UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA

LARISSA IULLE MOREIRA

Desenvolvimento de modelos preditores de resposta clínica em doentes com dor crônica com técnicas de inteligência artificial

São Paulo
2023

LARISSA IULLE MOREIRA

Desenvolvimento de modelos preditores de resposta clínica em doentes com dor crônica com técnicas de inteligência artificial

Versão Original

Dissertação apresentada à Faculdade de Medicina da Universidade de São Paulo para obtenção do título de Mestre em Ciências

Programa de Neurologia

Orientador: Prof. Dr. Daniel Ciampi Araujo de Andrade

(Versão corrigida. Resolução CoPGr 6018/11, de 1 de novembro de 2011. A versão original está disponível na Biblioteca da FMUSP)

São Paulo
2023

LARISSA IULLE MOREIRA

Development of predictive models of clinical response in patients with chronic pain with artificial intelligence techniques

Dissertation presented to the School of Medicine of the University of São Paulo to obtain the degree of Master in Science Graduate

Neurology Program

Advisor: Prof. Dr. Daniel Ciampi Araujo de Andrade

(Versão corrigida. Resolução CoPGr 6018/11, de 1 de novembro de 2011. A versão originalestá disponível na Biblioteca da FMUSP)

São Paulo
2023

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da
Faculdade de Medicina da Universidade de São Paulo

Moreira, Larissa Iulle
Desenvolvimento de modelos preditores de resposta clínica em doentes com dor crônica com técnicas de inteligência artificial / Larissa Iulle Moreira. -- São Paulo, 2023.
Dissertação(mestrado)--Faculdade de Medicina da Universidade de São Paulo.
Programa de Neurologia.
Orientador: Daniel Ciampi Araujo de Andrade.

Descritores: 1.Dor crônica 2.Inteligência artificial 3.Tratamento 4.Aprendizado de máquina 5.Controle da dor

USP/FM/DBD-003/23

Responsável: Erinalva da Conceição Batista, CRB-8 6755

À minha mãe, minha maior incentivadora e meu propósito para seguir sempre em frente. Que sempre valorizou o estudo e abdicou da própria vida para me oferecer as melhores oportunidades. Todo meu amor e gratidão.

AGRADECIMENTO

Ao Prof. Dr. Daniel Ciampi de Andrade, idealizador deste projeto, pela orientação, paciência e contribuições ao trabalho. Agradeço por ter transmitido uma parcela da sua vasta sabedoria acadêmica e pela confiança depositada em mim como pesquisadora.

Ao Prof. Dr. Manoel Jacobsen Teixeira, exemplo de dedicação à ciência, à medicina e tantas outras áreas que se complementam. Excelente professor e profissional. Pioneiro nos estudos de dor, defensor do exame físico e da avaliação sistematizada do doente. Um homem à frente do seu tempo, com a mente aberta para novos conhecimentos e descobertas científicas e que sempre transmite todo o conhecimento que possui. Minha eterna gratidão por contribuir para minha formação acadêmica, profissional e pessoal, pela amizade, pelo contínuo estímulo ao aprendizado, por acreditar e confiar diariamente em mim. Obrigada pelas horas dispensadas em entender este trabalho, pela visão crítica e correções do texto.

Ao Prof. Adriano Veloso, pelo incansável comprometimento com este trabalho. Esteve sempre disponível em nos ajudar, abriu as portas do seu laboratório na UFMG, abraçou nossa proposta e enriqueceu o projeto com o uso da ciência de dados. Agradeço por todo ensinamento, pela paciência e parceria.

Ao Anderson Bessa, pela paciência em me ensinar uma pequena fração do que é o mundo da ciência de dados e pela parceria em executar esse longo e complexo projeto, que é tão dele quanto meu.

À Profa. Dra. Lin Tchia Yeng, pela generosidade, companheirismo, amizade, apoio emocional e profissional, por todas as oportunidades e incentivo constante. Agradeço pelo seu olhar crítico, como médica e pesquisadora, na qualificação e defesa, que enriqueceu o trabalho e permitiu o aperfeiçoamento do texto.

Ao Prof. Dr. Eberval Gadelha Figueiredo, agradeço pelas contribuições e sugestões que enriqueceram este projeto, desde meu ingresso no programa de pós-graduação, depois no exame de qualificação e, finalmente, na defesa da dissertação.

