification model.

The paper is organized as follows: Section II focuses on the methodology. Section III introduces the proposed AI-based approaches for estimating pain levels. Section IV presents the findings. Finally, the discussion and conclusions, are encapsulated in Sections V and VI.

II. PAIN ASSESSMENT IN PACU

Recently, a prototype monitor, namely Anspec-Pro, for pain assessment [16]–[18] has been proposed and evaluated in awake communicating patients during PACU recovery time [19]. It uses time-frequency signals and spectrogram image-based information for delivering a non-parametric model of relative changes in tissue electrical impedance measured non-invasively via standard 3M patch sensors in the palm of the patient. The prototype Anspec-PRO (Ghent University, Ghent, Belgium) was approved for research by the Federal Agency for Medicines and Health Products Belgium (FAGG) no. AFMPS/80M0707, 2018. It was successfully tested in an experimental setup on awake healthy subjects having mechanically-induced acute pain [17], the monitor has also been successfully validated to detect clinical postsurgical pain [19]. The data presented and used in this paper are recorded using the Anspec-Pro monitor.

The datasets used here were signals recorded in the observational trial approved by the Ethics Committee of Ghent University Hospital (Protocol code: EC/2017/1517). The study was registered before patient enrolment at clinicaltrials.gov (Identifier: NCT03832764, Principal Investigator: MD M. Neckebroek, 2019), and written informed consent was obtained from each patient. The pain measured by Anspec-PRO was also correlated with the standard tool (i.e., numerical rating scale – NRS). Being observational, standard care and medication were given pre-, peri-, and post-operative, estimated through the expert skills and senior experience of the principal investigator. All patients underwent general anesthesia during surgery, following standard care and the anesthetists selected the anesthetics and analgesics during the operation. All medications were registered in the digital medical file of the patient. The study design, clinical investigation protocol, inclusion/ exclusion criteria, and results were reported in our previous publication [19]. For the present work, only data from the patients monitored with Anspec-PRO was used.

III. PROPOSED ESTIMATION METHODOLOGY

A. Data Processing

The data used in this example are from a patient recovering in PACU after gynecologic surgery, a female aged 33, with a height of 166 cm, weight of 75 kg, resulting in a body mass index of 27.22. Within the two approaches proposed in this study, the choice of HR as the input for the AI-based models is predicated upon its strong correlation with the NRS index, with a coefficient of 71.22%. Consequently, this correlation score serves as a vital metric for assessing the performance of the model. From a medical perspective, the relationship between HR and NRS is gaining recognition for its clinical significance. The study conducted by Ho et al. [20] demonstrates that the variations in HR are closely associated with self-reported pain levels in patients following surgery. Their findings suggest that physiological measures, such as HR, can serve as predictive indicators of postoperative pain levels. In PACU, the 20 NRS values were reported by the patient every 7 minutes, for 140 minutes as depicted in Fig. 1. and HR was registered from the electronic medical record every 5 seconds.

The NRS data has been sampled into 1680 segments of duration 5 seconds using a cubic interpolation procedure [21] as presented in Fig. 2. The form of this procedure is a third-degree polynomial, defined by:

\[
S_i(x_i) = a_i + b_i x_i + c_i x_i^2 + d_i x_i^3, \quad (1)
\]

where the subscript \(i\) corresponds to each segment within the range \([1 : n - 1]\), \(n\) denotes the total number of samples. \(x_i\) represents the interpolated value within the specific segment, constrained by \(x_{i-1}\) and \(x_{i+1}\). It is noteworthy that the coefficients \(a_i, b_i, c_i, d_i\) of this polynomial change every 5 seconds, contingent upon the interpolation interval.

The use of interpolation, specifically cubic, over directly using subjectively collected data points is primarily motivated by the need to densify the dataset for AI training. AI models thrive on detailed and continuous datasets to learn and identify patterns effectively. Given the sparse nature of subjectively reported NRS values, interpolation serves to fill in the gaps, providing a more continuous and comprehensive view of pain levels.
functions, achieving convergence to the global minimum relies on effectively scaling both the NRS and HR variables. To address this requirement, min-max normalization has been applied as an essential step in data preprocessing. This process ensures that all features have the same weight in the cost function on the AI model.

