

Article

Chronic Pain Diagnosis Using Machine Learning, Questionnaires, and QST: A Sensitivity Experiment

Alex Novaes Santana *, Charles Novaes de Santana and Pedro Montoya *

Research Institute of Health Sciences (IUNICS-IdISBa), University of the Balearic Islands, 07120 Palma de Mallorca, Spain; charles.santana@gmail.com

* Correspondence: alex.santana@gmail.com (A.N.S.); pedro.montoya@uib.es (P.M.)

Received: 9 November 2020; Accepted: 13 November 2020; Published: 17 November 2020



Abstract: In the last decade, machine learning has been widely used in different fields, especially because of its capacity to work with complex data. With the support of machine learning techniques, different studies have been using data-driven approaches to better understand some syndromes like mild cognitive impairment, Alzheimer’s disease, schizophrenia, and chronic pain. Chronic pain is a complex disease that can recurrently be misdiagnosed due to its comorbidities with other syndromes with which it shares symptoms. Within that context, several studies have been suggesting different machine learning algorithms to classify or predict chronic pain conditions. Those algorithms were fed with a diversity of data types, from self-report data based on questionnaires to the most advanced brain imaging techniques. In this study, we assessed the sensitivity of different algorithms and datasets classifying chronic pain syndromes. Together with this assessment, we highlighted important methodological steps that should be taken into account when an experiment using machine learning is conducted. The best results were obtained by ensemble-based algorithms and the dataset containing the greatest diversity of information, resulting in area under the receiver operating curve (AUC) values of around 0.85. In addition, the performance of the algorithms is strongly related to the hyper-parameters. Thus, a good strategy for hyper-parameter optimization should be used to extract the most from the algorithm. These findings support the notion that machine learning can be a powerful tool to better understand chronic pain conditions.

Keywords: chronic pain; machine learning; classification; questionnaires; QST

1. Introduction

Pain is a subjective perceptual phenomenon, which is mainly determined by emotional, cognitive, and sociocultural factors (e.g., mood, learning, attention, and beliefs) [1,2]. More importantly, pain is an emergent property of brain activity, in which learning and memory processes, along with associated plastic brain changes, may play a very important role, particularly when pain persists over time [3]. The International Association for the Study of Pain (IASP) defines chronic pain as pain that lasts more than three or six months [4,5]. Other symptoms, such as sleep disturbance, mood changes, and fatigue, are also associated with chronic pain syndromes [6,7]. These physical, cognitive, and emotional alterations clearly affect patients’ daily routines, leading to impairments of quality of life and disability.

Hence, the assessment and treatment of pain require a multidimensional approach that takes into account neurophysiological, psychological, and cultural aspects related to pain perception. As a result, several techniques and tools have been specifically developed or incorporated for measuring all of these multidimensional aspects of pain. Thus, for instance, surveys and self-report questionnaires [8,9], Quantitative Sensory Tests (QSTs) [10–13], genetic factors [14–16], physical activity patterns [17–20], Electroencephalography (EEG) [21], neuroimaging [22–24], and, more recently, functional near-infrared spectroscopy (fNIRS) [25] have been incorporated into studies of the emotional and cognitive factors

that modulate pain. Eventually, these approaches are used to classify or differentiate groups, comparing patients with one chronic pain syndrome against pain-free controls or other chronic pain syndromes. Often, one study can present results for more than one tool or method, testing different cut-offs or applying it in different body regions. Usually, the different chronic syndromes do not share the same diagnostic approaches and, in most cases, the diagnosis of chronic pain is solely based on individuals' self-reports. All these studies have expanded our comprehension of the neurophysiological mechanisms (central sensitization, brain plasticity) and psychosocial factors involved in the origin and maintenance of chronic pain. Nevertheless, more than 75% of patients do not receive an accurate diagnosis [26,27]. In addition, primary care providers show inappropriate attitudes and beliefs about pain and its treatment even after participating in continuous education programs [28].

In this context of inaccurate diagnosis and treatment of chronic pain, clinicians' decisions could benefit from objective methods and criteria for better understanding and treatment of patients with chronic pain. In this regard, the US Food and Drug Administration (FDA) and the American Pain Society (APS) have proposed a multidimensional framework and operational diagnostic criteria for the major chronic pain conditions. Those criteria are composed of the patient's historical data, questionnaires (self-reported), and psychophysical tests that determine pain features and pain thresholds. Furthermore, the Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks (ACTTION) Framework [29] has suggested taking into account at least the following five dimensions: (1) core diagnostic criteria; (2) common features; (3) common medical and psychiatric comorbidities; (4) neurobiological, psychosocial, and functional consequences; and, finally, (5) putative mechanisms, risk factors, and protective factors. Expanding this framework, other studies have proposed evidence-based diagnostic criteria for specific chronic pain conditions [30–36].

Chronic lower back pain (CLBP) is one of the most frequently investigated syndromes and a major cause of work disability. The majority of the studies about CLBP use methods related to mobility measures [19,37–40] or electromyography [41–43]. Using the ACTTION framework, Markman and colleagues listed four items as the core diagnostic criteria for CLBP: (i) pain restricted to the lower back or with a referral pattern limited to the proximal legs, (ii) pain in the lower back on most days for at least three months and at least half of the days in the past six months, (iii) absence of neurological symptoms and deficit or symptoms in the lower extremities, and (iv) absence of tumors, infections, spondylolisthesis Grade 2 or higher, acute vertebral fractures, or other identifiable primary causes of lower back pain. This core is complemented with the other dimensions, which include a set of comorbidities and other syndromes that may be associated with CLBP, including other common chronic pain syndromes, such as fibromyalgia (FM).

The American College of Rheumatology (ACR) recommends a combination of questionnaire data (the Symptom Severity Score) and the Widespread Pain Index (WPI) as classification criteria of fibromyalgia (FM) [6]. Over the years, the ACR's criteria were updated and adapted for different languages [44,45], but in addition to their consolidation, new alternatives have been proposed [46]. Using the ACTTION framework, a Fibromyalgia Working Group defined three items as the core diagnostic criteria for FM: (i) multi-site pain, defined as six or more pain sites from a total of nine possible sites, (ii) moderate to severe sleep problems or fatigue, and (iii) multi-site pain plus fatigue or sleep problems must have been present for at least three months [32]. Complementary tenderness, discognition, and musculoskeletal stiffness are common features experienced by FM patients. In both FM and CLBP, patients may face pain sensitivity alterations, as well as mood disorders such as depression and anxiety.

Finally, it seems clear that complex and multidimensional classification problems could take advantage of machine learning techniques applied to clinical data for supporting clinical decisions [47–51]. In the context of pain and pain chronification, machine learning approaches have recently been applied to several pain syndromes [24,52–54], including fibromyalgia [55–57] and chronic lower back pain [58–61]. While traditional statistical analyses commonly make some a priori assumptions about the data model (e.g., normality) and about the relationships among variables (e.g., linearity), machine learning prioritizes a "distribution-free" context. Thus, machine learning algorithms can learn through the data, identifying

more complex relationships among variables and selecting the model that best describes a problem [62]. The design of a machine learning application may include several steps, starting from the data selection and preprocessing until the correct evaluation and validation. The application's performance is dependent on the correct implementation of these steps [63,64], especially the training process, where the model's hyper-parameters are optimized [65,66]. In this current study, we assessed the sensitivity of different machine learning algorithms and datasets for classifying chronic pain syndromes along with investigating the effect of hyper-parameter optimization.

2. Materials and Method

2.1. Participants

The participants of the study were 338 pain-free controls (HC) (age: 40.66 \pm 15.46, females: 258) and 659 chronic pain patients (age: 50.55 \pm 10.50, females: 567). The chronic pain group (CP) was composed of 440 fibromyalgia patients and 219 chronic back pain patients. There was no significant difference in age between the groups. All chronic pain participants matched the IASP's criteria [5] and/or the ACR's criteria for fibromyalgia [6]. A relatively small number of patients reported the use of non-steroidal anti-inflammatory drugs (NSAIDs) or paracetamol ($n = 4$), benzodiazepine (1–5 mg per day) ($n = 3$), and serotonin re-uptake inhibitors ($n = 1$). None of the selected participants used opiates, gabapentin, or pregabalin for pain treatment. The study was approved by the Research Ethics Committee of the Balearic Islands with the following codes: IB833/07PI and IB2268/14PI on 25 February 2015. The participants provided their written informed consent to participate in this study.

2.2. Data Acquisition

The Beck depression inventory II (BDI) [67], the State–Trait Anxiety Inventory (STAI) [68], and nine measures related to quantitative sensory tests (QSTs) of pain were analyzed in this study. The BDI and STAI are two questionnaires associated with depression and anxiety, respectively. These are disorders commonly correlated with chronic pain conditions [69–73]. The QST measures and body location included: pressure thresholds (index finger and wrist), heat threshold (index finger), and cold threshold (index finger, wrist, and elbow). In addition, the ratio between pain ratings and stimulus intensity applied was computed for pressure stimulus (index finger and wrist) and heat stimulus (index finger). All tests were applied on the dominant side. Heat and cold pain sensitivity were measured with a computer-controlled contact thermal stimulator (cold/warm plate AHP-301CPV, Teca, Schubert, IL, USA), while the pressure pain sensitivity was measured with a digital dynamometer using a flat rubber tip (1 cm²; Force One, Wagner Instruments, Greenwich, CT, USA).

2.3. Data Preprocessing

All data were previously normalized, and missing values were replaced by using multiple imputation by chained equation (MICE) [74]. Other imputation methods, such as the k-nearest neighborhood (k-NN) imputer, mean, and median, were also tested, but MICE was chosen because it preserved characteristics such as variance and correlation among variables. Subjects with more than one missing value were excluded, resulting in the number of participants cited previously. Finally, to avoid any kind of bias due to the position of the subjects inside the data, the subjects' positions were shuffled.

