

# Facial Expression and Electrodermal Activity Analysis for Continuous Pain Intensity Monitoring On The X-ITE Pain Database

Ehsan Othman (✉ [ehsan.othman@ovgu.de](mailto:ehsan.othman@ovgu.de))

Information Technology and Communications, Otto-von-Guericke University Magdeburg  
<https://orcid.org/0000-0002-4738-1897>

**Philipp Werner**

Department of Neuro-Information Technology, Institute for Information Technology and communications, Otto-von-Guericke University Magdeburg

**Frerk Saxen**

Department of Neuro-Information Technology, Institute for Information Technology and communications, Otto-von-Guericke University Magdeburg

**Ayoub Al-Hamadi**

Department of Neuro-Information Technology, Institute for Information Technology and communications, Otto-von-Guericke University Magdeburg

**Sascha Gruss**

Medical Psychology Division, University Hospital of Ulm, ULM

**Steffen Walter**

Medical Psychology Division, University Hospital of ULM, ULM

---

## Research

**Keywords:** continuous pain intensity recognition, facial expression, electrodermal activity, modalities, Random Forest, Long-Short Term Memory Network, sample weighting, fusion

**Posted Date:** October 5th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-927204/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

## RESEARCH

# Facial Expression and Electrodermal Activity Analysis for Continuous Pain Intensity Monitoring on the X-ITE Pain Database

Ehsan Othman<sup>1\*</sup>, Philipp Werner<sup>1</sup>, Frerk Saxen<sup>1</sup>, Ayoub Al-Hamadi<sup>1</sup>, Sascha Gruss<sup>2</sup> and Steffen Walter<sup>2</sup>

\*Correspondence:

[Ehsan.Othman@ovgu.de](mailto:Ehsan.Othman@ovgu.de)

<sup>1</sup>Department of Neuro-Information Technology, Institute for Information Technology and Communications, Otto-von-Guericke University Magdeburg, 39106 Magdeburg, Germany

Full list of author information is available at the end of the article

## Abstract

Automatic systems enable continuous monitoring of patients' pain intensity as shown in prior studies. Facial expression and physiological data such as electrodermal activity (EDA) are very informative for pain recognition. The features extracted from EDA indicate the stress and anxiety caused by different levels of pain. In this paper, we investigate using the EDA modality and fusing two modalities (frontal RGB video and EDA) for continuous pain intensity recognition with the X-ITE Pain Database. Further, we compare the performance of automated models before and after reducing the imbalance problem in heat and electrical pain datasets that include phasic (short) and tonic (long) stimuli. We use three distinct real-time methods: A Random Forest (RF) baseline methods [Random Forest classifier (RFc) and Random Forest regression (RFr)], Long-Short Term Memory Network (LSTM), and LSTM using sample weighting method (called LSTM-SW). Experimental results (1) report the first results of continuous pain intensity recognition using EDA data on the X-ITE Pain Database, (2) show that LSTM and LSTM-SW outperform guessing and baseline methods (RFc and RFr), (3) confirm that the electrodermal activity (EDA) with most models is the best, (4) show the fusion of the output of two LSTM models using facial expression and EDA data (called Decision Fusion = DF). The DF improves results further with some datasets (e.g. Heat Phasic Dataset (HTD)).

**Keywords:** continuous pain intensity recognition; facial expression; electrodermal activity; modalities; Random Forest; Long-Short Term Memory Network; sample weighting; fusion

## 1 Introduction

A reliable assessment of pain is necessary to determine appropriate and prompt treatment for vulnerable patients who are unable to self-report their pain, such as intensive care patients, people with dementia, or adults with cognitive impairment. To make the clinical observations go well, it is promising to provide an automated system due to its possibility for an objective and robust measurement and monitoring of pain [1]. Othman et al. [2] reported that machines are much better than human observers in recognising pain intensity using facial expression. Further, automatic pain recognition when analysing electrodermal activity (EDA) data improves pain assessment [3]. This work focuses on the EDA modality and on combining the facial expression and EDA modalities for continuously recognising pain intensity, and then compares the obtained results to facial expression modality results pre-

sented in [4, 5]. Such automated monitoring systems could be highly beneficial for reliable and economical pain intensity assessment.

In the study of pain recognition, the X-ITE Pain Database has been introduced to evaluate different proposed methods. It comprises reactions to pain stimuli in four different qualities: phasic (short) and tonic (long) variants of each, heat and electrical stimuli. The X-ITE Pain Database includes 134 healthy participants (subjects) between 18 and 50 years who were stimulated with heat and electricity; it involves data from different sensors such as frontal RGB camera video, audio signals, physiological signals (ECG, EMG, and EDA), for more details on the database see Gruss *et al.* [6]. Conducting experiments with healthy subjects helps developing the pain assessment technology, which later can be validated and applied with vulnerable patients, who are unable to self-report. In this work, the selected database contains 127 participants (subjects), for which data were available from all sensors. The database is extremely imbalanced. Thus, to reduce the impact of the imbalanced database, we used the same 11 datasets proposed by Othman *et al.* [4, 5]: (1) Phasic Dataset (PD), (2) Heat Phasic Dataset (HPD), (3) Electrical Phasic Dataset (EPD), (4) Tonic Dataset (TD), (5) Heat Tonic Dataset (HTD), (6) Electrical Tonic Dataset (ETD), (7) Reduced Phasic Dataset (RPD), (8) Reduced Heat Phasic Dataset (RHPD), (9) Reduced Electrical Phasic Dataset (REPD), (10) Reduced Tonic Dataset (RTD), and (11) Reduced Electrical Tonic Dataset (RETD). Reduced datasets were obtained after using a reduction strategy (see Section 2.1), which addresses the problem of imbalanced data. We focus in this study on facial expression and EDA data involving the phasic and tonic pain intensity during the application of the thermal and electrical pain stimuli and no pain.

Many studies in the past years have focused on the temporal integration of frame-level features to recognise pain intensity because it is good in describing dynamic among neighboring frames. In this work, and in line with Othman *et al.* [4, 5] who analysed facial expression to recognise continuous pain intensity, the temporal integration of EDA features is done using a time series statistics descriptor. Three different methods are applied to discriminate between no pain, three pain intensities (low, moderate, and severe) and qualities (heat and electrical stimuli) in regarding to sequence classification and regression: (1) Random Forest (RF) as baseline methods [Random Forest classifier (RFc) and Random Forest regression (RFR)], (2) Long-Short Term Memory (LSTM), and (3) LSTM using sample weighting method (called LSTM-SW).

Othman *et al.* [4, 5] reported that RF, LSTM, and LSTM-SW methods with a time series statistics descriptor called Facial Activity Descriptor (FAD) and using the reduced datasets from the X-ITE Pain Database help with the imbalanced database problem and improve the results for both classification and regression. Further, LSTM, and LSTM-SW methods were better than RF (baseline methods: RFc and RFR) to recognise continuous pain intensity. The exceptions were small datasets, for which RFs were the best but the performance of the models was still poor. This paper extends [4, 5] by investigating the same methods with EDA data and comparing the results with results when using facial expression data. In addition, we present the results of fusing the output of models based on facial expression and EDA data with mean-score mapping approach (called Decision Fusion = DF) to show whether recognition performance of continuous pain intensity improves.

The current work is organised as follows. Section 1.1 provides an overview of pain recognition methods based on facial expression and EDA data and then describes their relevance to this paper, while Section 1.2 describes the contribution of this work. In Section 2 the used material and methods are presented for automatic recognition of continuous pain intensity using electrodermal activity sensors. The X-ITE Pain Database preprocessing, feature extraction preprocessing, the experimental setup, Random Forest (RF) methods, Long-Short Term Memory (LSTM), and LSTM using sample weighting (called LSTM-SW) methods for classification and regression task are described in detail. Section 3 presents results and compares with previous work, followed by a discussion in Section 4. We conclude the results and describe the potential future works in Section 5.

### 1.1 Related Work

Many studies in the past years have focused on facial expression, which is very informative for pain recognition [7, 8]. Ekman and Friesen [9] decompose facial expression into individual facial Action Units (AUs) with the Facial Action Coding System (FACS). Several automatic systems analyse AUs and their combinations for recognising frame-level and sequence (or video)-level pain intensity. Frame-level methods such as Prkachin and Solomon Pain Intensity (PSPI) [10] have their limitation in describing relevant dynamic information that is beneficial for pain intensity recognition. Thus, many recent works focus on video-level pain recognition because it is more effective in describing such information [11, 12]. It often uses temporal integration of frame-level features. For example, the video content can be condensed to high-level features by using a time series statistics descriptor that consists of several statistical measures of the time series. According to the ability of Random Forest (RF) [13] for pain detection using facial expression [3, 12, 14], Othman et al. [2, 15] introduced RFc, with time series statistics descriptor by calculating several statistical measures with their first and second derivatives per time series, as baseline method and compare its performance to the proposed deep learning methods that analyse and RGB image encoding temporal information. The performance of reduced MobileNetV2 [15] and simple Convolutional Neural Network (CNN) [2] was better than RFc. CNN accuracy improved when using the sample weighting method by about 1% [2]. The sample weighting method is suggested to reduce the weight of misclassified samples by duplicating some training samples with more facial responses if their classification scores are above 0.3 to improve the pain intensity recognition performance. Further, LSTM [16] is an effective method for better handling time series prediction. It was proposed to learn long-term dependency among longer time periods by storing information from previous periods. Thus, Othman et al. [4, 5] utilise long-term memory (LSTM) and LSTM using sample weighting (LSTM-SW), which are significantly better than RF for recognising continuous pain intensity when using facial expressions with classification and regression. These results are used in this work for the comparison.