**B. The investigated artificial intelligence tools**

**Approach I: Interpolation-Based Training**

In this approach, a sliding window training approach was implemented to maximize the use of the available dataset. Each training window consisted of 840 samples, with a sliding step of one sample, equivalent to 5 seconds of data. To illustrate this process further Fig. 3 is presented.

Consider the following example:

Training Window 1 (W1): Includes samples from 0 to 840 and is used to predict sample number 841. Training Window 2 (W2): Encompasses samples from 1 to 841, incorporating the previously predicted sample, and is used to predict sample number 842. This sequence continues until Training Window 840 (W840), which covers samples from 839 to 1679 and is used to predict sample number 1680. Through this sliding window training approach, the model was effectively trained to predict the latter 50% of the NRS values. It is important to note that the interpolated NRS data was employed for training.

This approach was implemented using three machine learning models: i) lasso regression (LR), ii) polynomial regression (PL), and iii) ridge regression (RR).

The least absolute shrinkage and selection operator (lasso) is a popular machine learning model. Lasso regression performs L1 regularization, which reduces the coefficient of the least relevant characteristic to zero. The method uses shrinkage, which is the process of narrowing the data values to a central point such as the mean [22]. The motivation for using this model is its potential to enhance both prediction accuracy and interpretation thanks to the implied regularization and variable selection.

Polynomial regression [23] is a type of linear regression in which the relationship between the independent variable (HR) and dependent variable (NRS) is modeled as a 7-th degree polynomial:

\[ y = f(x) = b_0 + b_1x + b_2x^2 + \ldots + b_7x^7, \]  

where \( x \) and \( y \) denote HR and NRS, respectively. Ridge regression is used to analyze multiple regression data. It implies the L2 norm regularization technique, which prevents over-fitting problems in AI models, by appending a penalty term to the loss function, this technique avoids the sensitivity of the model to individual or noisy data [24].

**Approach II: LSTM-Based Modeling with Collected NRS Data**

To enhance pain score accuracy, this approach employs LSTM-based classification. The choice of this model stems from its proficiency in capturing temporal dependencies within the data, rendering it a well-suited option for the task [25]. The choice of classification over prediction was driven by the discrete nature of NRS values (from 0 to 10). Unlike continuous prediction, classification allowed us to categorize the result values into these predefined classes, ensuring that the output was within the desired range.

Here we employed an LSTM-based classification model. The Adam optimizer with a set maximum of 500 epochs was used for training. The mini-batch size was adjusted to 64 for efficient training, and we closely monitored the training progress using the 'Verbose' option. These choices trade-off model complexity and training efficiency, while ensuring that the model had enough capacity to capture the temporal dependencies in the data.

Data processing is initiated with the extraction and normalization of input data. Subsequently, the data was divided into two distinct sets: the labeled dataset and the unlabeled dataset. The labeled dataset comprised NRS values collected every 75 minutes and the correspondent HR data. This dataset contained 20 samples which is the training size. Conversely, the unlabeled dataset consisted solely of HR data, where NRS values were not available. This set encompassed a total of 1680 samples.

During the training phase, the HR and the collected NRS were the inputs to the LSTM network. This training process depicted in Fig. 4 enabled the LSTM model to learn intricate patterns and relationships between HR and NRS scores. The nonlinear regression model operates on the training data set of HR \( \{x_t\} \) and corresponding NRS labels \( \{y_t\} \), with \( t \) denoting time steps. The LSTM model consists of the following layers:

1. **Sequence Input Layer**: This layer accepts the input data and initializes the sequence for LSTM processing. Each time step \( t \) passes the feature vector \( x_t \) to the subsequent layers.

2. **LSTM Layer**: The core of the model, the LSTM layer, operates with mathematical equations. It maintains internal variables, including:

   - **Input Gate** \( (i_t) \): The input gate controls the flow of new information into the cell state. It is computed using the sigmoid activation function \( \sigma \) and is defined as:

     \[ i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i). \]  

   - **Forget Gate** \( (f_t) \): The forget gate controls the retention of past information in the cell state and acts as a weighting factor of past-to-new data. It is computed similarly to the input gate and is defined as:

     \[ f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f). \]  

Fig. 3. Conceptual representation of how a moving window estimator uses new data for training.
- Candidate Cell State ($\tilde{C}_t$): The candidate cell state represents the new information that could be stored in the cell state. It is computed using the hyperbolic tangent activation function, defined as:

$$\tilde{C}_t = \tanh(W_{cx}x_t + W_{hh}h_{t-1} + b_c).$$  

(5)

