Journal of Thermal Biology 113 (2023) 103523

Contents lists available at ScienceDirect

Journal of Thermal Biology

journal homepage: www.elsevier.com/locate/jtherbio

Application of machine learning algorithms in thermal images for an automatic classification of lumbar sympathetic blocks

Mar Cañada-Soriano [a], Maite Bovaira [b], Carles García-Vitoria [b], Rosario Salvador-Palmer [c], Rosa Cibrián Ortiz de Anda [c], David Moratal [d,**], José Ignacio Priego-Quesada [c,e,*]

[a] Applied Thermodynamics Department (DTRA), Universitat Politècnica de València, Valencia, Spain
[b] Anaesthesia Department, Hospital Intermutual de Levante, Sant Antoni de Benaixeve, Valencia, Spain
[c] Research Group in Medical Physics (GIFIME), Department of Physiology, University of Valencia, Valencia, Spain
[d] Center for Biomaterials and Tissue Engineering, Universitat Politècnica de València, Valencia, Spain
[e] Research Group in Sports Biomechanics (GIBD), Department of Physical Education and Sports, University of Valencia, Valencia, Spain

ARTICLE INFO

Keywords:
Infrared thermography
Medicine
Complex regional pain syndrome
Sympathetic ganglia

ABSTRACT

Purpose: There are no previous studies developing machine learning algorithms in the classification of lumbar sympathetic blocks (LSBs) performance using infrared thermography data. The objective was to assess the performance of different machine learning algorithms to classify LSBs carried out in patients diagnosed with lower limbs Complex Regional Pain Syndrome as successful or failed based on the evaluation of thermal predictors.

Methods: 66 LSBs previously performed and classified by the medical team were evaluated in 24 patients. 11 regions of interest on each plantar foot were selected within the thermal images acquired in the clinical setting. From every region of interest, different thermal predictors were extracted and analysed in three different moments (minutes 4, 5, and 6) along with the baseline time (just after the injection of a local anaesthetic around the sympathetic ganglia). Among them, the thermal variation of the ipsilateral foot and the thermal asymmetry variation between feet at each minute assessed and the starting time for each region of interest, were fed into 4 different machine learning classifiers: an Artificial Neuronal Network, K-Nearest Neighbours, Random Forest, and a Support Vector Machine.

Results: All classifiers presented an accuracy and specificity higher than 70%, sensitivity higher than 67%, and AUC higher than 0.73, and the Artificial Neuronal Network classifier performed the best with a maximum accuracy of 88%, sensitivity of 100%, specificity of 84% and AUC of 0.92, using 3 predictors.

Conclusion: These results suggest thermal data retrieved from plantar feet combined with a machine learning-based methodology can be an effective tool to automatically classify LSBs performance.

1. Introduction

Complex regional pain syndrome (CRPS) is a chronic pain condition that commonly affects one limb, and it is characterized by disproportionate and prolonged intense pain (Harden et al., 2013; Shim et al., 2019). The afflicted extremity often presents increased sensitivity, and

Abbreviations: ΔTMax, The variation for ipsilateral foot difference in maximum temperature; ΔTMean, The variation for ipsilateral foot difference in mean temperature; ΔSD, The variation for ipsilateral foot difference in standard deviation temperature; ΔAsymMax, The asymmetry variation between ipsilateral and contralateral foot in maximum temperature; ΔAsymMean, The asymmetry variation between ipsilateral and contralateral foot in mean temperature; ΔAsymSD, The asymmetry variation between ipsilateral and contralateral foot in standard deviation temperature; ANN, Artificial neural networks; AUC, Area under the curve; CRPS, Complex regional pain syndrome; ES, Cohen effect size; FOV, Field of view; GUI, Graphical user interface; IRT, Infrared thermography; IFOV, Instantaneous field of view; KNN, K-Nearest neighbours; LSB, Lumbar sympathetic block; ML, Machine learning; NETD, Noise equivalent temperature difference; RF, Random forest; ROI, Region of interest; SVM, Support vector machine.

* Corresponding author. Department of Physical Education and Sports, Faculty of Physical Activity and Sport Sciences, Universitat de València, C/ Gascó Oliag, 3, 46010, Valencia, Spain.
** Corresponding author. Center for Biomaterials and Tissue Engineering, Universitat Politècnica de València, Camí de Vera, s/n, 46022, Valencia, Spain.
E-mail addresses: dmoratal@eln.upv.es (D. Moratal), j.ignacio.priego@uv.es (J.I. Priego-Quesada).

https://doi.org/10.1016/j.jtherbio.2023.103523
Received 8 November 2022; Received in revised form 30 January 2023; Accepted 13 February 2023
Available online 17 February 2023

most patients suffer from vasomotor, sudomotor and trophic disturbances (Marinus et al., 2011; Harden et al., 2010). The main symptom of vascular dysfunction is edema but abnormalities in coloration and in skin temperature are also observed (Stanton-Hicks et al., 2018). CRPS is usually precipitated by fractures or surgery (Borchers and Gershwin, 2014), and its aetiology is related to inflammatory mechanisms, sympathetic and somatosensory nervous system dysfunction, or central pain perception changes (Harden et al., 2013). It is worth noting that patients presenting CRPS have notable quality of life impairment, therefore, early treatment initiation in the clinical course of this condition is of paramount importance (Money, 2019).

Interventional approaches include sympathetic blocks which are performed to reduce symptoms and to achieve pain relief (Day, 2008). When lower extremities are affected, sympathetic blocks are performed between lumbar vertebral levels L2 and L4 (Qian et al., 2019; An et al., 2016) by injecting a local anaesthetic around the lumbar sympathetic ganglia (Gofeld et al., 2018). Although there is no gold standard to perform the guidance of a lumbar sympathetic block (LSB), fluoroscopic guidance is the imaging technique most frequently used (Zhu et al., 2019; Ryu et al., 2018). To check whether the needle tip is placed in the precise point, i.e., reaching the sympathetic chain (Gofeld et al., 2018), the proper contrast dye spread must be confirmed within the radioscopic images. Nevertheless, this assumption does not always ensure an exact performance, since the two-dimensional nature of the radioscopic images may lack of information regarding depth level, and consequently, may result in an inadequate technical block (Cañada-Soriano et al., 2021).

On the other hand, a skin temperature warming in the affected extremity after administering the local anaesthetic is expected. To date, in order to verify whether the intervention was successful or not, palpation based on the warming sensation has been widely used in the clinical setting. However, this technique may fail to distinguish subtle temperature differences. As a general example in the medical field, it has been reported that palpation cannot accurately detect the presence of a fever (Singh et al., 2003) or differences in legs' temperature up to 4.3 °C. For this reason, infrared thermography (IRT) was used in this work to monitor the thermal alterations. Although IRT is also 2D, it has been proved as a valuable technique in evaluating the performance of LSBs based on thermal alteration within the ipsilateral (Cañada-Soriano et al., 2021). Specifically, a previous study showed that in 32% of the total LSBs, the needle reposition was necessary since no temperature alterations in the affected extremity were identified (Cañada-Soriano et al., 2021). Therefore, since the palpation for detecting the cut-off temperature values of a successful LSB may be questionable, the use of IRT as a supplementary technique to radioscopic images has been preferred.