2.4. Classifiers

Eight classifiers were compared to determine the one that presented the best results in a binary classification problem: chronic pain patients against controls. This set of classifiers includes simple linear classifiers as well as ensemble-based classifiers. Logistic regression (LR) and support vector classifiers (SVC) are two linear models. LR uses a logistic function to model the probabilities of an object being part of a specific class [75], while the SVC algorithm searches for a hyper-plane capable of classifying the objects by maximizing the distance between the hyper-plane and the data points.

The k-NN classifier was also tested. The k-NN is a nonparametric technique where the class of an object is defined by the classes of k-other objects near it [76]. Another nonparametric method used was the dynamic tree classifier (DTC), which constructs a series of rules inferred from the data features [77]. The random forest classifier (RFC) and extra trees classifier (ETC) are ensembles of multiple DTCs in a unique classifier, where each tree predicts one class for the object, and the most predicted class is the final prediction made by the RFC [78] or ETC. The other classifiers used a multiple layer perceptron classifier (MPL), which is also known as neural network, and an extreme gradient-boosting classifier (XGBoost) [79]. Table 1 presents a list of all classifiers and the set of hyper-parameters.

In order to have a reference base for the chosen classifiers, dummy classifiers were used in this study. These classifiers do not perform any kind of learning. Their predictions are based on five strategies: a completely random guess following a uniform distribution (uniform); always predicting the most frequent class (most_frequent); always predicting the class that maximizes the class prior (prior); always predicting the HC class (constant = 0); always predicting the CP class (constant = 1); and predicting randomly but keeping the proportion of classes observed in the training data (stratified).

Table 1. List of tested classifiers and their corresponding sets of hyper-parameters and values. For more information about each hyper-parameter, please visit <https://scikit-learn.org/stable/modules/classes.html>.

Classifier	Set of Hyper-Parameters	Number of Combinations Evaluated
Dummy Classifier	strategy = ['constant', 'uniform', 'stratified', 'prior', 'most_frequent']; constant = [0, 1];	10
Logistic Regression	solver = ['liblinear']; penalty = ['l1', 'l2']; C = [1×10^{-4} , 1×10^{-3} , 1×10^{-2} , 1×10^{-1} , 5×10^{-1} , 1.0, 5.0, 1×10^1 , 1.5×10^1 , 2×10^1 , 2.5×10^1]; dual = [True, False];	22
SVC	kernel = ['rbf', 'poly']; tol = [1×10^{-5} , 1×10^{-4} , 1×10^{-3} , 1×10^{-2} , 1×10^{-1}]; C = [1×10^{-3} , 1×10^{-2} , 1×10^{-1} , 5×10^{-1} , 1.0, 5.0, 1×10^1 , 1.5×10^1 , 2×10^1 , 2.5×10^1];	110
K Neighbors Classifier	n_neighbors = range(1, 101, 1); weights = ['uniform', 'distance']; p = [1, 2];	2000
Decision Tree Classifier	criterion = ['gini', 'entropy']; max_depth = range(1, 19, 1); min_samples_split = range(2, 21, 1); min_samples_leaf = range(1, 21, 1);	1000
Random Forest Classifier	n_estimators = [3, 6, 9, 12, 15, 18, 100]; criterion = ['gini', 'entropy']; max_features = range(5×10^{-2} , 1, 5×10^{-2}); min_samples_split = range(2, 21, 1); min_samples_leaf = range(1, 21, 1); bootstrap = [True, False];	1000
Extra Trees Classifier	n_estimators = range(100, 500, 50); criterion = ['gini', 'entropy']; max_features = range(5×10^{-2} , 1, 5×10^{-2}); min_samples_split = range(2, 21, 1); min_samples_leaf = range(1, 21, 1); bootstrap = [True, False];	1000
MPL	hidden_layer_sizes = range(5, 100, 5); solver = ['lbfgs', 'adam', 'sgd']; learning_rate = ['adaptive', 'invscaling', 'constant']; learning_rate_init = [1×10^{-1} , 1×10^{-2} , 1×10^{-3}]	684
XGB Classifier	n_estimators = range(100, 500, 50); max_depth = range(1, 11, 1); learning_rate = [1×10^{-3} , 1×10^{-2} , 1×10^{-1} , 5×10^{-1} , 1.0]; subsample = range(5×10^{-2} , 1, 5×10^{-2}); min_child_weight = range(1, 21, 1); nthread = [1];	1000

2.5. Training and Evaluation

In this study, we evaluated five datasets with the following compositions: (1) *age* dataset with only the information about participants' ages; (2) *basic-wo-age* dataset with data from the BDI, STAI State (STAI-S), and STAI Trait (STAI-T); (3) *basic* dataset, adding the age information to *basic-wo-age*; (4) *qst* dataset with the nine QST measures; and, finally, (5) *all* dataset with all available data.

For all classifiers, we applied a stratified k-fold cross-validation approach with k equals 5. In addition, we forced the cross-validation to select the same folds for all classifiers. All scores presented in this paper are related to the average of the cross-validation folds. Figure 1 describes the entire process of acquisition, preprocessing, processing, learning, validation, and evaluation. During the cross-validation process, we selected different combinations of hyper-parameters using a randomized search [80] until a maximum of one thousand combinations. This selection was replicated for each classification method and used during the training, validation, and evaluation processes.

We used three scores to evaluate and compare the experiments. First, we used the balanced accuracy (BACC), defined as the average of recall obtained on each class, which, in turn, is the proportion of actual positives that are predicted as positives. Imbalanced groups do not affect this accuracy score. Then, we applied the area under the receiver operating curve (AUC) of a classifier, which is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance [81]. For the last, we used the cross-entropy loss or log-loss (LLOSS), which is defined as the negative log-likelihood of the true labels given the probabilistic classifier's predictions. It can be interpreted as a measure of certainty, where a classifier that predicts a wrong class with a high probability is punished [82]. We calculate it using the probability of an instance belonging to a target class, and the values can range from 0 to $+\infty$, where values close to 0 are better scores. Oppositely, for both the BACC and AUC, the values can range between 0 and 1, and values close to 1 indicate better classifiers. For all scores, we used CP as the predicted class.

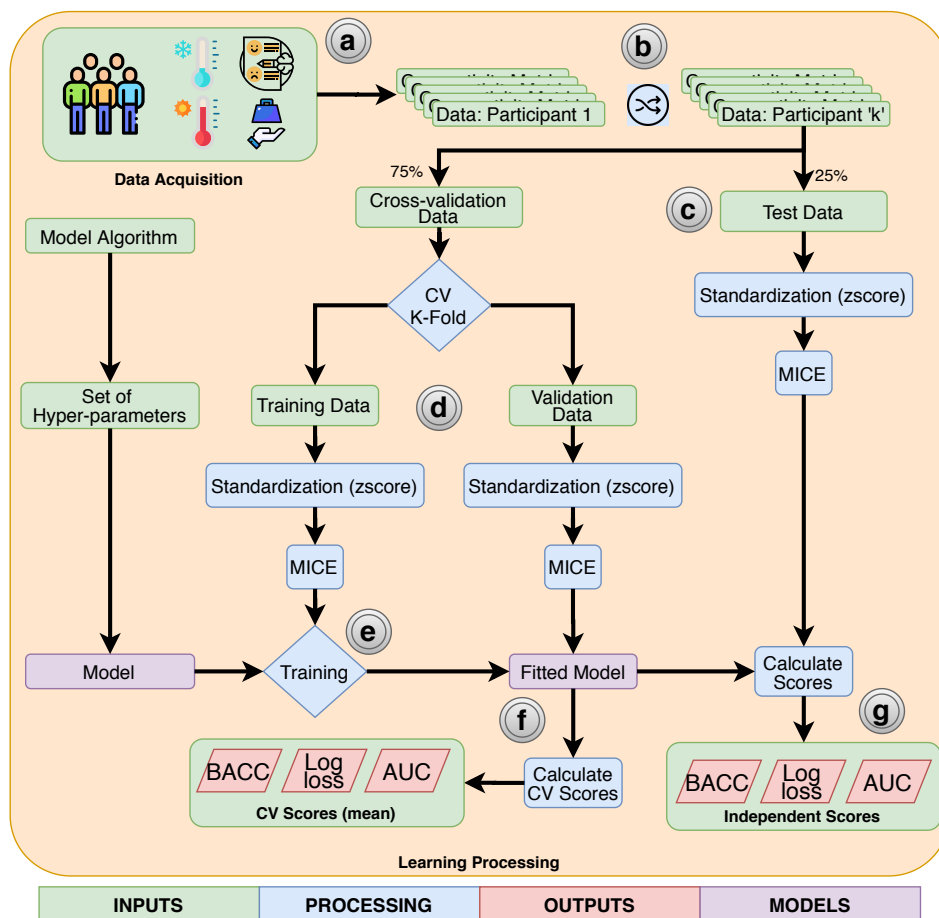


Figure 1. Flowchart describing the entire process of acquisition, preprocessing, processing, learning, and evaluation. (a) Participants answered the questionnaires and were evaluated using the sensorial tests of cold, heat, and pressure. (b) The learning process started shuffling the list of subjects' data. (c) An independent test was selected to be used to compare the best-fitted models after hyper-parameter optimization; this set represents 25% of the total dataset and preserves the proportion of target classes. With the remaining 75%, we applied a five-fold cross-validation process to optimize the hyper-parameters (d). The cross-validation applied two equal pipelines for the training and validation data. These pipelines sequentially apply a standardization method, an imputation method (multiple imputation by chained equation (MICE)), and the model in a training context (e) or the fitted model in order to obtain the cross-validation scores (f). Finally, (g) the fitted model is evaluated using the independent test and the same pipeline used during the cross-validation.