A promising technology used in automatic pain recognition is electrodermal activity (EDA) sensors [17]. EDA can be measured on superficial muscles of the skin of hands, which is controlled by the autonomic nervous system [18, 19]. The sweat on the skin surface changes the electrical conductivity of the skin. E.g. people sweat

when they are scared, nervous, and in pain. Thus, in earlier works of automatic pain recognition [2, 17, 20], EDA proved that it carries salient information about different pain level in people. EDA is composed of phasic and tonic signals. The phasic signal is a quick response caused by external stimuli such as pain stimuli, and the tonic signal is a slower component of the signal including the baseline of signal due to unconscious activities [19]. Werner et al. [3] trained an RF using features extracted from different physiological signals [6]. They reported that the EDA sensor is the best in recognising pain levels when using the X-ITE Pain Database. However, in contrast to our work they did not investigate continuous pain monitoring but classification of pre-segmented time windows. Further, prior works in emotion recognition such as [19] using EDA sensors were promising when using deep learning networks [Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs)]. The deep learning networks were utilised due to their ability to mine the sequential relationships between different periods of EDA signals. Thus, we train our RF and LSTM methods, which were proposed in [4, 5] with features extracted from EDA.

In [3], the authors reported that the fusion with modalities (frontal RGB camera, audio, ECG, EMG, EDA) can improve the results: first, they individually trained RF using the features of each modality; second, they concatenated the feature vectors of all modalities and trained and tested RF (called feature fusion); third, they applied decision fusion by training the RF in the individual modalities, then aggregating the RF scores into the final decisions. They used two types of aggregation: a fixed mapping and trained mapping approaches. In this work, we use fixed mapping approach on two modalities (facial expression and EDA), and we call it Decision Fusion (DF). DF for each trained model (RF, LSTM, and LSTM-SW) is based on calculating the mean of output scores per class from the two modalities and selecting the class with the highest score (classification task). In regards to regression, DF for each trained model is based on calculating the average of the predicted probabilities from the two modalities. Our main aim is to provide a highly valuable automated real-time system for clinical settings to quickly, accurately and objectively monitor the pain levels of patients.

## 1.2 Contributions

Werner et al. [3] and Walter et al. [21] report the results of using phasic (short) and tonic (long) stimulation samples from frontal RGB camera, audio, and psychological data of 7 seconds, which have been cut out from the continuous recording of the main stimulation phase in X-ITE Pain Database. This paper goes beyond their studies, as it reports the first continuous pain monitoring results based on analysing electrodermal activity (EDA) signal from X-ITE Pain Database. We use the continuous recording of the EDA signal of most of the experiment, which is about 1 hour and a half per subject. This paper additionally includes classification with discrete output categories and the comparison of obtained results with the results of continuous-valued outputs provided by regression. We report the Decision Fusion (DF) results when using facial expression and EDA modalities. We applied three automatic methods for classification and regression pain recognition EDA data sequences: (1) Random Forest classifier (RFc) and Random Forest regression (RFr) (as baseline methods), (2) Long-Short Term Memory (LSTM), and (3) LSTM using

sample weighting method (LSTM-SW). Further, we compare model performances achieved using facial expression (the results obtained from [4, 5]) with EDA and DF. All LSTMs models perform better than guessing and baseline methods (RFc and RFr) to recognise pain intensity for the two pain stimulus types (phasic and tonic) in three pain intensities (low, moderate, and severe) and two qualities (heat and electrical stimuli). In line with our recent studies, the used datasets proposed in [4, 5] to reduce the impact of imbalanced data. We show that LSTMs using EDA perform the best with most datasets and LSTMs using DF was fair with Heat Phasic Dataset (HTD) dataset.

## 2 Materials and Methods

In this section, we introduce the system structure to automatically recognise the continuous pain intensity using the facial expression in video and electrodermal activity (EDA) data on the X-ITE Pain Database. The X-ITE Pain Database preprocessing is described in Section 2.1. Our prior works focused on facial expression, see [4, 5]. The focus of this paper is on the electrodermal activity (EDA) and Decision Fusion (DF) to recognise continuous pain intensity, for more details see Section 2.2 and Section 2.3. Figure 1 shows an overview of the automatic system using frontal facial RGB video and EDA to recognise continuous pain intensity. We determined temporal integration in time window from the extracted facial features (FF) and EDA time series by using a time series statistics descriptor [called Facial Activity Descriptor (FAD) and electrodermal activity Descriptor (EDA-D)]. We moved the labels three seconds forward and then used sliding window with a time length of ten seconds ago. More details about the processing of facial video and EDA data are described in Section 2.2. For recognising continuous pain intensities, we used Random Forest (RF) as an automatic baseline method and two Long-Short Term Memory (LSTM) methods (one uses the sample weighting method and the other does not). Further, we applied Decision Fusion (DF) method, in which individual RF and LSTMs are trained for FF and EDA modalities; for more details, see Section 2.3.

### 2.1 Database preprocessing

In this section, we give an overview of the multimodal Experimentally Induced Thermal and Electrical (X-ITE) Pain Database [6], which we use to validate the performance of different automatic methods for continuous pain intensity recognition. In this database, only 127 participants (subjects) subset have data available from all sensors (frontal RGB camera, audio, ECG: electrocardiogram, EMG: surface electromyography, EDA: electrodermal activity). Alongside to Werner et al. [3] and Othman et al. [2, 4], we use this subset and focus on analysing the facial expression and EDA data from time series involving the phasic and tonic pain intensity in 3 intensities (low, medium, and high) during the application of the thermal (Medoc PATHWAY Model ATS) and electrical pain stimuli (Digitimer DS7A) and no pain. The 5 second phasic stimuli of each modality (heat and electrical pain) and intensity were repeated 30 times in randomised order with pauses of 8–12 seconds. The tonic stimuli were applied once for one minute per intensity and modality followed by a pause of five minutes. For more details about the data collection experiment, see

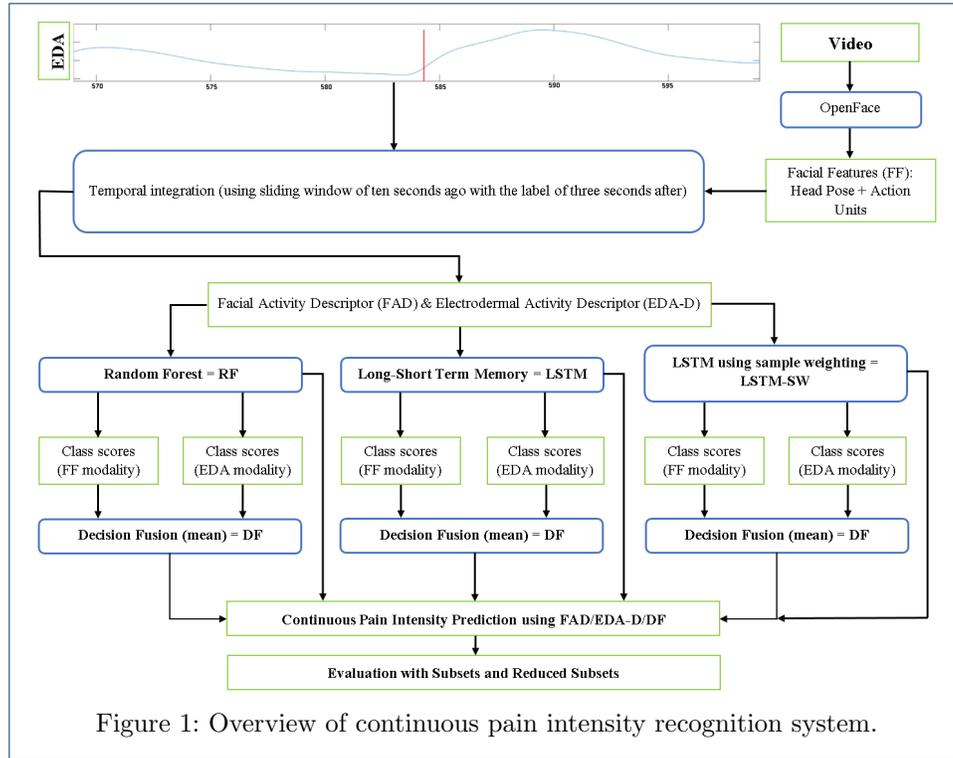
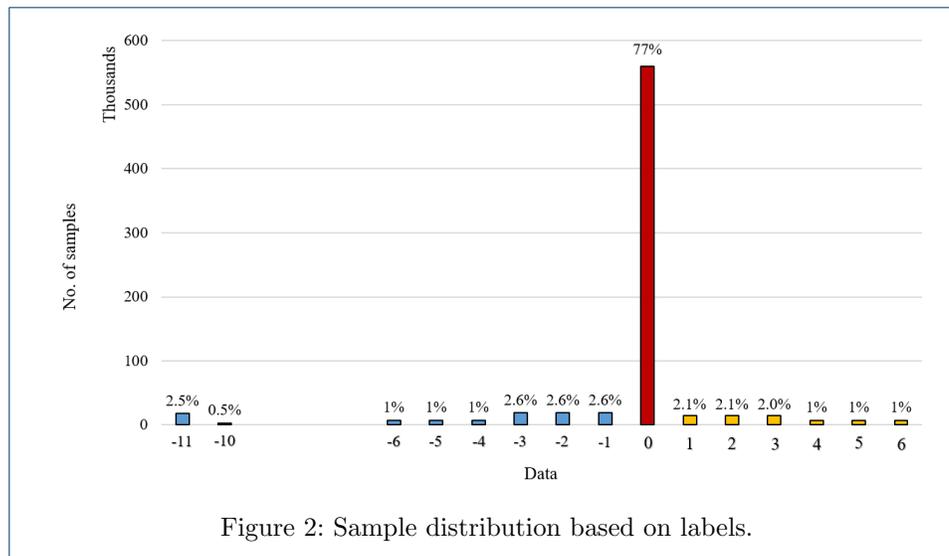


Figure 1: Overview of continuous pain intensity recognition system.

Gruss et al. [6]. Further, we explain in this paper the steps in X-ITE Pain Database preprocessing to reduce the impact of imbalanced database problem. Automatic methods should be able to recognise pain intensity from video and EDA data time series. However, we noticed that the distribution of samples for pain intensity labels is extremely unbalanced as shown in Figure 2.