Over the last years, machine learning (ML) has become broadly used in the analysis of medical images (Magalhaes et al., 2021; Seo et al., 2020). ML is regarded as an artificial intelligence field which is capable of learning from data samples through mathematical and statistical techniques (Houssein et al., 2021). There are different ML techniques available, and their choice mainly relies on its performance. Infrared data combined with ML algorithms have been evaluated in previous studies concerning different biomedical applications (Magalhaes et al., 2021). Breast cancer diagnosis is one of the most investigated topics applying thermal imaging in conjunction with different ML classifiers, such as Artificial Neural Networks (ANN) (Roslidar et al., 2020), Decision Trees (Raghavendra et al., 2016), K-Nearest Neighbours (KNN) (Araújo et al., 2014), Random Forest (RF) or Support Vector Machines (SVM) (Gogoi et al., 2019). In diabetic foot detection, the combination of IRT with ML algorithms has also proved to be useful (Maldonado et al., 2020; Adam et al., 2018) and other conditions such as rheumatoid arthritis (Umapathy et al., 2018), cardiovascular disease (Jayanthi and Anburajan, 2019), or prediction of different stages of cellulite (Bauer et al., 2020) have been also evaluated with ML algorithms. Specifically, in the evaluation of the LSBs' performance, ML can be used to achieve more objective classifications, to assess the learning technique process of the medical staff or even to reduce the time required for the procedures' classification in the clinical setting. In this regard, the authors have observed that, although the visualization of thermal images in real-time allows medical staff to evaluate the procedure, those with less experience take longer and have doubts about it. Conversely, those more experienced may make mistakes trying to both reduce time and interpret the existence of thermal changes when these are not yet significant. For this reason, an automatic classification through ML may be necessary.

The aim of this study was to assess the performance of different ML algorithms to automatically classify LSBs performed by clinicians as successful or failed based on IRT predictors extracted from the plantar of the foot. According to previous LSBs quantification by means of IRT (Cañada-Soriano et al., 2021), it was hypothesized a high accuracy of ML methods to classify LSBs performance. To the authors' knowledge, this study provides the first investigation that uses ML along with IRT to monitor and classify LSBs performance.

2. Materials and methods

2.1. Patients

The thermal recordings corresponding to 24 patients who underwent a set of three LSBs were analysed. The inclusion criteria included patients who: presented signs and symptoms in only one lower limb and met the Budapest clinical diagnostic criteria (Harden et al., 2007, 2010).

With a clinical presentation less than a year from the initial injury, and still presented pain rates greater than 5 (in a 10-rate-scale) a month after the standard therapy (both physical rehabilitation and pharmacological treatment). Budapest clinical diagnostic criteria is considered the most accepted diagnostic approach of CRPS in which symptoms and signs must be reported on several categories (Harden et al., 2007, 2010). The procedures were conducted in Hospital Intermutual de Levante (Valencia, Spain) from November 2019 to May 2021 in 24 patients with an age of 41 ± 9 years old (mean ± standard deviation), out of which 18 were men and 6 were women. Hospital Intermutual de Levante is an insurance company's hospital and for this reason, the number of patients diagnosed with CRPS is usually greater than in other more general hospitals. The study complied with the Declaration of Helsinki and was approved by the Ethics Committee of the Universitat de València (Valencia, Spain) (ref. 1 250779), and patients signed an informed written consent before starting the procedures.

2.2. Lumbar sympathetic block procedures

LSBs were performed by a skilled clinician team consisted of one or two pain medicine physicians, one or two nurses and an X-ray technician. Before undergoing a LSB, patients were asked to fast for 6 h and to avoid smoking during the previous hour. Prior the procedures, patients lied on a stretcher with surgical booties placed on their feet for 15 min. Hereinafter, patients were placed in prone position with bare feet, their backs were sterilely prepared, and they stayed in that position for 10 min. After that, procedures were performed using a needle 15 cm, 22-gauge, aiming the fourth lumbar. The technique was performed under fluoroscopic guidance using a C-arm (Flexiview, General Electrical Medical System, Salt Lake City, UT) and oblique, lateral, and anteroposterior view images were obtained to ensure the needle's proper site of entry. After the confirmation of correct spread agent (1.5 ml Omnipaque®) on the radioscopic images, a local anaesthetic (2 ml lidocaine 2%) was injected. Since it induces vasodilation, and based on previous results, thermal changes in the affected plantar foot were related to the proper needle placement, and therefore, to a successful LSB (e.g., an increase between 1.4 and 2.1 °C, 4 min after the anaesthetic injection, of the mean skin temperature of the ipsilateral foot) (Cañada-Soriano et al., 2021).

As mentioned in the previous section, although 24 patients were enrolled to the study, not everyone underwent a complete set of three LSBs because of different reasons (e.g., the COVID lockdown or technical problems during the thermal acquisition). Thus, of the initial 72 LSBs

that initially would have been performed, only 66 could be collected and analysed.

2.3. Thermal data acquisition

The acquisition of infrared images was performed in the same operating room with a controlled ambient temperature of 22 ± 0.5 °C and relative humidity 47 ± 5%, checked with a digital weather station Testo 623 (Testo SE&Co, Lenzkirch, Germany). During the acclimatization period of the patient (10 min), the thermal acquisition set up was prepared. To monitor thermal changes on the plantar surface, a FLIR E60 camera (FLIR Systems, Inc, Wilsonville, OR) was used with a pixel infrared resolution of 320 × 240, a field of view (FOV) of 25° × 19°, an instantaneous field of view (IFOV) of 1.36 mrad, a thermal resolution (NETD; noise equivalent temperature difference) of <50 mK at 30 °C and measurement uncertainty of ±2 °C of the overall temperature reading. The camera was mounted on a tripod at a distance of 1.2 m from the participants' feet and perpendicular to them so that both plantar feet were included within the image. The acquisition of infrared images was manually started right after the lidocaine test, and from that moment, they were automatically recorded every 10 s with the software FLIR Tools + (FLIR Systems, Inc, Wilsonville, OR), and with the emissivity fixed at 0.98 for skin measurements (Steketee, 1973). The thermal images evaluated were retrieved from 6 min-acquisitions starting from the lidocaine test for both failed and successful procedures.

2.4. Medical classification of lumbar sympathetic blocks based on thermal images

The medical procedure followed by the medical team in each patient was based both on the anaesthesiologists' previous experience assessing LSBs under fluoroscopic guidance and on previous studies (Ryu et al., 2018; Park et al., 2010). LSBs performance classification (successful or failed; outcome input of the training data) was performed by the medical team in real time observing the IRT images during the procedure in a qualitatively way. Hence, when thermal patterns (consisting in isolated warm small spots which became enlarged over time along with a progressively increase of their temperatures) were observed in the ipsilateral foot within the first minutes after the injection of the anaesthetic, LSBs were classified as successful, and the medication (levopuvicaine 0.25% 10 ml with 80 mg of triamicolone) was injected. Conversely, when no thermal patterns were observed (failed LSBs), the repositioning of the needle was carried out, and the process starting from the lidocaine, injection was repeated. Among those cases, if after the needle reposition, the thermal patterns on the ipsilateral were detected, the medication was finally injected, and the LSB was considered successful. On the other hand, when no thermal changes were still observed, the previous steps were repeated. Considering the anatomy of the lumbar ganglia where the procedures were performed and in order to avoid possible complications in the patient related to the needle placement, only three consecutive repositioning manoeuvres were carried out at most.