3. Results

In total, this study trained and evaluated 26,130 models, including dummy classifiers. From this set of models, we excluded all models that displayed an LLOSS higher than 0.69. In a binary classification problem, this cut-off represents the value of LLOSS for a model that predicts all elements with probability equal to 0.5 (randomly). Values higher than this cut-off mean that the model is less reliable than a prediction by chance.

The dummy classifiers served as a baseline for the other classifiers. Figure 2 presents the values of BACC and AUC obtained for all dummy classifiers. As expected, the values orbited around the selection by pure chance (0.5). The other classifiers presented a wider range of values, varying from 0.455 to 0.747 for BACC and between 0.403 and 0.828 for AUC. Figure 3 shows the dispersion of results for all 26,130 models. In that figure, we can observe that the dataset *all* prevails among the high values. Focusing on the models of the top-right corner (BACC > 0.7 and AUC > 0.8), Figure 4 reveals that only this dataset is represented.

Figure 5 exhibits more details about the dataset selection in the classifier performance. In addition to the greater number of features, the *qst* had a worse performance compared to the *basic* dataset, with values closer to the *only-age* dataset. In addition, we can observe a very similar result when comparing the *basic* and *basic-wo-age* datasets, which indicates that age does not contribute with relevant information to the classification problem. On the contrary, when we evaluate the *all* dataset, the inclusion of *qst* features seems to import useful information, resulting in a notable increase in the classification performance of all classifiers.

Excluding the baseline, we can notice that while DTC had the worst performance, the classifiers based on an ensemble of classification trees presented the best performance overall. The superior performance of the ensemble-based classifier is reinforced when the number of features per dataset is increased. Another interesting point of this result is the spread of the ensemble-based classifiers, which indicates that these classifiers are more sensitive to the hyper-parameter selection. Due to the clear superiority of the *all* dataset, henceforth, we will only show results for that dataset.

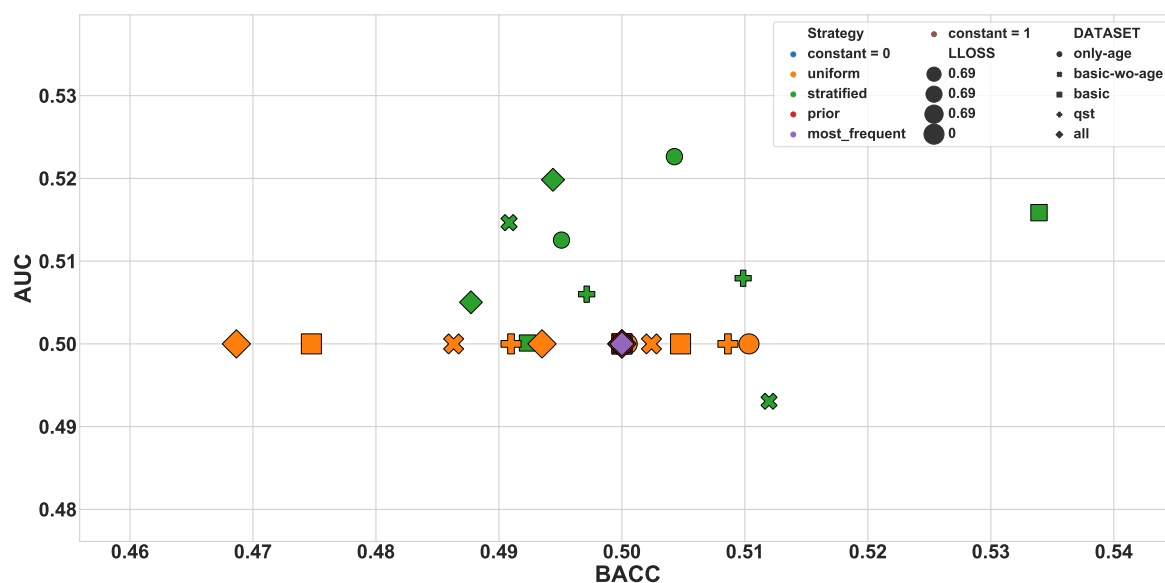


Figure 2. Baseline based on dummy classifiers.

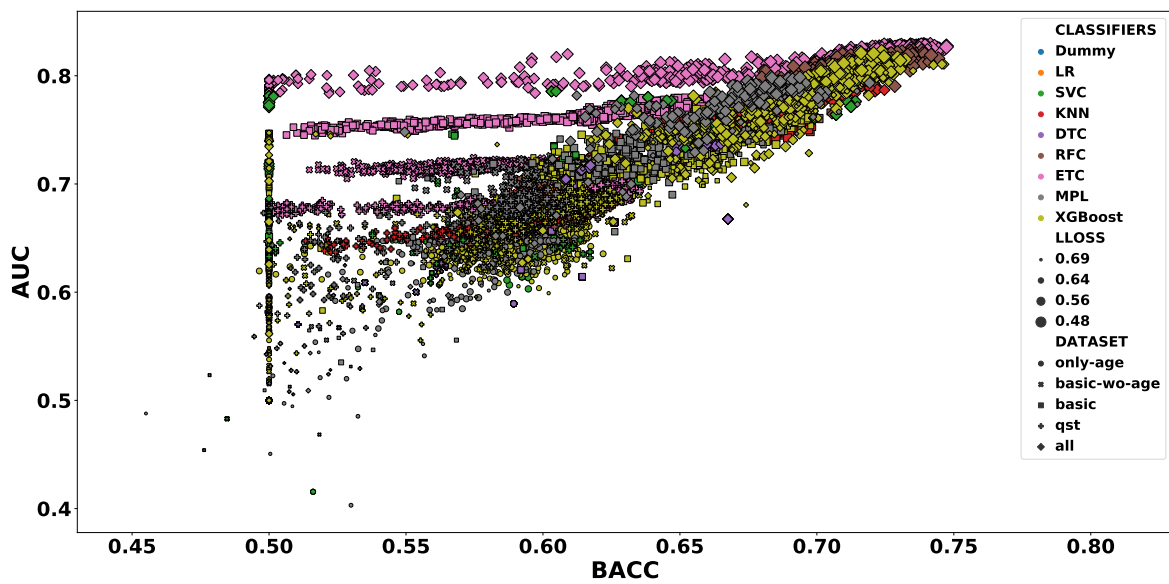


Figure 3. Dispersion (mean balanced accuracy (BACC) vs. mean area under the receiver operating curve (AUC)) of cross-validation scores. Each color represents one type of classifier, while the shape corresponds to the evaluated dataset. The size of each marker is inversely proportional to the value of log-loss (LLOSS; cross-validation mean); thus, bigger markers indicate more reliable classifiers.

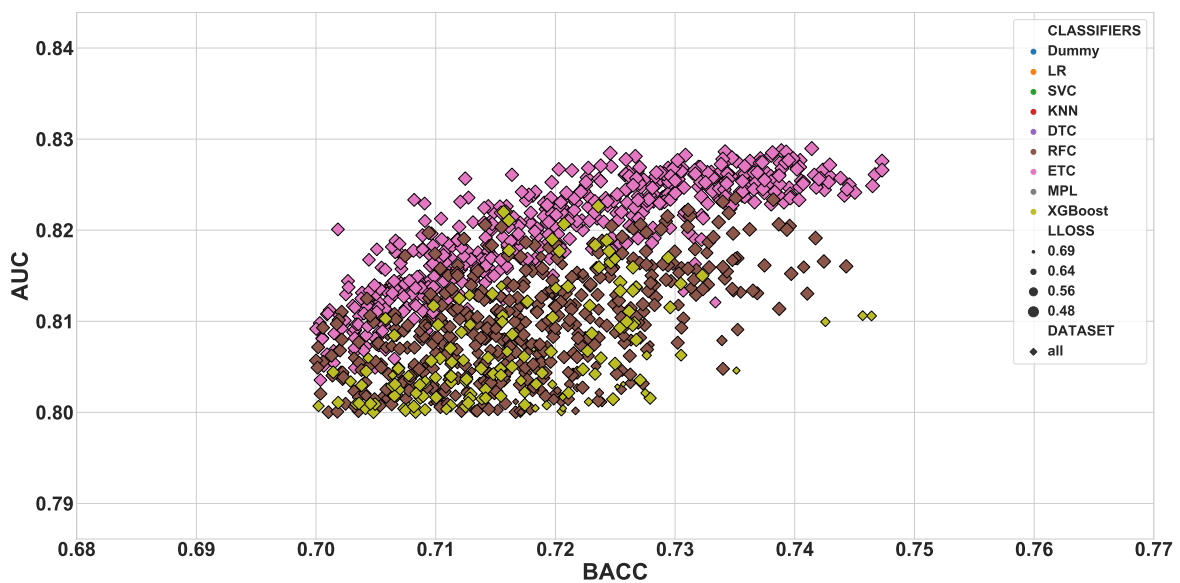


Figure 4. Close-up in the dispersion (mean BACC vs. mean AUC) of the cross-validation scores.