In Othman et al. [4], we proposed 11 datasets obtained from the X-ITE Pain Database to reduce the imbalance database problem, which are also used in this work. First, we investigated the intensity of facial expressions for most samples when expressing pain intensity, then we assigned all subjects into four categories based on how they expressed pain intensity. Second, we suggested to split the database into 80% of data for training (100 subjects = 572696 samples), 10% for validation (13 subjects = 75537 samples), and 10% for testing (14 subjects = 79485 samples); each split contains samples from all intensity categories. We selected the subjects randomly from each category based on the proposed percentage (see Figure 3). We noticed that the distribution of samples for pain intensity labels is extremely unbalanced as shown in Figure 2. Third, we processed the database: (a) we excluded all sequences of samples with labels -10, -11 and no pain samples sequence before and after these samples to simplify the problem and reduce the impact of imbalance in database; (b) we split the obtained dataset into 6 datasets (called Subsets) to evaluate the proposed methods, (c) we reduced each proposed subset by removing some no pain samples prior to pain intensity frames in a time series for each subject, these datasets are called Reduced Subsets. The applied Subsets are (1) Phasic Dataset (PD): Exclude tonic samples (labeled 4, 5, 6, -4, -5, -6, -10, -11) and no-pain samples before these samples and also after samples with -10, -11 labeled, (2) Heat Phasic Dataset (HPD): Exclude electrical samples (labeled -1, -2, -3) from



PD and no-pain samples before these frames, (3) Electrical Phasic Dataset (EPD): Exclude heat samples (labeled 1, 2, 3) from PD and no pain frames before these frames, (4) Tonic Dataset (TD): Exclude phasic samples (labeled 1, 2, 3, -1, -2, -3, -10, -11) and no pain samples before these samples and also after samples with -10, -11 labeled, (5) Heat Tonic Dataset (HTD): Exclude electrical samples (labeled -1, -2, -3) from TD and no pain frames before these frames, and (6) Electrical Tonic Dataset (ETD): Exclude heat samples (labeled 1, 2, 3) from TD and no pain frames before these frames. The Reduced Subsets are (7) Reduced Phasic Dataset (RPD): Reduce the no pain frames in PD to about 50%, (8) Reduced Heat Phasic Dataset (RHPD): Reduce the no pain frames in HPD to about 50%, (9) Reduced Electrical phasic Dataset (REPD): Reduce the no pain frames in EPD to about 50%, (10) Reduced Tonic Dataset (RTD): Reduce the no pain frames in TD to about 38%, (11) Reduced Electrical Tonic Dataset (RETD): Reduce the no pain frames in ETD to about 49%. Our reduction strategy focuses on reducing some no pain samples prior to each pain intensity sequence by preserving different numbers of no pain samples that are directly adjacent to each pain intensity sequence, this number is assigned based on the number of samples in each pain intensity sequence, e.g. for sequence of phasic electrical pain intensity that contains five samples, we keep the previous five no pain samples and delete the rest before.

Samples with label 0 indicate no pain; samples with label -10 or -11 were not used, which indicate some experiment problems. Samples with label 3 & -3 (phasic pain stimulus) and 6 & -6 (tonic pain stimulus) indicate severe heat and electrical stimuli, respectively; samples with label 2 & -2 (phasic pain stimuli) and 5 & -5 (tonic pain stimulus) indicate moderate heat and electrical stimulus, respectively; samples with label 1 & -1 (phasic pain stimuli) and 4 & -4 (tonic pain stimuli) indicate low heat and electrical stimulus, respectively. The pain intensity labels are conditioned into the right format to automatic models by: (1) converting the negative labels (-1, -2, -3) to positive (4, 5, 6) in PD, the obtained labels are 1, 2, 3, 4, 5, 6, (2) converting the labels 4, 5, 6, -4, -5, -6 to 1, 2, 3, 4, 5, 6 in TD, (3) converting the negative labels (-1, -2, -3) to positive (1, 2, 3) in EPD, (4) converting labels 4, 5, 6 to 1, 2, 3

Intensity: 1 Subjects: S002, S010, S019, S020, S022, S033, S042, S065, S079, S082, S093, S107, S113	Intensity: 3 Subjects: S003, S004, S005, S006, S008, S009, S012, S013, S015, S017, S018, S029, S031 S032, S034, S035, S036, S038, S039, S040, S043, S045, S049, S052, S053, S054 S055, S057, S058, S061, S062, S063, S064, S066, S067, S068, S069, S070, S071 S072, S073, S075, S076, S078, S083, S084, S085, S086, S087, S088, S090, S094 S096, S097, S098, S099, S101, S105, S106, S109, S110, S111, S112, S114, S116 S117, S119, S125, S126, S127, S129, S131, S133
Intensity: 2 Subjects: S007, S011, S016, S026, S027, S028, S037, S041, S044, S046, S047, S048, S050 S051, S056, S074, S077, S080, S081, S089, S091, S092, S095, S100, S102, S103 S118, S120, S122, S123, S124, S128, S130, S132	Intensity: 4 Subjects: S021, S060, S104, S108, S115, S134

Figure 3: Assignment of subjects to categories of facial response intensities. Intensity 1= lack of facial responses to pain, Intensity 2, 3= moderate intensity of facial responses to pain, and intensity 4= intensive facial responses to pain.

in HTD, and (5) converting the labels -4 , -5, -6 to 1, 2, 3 in ETD. With regression models, no pain and pain intensity labels were normalised to range 0-1.

## 2.2 Processing of facial expression and electrodermal activity (EDA) Data

After preprocessing database in Section 2.1, we processed the data of both modalities (facial RGB video and EDA) to extract features for recognising continuous pain intensity. First, OpenFace [22] was used to extract Facial Features (FF) from each frame for each video (subject), the average length of videos is about one and a half hours. OpenFace detects the face, facial landmarks, extracts Action Units (AUs), and estimates head pose. The FF we use include 21 features: 3 head pose (Yaw, Pitch and Roll), AU1 (binary occurrence output), and 17 AU intensity outputs of OpenFace which are AU1, AU2, AU4, AU5, AU6, AU7, AU9, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26, and AU45. The FF were recorded at 25 frames per second (fps). In line with the FF extraction process, we use only the EDA data at the same time series sampling rate (1/25 seconds). We calculated temporal integration features from the 1-dimensional EDA time series and 21-dimensional facial expression time series. Each time series was summarised by four statistics of the time series itself and its first and second derivative, including minimum, maximum, mean, and standard deviation, yielding a  $12 \times 21$ -dimensional and  $12 \times 1$ -dimensional descriptor per time series for FF and EDA features, respectively. A person-specific standardisation of the features [11] was applied with Facial Activity Descriptor (FAD) and EDA Descriptor (EDA-D) in order to focus on the within-subject response variation rather than the differences between subjects. The labels of each subject were moved 3 seconds after because the facial pain responses typically are delayed by 2-3 seconds compared to stimulus. Further, We applied the sliding time window with length ten seconds once per second by combining the FAD/EDA-D of ten seconds ago to predict the next time step of pain intensity labels. The data for the first ten seconds are removed because there are no prior observations to use.

## 2.3 Classification, Regression and Fusion

The descriptor (EDA-D) is used as features for recognising continuous pain intensity (no pain, low, moderate, and severe) and modality (heat and electrical pain stimuli)

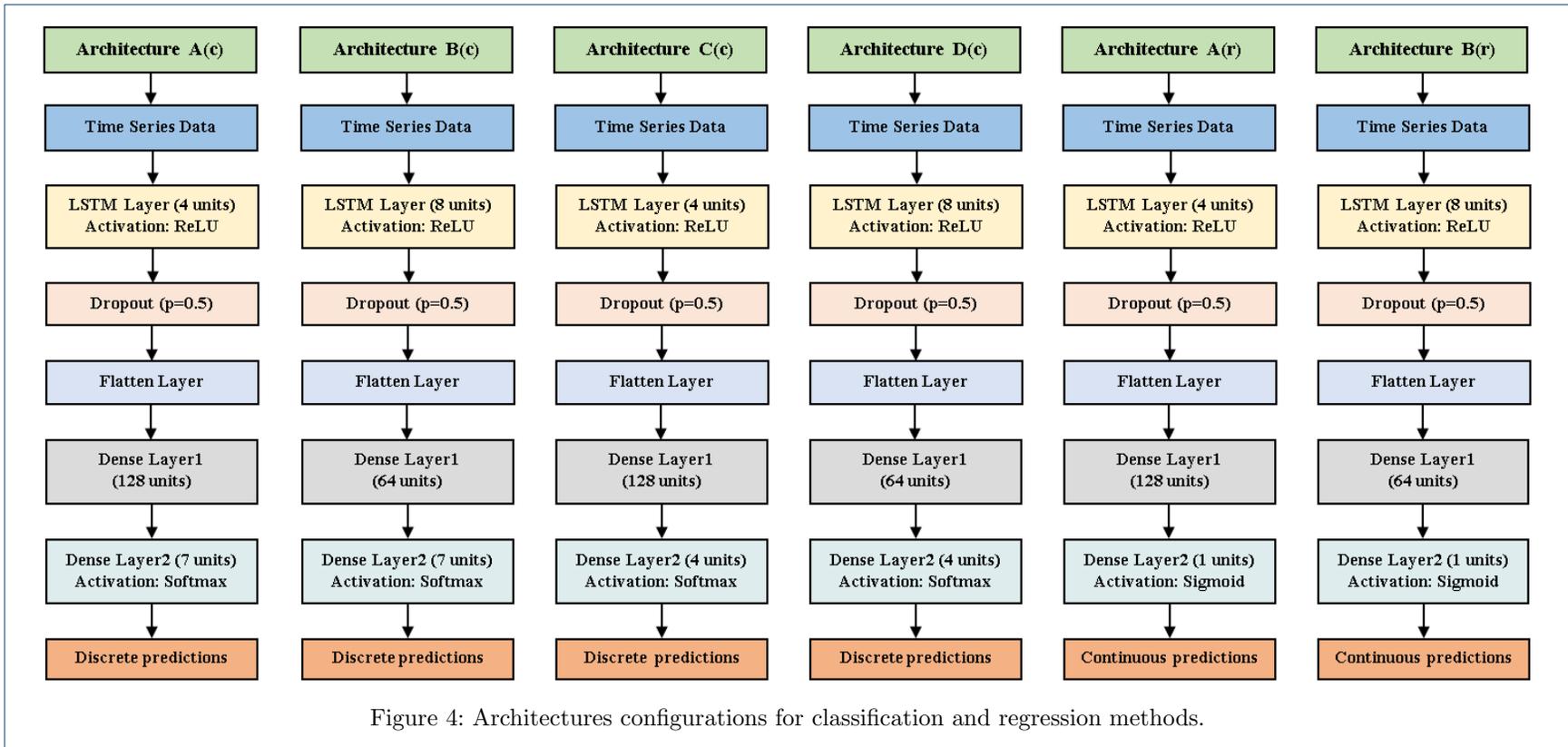
using Random Forest (RF), Long-Short Term Memory (LSTM), and LSTM using sample weighting method (LSTM-SW). We presented the results of using the above methods with FAD in [4, 5].