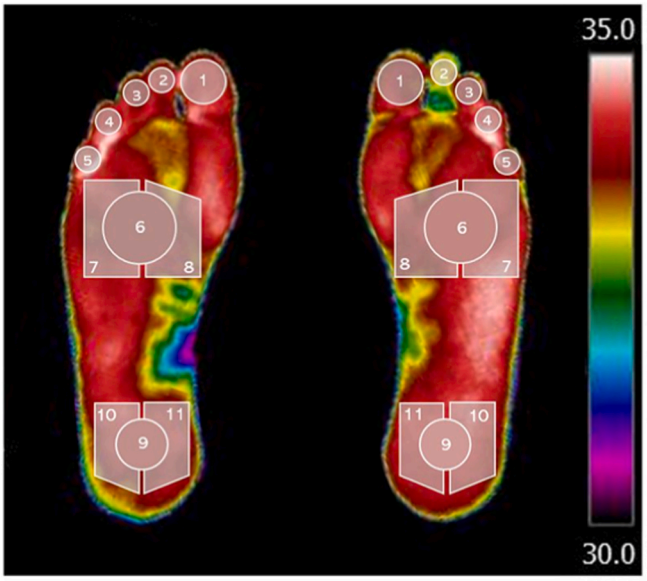

Fig. 2. Regions of interest indicated in an IR image selected for segmentation and extraction of relevant predictors. Regions were: 1) toe 1, 2) toe 2, 3) toe 3, 4) toe 4, 5) toe 5, 6) central metatarsal, 7) lateral metatarsal, 8) medial metatarsal, 9) central heel, 10) lateral heel, and 11) medial heel.

Table 1
The 66 predictors extracted.

Predictor	ROIs evaluated	Extremity evaluated	Description
ΔTMean	11	Ipsilateral	$Tmean_t$-$Tmean_0$
ΔTMax	11		$Tmax_t$ -$Tmax_0$
ΔSD	11		SD_t -SD_0
ΔAsymMean	11	Ipsilateral and contralateral	$\Delta TMean_{ipsi}$-$\Delta TMean_{contra}$
ΔAsymMax	11		$\Delta TMax_{ipsi}$-$\Delta TMax_{contra}$
ΔAsymSD	11		ΔSD_{ipsi}- ΔSD_{contra}

NOTE: "t" being the moment assessed: minute 4, minute 5 or minute 6. "0" being the baseline time (just after the lidocaine injection).

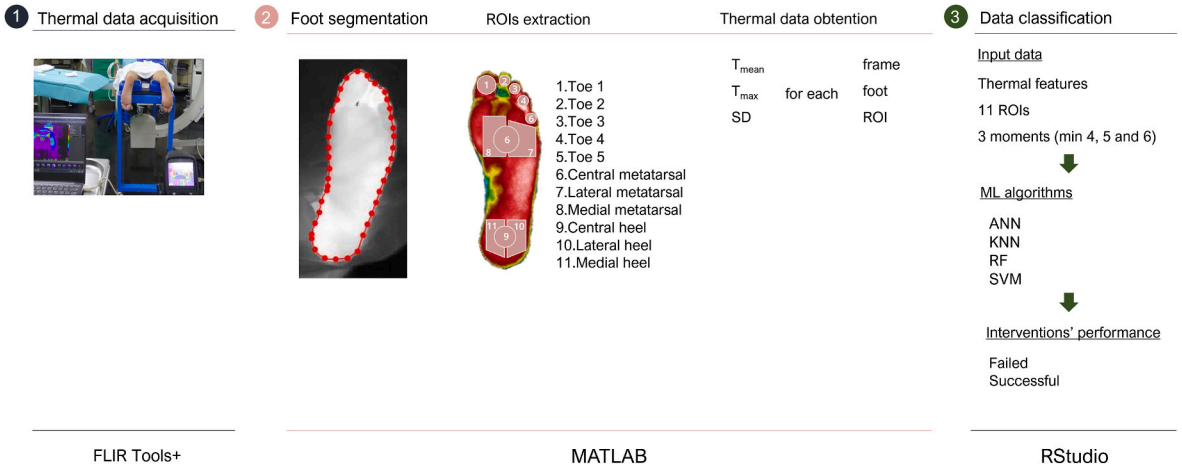

Fig. 1. Flowchart of the procedure from the lidocaine test to the LSBs performance classification using machine learning algorithms.

2.5. Thermal predictors' extraction

As depicted in Fig. 1, several steps were involved in the LSBs performance classification using ML algorithms. Once thermal data were acquired in the clinical setting, the segmentation of both plantar feet was conducted through a semi-automatic software tool developed under MATLAB (The MathWorks, Inc, Natick, MA) with a GUI (Graphical User Interface) to guide the user through the analysis steps (Rubio Mayo, 2021). First, the segmentation of both plantar feet using thresholding was performed. Then, a MATLAB code developed from Gauci et al. (2018) was used to extract the regions of interest (ROIs), within the plantar foot that is, 1 to 5 being the toes, 6 to 8 being the central, lateral, and medial metatarsal areas of the foot, and 9 to 11 being the central, lateral, and medial heel (see Fig. 2). To avoid the impact of eventual variations in the position of patients' feet, the user was able to perform modifications during the process such as resizing and/or repositioning any ROI every several images in a manual way through the GUI. Same size and position of the ROI within the foot was used for each participant in their images. Finally, the mean, maximum, as well as standard deviation from each ROI were obtained.

According to an authors' previous study (Cañada-Soriano et al., 2021; Cañada Soriano, 2022), three different assessing times (minutes 4, 5 and 6) after the baseline time (lidocaine injection) have been considered, because the first moment when the ipsilateral mean temperature presented an increase of moderate effect was at minute 4 from the lidocaine test, whereas the maximum temperature and SD presented moderated effect increases at minutes 5 and 6 respectively.

Thus, from the thermal parameters extracted, 66 predictors were then obtained at three-time points (Table 1): the variation for ipsilateral foot difference in ROI n (for n = 1 to 11 between the minute measured (4, 5 and 6) and at starting time in mean temperature (ΔTMean), maximum temperature (ΔTMax) and standard deviation (ΔSD). Additionally, the asymmetry variation between ipsilateral and contralateral foot at minute measured (4, 5 and 6) and the starting time for the 11 ROIs in mean temperature (ΔAsymMean), maximum temperature (ΔAsymMax) and standard deviation (ΔAsymSD) were obtained.

2.6. Machine learning algorithms and classification methods

ML algorithms were used to predict LSBs classification using the 66 plantar thermal predictors previously extracted. The evaluation of the classification methods was performed using the Caret package (Kuhn, 2008) in RStudio (Version 1.2.5033). The performance of four different machine learning classifiers was compared: Artificial Neuronal Network (ANN), K-Nearest Neighbours (KNN), Random Forest (RF), and Support Vector Machine (SVM).