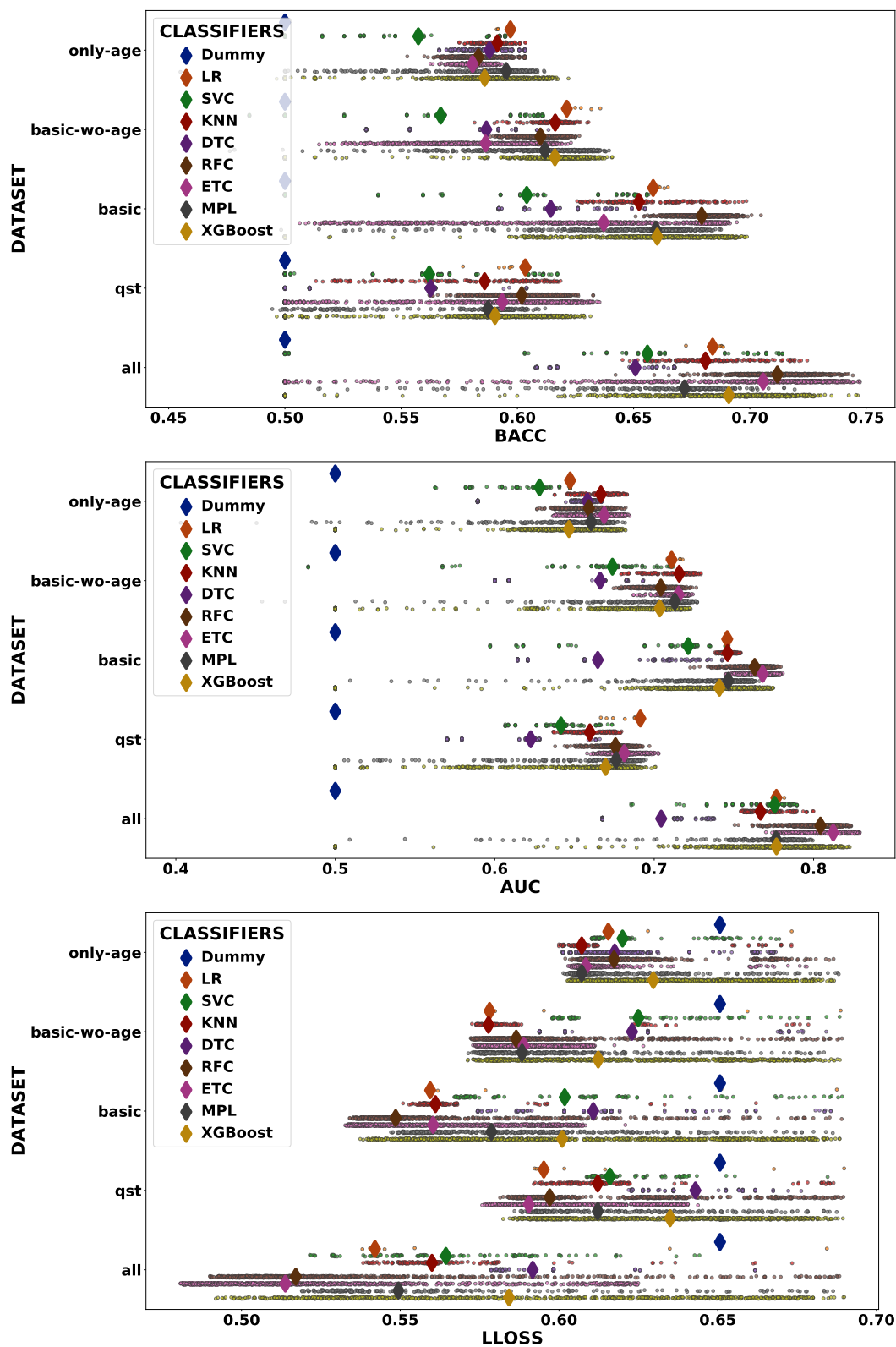


Figure 5. Strip plot for the BACC (top), AUC (middle), and LLOSS (bottom) grouped by dataset (vertical axis) and classifier type (colors). Each point represents one classifier with a specific hyper-parameter configuration, while the diamonds represent the medians of these points.

3.1. Independent Test Scores

After the hyper-parameter optimization, we selected the best classifier of each type based on the three scores: AUC, BACC, and LLOSS. Then, this set of 24 well-tuned classifiers was evaluated using the independent test to determine which of them performed the classification task better. In Figure 6, we present the values of BACC versus AUC for each of the best classifiers. In addition, the marker size is inversely proportional to the value of LLOSS. This result confirms the pattern found during the hyper-parameter optimization phase: The ensemble-based classifiers have the best performance in the classification problem. The ETC classifier with id 22353 provided the highest values of BACC (0.793), while an XGBoost classifier (id: 24836) presented the best AUC (0.876), as well as the lowest value of LLOSS (0.423). In addition, the classifiers with the higher AUC values (>0.85) presented good values of precision (0.81) and recall (0.90). Finally, analyzing the given probabilities for the independent test, we can also observe that only a small number (2%) of subjects were wrongly classified with a probability higher than 0.8.

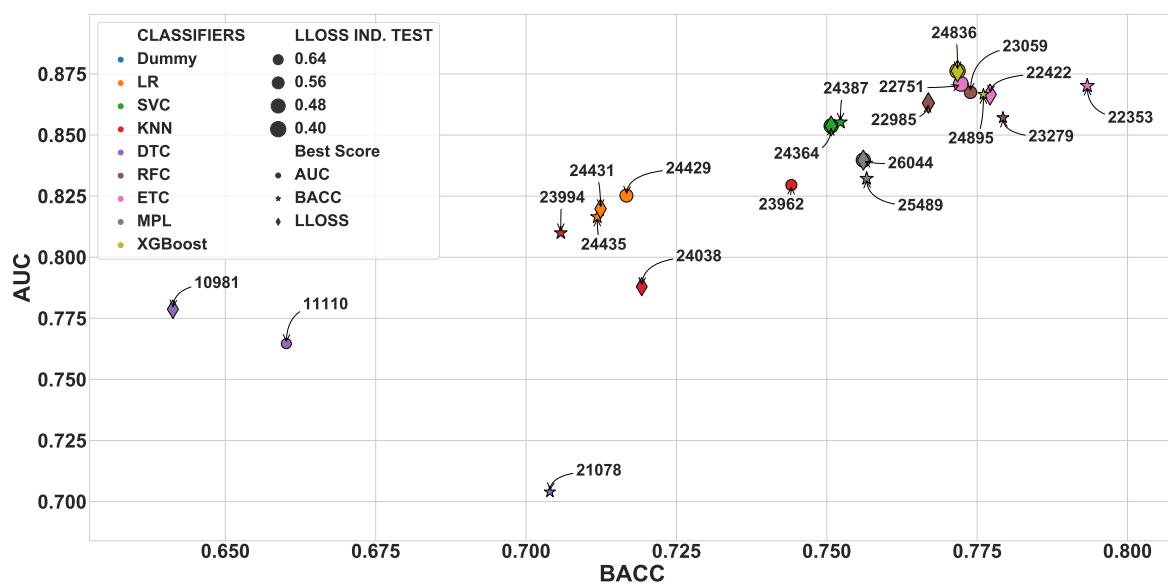


Figure 6. Independent test scores (BACC vs. AUC) for the best classifiers selected after the hyper-parameter optimization. The markers' shapes represent the score used to select the best classifier, while the colors represent the classifier type. Similarly to the other figures, the marker size is inversely proportional to the LLOSS. The annotated numbers are the unique identifiers of each classifier. In this evaluation, only the *all* dataset was evaluated.

3.2. Model Interpretation

One of the main challenges in machine learning solutions is to understand how the algorithms make predictions. Predictions based on factors that are supported by a knowledge-based theory are more reliable and preferable in the context of problems involving humans. In that direction, we used shapely additive explanations (SHAP) to obtain an approximation of how our trained algorithms predict chronic pain conditions. In Figure 7, we display a summary of the best ETC algorithm using SHAP. The explanation shows that the algorithm associates higher values of age, BDI, STAI-S, STAI-T, and pain ratios (any location) with CP patients. In the same way, higher values of pressure thresholds are associated with HC participants. These findings provide more reliance on the appropriateness of the machine learning algorithm to separate participants into healthy subjects and patients with chronic pain.

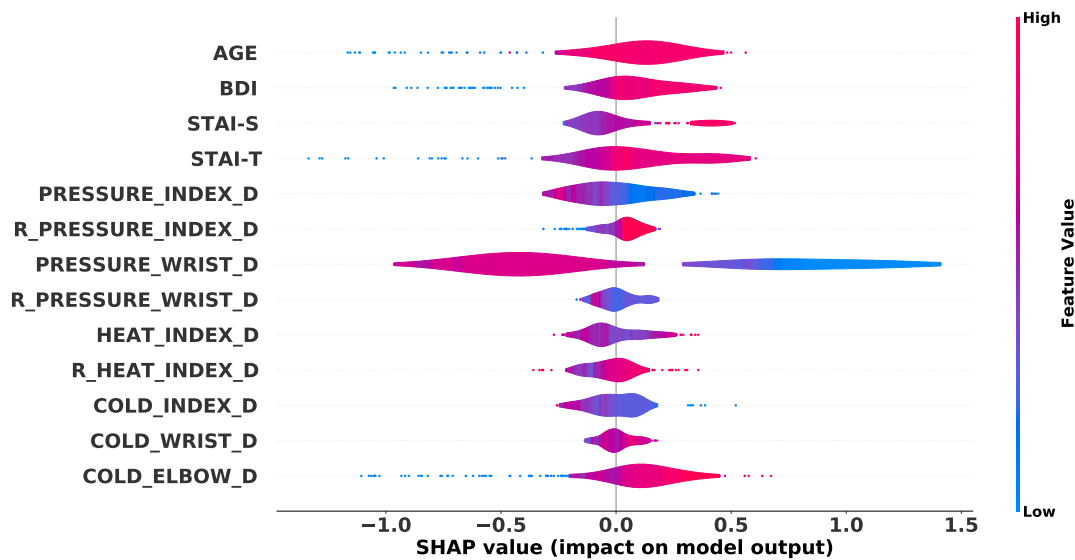


Figure 7. Summary explanation for the best extra trees classifier (ETC). Each dot represents one of the samples contained in the independent test. On the horizontal axis, the values indicate how much each feature (on the vertical axis) contributes for an HC or CP classification with negative and positive values, respectively. The color represents the feature value. The prefix “R_” among the quantitative sensory test (QST) variables indicates that they correspond to the ratio between pain ratings and stimulus intensity.