Ao Prof. Luiz Henrique Martins Castro, coordenador do programa de pós-graduação do Departamento de Neurologia, pela oportunidade de realizar esta pesquisa.

À Profa. Suely Kazue Nagahashi Marie, pelo comprometimento, atenção e cuidado com os alunos de pós-graduação do Departamento de Neurologia. Seu amor pelo ensino é evidente, minha eterna gratidão.

À Thais Figueira, secretária do programa de pós graduação do Departamento de Neurologia, pelo profissionalismo e atenção dispensada aos alunos. Agradeço também pela amizade, companheirismo e apoio diário.

Ao Dr. Jefferson Rosi Junior, pela amizade e por sempre me incentivar a crescer profissionalmente. Obrigada pelas contribuições na banca da defesa da dissertação.

Ao Dr. Gabriel Taricani Kubota pela amizade e confiança e a toda equipe do Grupo de dor e docentes que participaram da minha formação em dor no HCFMUSP, não cito nome pois são muitos, mas vocês sabem quem são! Obrigada por me acolherem no ambulatório e me ensinarem tanto sobre dor

A toda a equipe da UFMG e Kunumi pela dedicação com nossos projetos, em especial à Daniella Castro e Ana Macedo, obrigada pelo acolhimento, ajuda e apoio, mesmo com a distância física me ensinaram muito sobre esse mundo da ciência de dados.

Aos funcionários e colaboradores do Departamento de Neurologia do HC, Elizabeth Lauritano, Ester Tomazini, Maria Cecília Milani, Reiko Uchizono Simomura, Rondon Gustavo S. Moura, Vanessa Ramos, Erli Vieira Soares, Maria do Socorro, Sueli Martins, Valdenira dos Santos e Vera Lucia Pereira. Obrigada!

Ao Davi Strauss, meu professor de inglês, por me ajudar com a apresentação deste trabalho em congressos internacionais.

À minha família, por entenderem minha ausência para realizar e concretizar este projeto e serem meu suporte emocional e escape sempre que preciso.

À minha Mãe, meu porto seguro e razão pela minha busca em ser uma pessoa e profissional melhor. Que sempre batalhou para que eu tivesse a melhor educação e as melhores oportunidades. Sempre valorizou o estudo e o conhecimento e abdicou do próprio tempo e da própria vida para me dar todas as oportunidades. Obrigada por acreditar em mim, por apoiar minhas decisões de vida, pelo amor em cuidar de mim, por ter me dado educação, senso moral e por me fazer ser quem sou hoje, espero poder retribuir!

Ao Leandro, meu companheiro de vida, pelo apoio incondicional em todos os momentos, tanto de alegrias como de dificuldades. Agradeço pela compreensão com minhas loucuras profissionais e ajuda diária, seu apoio foi fundamental para que esse trabalho pudesse existir. Obrigada pela paciência, cuidado e amor!

À Cidinha (*in memoriam*), minha conselheira, amiga e tia, por sempre me incentivar, por me apoiar quando decidi iniciar o mestrado e ser sempre presente em meu caminho. Obrigada por me ensinar tanto da vida. Saudades!

Ao Renan, meu primo-irmão, pelo apoio emocional e ajuda incansável na formatação das várias versões deste trabalho.

Aos meus amigos que alimentaram minha alma durante este período.

À CAPES pelo apoio financeiro para o desenvolvimento deste projeto.

RESUMO

Moreira LI. Desenvolvimento de modelos preditores de resposta clínica em doentes com dor crônica com técnicas de inteligência artificial [dissertação]. São Paulo: Faculdade de Medicina de São Paulo, Universidade de São Paulo; 2023.