The classification methods were applied with the variation of the data in three different moments: minute 4, minute 5 and minute 6. The data of successful and failed LSBs, were randomly split preserving relative class sizes in each training and testing sample, using 60% for training and 40% for testing. This percentage of testing cases was determined to ensure a minimum number of failed cases (6 cases) in the testing dataset. Random splits were the same for the four supervised classification algorithms. Pre-process was performed to the predictors consisting of centering (subtracts the mean) and scaling (divides by the standard deviation). Before applying the classification algorithms, for each classification algorithm and minute assessed, a recursive feature selection (RFE) algorithm was performed to select the predictors used. The classification process was performed using a cross-validation structure with 10 repetitions for the split training data. L2 regularization was implemented for all the models. Moreover, hyperparameters for ANN, KNN, RF and SVM were optimized. For KNN, the pre-set number of considered neighbours K values were $K \in \{1, 3, \ldots, 15\}$. For RF the hyperparameters were mtry $\in \{2, 4, \ldots, 14\}$. For SVM, the hyperparameters were $C \in \{2^{-2}, 2^{-1}, \ldots, 2^{2}\}$ and $\Sigma \in \{10^{-2}, 10^{-1}, \ldots, 10^{2}\}$. SVM was performed with Radial Basis Function Kernel. Finally, for ANN, after a preliminary analysis, it was defined a size of 10 neurons and the decay was of 0.0001, 0.1 and 0.5. The ANN used a logistic classification method with a Multilayer Perceptron structure with 1 hidden layer composed by the 10 neurons.

2.7. Statistical analysis and predictor evaluation

Firstly, differences between failed and successful LSBs were assessed at each variable (ΔTMean, ΔTMax, ΔSD, ΔAsymMean, ΔAsymMax, and ΔAsymSD) and at each moment without considering the ROIs. As non-normal distribution was observed (Shapiro-Wilk test, $p < 0.05$), Friedman tests of two factors (moment [minute 4, 5 and 6] and performance [failed vs. successful]) with Mann-Whitney U post-hoc tests were performed. For significant differences (p significance stablished $\alpha = 0.05$), the Cohen effect size (ES) was calculated to determine the effects size and they were classified as small (0.2–0.5), moderate (0.5–0.8) or large (>0.8).

Classification performance of all methods and moments assessed was quantified by the accuracy, the sensitivity, the specificity, the Kappa coefficient, and the area under the curve (AUC) (Tharwat, 2021). Finally, the contribution of the predictor for the best models (based on AUC) was quantified using the SHAP value, which identifies for each predictor whether it has a positive or negative contribution (Mangalathu et al., 2020). Confusion matrixes and SHAP values analysis was performed on the models with the higher AUC values obtained.

3. Results

All the variables assessed (ΔAsymMax, ΔAsymMean, ΔAsymSD, ΔTMax, ΔTMean, ΔSD) presented higher values at successful cases ($p <$

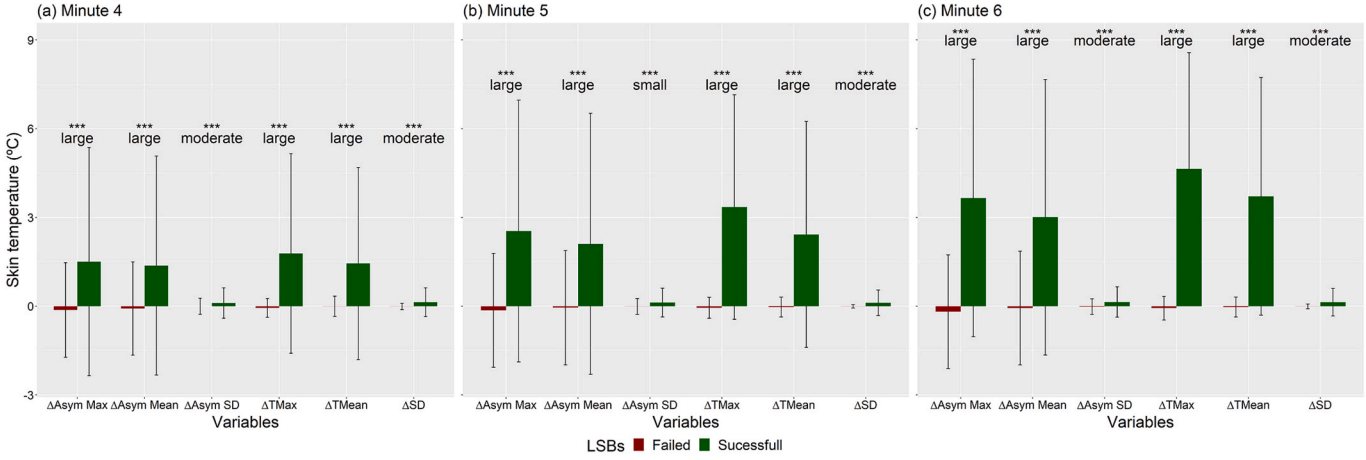

Fig. 3. Median and standard deviation of the variables assessed for minutes 4 (a), 5 (b), and 6 (c) after the lidocaine injection. Differences between successful and failed Lumbar Sympathetic Blocks (LSBs) are shown by symbols (***$p < 0.001$) and the magnitude of the effect size (large, moderate, and small).

Table 2
Performance parameters for the classification methods assessed (ANN: Artificial Neuronal Network, KNN: K-Nearest Neighbours, RF: Random Forest, SVM: Support Vector Machine) at the three variations moments between baseline and minutes 4, 5, and 6.

	ANN	KNN	RF	SVM
t = 4 min				
No. Of predictors	8	25	21	15
Accuracy	0.84	0.84	0.84	0.80
Kappa value	0.64	0.61	0.64	0.48
Sensitivity	1.00	0.83	1.00	0.67
Specificity	0.79	0.84	0.79	0.84
AUC	0.89	0.84	0.89	0.75
t = 5 min				
No. Of predictors	3	11	6	13
Accuracy	0.88	0.84	0.80	0.76
Kappa	0.72	0.61	0.57	0.41
Sensitivity	1.00	0.83	1.00	0.67
Specificity	0.84	0.84	0.74	0.79
AUC	0.92	0.84	0.87	0.73
t = 6 min				
No. Of predictors	4	5	7	17
Accuracy	0.88	0.80	0.80	0.80
Kappa	0.72	0.57	0.53	0.53
Sensitivity	1.00	1.00	0.83	0.83
Specificity	0.84	0.74	0.79	0.79
AUC	0.92	0.87	0.81	0.81

0.05; Fig. 3), increasing its differences over time, except in the SD variables (mean difference at min 4, 5 and 6: ΔAsymMax 3.1 °C vs. 4.0 °C vs. 4.7 °C, ΔAsymMean 2.8 °C vs. 3.7 °C vs. 4.3 °C, and ΔAsymSD 0.2 °C vs. 0.2 °C vs. 0.2 °C, ΔTMax 3.1 °C vs. 4.1 °C vs. 4.9 °C, ΔTMean 2.8 °C vs. 3.7 °C vs. 4.5 °C, ΔSD 0.3 °C vs. 0.2 °C vs. 0.2 °C). Specifically, for minutes 4 and 6, the differences obtained between successful and failed LSBs in all variables assessed were moderate to large. In minute 5, in turn, the magnitude of the effect size was moderate to large in five out of the six variables. In fact, the SD variables (ΔSD and ΔAsymSD) presented the lower effect size of the differences between failed and successful LSBs in all moments assessed.

On the other hand, the performance metrics for the four classification algorithms at the three different moments assessed are depicted in Table 2. This performance metrics were obtained applying the algorithms obtained with the testing sample. RFE determined the different number of predictors for each classification method and moment assessed, resulting, that at minute 4, the number of predictors necessary was higher in comparison with minute 5 and 6 in all methods evaluated except SVM (Table 2). The best classification methods according to their performance were ANN and RF at minute 4 (both with an AUC = 0.89) and ANN at every evaluated time (AUC = 0.89 at min 4, and AUC = 0.92 at min 5 and 6).