Finally, due to the fact that our target groups were not well balanced regarding gender (with a predominance of females in the CP group), we evaluated the best set of algorithms to determine if they were classifying participants’ sex instead of pain vs. healthy group. This could occur if the algorithms learned based on some sex information hidden in the features. To check this hypothesis, for the independent test, we calculated all the scores (AUC, BACC, LLOSS) using the sex information, encoded as 0 (male) and 1 (female) instead of the groups HC (0) and CP (1). In case our hypothesis is correct, we expect to have an equal or better performance of the algorithms using this new reference. In Figure 8, we show the results for both references.

We notice that AUC and BACC values are reduced when sex is used as a reference. This result becomes clearer when looking at the ensemble-based classifiers, the ones that had the best results overall. Another important point to note is the LLOSS score; we can see that the reliability of the classifiers degraded with an increase in the values of LLOSS for this new reference. All these factors gave us the confidence to exclude the cited hypothesis in all scenarios.

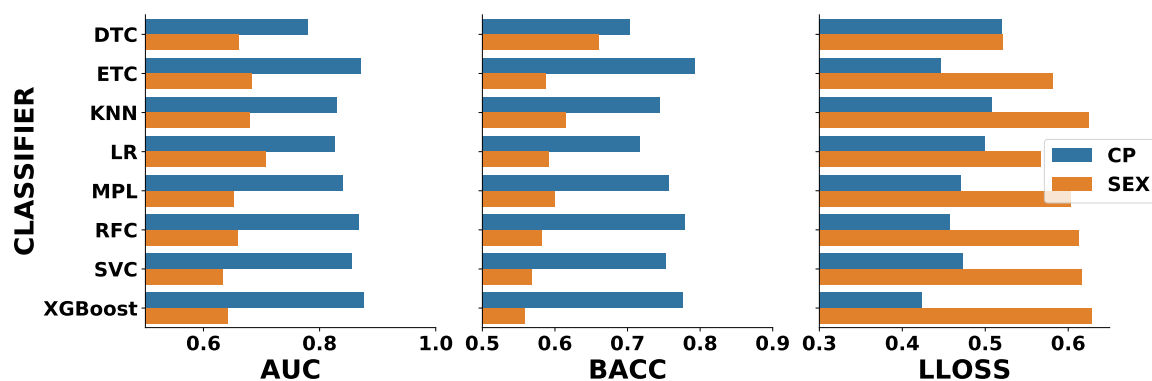


Figure 8. Independent test scores for the best classifiers compared against a hypothetical case where the classifiers were predicting the sex and not the chronic pain.

4. Discussion

In this paper, we assessed different machine learning algorithms to classify participants into chronic pain patients or healthy controls based on self-report and pain sensitivity data. In contrast to previous studies (e.g., Nijeweme-d'Hollosy and colleagues [83]), here, we compared different machine learning algorithms in a binary classification problem. From a methodological point of view, our study expands the study of Nijeweme-d'Hollosy by including more algorithms and a hyper-parameter optimization analysis, thus demonstrating that the success of classifiers depends on this optimization process.

The findings of this study provide evidence that the hyper-parameter optimization is an important part of the work when using machine learning algorithms. The results are very sensitive to the set of hyper-parameters chosen, which can result in a variation of more than 10% in BACC. Based on the studies presented at the Conference of Neural Information Processing Systems meeting in 2019 (NeurIPS2019) and the International Conference on Learning Representations (ICLR), grid-search and manual optimization are the most common approaches for selecting the best hyper-parameters [84]. While grid-search can consume too many resources, manual optimization can be limited by the researcher experience. In that context, randomized search has been proven to be a good choice, allowing a more broad search compared with manual optimization, but with a reduced cost compared with a grid-search [80].

In the present study, the *qst* dataset was the biggest dataset (isolated) with nine features. Nevertheless, this dataset had a poor performance if compared with the *basic* dataset. Age, BDI, STAT-T, and STAI-S were the four features comprising the *basic* dataset; this collection of features had three types of information (age, depression, and anxiety), while the *qst* features described the same kind of information: pain sensitivity. Thus, another important finding of this study was that a more divergent dataset could improve the chronic pain classification and that the number of features by itself could not correspond directly with performance. On the other hand, a dataset with a poorer performance would still improve a superior dataset, which was the case when the *basic* dataset was unified with *qst*.

Finally, only information about patients' age did not provide better performance than the baseline, and did not change the performance of the *basic* dataset. However, age information had some impact on the model outputs when the *all* dataset was considered. Thus, similarly to the hyper-parameter optimization problem, the training features should be chosen carefully. This feature selection can be based on previous knowledge of the problem, usually provided by some experts or by other studies. In addition, a sensitive experiment or other feature selection technique should be applied to select or confirm the most informative set of features.

In the present study, the best algorithms presented a fair to good performance overall, with values of balanced accuracy and AUC of around 0.75 and 0.85, respectively. Furthermore, the values of precision and recall indicated that the model produces more false negatives than false positives. In the context of our classification problem, the false positives were more acceptable, since they would further pass by a more complex exam to confirm or reject the prediction.

In the previous literature, other studies for classifying into chronic pain patients and healthy controls have shown higher performance rates. For instance, Hidalgo and colleagues presented a study using eight cameras and kinematic analysis to differentiate CLBP patients. In that study, the best scenario reached an AUC of 0.95 [18,19]. With similar performance, Jones and colleagues used RNA analysis and identified 10 candidate genes that could be used to predict FM with an AUC equal to 0.931 [85]. These studies presented more sophisticated and complex methods that are not commonly available in primary care. On the other hand, they could be used as complementary exams. Other studies also used simpler methods like questionnaires [86] or cost-saving sensorial tests [11].

Using common pegs and comparing against standardized algometers, Camara and colleagues also found that the pegs can have a higher AUC (0.81) in classifying non-malignant musculoskeletal pain [11]. Based on a Multidimensional Health Assessment Questionnaire (MDHAQ), Gibson and

colleagues developed a simple fibromyalgia assessment screening tool to support the identification of FM in patients with other rheumatic diseases [87]. Different subsets of the MDHAQ were analyzed, and the *symptom checklist* exhibited the best AUC (0.926). This checklist has historical information about the patient in the last month. Among the sixty items, some variables were correlated with FM symptoms, such as sleeping disorders and, more directly, pain at different body sites. Isolating the variables related to anxiety and depression, the values of AUC decreased to 0.716 and 0.745, respectively.

The feature set of our study did not include any information about non-induced pain. Our objective was to avoid any kind of data leakage. Variables about persistent pain, such as the pain felt in the last week or daily, can be directly correlated with the predicted variable (CP or HC). The algorithms could take advantage of that data leakage and focus only on the pain variable during the learning process. That could affect our evaluation of the learning capacity of the algorithms. In future analyses, we understand that some information about persistent pain should be included.

Excluding the data leakage hypothesis and based on the shapely additive explanations (SHAP), we can perceive how each feature contributes to the prediction. Besides the fact that the *only-age* dataset did not present good results, the SHAP analysis indicates that younger subjects have a higher chance of being classified in the healthy control group. Following the same analyses, the algorithm learned that subjects with depression (high BDI values) had a higher chance of belonging to the chronic pain group than participants without depression. This feature seems to be supported by previous literature on chronic pain, where several studies have linked chronic pain to depression [69–73]. A similar interpretation could be observed on anxiety (STAI variables), with high values of anxiety being related to chronic pain conditions [88,89]. Finally, data about pain sensitivity variables seem to indicate two relevant issues: (1) Subjects with higher pressure thresholds had a higher probability of being a healthy control; and (2) when a participant presented a high ratio between the subjective pain reported and the objective pressure applied during the QST, there was also a high probability of being classified as a chronic pain patient. These two facts are also supported by previous studies, indicating greater sensitivity to pain in chronic pain patients than in healthy controls [13,90–92]. The other QST features presented the same pattern [93–95], but with a smaller contribution to the prediction. Summarizing, the findings from the present study revealed that: (a) Ensemble-based classifiers outperformed the other classifiers, (b) the selection of the hyper-parameter values and features plays an important role in the learning process, and, finally, (c) the classifiers presented in this study were able to learn and make decisions based on information that is supported by the literature of chronic pain.

In a clinical context, the findings of the present study reinforce the reliability of decision-support systems based on machine learning approaches for chronic pain classification. This approach could provide faster, cheaper, and, consequently, more accessible diagnoses for chronic pain patients.

Nevertheless, this study has some limitations. The sample of patients was restricted to a small part of the Spanish population, making necessary the validation of the method for other populations where the psychological response to chronic pain may differ [96,97], affecting the score of the questionnaires used in this study. Moreover, as shown by Botvinik-Nezer and colleagues, different data analysis pipelines could generate different results using the same data [98]. Thus, more validations are necessary to validate the machine learning approach in a more extensive population and with different data analysis pipelines before a translation to clinical practice can be made.

5. Conclusions

Machine learning techniques have been presented as an efficient option to support the solution of complex problems. Ultimately, the correct application and success of such techniques rely on the quality of the data passed to the algorithm as well as some learning process parameters. In health sciences, these applications have helped researchers to understand and to identify complex syndromes such as neoplasms, Alzheimer's disease, and schizophrenia. Influenced by biological, social, and psychological factors, chronic pain is a complex syndrome that could take advantage of machine learning algorithms. As shown in this study on chronic pain classification, the success of these machine learning algorithms

is tightly correlated with the amount of processed information, the hyper-parameter optimization, and the class of algorithms used. In brief, well-tuned algorithms based on an ensemble of classifiers present a higher chance of success. Nevertheless, this study has some limitations: To expand and reinforce the results, data from different sources and other preprocessing techniques must be evaluated.