In line with Werner *et al.* [3] and Othman *et al.* [2, 4, 5], we trained Random Forest classifier (RFc) and Random Forest regression (RFR) with 100 trees and a maximum depth of 10 nodes for classification and regression tasks. Both RFc and RFR are the baseline methods to compare them with other deep learning methods in this study.

Figure 4 shows the six LSTM architectures used in this work: four for classification (A(c), B(c), C(c), D(c)) and two for regression (A(r), B(r)), input size is  $10 \times 252$  (10 = timesteps and 252 = features). A(c) and C(c) classification architectures comprise a single LSTM layer with 4 units activated by ReLU and followed by a flatten layer, and then one dense layer with 128 neurons activated by ReLU. The final dense output layer has 7 neurons in A(c) and 4 neurons in C(c). B(c) and D(c) classification architectures comprise a single LSTM layer with 8 units activated by ReLU and followed by a flatten layer, and then one dense layer with 64 neurons activated by ReLU. The final dense output layer has 7 neurons in B(c) and 4 neurons in D(c). The output layer is activated using Softmax function, and we use Categorical Cross-Entropy (CCE) loss function. The LSTM with this loss called LSTM-CCE. The configurations of A(r) regression architecture is similar to A(c) & C(c) and the configurations of B(r) regression architecture is similar to B(c) & D(c) except the final dense output layer with 1 neuron is activated using sigmoid function and the used loss function is the Binary Cross-Entropy (BCE). The LSTM with this loss called LSTM-BCE. The obtained models are trained for 2000 epochs with  $10^{-4}$  or  $10^{-5}$  or  $10^{-6}$  learning rate.

The applied LSTM-SW was introduced by our recent studies [2, 4, 5]. In LSTM-SW, the samples were trained on LSTM after increasing some training samples using sample weighting method [2]. We used RF with FAD to determine samples with prediction scores higher than 0.3 in training datasets and then replicated these samples once. LSTM-SW using CCE is called LSTM-SW-CCE and LSTM-SW using BCE is called LSTM-SW-BCE.

In this work, we applied Decision Fusion (DF) on the outputs from individual trained models that using facial expression and EDA modalities. The classification models (RFc, LSTM, and LSTM-SW) yield a score for each possible class and the regression models (RFR, LSTM, and LSTM-SW) yield predict a continuous value. We aggregated the classifier scores and regression outputs individually into a final decision by using a fixed mapping approach. In regards to classification, DF is implemented by calculating the mean of output scores per class of both models using FAD and EDA-D and selecting the class with the highest score. In regards to regression, all RFR, LSTM and LSTM-SW predictions were averaged individually in terms of calculating DF.



### 3 Results

This section shows the results of recognizing pain intensity and modality with 7-class and 4-class recognition tasks. 7-Class pain intensity recognition for pain stimuli modality (Heat= H and Electrical= E) are considered: (1) BL, PH1, PH2, PH3, PE1, PE2, and PE3 for phasic recognition task, (2) BL, TH1, TH2, TH3, TE1, TE2, and TE3 for tonic recognition task. Further, 4-Class pain intensity recognition for pain stimulus modality (Heat/ Electrical) are considered: (1) BL, PH1, PH2, and PH3 for phasic heat recognition task, (2) BL, PE1, PE2, and PE3 for phasic electrical recognition task, (3) BL, TH1, TH2, and TH3 for tonic heat recognition task, (4) BL, TE1, TE2, and TE3 for tonic electrical recognition task. In our experiments, we compare our methods (LSTM-CCE and LSTM-SW-CCE) with the Trivial and baseline methods (RFc and RFr) on all proposed datasets (Section 2.1) from X-ITE Pain Database regarding the classification (see Section 3.1) and regression (see Section 3.2).

#### 3.1 Classification

In regards to classification, the accuracy is used to measure and compare the performance of our models (See Section 2.3 and Figure 4) with Trivial and baseline model (RFc) in recognising continuous pain intensity, see Table 1 and Figure 5.

Table 1: Comparison of accuracy for Trivial, baseline (RFc), LSTM-CCE and LSTM-SW-CCE models using FAD (cyan), EDA-D (yellow), and DF (orange) to predict continuous pain intensity. \*  $p < 0.05$  when using paired t-test.

Datasets		Models				
		n-Class	Trivial	RFc		
				FAD	EDA-D	DF
Subsets	PD	7	77.7	76.6	75.8	78.1
	HPD	4	78.5	77.8	78.1	78.2
	EPD	4	86.1	85.4	86.1	47.7
	TD	7	70.3	68.5	50.0	67.5
	HTD	4	20.0	29.1	41.0	40.9
	ETD	4	82.0	80.7	74.3	81.9
Reduced Subsets	RPD	7	50.0	47.6	43.2	51.7
	RHPD	4	50.1	49.0	45.1	52.0
	REPD	4	50.0	49.2	55.5	58.9
	RTD	7	38.1	33.9	30.4	36.7
	RETD	4	49.0	43.1	48.7	50.4

Datasets		Models								lr
		n-Class	Archit.	LSTM-CCE			LSTM-SW-CCE			
				FAD	EDA-D	DF	FAD	EDA-D	DF	
Subsets	PD	7	A(c)	78.2*	78.6*	78.5*	77.5*	78.4*	78.8*	$10^{-5}$
	HPD	4	C(c)	78.9*	79.8	79.9*	78.0	79.0	79.9*	
	EPD	4	C(c)	86.6*	87.1*	87.2*	85.4	87.0*	87.2*	
	TD	7	A(c)	70.7*	70.7*	70.4*	70.0*	70.6*	70.9*	$10^{-6}$
	HTD	4	C(c)	32.8	48.4	45.5	33.7	47.7	43.2	
	ETD	4	C(c)	81.7	82.4*	82.4	82.4*	82.0*	81.0	
Reduced Subsets	RPD	7	A(c)	57.4*	67.3*	66.2*	54.7*	66.9*	65.5*	$10^{-4}$
	RHPD	4	C(c)	61.0*	68.6*	68.0*	58.7*	68.6*	67.2*	
	REPD	4	C(c)	61.5*	75.8*	74.9*	58.9*	75.1*	73.9*	
	RTD	7	B(c)	39.2*	42.2*	41.6*	37.4	42.7*	42.2*	$10^{-6}$
	RETD	4	D(c)	50.1*	57.7*	54.8*	47.5	57.5*	54.3*	

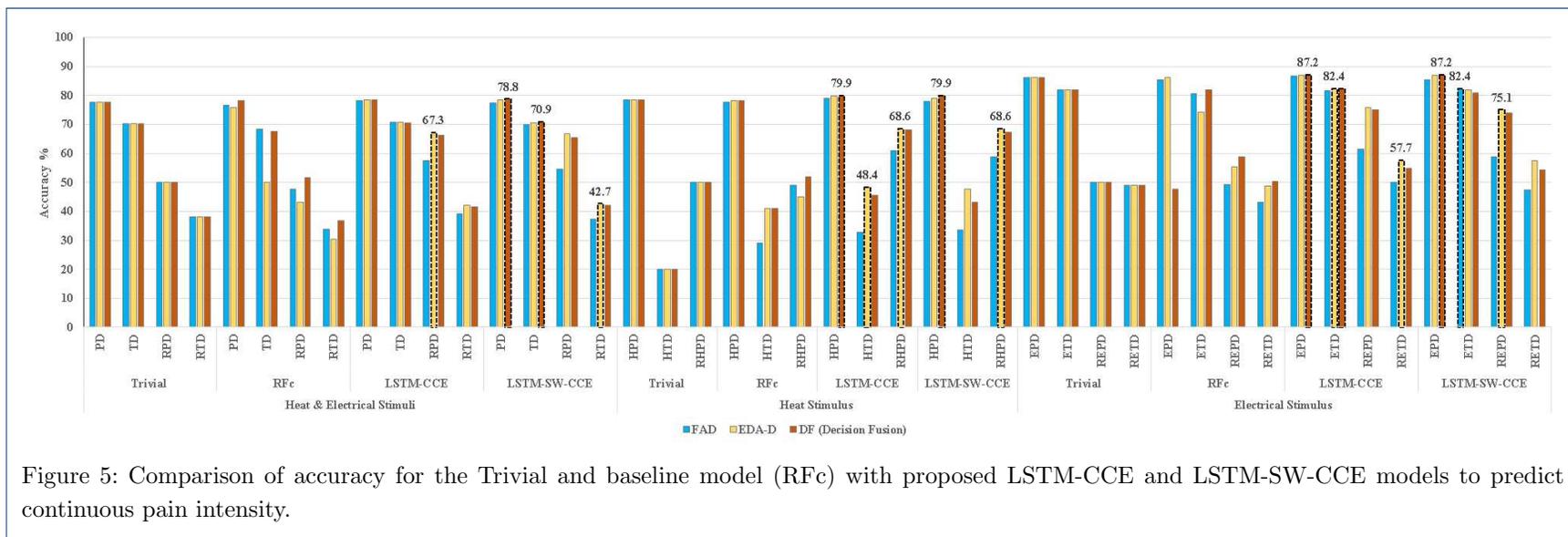


Figure 5: Comparison of accuracy for the Trivial and baseline model (RFe) with proposed LSTM-CCE and LSTM-SW-CCE models to predict continuous pain intensity.