In order to assess the models' both performance and predictors contribution in more detail, the three ML algorithms with the best results from Table 2 (ANN at minutes 4 and 5 and RF at minute 4) were selected. As an exemplification of the models' performance, Fig. 4 shows the Confusion matrix of the RF and ANN at minute 4 (Fig. 4a) and ANN at minute 5 (Fig. 4b). Procedures classified by physicians as successful and failed, were also so classified by the models in 60% (RF and ANN at minute 4) and 64% (ANN at minute 5) and in 24% (RF and ANN) respectively. On the other hand, the models classified procedures as failed having them previously medically classified as successful in 16% (RF and ANN at minute 4) and 12% (ANN at minute 5) of the cases. None of the three models evaluated classified the procedures as successful having them previously considered by the physicians as failed ones.

As an exemplification of the models' predictors contribution, Fig. 5 shows the SHAP values obtained at minute 4 for RF (Fig. 5a) and ANN (Fig. 5b), and for ANN at minute 5 (Fig. 5c). The SHAP value allows to compare the individual strength of each predictor in the model. In addition, it allows the predictors to be considered as excitatory (positive values) or inhibitory (negative values), thus their behaviour facilitates the LSBs prediction as successful or failed, respectively. Thus, as it can be observed in Fig. 5, the ΔAsymMean and ΔAsymMax skin temperatures of the central heel are the predictors with the highest contribution present in the three models.

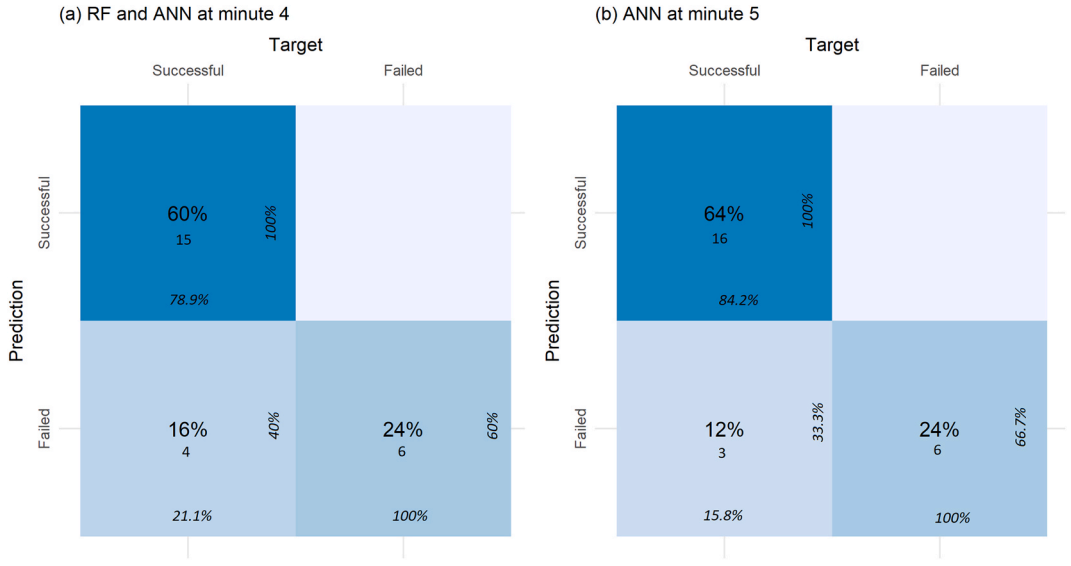

Fig. 4. Confusion matrix of the test dataset for the classification methods: Random Forest (RF) and Artificial Neuronal Network (ANN) at minute 4 (a); same matrix confusion for ANN at minute 5 (b).

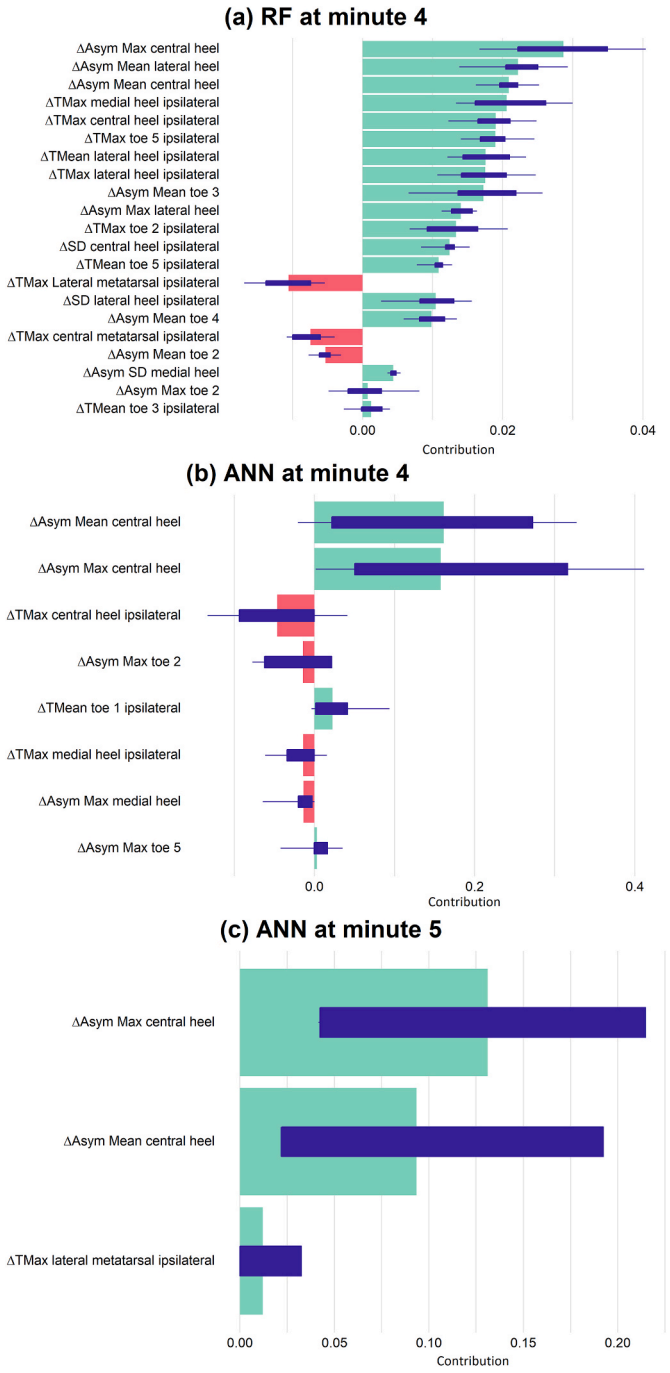

Fig. 5. The predictor's contribution (SHAP values) which obtained the best performance for the classification methods: Random Forest (RF) at minute 4 (a), and Artificial Neuronal Network (ANN) at minute 4 (b) and at minute 5 (c).

4. Discussion

The main purpose of this study was to assess the performance of different ML algorithms to automatically classify LSBs performed by clinicians as successful or failed based on IRT predictors extracted from the plantar of the foot. To the authors' knowledge, there are no previous studies developing ML algorithms in the classification of LSBs performance. Therefore, our study assessed the performance of four ML algorithms: ANN, KNN, RF, and SVM.