Author Contributions: Conceptualization, A.N.S. and P.M.; methodology, A.N.S. and P.M.; software, A.N.S.; formal analysis, A.N.S. and P.M.; writing—original draft preparation, A.N.S., C.N.d.S. and P.M.; visualization, A.N.S.; supervision, P.M. All authors have read and agreed to the published version of the manuscript.

Funding: A.N.S. would like to acknowledge the financial support of the CAPES Foundation, Brazil (proc. BEX 1703/2015-3). The research was also funded by several grants from ERDF/Spanish Ministry of Science, Innovation, and Universities—State Agency of Research (Grant Nos: PSI2017-88388-C4-1-R, PSI2013-48260-C3-1-R).

Acknowledgments: The icons in Figure 1 were freely provided by Eucalyp, Freepik, and pongsakornRed at www.flaticon.com.

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

References

- Albe-Fessar, D.; Berkley, K.; Kruger, L.; Ralston, H., III; Willis, W., Jr. Diencephalic mechanisms of pain sensation. *Brain Res. Rev.* **1985**, *9*, 217–296. [[CrossRef](#)]
- Montoya, P.; Larbig, W.; Braun, C.; Preissl, H.; Birbaumer, N. Influence of Social Support and Emotional Context on Pain Processing and Magnetic Brain Responses in Fibromyalgia. *Arthritis Rheum.* **2004**, *50*, 4035–4044. [[CrossRef](#)] [[PubMed](#)]
- Bevers, K.; Watts, L.; Kishino, N.D.; Gatchel, R.J. The biopsychosocial model of the assessment, prevention, and treatment of chronic pain. *Eur. Neurol. Rev.* **2016**, *12*, 98. [[CrossRef](#)]
- Treede, R.D.; Rief, W.; Barke, A.; Aziz, Q.; Bennett, M.I.; Benoliel, R.; Cohen, M.; Evers, S.; Finnerup, N.B.; First, M.B.; et al. Chronic pain as a symptom or a disease: The IASP Classification of Chronic Pain for the International Classification of Diseases (ICD-11). *Pain* **2019**, *160*, 19–27. [[CrossRef](#)]
- Merskey, H.; Addison, R.G.; Beric, A.; Blumberg, H.; Bogduk, N.; Boivie, J.; Bond, M.R.; Bonica, J.J.; Boyd, D.B.; Deathe, A.B.; et al. *Classification of Chronic Pain: Descriptions of Chronic Pain Syndromes and Definitions of Pain Terms*; IASP Press: Seattle, WA, USA, 1994.
- Wolfe, F.; Clauw, D.J.; Fitzcharles, M.A.; Goldenberg, D.L.; Häuser, W.; Katz, R.L.; Mease, P.J.; Russell, A.S.; Russell, I.J.; Walitt, B. 2016 Revisions to the 2010/2011 fibromyalgia diagnostic criteria. *Semin. Arthritis Rheum.* **2016**, *46*, 319–329. [[CrossRef](#)]
- Gatchel, R.J.; Peng, Y.B.; Peters, M.L.; Fuchs, P.N.; Turk, D.C. The Biopsychosocial Approach to Chronic Pain: Scientific Advances and Future Directions. *Psychol. Bull.* **2010**, *133*, 581–624. [[CrossRef](#)]
- Traeger, A.C.; Henschke, N.; Hübscher, M.; Williams, C.M.; Kamper, S.J.; Maher, C.G.; Moseley, G.L.; McAuley, J.H. Estimating the Risk of Chronic Pain: Development and Validation of a Prognostic Model (PICKUP) for Patients with Acute Low Back Pain. *PLoS Med.* **2016**, *13*, e1002019. [[CrossRef](#)]
- Pagé, I.; Abboud, J.; O’Shaughnessy, J.; Laurencelle, L.; Descarreaux, M. Chronic low back pain clinical outcomes present higher associations with the STarT Back Screening Tool than with physiologic measures: A 12-month cohort study. *BMC Musculoskelet. Disord.* **2015**, *16*, 201. [[CrossRef](#)]
- Cruz-Almeida, Y.; Fillingim, R.B. Can quantitative sensory testing move us closer to mechanism-based pain management? *Pain Med.* **2014**, *15*, 61–72. [[CrossRef](#)]
- Cámara, R.J.A.; Merz, C.; Wegmann, B.; Stauber, S.; Von Känel, R.; Egloff, N. Cost-saving early diagnosis of functional pain in nonmalignant pain: A noninferiority study of diagnostic accuracy. *Pain Res. Treat.* **2016**, *2016*, 1–7. [[CrossRef](#)]
- Gracely, R.H.; Grant, M.A.; Giesecke, T. Evoked pain measures in fibromyalgia. *Best Pract. Res. Clin. Rheumatol.* **2003**, *17*, 593–609. [[CrossRef](#)]
- Vaillant, J.; Pons, C.; Balaguier, R.; Dumolard, A.; Vuillermé, N. In patients with fibromyalgia, there are 18 tender points that are more sensitive than in healthy subjects. *Ann. Phys. Rehabil. Med.* **2017**, *60*, e95. [[CrossRef](#)]
- Ultsch, A.; Löttsch, J. Machine-learned cluster identification in high-dimensional data. *J. Biomed. Inform.* **2016**, *66*, 95–104. [[CrossRef](#)] [[PubMed](#)]

15. Ablin, J.N.; Buskila, D. Update on the genetics of the fibromyalgia syndrome. *Best Pract. Res. Clin. Rheumatol.* **2015**, *29*, 20–28. [[CrossRef](#)] [[PubMed](#)]
16. Diatchenko, L.; Slade, G.D.; Nackley, A.G.; Bhalang, K.; Sigurdsson, A.; Belfer, I.; Goldman, D.; Xu, K.; Shabalina, S.A.; Shagin, D.; et al. Genetic basis for individual variations in pain perception and the development of a chronic pain condition. *Hum. Mol. Genet.* **2005**, *14*, 135–143. [[CrossRef](#)]
17. Paraschiv-Ionescu, A.; Perruchoud, C.; Buchser, E.; Aminian, K. Barcoding human physical activity to assess chronic pain conditions. *PLoS ONE* **2012**, *7*, e32239. [[CrossRef](#)]
18. Hidalgo, B.; Gilliaux, M.; Poncin, W.; Detrembleur, C. Reliability and validity of a kinematic spine model during active trunk movement in healthy subjects and patients with chronic non-specific low back pain. *J. Rehabil. Med.* **2012**, *44*, 756–763. [[CrossRef](#)]
19. Hidalgo, B.; Nielens, H.; Gilliaux, M.; Hall, T.; Detrembleur, C.; Christine, P.; Gilliaux, M.; Hall, T.; Detrembleur, C. Use of kinematic algorithms to distinguish people with chronic non-specific low back pain from asymptomatic subjects: A validation study. *J. Rehabil. Med.* **2014**, *46*, 819–823. [[CrossRef](#)]
20. Costa, I.D.S.; Gamundí, A.; Miranda, J.G.V.; França, L.G.S.; De Santana, C.N.; Montoya, P. Altered functional performance in patients with fibromyalgia. *Front. Hum. Neurosci.* **2017**, *11*, 14. [[CrossRef](#)]
21. Pinheiro, E.S.D.S.; de Queirós, F.C.; Montoya, P.; Santos, C.L.; do Nascimento, M.A.; Ito, C.H.; Silva, M.; Nunes Santos, D.B.; Benevides, S.; Miranda, J.G.V.; et al. Electroencephalographic Patterns in Chronic Pain: A Systematic Review of the Literature. *PLoS ONE* **2016**, *11*, e0149085. [[CrossRef](#)]
22. Schmidt-Wilcke, T. Neuroimaging of chronic pain. *Best Pract. Res. Clin. Rheumatol.* **2015**, *29*, 29–41. [[CrossRef](#)] [[PubMed](#)]
23. Davis, K.D.; Racine, E.; Collett, B. Neuroethical issues related to the use of brain imaging: Can we and should we use brain imaging as a biomarker to diagnose chronic pain? *Pain* **2012**, *153*, 1555–1559. [[CrossRef](#)] [[PubMed](#)]
24. Santana, A.N.; Cifre, I.; De Santana, C.N.; Montoya, P. Using Deep Learning and Resting-State fMRI to Classify Chronic Pain Conditions. *Front. Neurosci.* **2019**, *13*, 1313. [[CrossRef](#)] [[PubMed](#)]
25. Lopez-Martinez, D.; Peng, K.; Lee, A.; Borsook, D.; Picard, R. Pain Detection with fNIRS-Measured Brain Signals: A Personalized Machine Learning Approach Using the Wavelet Transform and Bayesian Hierarchical Modeling with Dirichlet Process Priors. In Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII) Workshop on Recognition, Treatment and Management of Pain and Distress, Cambridge, UK, 3 September 2019.
26. Dodick, D.W.; Loder, E.W.; Manack Adams, A.; Buse, D.C.; Fanning, K.M.; Reed, M.L.; Lipton, R.B. Assessing barriers to chronic migraine consultation, diagnosis, and treatment: Results from the Chronic Migraine Epidemiology and Outcomes (CaMEO) study. *Head. J. Head Face Pain* **2016**, *56*, 821–834. [[CrossRef](#)] [[PubMed](#)]
27. Kress, H.G.; Aldington, D.; Alon, E.; Coaccioli, S.; Collett, B.; Coluzzi, F.; Huygen, F.; Jaksch, W.; Kalso, E.; Kocot-Kepska, M.; et al. A holistic approach to chronic pain management that involves all stakeholders: change is needed. *Curr. Med. Res. Opin.* **2015**, *31*, 1743–1754. [[CrossRef](#)]
28. Lalonde, L.; Leroux-Lapointe, V.; Choinière, M.; Martin, E.; Lussier, D.; Berbiche, D.; Lamarre, D.; Thiffault, R.; Jouini, G.; Perreault, S. Knowledge, attitudes and beliefs about chronic noncancer pain in primary care: A Canadian survey of physicians and pharmacists. *Pain Res. Manag.* **2014**, *19*, 241–250. [[CrossRef](#)]
29. Dworkin, R.H.; Bruehl, S.; Fillingim, R.B.; Loeser, J.D.; Terman, G.W.; Turk, D.C. Multidimensional Diagnostic Criteria for Chronic Pain: Introduction to the ACTTION—American Pain Society Pain Taxonomy (AAPT). *J. Pain* **2016**, *17*, T1–T9. [[CrossRef](#)]
30. Markman, J.D.; Czerniecka-Foxx, K.; Khalsa, P.S.; Hayek, S.M.; Asher, A.L.; Loeser, J.D.; Chou, R. AAPT Diagnostic Criteria for Chronic Low Back Pain. *J. Pain* **2020**. [[CrossRef](#)]
31. Ohrbach, R.; Dworkin, S.F. AAPT Diagnostic Criteria for Chronic Painful Temporomandibular Disorders. *J. Pain* **2019**. [[CrossRef](#)]
32. Arnold, L.M.; Bennett, R.M.; Crofford, L.J.; Dean, L.E.; Clauw, D.J.; Goldenberg, D.L.; Fitzcharles, M.A.; Paiva, E.S.; Staud, R.; Sarzi-Puttini, P.; et al. AAPT Diagnostic Criteria for Fibromyalgia. *J. Pain* **2019**, *20*, 611–628. [[CrossRef](#)]
33. Freeman, R.; Edwards, R.; Baron, R.; Bruehl, S.; Cruccu, G.; Dworkin, R.H.; Haroutounian, S. AAPT Diagnostic Criteria for Peripheral Neuropathic Pain: Focal and Segmental Disorders. *J. Pain* **2019**, *20*, 369–393. [[CrossRef](#)] [[PubMed](#)]