Most RFc models perform worse than guessing. In contrast, Trivial model totally failed to recognise continuous pain intensity because it always votes for the majority class (no pain in our experiment), when RFc is better. Thus, we compared the results of the proposed models (LSTM-CCE and LSTM-SW-CCE) if they are significantly better than the baseline model (RFc) by considering the p-value of the paired t-test. Decision Fusion (DF) of 7-Class modalities using LSTM-SW-CCE with Subsets datasets improves the performance significantly compared to the single modality or DF using RFc (78.8% for PD, 79.9% for HPD, 87.2% for EPD, and 70.9% for TD using decision fusion with mean-score aggregation). LSTM-CCE with DF yields the similar results as LSTM-SW-CCE to recognise phasic heat/electrical pain intensity in Subsets (79.9% and 87.2% for HPD and EPD, respectively). Further, LSTM-CCE with EDA-D performs the best to recognise tonic heat/electrical pain intensity in Subsets, 7-Class pain intensity in Reduced Subsets, heat & electrical pain intensity in Reduced Subsets for phasic stimuli, and electrical phasic intensity (48.4%, 82.4%, 67.3%, 68.6%, 75.8%, and 57.7% for HTD, ETD, RPD, RHPD, REPD, and RETD, respectively in this sequence). LSTM-SW-CCE with EDA-D performs the best in RTD by about 42.7%. LSTM-CCE with EDA-D and DF, and LSTM-SW-CCE with FAD are the best in the ETD dataset and obtain the same performance (82.4%).

### 3.2 Classification vs Regression

This section provides the comparison between Trivial and our methods with phasic and tonic pain stimuli datasets (7-Class, see Section 3.2.1), phasic and tonic heat pain stimuli datasets (4-Class, see Section 3.2.2), and phasic and tonic electrical pain stimuli datasets (4-Class, see Section 3.2.3). Mean Squared Error (MSE) and the Intraclass Correlation Coefficient (ICC) [23] are used to measure the performance of classification models versus regression models.

#### 3.2.1 Heat and Electrical Pain Intensity Recognition (7-Class)

Table 2 shows the results of classifying all 7 available classes (datasets PD, ED, RPD, EPD) for both phasic and tonic pain, combining the pain intensity and the stimuli modality (heat and electrical pain). In regards to classification and regression, all automatic models with EDA-D and DF are superior to those with FAD. In line with Othman et al. [4, 5], our models with EDA-D and DF outperform Trivial and baseline methods (RFr and RFc). In this experiment, the fusion using the average outputs from the FAD and EDA-D modalities (DF) improves performance significantly compared to the best FAD modalities as shown in Figure 6. Further, most EDA-D modalities outperform DF modalities to recognise continuous pain intensity. The best results based on the lowest MSE and highest ICC were obtained from: (1) LSTM-BCE with EDA-D models when using PD and TD datasets (MSE of 0.06, 0.08 and ICC of 0.43, 0.12, respectively in this sequence), (2) LSTM-SW-BCE with EDA-D model when using RPD dataset (MSE of 0.04 and ICC of 0.84), and (3) LSTM-SW-CCE with EDA-D model when using RTD dataset (MSE of 0.11 and ICC of 0.31).

#### 3.2.2 Heat Pain Intensity Recognition (4-Class)

Due to the results of 7-Class pain intensity recognition, our models were trained with the 4-Class classification and regression tasks to simplify the problem and increase

the performance. For each pain stimulation modality type (phasic and tonic), we excluded electrical pain intensities; the used combinations are: BL, PH1, PH2, and PH3 (datasets HPD, RHPD), as well as BL, TH1, TH2, and TH3 (dataset HTD). The performance of our automatic models with EDA-D and DF for continuous pain recognition with both combinations (see Table 2 and Figure 7) is significantly greater than Trivial and baseline method (RFR and RFC). The best models results: (1) LSTM-BCE-SW with DF (when using HPD dataset) yields the highest ICC (0.33) and MSE of 0.08, while LSTM-BCE with EDA-D yields the ICC of 0.31 and lowest MSE (0.07), (2) LSTM-CCE with EDA and DF (when using HTD dataset) perform well based on the ICC of 0.33, 0.35 and MSE 0.15, 0.16, respectively in this sequence, and (3) LSTM-BCE with EDA-D (when using RHPD dataset) obtains the highest ICC (0.81) and lowest MSE (0.05).

### 3.2.3 Electrical Pain Intensity Recognition (4-Class)

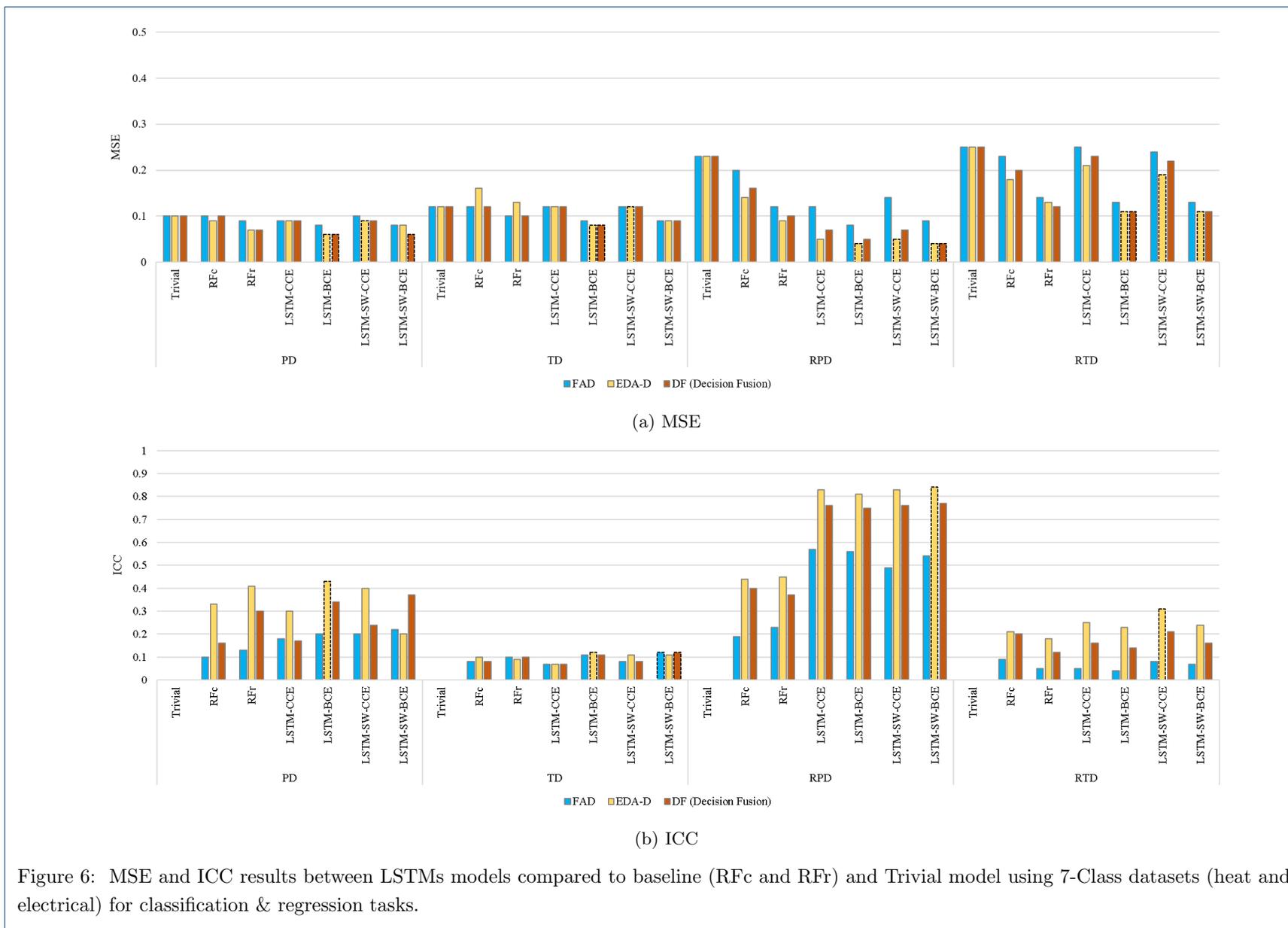
Table 2 and Figure 8 show the high performance in recognising continuous pain intensity of the 4-Class classification and regression tasks. 4 classes for electrical stimulation method are considered, we excluded heat stimuli: BL, PE1, PE2, and PE3 for phasic pain stimulation (datasets EPD, REPD) and BL, TE1, TE2, and TE3 for tonic pain stimulation (datasets ETD, RETD). The electrical pain recognition models with EDA-D and DF show better performance than Trivial and baseline method (RFR and RFC), same models with FAD provided the best as shown in Othman et al. [4, 5]. LSTM-SW-BCE with EDA-D model when using EPD, ETD, and REPD datasets performed the best based on the highest ICC (0.53, 0.21, 0.88) and MSE of 0.05, 0.07, 0.03, respectively in this sequence. LSTM-BCE with EDA-D model when using RETD yield the highest ICC (0.49) and lowest MSE (0.10). Further, RFC with DF when using EPD dataset performs the worse because most of no pain samples were labeled with pain after fusion of the FAD and EDA-D modalities.