Considering the performance metrics and the number of predictors obtained in this study, the best desirable model to automatically classify LSBs into successful or failed would be ANN at minute 5. Nevertheless, the model's performance may be affected by several factors such as the selected predictors, hyperparameters and so forth (Taha and Hanbury, 2015). For example, for an ANN, a high decay rate can lead to a stronger regularization effect, which can reduce the capacity of the model, and prevent overfitting, and a large number of neurons can increase the capacity of the model to fit the training data, but it also increases the risk of overfitting, which means that the model performs well on the training data but poorly on unseen data (LeCun et al., 2015; Hochreiter and Schmidhuber, 1997). According to a recent systematic review concerning the combined use of IRT and ML methods, the classifier most applied in terms of its performance is ANN followed by SVM, both of them mainly applied for breast diagnosis (Magalhaes et al., 2021). On the other hand, other conditions such as diabetes have been studied analysing temperatures on the plantar region in combination with ML methods (Maldonado et al., 2020; Vardasca et al., 2018). In a study to identify prediabetes from plantar thermograms, different classifiers, including SVM, KNN or RF among others were implemented in 60 subjects (Thirunavukkarasu et al., 2020). The SVM outperformed the other classifiers with an accuracy rate of 81.60%, although RF presented the best value of AUC reaching 0.87. In another research, ANN and SVM along with other deep learning methods were compared to classify diabetic foot thermograms, achieving for ANN an accuracy of 83.33%, sensitivity of 66.60% and AUC of 0.83 (Cruz-Vega et al., 2020). ANN has also been integrated successfully for the support of early diagnosis and follow-up of diabetic patients using plantar foot thermograms, with a classification rate of 94.33% (Hernandez-Contreras et al., 2015). The results obtained in the current study show that all algorithms would be effective (AUC>0.73) in classifying LSBs. For this reason, even though the ANN achieved the best performance, the other algorithms should also be considered for their specific advantages: SVM is particularly suited for binary classification problems, RF provides a good balance between bias and variance, and is relatively robust to overfitting, and KNN is relatively simple to implement and can work well on non-linear decision boundaries (Hastie et al., 2009; Cristianini and Shawe-Taylor, 2000; Breiman, 2001). Although the classification approach in the CRPS condition, and specifically in LSB procedures, has not been evaluated before, the results of his work are in line with previous studies regarding disease classification using ML algorithms (Ibrahim and Abdulazeez, 2021).

Currently, the employment of ML classifiers using thermal data is still unexplored in many biomedical applications, especially other than cancer, and this circumstance could occur because health professionals are not familiar with this imaging technique yet (Vardasca et al., 2019). To date, the evaluation of the predictors' contribution (SHAP value) has received little attention so far due to its recent development (Rodríguez-Pérez and Bajorath, 2020; Wang et al., 2021), and no previous studies implementing it in ML models using IR predictors have been found. However, its implementation in other topics shows promising outcomes, which is in agreement with our findings (Xie et al., 2021; Pan et al., 2020). The predictors having superior contribution were ΔAsymMean and ΔAsymMax (the asymmetry variations between ipsilateral and contralateral foot between the moment assessed and the baseline time in mean and maximum temperature) of the central heel in RF for minute 4 and ANN for both minutes 4 and 5. In this sense, the fact that the central heel presents the highest contribution may be attributed to the feet arterial vascularization, since the plantar blood suppliance has its origin primarily on the posterior tibial artery (Xu et al., 2021; Gil-Calvo et al., 2017).

The small sample size may be a limitation since it gives rise to whether the models were trained with a sufficient number of images or not, and if it may lead to bias in the results. In this sense, it should be remarked that the lockdown due to the COVID-19 took place within the infrared images acquisition period, so the workload in the hospital, hence, the number of patients treated, was lower than usual. Nevertheless, the number of procedures performed in 18 months is substantial in comparison with the vast majority of hospitals, where the number of

patients diagnosed with lower limbs CPRS undergoing LSBs is scarce (Shim et al., 2019). Moreover, considering the mean effect size of the differences observed between cases on the predictors of the models (ES = 1.2; Fig. 3a), the statistical power was 99% with an α error of the 5% for differences between means using Wilcoxon tests (G*Power 3.1.7, University of Düsseldorf, Germany). On the other hand, although only two different pain physicians performed the interventions, and both with the same experience, this may impact on the LSBs performance and, therefore, on the results. In some cases, it was observed that, the warming thermal patterns faded out short after having the intervention being classified as successful, which lead to false positives. The underlying cause is suspected to be the patients' anatomy (i.e., long vertebral body's apophysis, or presence of communicating branches in the sympathetic chain), although further analysis on these patients should be carried out.

Future studies should address whether the thermal regional alterations would be predictive of clinical outcomes since the results obtained in previous studies so far are not conclusive (van Eijs et al., 2012; Schürmann et al., 2001). In a study performed by van Eijs et al., 15 patients (31%) had good or moderate response. The response rate was not different in patients' group with cold or warm CRPS I or in those with more or less than 1.5 °C differential increase in skin temperature after LSB (van Eijs et al., 2012). In another study performed by Schurmann et al., it was suggested that the proof of sympathetically maintained pain based on pain relief after stellate ganglion blockade is not conclusive, since the observation of sympathicolysis (patients who developed an increase in temperature difference between the ipsilateral and contralateral hand of more than 1.5 °C) did not involve pain relief (Schürmann et al., 2001).

Concerning the thermal data acquisition, an IRT protocol should be developed to ensure medical staff unfamiliar with thermal equipment would follow it. Thus, untrained physicians who commence performing LSBs may take advantage of the infrared images helping them to decide whether the needle is on the proper site or not. Moreover, for trained physicians, the use of thermal images as a support tool for this purpose is advisable, thereby providing the medical staff an objective validation.

5. Conclusions

The results of this study reveal a high performance of different ML algorithms (AUC>0.73) to automatically classify LSBs performed by clinicians as successful or failed based on IRT predictors extracted from the plantar of the foot. Specifically, the methods with higher AUC values were ANN and RF at minute 4 (0.89) and ANN at every evaluated time (0.89 at min 4, and 0.92 at min 5 and 6). Therefore, 4 min may be enough for the ML approaches to give a good performance rating. Finally, the most important predictors for the procedures classification were the ΔAsymMean and ΔAsymMax of the central heel.

Credits authors statement

Conceptualization; all the authors. Data curation; M.C–S and J.I.P-Q. Formal analysis; J.I.P-Q. Investigation; M. C–S, M.B, and C.G-V. Methodology; all the authors. Project administration; D. M and J.I.P-Q. Resources; R.S–P and R.C.O.A Supervision; D. M and J.I.P-Q. Visualization; J.I.P-Q. Roles/Writing - original draft; M.C–S. Writing - review & editing; all the authors.

Funding sources

The study was performed without specific funding.

Declaration of competing interest

None.

Data availability

The dataset generated and analysed during the current study is available in the Mendeley repository (DOI:10.17632/2vsf9vb9z7.1).

Acknowledgements

The authors would like to thank researchers Dr. Silvia Ruiz-España (Universitat Politècnica de València, València, Spain), Dr. Marcio Goethel (Universidade do Porto, Porto, Portugal) and Daniel Andrés López (University of Mainz, Germany) for their support during the data analysis using machine learning techniques.