34. Zhou, Q.; Wesselmann, U.; Walker, L.; Lee, L.; Zeltzer, L.; Verne, G.N. AAPT Diagnostic Criteria for Chronic Abdominal, Pelvic, and Urogenital Pain: Irritable Bowel Syndrome. *J. Pain* **2018**, *19*, 257–263. [[CrossRef](#)]
35. Widerström-Noga, E.; Loeser, J.D.; Jensen, T.S.; Finnerup, N.B. AAPT Diagnostic Criteria for Central Neuropathic Pain. *J. Pain* **2017**, *18*, 1417–1426. [[CrossRef](#)] [[PubMed](#)]
36. Dampier, C.; Palermo, T.M.; Darbari, D.S.; Hassell, K.; Smith, W.; Zempsky, W. AAPT Diagnostic Criteria for Chronic Sickle Cell Disease Pain. *J. Pain* **2017**, *18*, 490–498. [[CrossRef](#)]
37. Mellor, F.E.; Thomas, P.W.; Thompson, P.; Breen, A.C. Proportional lumbar spine inter-vertebral motion patterns: A comparison of patients with chronic, non-specific low back pain and healthy controls. *Eur. Spine J.* **2014**, *23*, 2059–2067. [[CrossRef](#)]
38. Dankaerts, W.; O’Sullivan, P.B.; Burnett, A.F.; Straker, L.M.; Danneels, L.A. Reliability of EMG measurements for trunk muscles during maximal and sub-maximal voluntary isometric contractions in healthy controls and CLBP patients. *J. Electromyogr. Kinesiol.* **2004**, *14*, 333–342. [[CrossRef](#)]
39. Rantanen, P.; Nykvist, F. Optimal sagittal motion axis for trunk extension and flexion tests in chronic low back trouble. *Clin. Biomech.* **2000**, *15*, 665–671. [[CrossRef](#)]
40. Hoyer, D.; Kletzin, U.; Adler, D.; Adler, S.; Meissner, W.; Blickhan, R. Gait information flow indicates complex motor dysfunction. *Physiol. Meas.* **2005**, *26*, 545–554. [[CrossRef](#)]
41. Humphrey, A.R.; Nargol, A.V.; Jones, A.P.; Ratcliffe, A.A.; Greenough, C.G. The value of electromyography of the lumbar paraspinal muscles in discriminating between chronic-low-back-pain sufferers and normal subjects. *Eur. Spine J.* **2005**, *14*, 175–184. [[CrossRef](#)]
42. Elfving, B.; Dederig, A.; Nemeth, G. Lumbar muscle fatigue and recovery in patients with long-term low-back trouble—electromyography and health-related factors. *Clin. Biomech.* **2003**, *18*, 619–630. [[CrossRef](#)]
43. Neblett, R.; Brede, E.; Mayer, T.G.; Gatchel, R.J. What is the best surface EMG measure of lumbar flexion-relaxation for distinguishing chronic low back pain patients from pain-free controls? *Clin. J. Pain* **2013**, *29*, 334. [[CrossRef](#)] [[PubMed](#)]
44. Usui, C.; Hatta, K.; Aratani, S.; Yagishita, N.; Nishioka, K.; Kanazawa, T.; Ito, K.; Yamano, Y.; Nakamura, H.; Nakajima, T.; et al. The Japanese version of the 2010 American College of Rheumatology Preliminary Diagnostic Criteria for Fibromyalgia and the Fibromyalgia Symptom Scale: Reliability and validity. *Mod. Rheumatol.* **2012**, *22*, 40–44. [[CrossRef](#)] [[PubMed](#)]
45. Casanueva, B.; García-Fructuoso, F.; Belenguer, R.; Alegre, C.; Moreno-Muelas, J.V.; Hernández, J.L.; Pina, T.; González-Gay, M.Á. The Spanish version of the 2010 American College of Rheumatology Preliminary Diagnostic Criteria for fibromyalgia: Reliability and validity assessment. *Clin. Exp. Rheumatol.* **2016**, *34*, 55.
46. Stewart, J.A.; Mailler-burch, S.; Müller, D.; Studer, M.; Känel, R.V. Rethinking the criteria for fibromyalgia in 2019: The ABC indicators. *J. Pain Res.* **2019**, *12*, 2115–2124. [[CrossRef](#)] [[PubMed](#)]
47. Wiens, J.; Shenoy, E.S. Machine learning for healthcare: On the verge of a major shift in healthcare epidemiology. *Clin. Infect. Dis.* **2018**, *66*, 149–153. [[CrossRef](#)]
48. Munir, K.; Elahi, H.; Ayub, A.; Frezza, F.; Rizzi, A. Cancer diagnosis using deep learning: A bibliographic review. *Cancers* **2019**, *11*, 1235. [[CrossRef](#)] [[PubMed](#)]
49. Hashi, E.K.; Zaman, M.S.U.; Hasan, M.R. An expert clinical decision support system to predict disease using classification techniques. In Proceedings of the 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox’s Bazar, Bangladesh, 16–18 February 2017; pp. 396–400.
50. Pollettini, J.T.; Panico, S.R.; Daneluzzi, J.C.; Tinós, R.; Baranauskas, J.A.; Macedo, A.A. Using machine learning classifiers to assist healthcare-related decisions: Classification of electronic patient records. *J. Med. Syst.* **2012**, *36*, 3861–3874. [[CrossRef](#)]
51. Stafford, I.; Kellermann, M.; Mossotto, E.; Beattie, R.; MacArthur, B.; Ennis, S. A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. *NPJ Digit. Med.* **2020**, *3*, 1–11. [[CrossRef](#)]
52. Lötsch, J.; Ultsch, A. Machine learning in pain research. *Pain* **2018**, *159*, 623. [[CrossRef](#)]
53. Lötsch, J.; Alfredsson, L.; Lampa, J. Machine-learning-based knowledge discovery in rheumatoid arthritis-related registry data to identify predictors of persistent pain. *Pain* **2020**, *161*, 114–126. [[CrossRef](#)]
54. Battineni, G.; Sagaro, G.G.; Chinatalapudi, N.; Amenta, F. Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis. *J. Personal. Med.* **2020**, *10*, 21.
55. López-Solà, M.; Woo, C.W.; Pujol, J.; Deus, J.; Harrison, B.J.; Monfort, J.; Wager, T.D. Towards a neurophysiological signature for fibromyalgia. *Pain* **2017**, *158*, 34–47. [[CrossRef](#)] [[PubMed](#)]