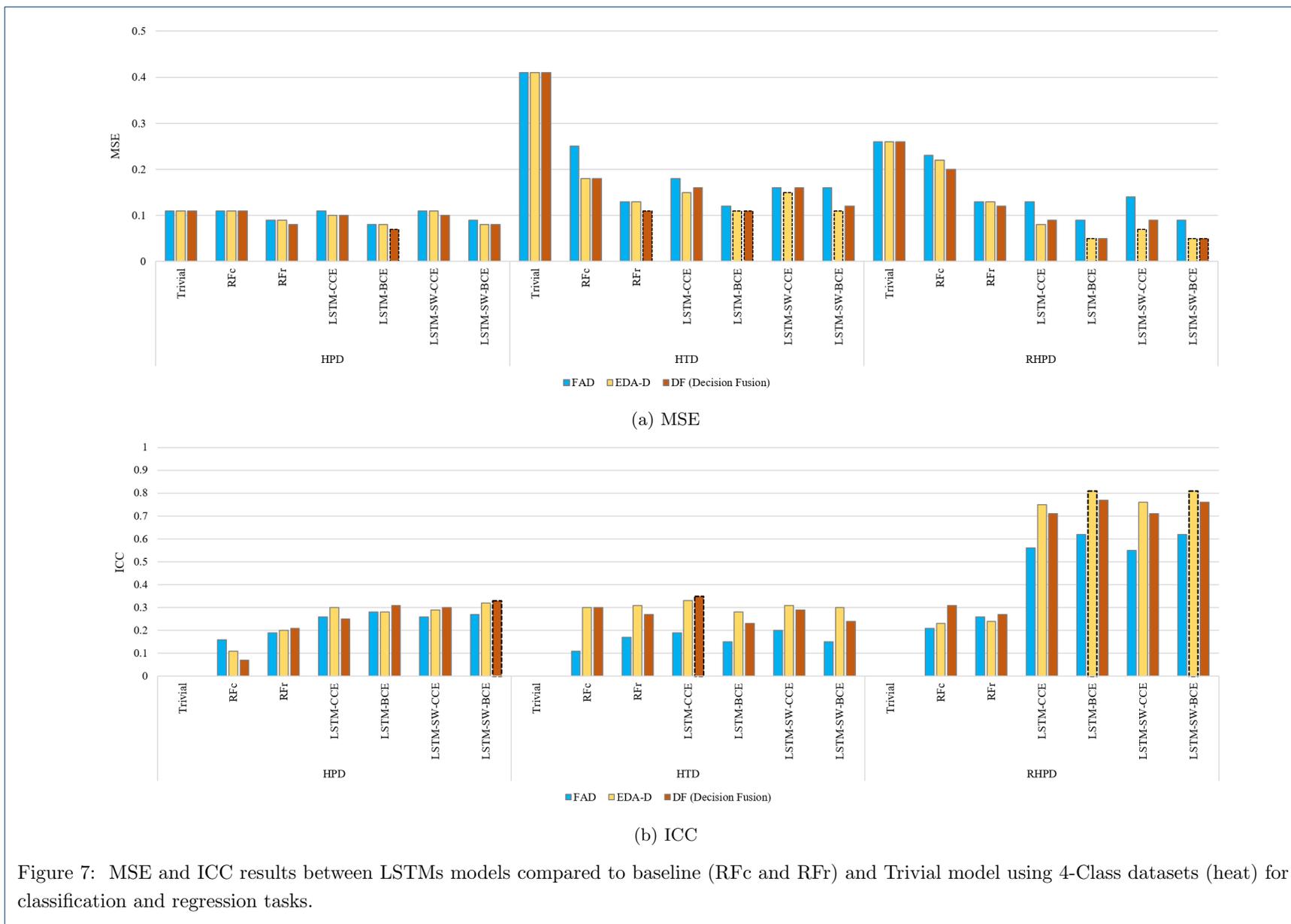
## 4 Discussion

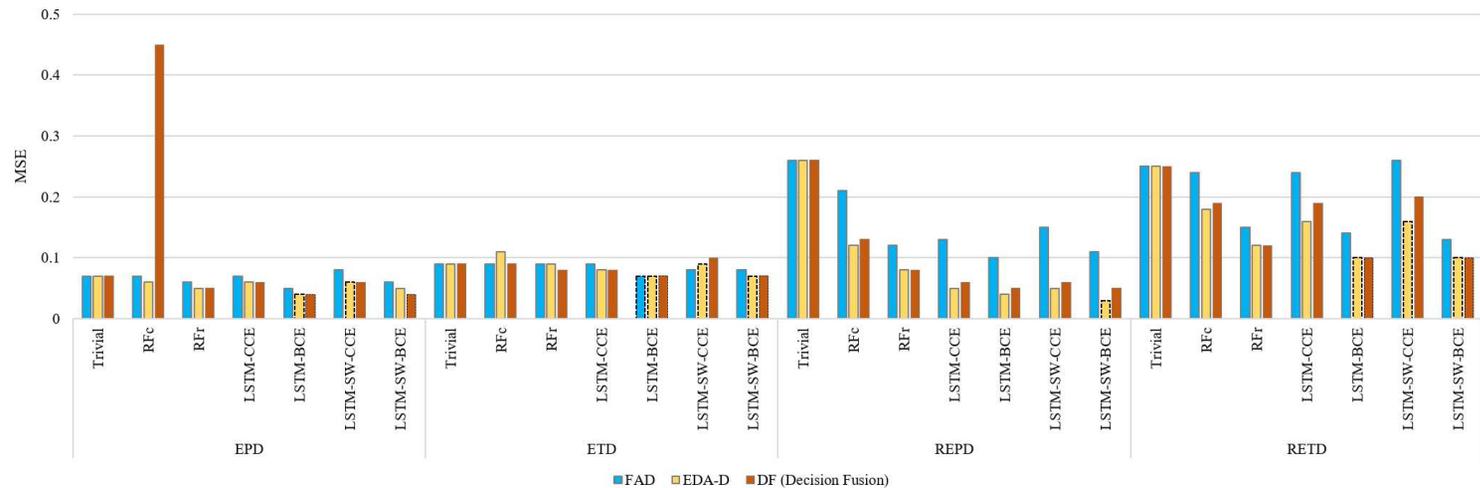
In this paper, we conducted experiments to compare the performance between different automatic methods in recognising continuous pain on the X-ITE Pain Database, when using two modalities (Frontal RGB camera and EDA), and the fusion of these two modalities, see Table 1, 2 and Figures 5, 6, 7, 8. The results in both phasic and tonic datasets show that it is possible to monitor the continuous pain intensity, which is in line with results of our recent works [4, 5] and many prior studies. The results of analysing the facial expression features were reported in Othman et al. [4, 5] studies and are used here for comparison. In this work, we trained RF (RFC and RFR as baseline models) and LSTMs (LSTM-CCE, LSTM-SW-CCE, LSTM BCE and LSTM- SW-BCE models) with EDA-D and DF on several datasets from X-ITE Pain Database (the used datasets and methods are described in Section 2.1 and Section 2.3). To achieve continuous pain intensity recognition, we used the sliding window strategy for obtaining 10s-length input samples. Further we move the labels for each subject 3 seconds after. The results show that (1) the models using networks A and C perform better than models using networks B and D, and (2) LSTMs models using EDA-D and DF are significantly better than guessing and most baseline models (RFC and RFR).

Table 2: Comparison the MSE and ICC measurements results for Trivial, baseline model (RFc and RFr), and LSTMs regarding classification and regression tasks. Measurements:Meas., Reduced Subsets: Red. Subsets, Architecture:Archit., learning rate:lr.