References

Adam, M., Ng, E.Y.K., Oh, S.L., Heng, M.L., Hagiwara, Y., Tan, J.H., et al., 2018. Automated characterization of diabetic foot using nonlinear features extracted from thermograms. Infrared Phys. Technol. 89, 325–337. https://doi.org/10.1016/j.infrared.2018.01.022.

An, J.W., Koh, J.C., Sun, J.M., Park, J.Y., Choi, J.B., Shin, M.J., et al., 2016. Clinical identification of the vertebral level at which the lumbar sympathetic ganglia aggregate. Korean Journal of Pain 29, 103–109. https://doi.org/10.3344/kjp.2016.29.2.103.

Araújo, M.C., Lima, R.C.F., de Souza, R.M.C.R., 2014. Interval symbolic feature extraction for thermography breast cancer detection. https://doi.org/10.1016/j.eswa.2014.04.027.

Bauer, J., Hoq, M.N., Mulcahy, J., Tofail, S.A.M., Gulshan, F., Silien, C., et al., 2020. Implementation of artificial intelligence and non-contact infrared thermography for prediction and personalized automatic identification of different stages of cellulite. EPMA J. 11, 17–29. https://doi.org/10.1007/s13167-020-00199-x.

Borchers, A.T., Gershwin, M.E., 2014. Complex regional pain syndrome: a comprehensive and critical review. Autoimmun. Rev. 13, 242–265. https://doi.org/10.1016/j.autrev.2013.10.006.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Cañada Soriano, M., 2022. Infrared Thermography for the Assessment of Lumbar Sympathetic Blocks in Patients with Complex Regional Pain Syndrome. Universitat Politècnica de València. https://doi.org/10.4995/Thesis/10251/181699.

Cañada-Soriano, M., Priego-Quesada, J.I., Bovaira, M., García-Vitoria, C., Salvador Palmer, R., Cibrián Ortiz de Anda, R., et al., 2021. Quantitative analysis of real-time infrared thermography for the assessment of lumbar sympathetic blocks: a preliminary study. Sensors 21. https://doi.org/10.3390/s21113573.

Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511801389.

Cruz-Vega, I., Hernandez-Contreras, D., Peregrina-Barreto, H., Rangel-Magdaleno, J. de J., Ramirez-Cortes, J.M., 2020. Deep learning classification for diabetic foot thermograms. Sensors 20. https://doi.org/10.3390/s20061762.

Day, M., 2008. Sympathetic blocks : the evidence. Pain Pract. 8, 98–109.

Gauci, J., Falzon, O., Formosa, C., Gatt, A., Ellul, C., Mizzi, S., et al., 2018. Automated region extraction from thermal images for peripheral vascular disease monitoring. J Healthc Eng 2018. https://doi.org/10.1155/2018/5092064.

Gil-Calvo, M., Jimenez-Perez, I., Pérez-Soriano, P., Quesada, J.I.P., 2017. Foot Temperature Assessment. Application of Infrared Thermography in Sports Science. Springer, pp. 235–263. https://doi.org/10.1007/978-3-319-47410-6_10.

Gofeld, M., Shankar, H., Benzon, H.T., 2018. Fluoroscopy and ultrasound-guided sympathetic blocks: stellate ganglion, lumbar sympathetic blocks, and visceral sympathetic blocks. In: Benzon, H.T., Raja, S.N., Liu, S.S., Fishman, S.M., Cohen SPBT-E of, P.M., Fourth, E. (Eds.), Essentials of Pain Medicine, fourth ed. Elsevier, pp. 789–804.e2. https://doi.org/10.1016/B978-0-323-40196-8.00084-X.

Gogoi, U.R., Majumdar, G., Bhowmik, M.K., Ghosh, A.K., 2019. Evaluating the efficiency of infrared breast thermography for early breast cancer risk prediction in asymptomatic population. Infrared Phys. Technol. 99, 201–211. https://doi.org/10.1016/j.infrared.2019.01.004.

Harden, R.N., Bruehl, S., Stanton-Hicks, M., Wilson, P.R., 2007. Proposed new diagnostic criteria for complex regional pain syndrome. Pain Med. 8, 326–331. https://doi.org/10.1111/j.1526-4637.2006.00169.x.

Harden, R.N., Bruehl, S., Perez, R.S.G.M., Birklein, F., Marinus, J., Maihofner, C., et al., 2010. Validation of proposed diagnostic criteria (the "budapest criteria") for complex regional pain syndrome. Pain 150, 268–274. https://doi.org/10.1016/j.pain.2010.04.030.

Harden, R.N., Oaklander, A.L., Burton, A.W., Perez, R.S.G.M., Richardson, K., Swan, M., et al., 2013. Complex Regional Pain Syndrome: Practical Diagnostic and Treatment Guidelines, fourth ed. Pain Med, pp. 180–229. https://doi.org/10.1111/pme.12033.14.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, NY. https://doi.org/10.1007/978-0-387-84858-7 vol. 27.

Hernandez-Contreras, D., Peregrina-Barreto, H., Rangel-Magdaleno, J., Ramirez-Cortes, J., Renero-Carrillo, F., 2015. Automatic classification of thermal patterns in diabetic foot based on morphological pattern spectrum. Infrared Phys. Technol. 73, 149–157. https://doi.org/10.1016/j.infrared.2015.09.022.