- Cell State Update ($C_t$): The cell state $C_t$ is updated by combining the previous cell state $C_{t-1}$ with the new information from the input gate and candidate cell state:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t.$$  

(6)

- Output Gate ($o_t$): The output gate controls what information from the cell state should be used to compute the output. It is defined as:

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o).$$  

(7)

Here, $\sigma$ represents the sigmoid activation function, $\tanh$ is the hyperbolic tangent activation function, $W$ are weight matrices, they are denoted as $W_x$ and $W_h$ to represent the learnable parameters governing connections between input data ($x_t$) and the previous hidden state ($h_{t-1}$) in different layers (e.g., $i$ for input gate, $f$ for forget gate), and $b$ are bias vectors.

3. Fully Connected Layer: Following the LSTM layer, the fully connected layer prepares the data for classification. It applies learned weights and biases to the output of the LSTM.

4. Softmax Layer: This layer computes class probabilities based on the fully connected layer’s output. It provides the possibilities for each NRS class from 0 to 10.

5. Classification Layer: The final classification layer assigns an NRS class label to each time step, ensuring accurate categorization of pain levels [26].

The training process optimizes the parameters of the LSTM, resulting in a model capable of precise NRS classification for patient monitoring and care in the PACU.

Once the LSTM model was trained, in the predicting phase, the predictive loop with 1680 iterations was employed. During each iteration, the model predicted a single NRS value. This iterative process allowed us to generate 1680 predicted NRS values.

C. Performance Metrics

To evaluate the performance of the proposed methodology and estimator output, we compare the estimated NRS values $\hat{y}_i$ against the real measured NRS $y_i$ by means of root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2},$$  

(8)

the mean absolute errors (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|,$$  

(9)

and the coefficient of determination $R^2$:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2},$$  

(10)

where $\bar{y}$ is the mean of the observed data and $N$ is the total number of predicted samples.

IV. RESULTS

The results presented in Figure 5 offer valuable insights into the impact of using interpolated NRS data in the training of machine learning models for pain assessment. The figure shows the illustration of interpolated NRS data (shown as a continuous black line), and the prediction outputs of three distinct models (LR, RR, and PR).

![Fig. 4. LSTM model workflow diagram.](image)

The trained LSTM classification model generates a single NRS value for each iteration. The predictive loop is a for loop, implemented to predict NRS values for individual data points. Instead of obtaining a single NRS value for the entire sequence from the classification model, this action is repeated for 1680 iterations. The reason for employing this predictive loop is to generate a full segment of the predicted NRS.

The interpolated NRS data, which was introduced into the training process, does not contribute substantially to the information contained within the dataset. Instead, it effectively creates dependent samples, which in turn can introduce bias into the results obtained through machine learning modeling. The initial analysis of the results indicates that all three models exhibit a qualitative ability to generate predicted NRS data that closely resembles the real data. However, further scrutiny reveals that the models do not offer any significant additional information about the pain levels between the two collected NRS values. In essence, the use of interpolated NRS data appears to be redundant in the context of this study, as the models were only able to replicate the curve...
observed in the interpolated data. In conclusion, the use of interpolated NRS data in training appears to introduce a source of dependency among samples without offering substantial advantages in terms of model performance. Hence, in the second approach, only the subjectively collected NRS scores were included in the training dataset.

In the second approach, the results of the LSTM model in predicting NRS values are given.

Fig. 6 presents the prediction results, to visually examine our findings. The blue line corresponds to the predicted NRS values, the green circles denote the subjectively collected NRS, and the red dashed line corresponds to the interpolated NRS values. These results reveal the performance of the model versus the collected and interpolated NRS data for the selected patients, and the sampling time is 5 seconds.

From Fig. 6, a difference between the interpolated NRS (the red dashed line) and the predicted NRS (the blue line) can be noticed. From a control engineering point of view, the sudden bursts in the predicted NRS might lead to aggressive control actions, whereas the interpolated NRS has a smoother, more transient change of values. Ideally, the combined result of the two could be useful information for the feedback loop in optimization procedures for pain management.