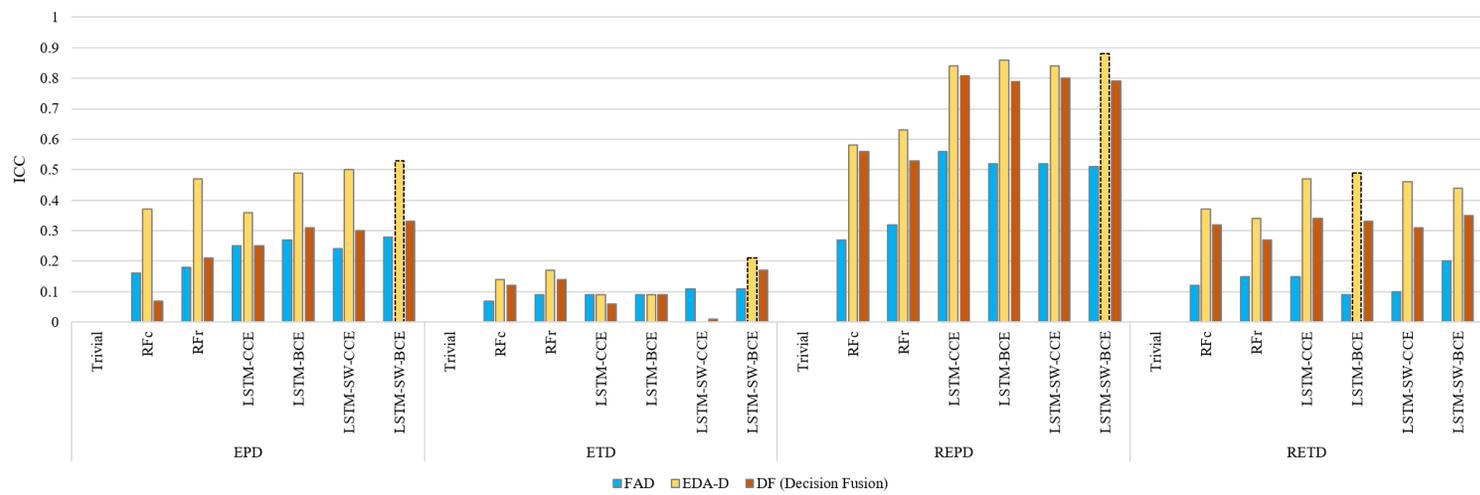
Meas.	Task		Classification										Regression						lr					
	Models	Trivial	RFc			Archit.	LSTM-CCE			LSTM-SW-CCE			RFr			Archit.	LSTM-BCE			LSTM-SW-BCE				
			FAD	EDA-D	DF		FAD	EDA-D	DF	FAD	EDA-D	DF	FAD	EDA-D	DF		FAD	EDA-D		DF	FAD	EDA-D	DF	
MSE	Subsets	PD	0.10	0.10	0.09	0.10	A(c)	0.09	0.09	0.09	0.10	0.09	0.09	0.09	0.07	0.07	A(r)	0.08	0.06	0.06	0.08	0.08	0.06	10 <sup>-5</sup>
		HPD	0.11	0.11	0.11	0.11	C(c)	0.11	0.10	0.10	0.11	0.11	0.10	0.09	0.09	0.08	C(r)	0.08	0.08	0.07	0.09	0.08	0.08	
		EPD	0.07	0.07	0.06	0.45	C(c)	0.07	0.06	0.06	0.08	0.06	0.06	0.06	0.05	0.05	C(r)	0.05	0.04	0.04	0.06	0.05	0.04	
	Subsets	TD	0.12	0.12	0.16	0.12	A(c)	0.12	0.12	0.12	0.12	0.12	0.12	0.10	0.13	0.10	A(r)	0.09	0.08	0.08	0.09	0.09	0.09	10 <sup>-6</sup>
		HTD	0.41	0.25	0.18	0.18	C(c)	0.18	0.15	0.16	0.16	0.15	0.16	0.13	0.13	0.11	C(r)	0.12	0.11	0.11	0.16	0.11	0.12	
		ETD	0.09	0.09	0.11	0.09	C(c)	0.09	0.08	0.08	0.08	0.09	0.10	0.09	0.09	0.08	C(r)	0.07	0.07	0.07	0.08	0.07	0.07	
	Red. Subsets	RPD	0.23	0.20	0.14	0.16	A(c)	0.12	0.05	0.07	0.14	0.05	0.07	0.12	0.09	0.10	A(r)	0.08	0.04	0.05	0.09	0.04	0.04	10 <sup>-4</sup>
		RHPD	0.26	0.23	0.22	0.20	C(c)	0.13	0.08	0.09	0.14	0.07	0.09	0.13	0.13	0.12	C(r)	0.09	0.05	0.05	0.09	0.05	0.05	
		REPD	0.26	0.21	0.12	0.13	C(c)	0.13	0.05	0.06	0.15	0.05	0.06	0.12	0.08	0.08	C(r)	0.10	0.04	0.05	0.11	0.03	0.05	
		RTD	0.25	0.23	0.18	0.20	B(c)	0.25	0.21	0.23	0.24	0.19	0.22	0.14	0.13	0.12	B(r)	0.13	0.11	0.11	0.13	0.11	0.11	10 <sup>-6</sup>
RETD		0.25	0.24	0.18	0.19	D(c)	0.24	0.16	0.19	0.26	0.16	0.20	0.15	0.12	0.12	D(r)	0.14	0.10	0.10	0.13	0.10	0.10		
ICC	Subsets	PD	0	0.10	0.33	0.16	A(c)	0.18	0.30	0.17	0.20	0.40	0.24	0.13	0.41	0.30	A(r)	0.20	0.43	0.34	0.22	0.20	0.37	10 <sup>-5</sup>
		HPD	0	0.16	0.11	0.07	C(c)	0.26	0.30	0.25	0.26	0.29	0.30	0.19	0.20	0.21	C(r)	0.28	0.28	0.31	0.27	0.32	0.33	
		EPD	0	0.16	0.37	0.16	C(c)	0.25	0.36	0.26	0.24	0.50	0.33	0.18	0.47	0.37	C(r)	0.27	0.49	0.42	0.28	0.53	0.47	
	Subsets	TD	0	0.08	0.10	0.11	A(c)	0.07	0.07	0.01	0.08	0.11	0.07	0.10	0.09	0.10	A(r)	0.11	0.12	0.12	0.12	0.11	0.11	10 <sup>-6</sup>
		HTD	0	0.11	0.30	0.30	C(c)	0.19	0.33	0.35	0.20	0.31	0.29	0.17	0.31	0.27	C(r)	0.15	0.28	0.23	0.15	0.30	0.24	
		ETD	0	0.07	0.14	0.12	C(c)	0.09	0.09	0.06	0.11	0.00	0.01	0.09	0.17	0.14	C(r)	0.09	0.09	0.09	0.11	0.21	0.17	
	Red. Subsets	RPD	0	0.19	0.44	0.40	A(c)	0.57	0.83	0.76	0.49	0.83	0.76	0.23	0.45	0.37	A(r)	0.56	0.81	0.75	0.54	0.84	0.77	10 <sup>-4</sup>
		RHPD	0	0.21	0.23	0.31	C(c)	0.56	0.75	0.71	0.55	0.76	0.71	0.26	0.24	0.27	C(r)	0.62	0.81	0.77	0.62	0.81	0.76	
		REPD	0	0.27	0.58	0.56	C(c)	0.56	0.84	0.81	0.52	0.84	0.80	0.32	0.63	0.53	C(r)	0.52	0.86	0.79	0.51	0.88	0.79	
		RTD	0	0.09	0.21	0.20	B(c)	0.05	0.25	0.16	0.08	0.46	0.21	0.05	0.18	0.12	B(r)	0.04	0.23	0.14	0.07	0.24	0.16	10 <sup>-6</sup>
RETD		0	0.12	0.37	0.32	D(c)	0.15	0.47	0.34	0.10	0.46	0.31	0.15	0.34	0.27	D(r)	0.09	0.49	0.33	0.20	0.44	0.35		







(a) MSE



(b) ICC

Figure 8: MSE and ICC results between LSTMs models compared to baseline (RFc and RFr) and Trivial model using 4-Class datasets (electrical) for classification and regression tasks.

In line with Werner *et al.* [3] results, the accuracy performance of RFc with DF instead of FAD (Facial Activity Descriptor) improved the almost balanced Heat Tonic Dataset (HTD) (only 20% of samples experience no pain). The possible reason for the poor performance of other models is that our datasets are highly imbalanced. However, RFc with DF increased pain intensity recognition performance but at the same time decreased no pain recognition performance when input data are sequences. Thus, we improved the continuous pain intensity recognition and its qualities (heat and electrical stimuli) using LSTMs methods because they work better with imbalanced datasets (see [4, 5]).

The LSTMs with DF and EDA-D outperform LSTMs using FAD, DF is best in most results of the Subsets (huge imbalanced datasets) in regards to classification (see Table 1). LSTMs with EDA-D perform the best when using Reduced Subsets, it seems that the amount of EDA data was appropriate with the complexity of the problem and informative when using LSTMs. In Reduced Subsets, we only preserved some adjacent no pain samples prior to each sub-sequence of pain intensity, so that the number of no pain samples is the same as the number of samples in the adjacent pain intensity sub-sequence, e.g. for a sub-sequence of tonic heat pain of low intensity that contains 5 samples, we keep the previous 5 no pain samples and delete the rest before. The reduced datasets (Reduced Subsets) were suggested to reduce the impact of huge imbalanced datasets which maybe include a lot of outlier samples. In Reduced Subsets, the obtained results show that LSTMs were significantly better than RFc by at least 10% in reduced phasic datasets and at least 5% in tonic phasic datasets. In contrast, the results of most datasets in Subsets show that LSTMs improve the accuracy of about 2.5% (maximum) compared to RFc except with Heat Tonic Dataset (HTD), which improved by up to 7%.

In regards to regression, the performance of most models with EDA-D outperform models with FAD and are better than several models with DF, see Table 2. Further, the performance of the LSTM-BCE model with EDA-D was very good for pain intensity monitoring when using Reduced Electrical Phasic Dataset (REPD) (ICC about 0.88). It might be due to reducing the noise or outlier data and more data of pain intensities. LSTM-SW-BCE model with DF also performs fair when using Heat Phasic Dataset (HPD), ICC about 0.33. Most of the LSTMs models using EDA-D achieve the best performance, moreover, LSTM-SW-BCE with EDA-D increased the performance compared to LSTM-BCE with EDA-D of several models and it performs the best with RPD, EPD, and ETD. This leads to the hypothesis that the success of LSTM-SW using the binary cross-entropy loss function is based on down-weighting samples in training set with less facial response using sample weighting method [2]. These samples might negatively affect the model performance. LSTM-CCE with EDA-D and DF models perform fairly and the best when using the HPD and RTD, about 0.35 and 0.31, respectively. The possible reason why the classification model performs well is that the HPD is almost balanced dataset, and the balanced data are good for classification. In accordance with Othman *et al.* [4, 5], we confirm that regression models are superior to classification models when using huge imbalanced datasets.

## 5 Conclusions

Prior works with the X-ITE Pain Database [3, 21] focused on parts (time windows) from the continuous recording of the main stimulation phase. These time windows have been cut out from the continuous recording frontal RGB camera, audio, and physiological data of 7 seconds. See [3] for more details about the time windows. So this paper advances by introducing continuous monitoring system using the frontal RGB camera and EDA data for pain intensity recognition in the continuous recording of the main stimulation phase of about one hour and a half per subject.

As we have seen in the promising results in the discussion section, an automated system can be provided which is able to continuously monitor the pain intensity of patients. In regards to classification and regression, we applied three methods [Random Forest (RF), Long-Short Term Memory (LSTM), and LSTM using sample weighting method (called LSTM-SW)] with Facial Activity Descriptors (FAD), Electrodermal Activity Descriptors (EDA-D), and Decision Fusion (DF) for continuous pain intensity recognition on several datasets from X-ITE Pain Database. The major difficulties for using machine learning classifier or regressor are the huge data imbalance and outliers. Thus, we used the reduced datasets to simplify and to reduce the impact of imbalanced database problem by reducing no pain samples. The applied strategy is reported in Section 2.1 and [4, 5]. The best results were obtained with LSTM models. Generally, they outperformed baseline models (RFc and RFr) and guessing (the majority of votes = no pain). LSTM models most clearly outperform other models when using the Reduced Subsets (see Table 1, 2 and Figure 5, 6, 7, and 8). Most LSTM models with EDA-D are the best for continuous monitoring pain intensity. LSTM model with DF was fair when using HPD because the fusion of facial expressions and EDA modalities was good with the complexity of the problem of the heat phasic dataset. Further, the results showed that regression is superior to classification for huge imbalanced data. The classification performed better than regression with Heat Tonic Dataset (HTD) because it is almost balanced, and the best result when using LSTM-CCE with DF. This result is consistent with Werner *et al.* [3] findings, who used a balanced database for classification of pre-segmented time windows. LSTM model with EDA-D using Reduced Tonic Dataset (RTD) is the best regarding classification. Although, the performance of tonic datasets improved compared to Othman *et al.* [4, 5] who used FAD only (see Table 1 and 2), the performance is still worse compared to the phasic datasets due to the small size of the data. Thus, we assume more data with more pain intensities will improve LSTMs or other deep learning methods. Further, finding a solution for imbalanced datasets and using other modalities (Audio, EMG, and ECG) could improve the results.