56. Davis, F.; Gostine, M.; Roberts, B.; Risko, R.; Cappelleri, J.C.; Sadosky, A. Characterizing classes of fibromyalgia within the continuum of central sensitization syndrome. *J. Pain Res.* **2018**, *11*, 2551. [[PubMed](#)]
57. Andrés-Rodríguez, L.; Borràs, X.; Feliu-Soler, A.; Pérez-Aranda, A.; Rozadilla-Sacanell, A.; Arranz, B.; Montero-Marin, J.; García-Campayo, J.; Angarita-Osorio, N.; Maes, M.; et al. Machine Learning to Understand the Immune-Inflammatory Pathways in Fibromyalgia. *Int. J. Mol. Sci.* **2019**, *20*, 4231.
58. Ung, H.; Brown, J.E.; Johnson, K.A.; Younger, J.; Hush, J.; Mackey, S. Multivariate Classification of Structural MRI Data Detects Chronic Low Back Pain. *Cereb. Cortex* **2014**, *24*, 1037–1044. [[CrossRef](#)]
59. Judd, M.; Zulkernine, F.; Wolfrom, B.; Barber, D.; Rajaram, A. Detecting low back pain from clinical narratives using machine learning approaches. In Proceedings of the International Conference on Database and Expert Systems Applications, Regensburg, Germany, 3–6 September 2018; pp. 126–137.
60. Shen, W.; Tu, Y.; Gollub, R.L.; Ortiz, A.; Napadow, V.; Yu, S.; Wilson, G.; Park, J.; Lang, C.; Jung, M.; et al. Visual network alterations in brain functional connectivity in chronic low back pain: A resting state functional connectivity and machine learning study. *NeuroImage Clin.* **2019**, *22*, 101775.
61. Mano, H.; Kotecha, G.; Leibnitz, K.; Matsubara, T.; Sprenger, C.; Nakae, A.; Shenker, N.; Shibata, M.; Voon, V.; Yoshida, W.; et al. Classification and characterisation of brain network changes in chronic back pain: A multicenter study. *Wellcome Open Res.* **2018**, *3*. [[CrossRef](#)]
62. Shalev-Shwartz, S.; Ben-David, S. *Understanding Machine Learning: From Theory to Algorithms*; Cambridge University Press: Cambridge, UK, 2014; pp. 6–7.
63. Burdack, J.; Horst, F.; Giesselbach, S.; Hassan, I.; Daffner, S.; Schöllhorn, W.I. Systematic Comparison of the Influence of Different Data Preprocessing Methods on the Performance of Gait Classifications Using Machine Learning. *Front. Bioeng. Biotechnol.* **2020**, *8*, 260.
64. Shin, H.; Markey, M.K. A machine learning perspective on the development of clinical decision support systems utilizing mass spectra of blood samples. *J. Biomed. Inform.* **2006**, *39*, 227–248.
65. Eggenesperger, K.; Lindauer, M.; Hoos, H.H.; Hutter, F.; Leyton-Brown, K. Efficient benchmarking of algorithm configurators via model-based surrogates. *Mach. Learn.* **2018**, *107*, 15–41.
66. Van Rijn, J.N.; Hutter, F. Hyperparameter importance across datasets. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data, Mining, London, UK, 19–23 August 2018; pp. 2367–2376.
67. Beck, A.T.; Steer, R.A.; Brown, G.K. *Manual for the Beck Depression Inventory-II*; Psychological Corporation: San Antonio, TX, USA, 1996.
68. Spielberger, C.D.; Gorsuch, R.L.; Lushene, R.; Vagg, P.R.; Jacobs, G.A. *Manual for the State-Trait Anxiety Inventory*; Consulting Psychologists Press: Palo Alto, CA, USA, 1983.
69. Breivik, H.; Collett, B.; Ventafridda, V.; Cohen, R.; Gallacher, D. Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment. *Eur. J. Pain* **2006**, *10*, 287–333. [[CrossRef](#)] [[PubMed](#)]
70. de Heer, E.W.; ten Have, M.; van Marwijk, H.W.; Dekker, J.; de Graaf, R.; Beekman, A.T.; van der Feltz-Cornelis, C.M. Pain as a risk factor for common mental disorders. Results from the Netherlands Mental Health Survey and Incidence Study-2: A longitudinal, population-based study. *Pain* **2018**, *159*, 712–718. [[CrossRef](#)] [[PubMed](#)]
71. Dersh, J.; Polatin, P.B.; Gatchel, R.J. Chronic pain and psychopathology: Research findings and theoretical considerations. *Psychosom. Med.* **2002**, *64*, 773–786. [[PubMed](#)]
72. Rapti, E.; Damigos, D.; Apostolara, P.; Roka, V.; Tzavara, C.; Lionis, C. Patients with chronic pain: Evaluating depression and their quality of life in a single center study in Greece. *BMC Psychol.* **2019**, *7*, 86. [[CrossRef](#)] [[PubMed](#)]
73. Villafaina, S.; Sitges, C.; Collado-Mateo, D.; Fuentes-García, J.P.; Gusi, N. Influence of depressive feelings in the brain processing of women with fibromyalgia: An EEG study. *Medicine* **2019**, *98*, e15564. [[CrossRef](#)]
74. White, I.R.; Royston, P.; Wood, A.M. Multiple imputation using chained equations: Issues and guidance for practice. *Stat. Med.* **2011**, *30*, 377–399. [[CrossRef](#)]
75. Yu, H.f.; Huang, F.l.; Lin, C.J. Dual coordinate descent methods for logistic regression and maximum entropy models. *Mach. Learn.* **2011**, *85*, 41–75. [[CrossRef](#)]
76. Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **1992**, *46*, 175–185.

77. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. Classification and regression trees. *Int. Group* **1984**, *432*, 151–166.
78. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32.
79. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
80. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
81. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
82. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006; p. 209.
83. Nijeweme-d'Hollosy, W.O.; van Velsen, L.; Poel, M.; Groothuis-Oudshoorn, C.G.; Soer, R.; Hermens, H. Evaluation of three machine learning models for self-referral decision support on low back pain in primary care. *Int. J. Med. Inform.* **2018**, *110*, 31–41. [[CrossRef](#)] [[PubMed](#)]
84. Bouthillier, X.; Varoquaux, G. *Survey of Machine-Learning Experimental Methods at NeurIPS2019 and ICLR2020*; Research Report; Inria Saclay Ile de France: Paris, France, 2020.
85. Jones, K.D.; Gelbart, T.; Whisenant, T.C.; Waalen, J.; Mondala, T.S.; Iklé, D.N.; Salomon, D.R.; Bennett, R.M.; Kurian, S.M. Genome-wide expression profiling in the peripheral blood of patients with fibromyalgia. *Clin. Exp. Rheumatol.* **2016**, *34*, 89.
86. Schmukler, J.; Jamal, S.; Castrejon, I.; Block, J.A.; Pincus, T. Fibromyalgia Assessment Screening Tools (FAST) based on only Multidimensional Health Assessment Questionnaire (MDHAQ) scores as clues to fibromyalgia. *ACR Open Rheumatol.* **2019**, *1*, 516–525. [[CrossRef](#)]
87. Gibson, K.A.; Castrejon, I.; Descallar, J.; Pincus, T. Fibromyalgia Assessment Screening Tool (FAST): Clues to fibromyalgia on a multidimensional health assessment questionnaire (MDHAQ) for routine care. *J. Rheumatol.* **2019**, *47*, 761–769.
88. Gerrits, M.M.; van Marwijk, H.W.; van Oppen, P.; van der Horst, H.; Penninx, B.W. Longitudinal association between pain, and depression and anxiety over four years. *J. Psychosom. Res.* **2015**, *78*, 64–70. [[CrossRef](#)]
89. Mills, S.E.; Nicolson, K.P.; Smith, B.H. Chronic pain: A review of its epidemiology and associated factors in population-based studies. *Br. J. Anaesth.* **2019**, *123*, e273–e283. [[CrossRef](#)]
90. Giesecke, T.; Gracely, R.H.; Grant, M.A.; Nachemson, A.; Petzke, F.; Williams, D.A.; Clauw, D.J. Evidence of Augmented Central Pain Processing in Idiopathic Chronic Low Back Pain. *Arthritis Rheum.* **2004**, *50*, 613–623. [[CrossRef](#)]
91. O'Neill, S.; Manniche, C.; Graven-Nielsen, T.; Arendt-Nielsen, L. Generalized deep-tissue hyperalgesia in patients with chronic low-back pain. *Eur. J. Pain* **2007**, *11*, 415–420. [[CrossRef](#)]
92. Wolfe, F.; Smythe, H.A.; Yunus, M.B.; Bennett, R.M.; Bombardier, C.; Goldenberg, D.O.N.L.; Tugwell, P.; Campbell, S.M.; Abeles, M.; Clark, P.; et al. The american college of rheumatology 1990 criteria for the classification of fibromyalgia report of the Multicenter Criteria Committee. *Arthritis Rheum.* **1990**, *33*, 160–172. [[PubMed](#)]
93. Georgopoulos, V.; Akin-Akinyosoye, K.; Zhang, W.; McWilliams, D.F.; Hendrick, P.; Walsh, D.A. Quantitative sensory testing and predicting outcomes for musculoskeletal pain, disability, and negative affect: A systematic review and meta-analysis. *Pain* **2019**, *160*, 1920–1932. [[CrossRef](#)] [[PubMed](#)]
94. Vierck, C.J.; Wong, F.; King, C.D.; Mauderli, A.P.; Schmidt, S.; Riley, J.L., III. Characteristics of sensitization associated with chronic pain conditions. *Clin. J. Pain* **2014**, *30*, 119. [[CrossRef](#)] [[PubMed](#)]
95. Slade, G.D.; Sanders, A.E.; Ohrbach, R.; Fillingim, R.B.; Dubner, R.; Gracely, R.H.; Bair, E.; Maixner, W.; Greenspan, J.D. Pressure pain thresholds fluctuate with, but do not usefully predict, the clinical course of painful temporomandibular disorder. *Pain* **2014**, *155*, 2134–2143. [[CrossRef](#)] [[PubMed](#)]
96. Meeus, M. Are pain beliefs, cognitions, and behaviors influenced by race, ethnicity, and culture in patients with chronic musculoskeletal pain: A systematic review. *Pain Phys.* **2018**, *21*, 541–558.
97. Rahavard, B.B.; Candido, K.D.; Knezevic, N.N. Different pain responses to chronic and acute pain in various ethnic/racial groups. *Pain Manag.* **2017**, *7*, 427–453. [[CrossRef](#)]

98. Botvinik-Nezer, R.; Holzmeister, F.; Camerer, C.F.; Dreber, A.; Huber, J.; Johannesson, M.; Kirchler, M.; Iwanir, R.; Mumford, J.A.; Adcock, R.A.; et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* **2020**, *582*, 1–7. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).