On the other hand, the predicted NRS introduces more details, as exemplified in the figure. For instance, between sample numbers 500 and 600, and again between sample numbers 800 and 900, the predicted NRS score (blue line) increases from 5 to 6 (indicating a shift from moderate to severe pain) as revealed by the LSTM result. However, this crucial information is not captured by the interpolation method. The NRS predictions provide a granular understanding of pain dynamics, enabling the detection of rapid changes in the pain levels of the patient. The interpolated NRS gives us a general sense of pain. The limitation of this method is that fast changes in the pain levels of the patient are not detected. The AI approach used in this work, reveals these changes in pain, improving the understanding of patients’ response to sedation in the PACU.

To quantitatively evaluate the results of the proposed approaches, the performance metrics, including MAE, MSE, and $R^2$, are provided in Table I. These metrics were calculated by comparing the interpolated and predicted values of NRS for each approach. Analysis of these metrics reveals a difference in the performance metrics results between Approach I, which used the interpolated NRS for training, and the LSTM-based method.

Approach I used three different AI models (PR, RR, and LR), and produced results with high $R^2$ values, indicating alignment with the interpolated NRS values. This suggests that the information derived from Approach I was consistent with the interpolation method.

In contrast, Approach II, which employed an LSTM-based model, produced results with slightly higher errors, particularly a lower $R^2$ value, indicating deviations from the interpolated NRS values. This suggests that Approach II captured changes in the predicted NRS values, providing a more detailed and objective assessment of pain levels within the PACU, and potentially providing valuable insights into patient care.

<table>
<thead>
<tr>
<th>Approach</th>
<th>AI model</th>
<th>MAE</th>
<th>MSE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>PR</td>
<td>0.491</td>
<td>0.639</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>RR</td>
<td>0.335</td>
<td>0.427</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.497</td>
<td>0.631</td>
<td>0.89</td>
</tr>
<tr>
<td>II</td>
<td>LSTM</td>
<td>0.524</td>
<td>0.795</td>
<td>0.72</td>
</tr>
</tbody>
</table>

**V. DISCUSSION**

In this study focused on PACU pain assessment using NRS levels, we compared two prediction approaches: one using interpolated data and another using an LSTM model without interpolation. Our findings showed that the LSTM approach was more effective, offering a deeper, more objective insight into pain level variations. Unlike the first method, which closely mirrored the interpolated values, the LSTM model revealed significant dynamics in pain levels, highlighting its potential to improve pain assessment.

In this research, the LSTM model and other explored AI models are not designed as direct decision-makers but as tools to enhance the data quality for further analysis or integration into control feedback loops, refining the interpretation of sparse captured pain level data for better decision-making. From a control engineering perspective, these AI approaches, particularly the LSTM model, hold promise for being incorporated into feedback loops to significantly improve real-time pain management protocols. By providing a continuous, objective, and precise estimation of the NRS, AI tools can serve as essential estimators for model-based control, potentially optimizing pain management in post-operative settings.

The use of a sliding window approach in interpolation-based training significantly impacts the estimation of pain levels. This paper introduces a dependency on the size and
step size of the window, which can influence the ability of the model to generalize and its responsiveness to fluctuations in heart rate data. A larger window might smooth out rapid changes in pain levels, potentially missing short-term variations, while a smaller window could capture these changes more accurately. However, there is a risk that the AI model might mistakenly interpret random data noise as meaningful patterns. Therefore, optimizing the sliding window parameters is crucial for achieving a balance between sensitivity and specificity in pain level estimation. This need highlights the importance of further research into the optimal configuration of the sliding window approach, including consideration of future data obtained during general anesthesia to refine and validate our estimations.

The limitation of this study is the reliance on data from a single patient with a single input feature (HR), potentially restricting the generalizability of the findings. Future research should explore larger patient populations, as well as incorporate additional physiological variables into a multivariate model. Additionally, the promise of LSTM-based models for pain assessment suggests the need for further development of personalized AI models attuned to individual pain responses.

VI. CONCLUSIONS

This paper highlights the importance of monitoring changes in pain levels in the PACU, suggesting that LSTM-based models may provide a more detailed and objective evaluation than traditional interpolation methods. This advancement could significantly enhance patient care and pain management in clinical settings.

ACKNOWLEDGEMENT

The authors thank Mihaela Ghita for her collaboration and discussion.

REFERENCES