Hochreiter, S., Schmidhuber, J., 1997. Flat minima. Neural Comput. 9, 1–42. https://doi.org/10.1162/neco.1997.9.1.1.
Houssein, E.H., Emam, M.M., Ali, A.A., Suganthan, P.N., 2021. Deep and machine learning techniques for medical imaging-based breast cancer: a comprehensive review. Expert Syst. Appl. 167, 114161 https://doi.org/10.1016/j.eswa.2020.114161.
Ibrahim, I., Abdulazeez, A., 2021. The role of machine learning algorithms for diagnosing diseases. Journal of Applied Science and Technology Trends 2, 10–19.
Jayanthi, T., Anburajan, M., 2019. Model-based computer-aided method for diagnosis of cardiovascular disease using IR thermogram. Biomed. Res. 30, 95–101. https://doi.org/10.35841/biomedicalresearch.30-19-004.
Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Software 28, 1–26.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.
Magalhaes, C., Mendes, J., Vardasca, R., 2021. Meta-analysis and systematic review of the application of machine learning classifiers in biomedical applications of infrared thermography. Appl. Sci. 11 https://doi.org/10.3390/app11020842.
Maldonado, H., Bayareh, R., Torres, I.A., Vera, A., Gutiérrez, J., Leija, L., 2020. Automatic detection of risk zones in diabetic foot soles by processing thermographic images taken in an uncontrolled environment. Infrared Phys. Technol. 105, 103187 https://doi.org/10.1016/j.infrared.2020.103187.
Mangalathu, S., Hwang, S.-H., Jeon, J.-S., 2020. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. Eng. Struct. 219, 110927 https://doi.org/10.1016/j.engstruct.2020.110927.
Marinus, J., Moseley, G.L., Birklein, F., Baron, R., Maihöfner, C., Kingery, W.S., et al., 2011. Clinical features and pathophysiology of complex regional pain syndrome. Lancet Neurol. 10, 637–648. https://doi.org/10.1016/S1474-4422(11)70106-5.
Money, S., 2019. Pathophysiology of complex regional pain syndrome and treatment: recent advancements. Curr Emerg Hosp Med Rep 7, 203–207. https://doi.org/10.1007/s40138-019-00198-x.
Pan, P., Li, Y., Xiao, Y., Han, B., Su, L., Su, M., et al., 2020. Prognostic assessment of COVID-19 in the intensive care unit by machine learning methods: model development and validation. J. Med. Internet Res. 22, e23128 https://doi.org/10.2196/23128.
Park, S.Y., Nahm, F.S., Kim, Y.C., Lee, S.C.J.C., Sim, S.E., Lee, S.C.J.C., 2010. The cut-off rate of skin temperature change to confirm successful lumbar sympathetic block. J. Int. Med. Res. 38, 266–275. https://doi.org/10.1177/147323001003800131.
Qian, S., Sengupta, V., Urbiztondo, N., Haider, N., 2019. Lumbar sympathetic block. In: Deer, T.R., Pope, J.E., Lamer, T.J., Provenzano, D. (Eds.), Deer's Treatment of Pain: an Illustrated Guide for Practitioners. Springer International Publishing, Cham, pp. 467–475. https://doi.org/10.1007/978-3-030-12281-2_57.
Raghavendra, U., Rajendra Acharya, U., Ng, E.Y.K., Tan, J.-H., Gudigar, A., 2016. An integrated index for breast cancer identification using histogram of oriented gradient and kernel locality preserving projection features extracted from thermograms. Quant. InfraRed Thermogr. J 13, 195–209. https://doi.org/10.1080/17686733.2016.1176734.
Rodríguez-Pérez, R., Bajorath, J., 2020. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. J. Comput. Aided Mol. Des. 34, 1013–1026. https://doi.org/10.1007/s10822-020-00314-0.
Roslidar, R., Rahman, A., Muharar, R., Syahputra, M.R., Arnia, F., Syukri, M., et al., 2020. A review on recent progress in thermal imaging and deep learning approaches for breast cancer detection. IEEE Access 8, 116176–116194. https://doi.org/10.1109/ACCESS.2020.3004056.
Rubio Mayo, P., 2021. Desarrollo de una herramienta software para la cuantificación de secuencias de vídeo e imágenes de termografía infrarroja en el diagnóstico y seguimiento del Síndrome de Dolor Regional Complejo. Universitat Politècnica de València.
Ryu, J., Lee, C., Kim, Y., Lee, S., Shankar, H., Moon, J., 2018. Ultrasound-assisted versus fluoroscopic-guided lumbar sympathetic ganglion block: a prospective and randomized study. Chronic Pain Medicine 126, 1362–1368. https://doi.org/10.1213/ANE.0000000000002640.
Schürmann, M., Gradl, G., Wizgal, I., Tutic, M., Moser, C., Azad, S., et al., 2001. Clinical and physiologic evaluation of stellate ganglion blockade for complex regional pain syndrome type I. Clin. J. Pain 17.
Seo, H., Badiei Khuzani, M., Vasudevan, V., Huang, C., Ren, H., Xiao, R., et al., 2020. Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state-of-art applications. Med. Phys. 47, e148–e167. https://doi.org/10.1002/mp.13649.
Shim, H., Rose, J., Halle, S., Shekane, P., 2019. Complex regional pain syndrome: a narrative review for the practising clinician. Br. J. Anaesth. 123, e424–e433. https://doi.org/10.1016/j.bja.2019.03.030.
Singh, M., Pai, M., Kalantri, S.P., 2003. Accuracy of perception and touch for detecting fever in adults: a hospital-based study from a rural, tertiary hospital in Central India. Trop. Med. Int. Health 8, 408–414. https://doi.org/10.1046/j.1365-3156.2003.01049.x.
Stanton-Hicks, M., 2018. Complex regional pain syndrome. In: Cheng, J., Rosenquist, R. W. (Eds.), Fundamentals of Pain Medicine. Springer International Publishing, Cham, pp. 211–220. https://doi.org/10.1007/978-3-319-64922-1_23.
Steketee, J., 1973. Spectral emissivity of skin and pericardium. Phys. Med. Biol. 18, 686–694.
Taha, A.A., Hanbury, A., 2015. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med. Imag. 15, 29. https://doi.org/10.1186/s12880-015-0068-x.
Tharwat, A., 2021. Classification assessment methods. Applied Computing and Informatics 17, 168–192. https://doi.org/10.1016/j.aci.2018.08.003.
Thirunavukkarasu, U., Umapathy, S., 2020. Classification of prediabetes and healthy subjects in plantar infrared thermal imaging using various machine learning algorithms. In: Sharma, D.K., Balas, V.E., Son, L.H., Sharma, R., Cengiz, K. (Eds.), Micro-Electronics and Telecommunication Engineering. Springer Singapore, Singapore, pp. 85–96.
Umapathy, S., Vasu, S., Gupta, N., 2018. Computer aided diagnosis based hand thermal image analysis: a potential tool for the evaluation of rheumatoid arthritis. J. Med. Biol. Eng. 38, 666–677. https://doi.org/10.1007/s40846-017-0338-x.
van Eijs, F., Geurts, J., van Kleef, M., Faber, C.G., Perez, R.S., Kessels, A.G.H., et al., 2012. Predictors of pain relieving response to sympathetic blockade in complex regional pain syndrome type 1. Anesthesiology 116, 113–121. https://doi.org/10.1097/ALN.0b013e31823da45f.
Vardasca, R., Vaz, L., Magalhaes, C., Seixas, A., Mendes, J., 2018. Towards the diabetic foot ulcers classification with infrared thermal images. In: Berlin, G. (Ed.), Organizing Committee of the Conference QIRT 2018. QIRT Council, pp. 293–296. https://doi.org/10.21611/qirt.2018.008. Proceedings Quantitative InfraRed Thermography Conference (QIRT 2018).
Vardasca, R., Magalhaes, C., Mendes, J., 2019. Biomedical applications of infrared thermal imaging: current state of machine learning classification. Proc West Mark Ed Assoc Conf 27. https://doi.org/10.3390/proceedings2019027046.
Wang, K., Tian, J., Zheng, C., Yang, H., Ren, J., Liu, Y., et al., 2021. Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. Comput. Biol. Med. 137, 104813 https://doi.org/10.1016/j.compbiomed.2021.104813.
Xie, P., Li, Y., Deng, B., Du, C., Rui, S., Deng, W., et al., 2021. An explainable machine learning model for predicting in-hospital amputation rate of patients with diabetic foot ulcer. Int. Wound J. https://doi.org/10.1111/iwj.13691 n/a.
Xu, X., Kim, S., Clune, J.E., Narayan, D., 2021. Upper and lower extremity vascular variations. In: Narayan, D., Kapadia, S.E., Kodumudi, G., Vadivelu, N. (Eds.), Surgical and Perioperative Management of Patients with Anatomic Anomalies. Springer International Publishing, Cham, pp. 437–466. https://doi.org/10.1007/978-3-030-55660-0_19.
Zhu, X., Kohan, L.R., Morris, J.D., Hamill-Ruth, R.J., 2019. Sympathetic blocks for complex regional pain syndrome: a survey of pain physicians. Reg. Anesth. Pain Med. 44, 736–741. https://doi.org/10.1136/rapm-2019-100418.