**Author Contributions:** Methodology, software, validation, investigation, and visualization: E. Othman; formal analysis, writing—original draft preparation: E. Othman, P. Werner; data gathering: S. Gruss, S. Walter; conceptualisation, writing—review and editing: all authors; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is supported by German Academic Exchange Service (DAAD), and the Federal Ministry of Education and Research of Germany (BMBF);

projects HuBA no. 03ZZ0470, RoboAssist no. 03ZZ0448L, and Robo-Lab no. 03ZZ04X02B within the Zwanzig20 Alliance 3Dsensation for support.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of the Ulm University (protocol code: 372/16, date of approval: 5 January 2017).

**Conflicts of Interest:** The authors declare no conflict of interest.

### Abbreviations:

The following abbreviations are used in this manuscript:

Archit.	architecture
CNN	simple Convolutional Neural Network
ECG	Electrocardiogram
EDA	Electrodermal activity
EDA-D	Electrodermal activity Descriptor
EMG	Electromyography
EPD	Heat Phasic Dataset
ETD	Electrical Tonic Dataset
FACS	Facial Action Coding System
FAD	Facial Activity Descriptor
HPD	Heat Phasic Dataset
HTD	Heat Tonic Dataset
ICC	Intraclass Correlation Coefficient
K-NN	K-Nearest Neighbors
LSTM	Long-Short Term Memory
LSTM-BCE	LSTM using Binary Cross-Entropy loss function
LSTM-CCE	LSTM using Categorical Cross-Entropy loss function
LSTM-SW	Long-Short Term Memory using sample weighting
LSTM-SW-BCE	LSTM-SW using Binary Cross-Entropy loss function
LSTM-SW-CCE	LSTM-SW using Categorical Cross-Entropy loss function
lr	learning rate
Meas.	Measurements
MSE	Mean Squared Error
PD	Phasic Dataset
PE1	Phasic electrical stimulus with low pain intensity
PE2	Phasic electrical stimulus with moderate pain intensity
PE3	Phasic electrical stimulus with severe pain intensity
PH1	Phasic heat stimulus with low pain intensity
PH2	Phasic heat stimulus with moderate pain intensity
PH3	Phasic heat stimulus with severe pain intensity
Red. Subsets	Reduced Subsets
REPD	Reduced Electrical Phasic Dataset
RETD	Reduced Electrical Tonic Dataset
RHPD	Reduced Heat Phasic Dataset
RF	Random Forest
RFc	Random Forest classifier
RFr	Random Forest regression
RPD	Reduced Phasic Dataset
RTD	Reduced Tonic Dataset
SVM	Support Vector Machine
TD	Tonic Dataset
TE1	Tonic electrical stimulus with low pain intensity
TE2	Tonic electrical stimulus with moderate pain intensity
TE3	Tonic electrical stimulus with severe pain intensity
TH1	Tonic heat stimulus with low pain intensity
TH2	Tonic heat stimulus with moderate pain intensity
TH3	Tonic heat stimulus with severe pain intensity

### Author details

<sup>1</sup>Department of Neuro-Information Technology, Institute for Information Technology and Communications, Otto-von-Guericke University Magdeburg, 39106 Magdeburg, Germany. <sup>2</sup>Medical Psychology Division, Department of Psychosomatic Medicine and Psychotherapy, University Hospital of Ulm, 89081 Ulm, Germany.

## References

1. R. Mieronkoski, E. Syrjälä, M. Jiang, A. Rahmani, T. Pahikkala, P. Liljeberg, S. Salanterä, Developing a pain intensity prediction model using facial expression: A feasibility study with electromyography. *PLoS ONE* 15(7), (2020). DOI:10.1371/journal.pone.0235545
2. E. Othman, P. Werner, F. Saxen, A. Al-Hamadi, S. Gruss, S. Walter, Automatic vs. Human Recognition of Pain Intensity from Facial Expression on the X-ITE Pain Database. *Sensors* 21(9), 1-19 (2021). DOI:10.3390/s21093273
3. P. Werner, A. Al-Hamadi, S. Gruss, S. Walter, Twofold-Multimodal Pain Recognition with the X-ITE Pain Database, in: The international Conference on Affective Computing and Intelligent Interaction Cambridge, UK, 3-6 September (2019). DOI:10.1109/ACIIW.2019.8925061
4. E. Othman, P. Werner, F. Saxen, A. Al-Hamadi, S. Gruss, S. Walter, Regression Networks for Automatic Pain Intensity Recognition in Video using Facial Expression on the X-ITE Pain Database, in: The 25th Int'l Conf on Image Processing, Computer Vision, Pattern Recognition (ICIP'21), Las Vegas, USA, 26-29 July (2021).
5. E. Othman, P. Werner, F. Saxen, A. Al-Hamadi, S. Gruss, S. Walter, Classification vs Regression Networks for Continuous Automatic Pain Intensity Recognition in Video using Facial Expression on the X-ITE Pain Database. *Computers in Biology and Medicine journal*. Manuscript submitted for publication.
6. S. Gruss, M. Geiger, P. Werner, O. Wilhelm, H. Traue, A. Al-Hamadi, S. Walter, Multi-Modal Signals for Analyzing Pain Responses to Thermal and Electrical Stimuli. *J Vis Exp*. 146, e59057 (2019). DOI:10.3791/59057
7. A. C. d. C Williams, Facial expression of pain: An evolutionary account. *Behavioral and Brain Sciences* 25(4), 439-455 (2002). DOI:10.1017/S0140525X02000080
8. P. Werner, D. Lopez-Martinez, S. Walter, A. Al-Hamadi, S. Gruss, R. W. Picard, Automatic Recognition Methods Supporting Pain Assessment: A Survey. *Transactions on Affective Computing*, 1-23 (2019). DOI:10.1109/TAFFC.2019.2946774
9. P. Ekman, W. V.Friesen, Facial action coding system: A technique for the measurement of facial movement. Consulting Psychologists Press: Palo Alto, CA, USA, (1978). DOI:10.2217/nmt-2015-0006
10. K. M. Prkachin, P. E. Solomon, The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. *Pain* 139(2), 267-274 (2008). DOI:10.1016/j.pain.2008.04.010
11. P. Werner, A. Al-Hamadi, K. Limbrecht-Ecklundt, S. Walter, H. C. Traue, Automatic Pain Assessment with Facial Activity Descriptors, in *Transactions on Affective Computing* 8(3), 286-299 (2017). DOI:10.1109/TAFFC.2016.2537327
12. M. Kächele, P. Thiam, M. Amirian, P. Werner, S. Walter, F. Schwenker, G. Palm, Multimodal Data Fusion for Person-Independent, Continuous Estimation of Pain Intensity, in: *Engineering Applications of Neural Networks: 16th International Conference, Rhodes, Greece, 25 - 28 September (2015)*. DOI:10.1007/978-3-319-23983-5\_26
13. L. Breiman, Random Forests. *Machine Learning* 45(1), 5-32 (2001). DOI:10.1023/A:1010933404324
14. P. Thiam, V. Kessler, S. Niese, M. Amirian, P. Bellmann, G. Layher, Y. Zhang, M. Velana, S. Gruss, S. Walter, H. C. Traue, J. Kim, D. Schork, E. Andre, H. Neumann, F. Schwenker, Multi-modal Pain Intensity Recognition based on the SenseEmotion Database. *IEEE Transactions on Affective Computing*. (2019). DOI:10.1109/TAFFC.2019.2892090
15. E. Othman, P. Werner, F. Saxen, A. Al-Hamadi, S. Walter, Cross-Database Evaluation of Pain Recognition from Facial Video, in *International Symposium on Image and Signal Processing and Analysis (ISPA)*, Dubrovnik, Croatia, 23-25 September (2019). DOI:10.1109/ISPA.2019.8868562
16. S. Zhou, J. Schmidhuber, Long Short-Term Memory. *Neural Computation* 9(8), 1735-1780 (1997). DOI:10.1162/neco.1997.9.8.1735
17. B. T. Susam, M. Akcakaya, H. Nezamfar, D. Diaz, X. Xu, V. R. d. Sa, K. D. Craig, J. S. Huang, M. S. Goodwin, Automated Pain Assessment using Electrodermal Activity Data and Machine Learning, in: *Conf Proc IEEE Eng Med Biol Soc.*, Honolulu, Hawaii 17-21 July (2018). pp 372-375. DOI:10.5244/C.27.119
18. Christie, M.J., Electrodermal activity in the 1980s: a review. *J R Soc Med*. 74(8), 616-622 (1981). DOI:10.1023/A:1010933404324
19. D. Yu, S. Sun, A Systematic Exploration of Deep Neural Networks for EDA-Based Emotion Recognition. *Information* 11(4), (2020). DOI:10.3390/info11040212
20. S. A. H. Aqajari, R. Cao, E. K. Naeini, M.-D. Calderon, K. Zheng, N. Dutt, P. Liljeberg, S. Salanterä, A. Nelson, A. M. Rahmani, Pain Assessment Tool with Electrodermal Activity for Post-Operative Patients: A Method Validation Study. *JMIR mhealth and uhealth* 9(5), (2020). DOI:10.2196/25258
21. S. Walter, A. Al-Hamadi, S. Gruss, S. Frisch, H. C. Traue, P. Werner, Multimodale Erkennung von Schmerzintensität und-modalität mit maschinellen Lernverfahren. *Der Schmerz* 34(5), 400-409 (2020). DOI:10.1007/s00482-020-00468-8
22. T. Baltrusaitis, P. Robinson, L.-P. Morency, OpenFace: An open source facial behavior analysis toolkit. *Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid, NY, USA, 7-10 March (2016). DOI:10.1109/WACV.2016.7477553
23. P. E. Shrout, J. L. Fleiss, Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 86(2), 420-428 (1979). DOI:10.1037//0033-2909.86.2.420

## Supplementary Files

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

- [ProbandeninformationEmailSchmerzstudie.pdf](#)