Introdução: A dor é o sintoma mais prevalente no ser humano, está entre as três principais causas globais de anos vividos com incapacidade. O tratamento inadequado da dor é grave problema de saúde pública. O uso de abordagens de inteligência artificial (IA) nas áreas da saúde possibilita melhora da prevenção, detecção, diagnóstico, tratamento das doenças e utilização de recursos de saúde e pode transformar os modelos de prestação de cuidados à saúde. Objetivo: se os modelos de aprendizado de máquina são capazes de predizer melhora significativa ou não em doentes com dor com base nas informações da primeira consulta. Método: Foram analisados dados de 506 doentes atendidos no Ambulatório do Centro de Dor da Clínica Neurológica do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, coletados por uma interface sistematizada e padronizada de avaliação no período de 1 ano. Os desfechos foram a melhora de dor com base na escala verbal analógica (EVA) e impressão global de mudança pela versão do médico e do doente (IGM). A análise descritiva foi realizada por estatística simples. Uma abordagem de aprendizado de máquina (AM) supervisionada foi realizada a partir de um algoritmo desenvolvido e um total de 338 atributos relacionados a dor foram incluídos no estudo. O algoritmo foi treinado a partir de algoritmos de *Random Forest* e *XGBoost* disponíveis. O desempenho foi avaliado pela métrica da Área Sobre a Curva (AUC – ROC) e os fatores explicativos apresentados como gráficos de resumo SHAP. Resultados: Os resultados fornecem evidência de que o AM tem potencial de auxiliar no manejo da dor e a tomada de decisões clínicas. Os modelos gerados originaram um questionário com 12 questões com as melhores variáveis. Conclusão: Os resultados deste estudo sugerem que o uso de IA tem efeito positivo no gerenciamento do doente com dor e são necessários mais estudos com abordagens de IA nos doentes com dor crônica.

Palavras-chaves: Dor crônica. Inteligência artificial. Tratamento. Aprendizado de máquina. Controle da dor.

ABSTRACT

Moreira LI. Development of predictive models of clinical response in patients with chronic pain with artificial intelligence techniques [dissertation]. São Paulo: “Faculdade de Medicina, Universidade de São Paulo”; 2023.

Introduction: Pain is the most prevalent symptom in humans; it is among the three main global causes of years lived with disability. Inadequate pain treatment is a severe public health problem. Artificial intelligence (AI) approaches in healthcare enable improved prevention, detection, diagnosis, treatment of diseases, and utilization of healthcare resources and can transform healthcare delivery models. Objective: whether machine learning models can predict significant improvement or not in patients with pain based on information from the first consultation. Method: Data from 506 patients who attended the Ambulatory of the Pain Center of the Neurological Clinic of the Hospital das Clínicas of the Faculty of Medicine of the University of São Paulo were analyzed, collected by a systematized and standardized interface of evaluation in 1 year. Outcomes were improvement in pain based on the verbal analog scale (VAS) and global impression of change by the doctor and patient version (GIC). Descriptive analysis was performed using simple statistics. A supervised machine learning (ML) approach was performed using a developed algorithm, and 338 pain-related attributes were included in the study. The algorithm was trained from available Random Forest and XGBoost algorithms. Performance was evaluated by the Area Under Curve (AUC – ROC) metric, and the explanatory factors were presented as SHAP summary graphs. Results: The results provide evidence that BF has the potential to help with pain management and clinical decision-making. The generated models originated a questionnaire with 12 questions with the best variables. Conclusion: The results of this study suggest that the use of AI has a positive effect on the management of patients with pain, and further studies are needed with AI approaches in patients with chronic pain.

Keywords: Chronic pain. Artificial intelligence. Treatment. Machine learning. Pain control.

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS LIST

PCNC	Pain Center of the Neurological Clinic
EDA	Exploratory data analysis
ML	Machine Learning
AUC	Area under the curve
DN-4	Douleur Neuropathique Pain 4 Questions
VAS	Verbal Analog Scale
FMUSP	Faculty of Medicine of the University of São Paulo
GBD	Global Burden of Disease Study
GIC	Global Impressions Clinical
HCFMUSP	Hospital das Clínicas of the Faculty of Medicine of the University of São Paulo
AI	Artificial intelligence
IASP	International Association for the Study of Pain
IBM	International Business Machines
GIC	Global Impression Clinical
IEAA	Interface Eletrônica de Atendimento Ambulatorial
MAD	mean absolute deviation
ML	Machine Learning
"n"	Numbers
WHO	World Health
PPSUS	Research Program for the SUS
"%"	percentages
ROC	receiver operator characteristic curve
SHAP	SHapley Additive exPlanations
SP	São Paulo
SPSS	Statistical Package for the Social Sciences
UFMG	University Federal of Minas Gerais
USP	University of Sao Paulo

DEFINITIONS

Accuracy	Refers to how close to reality the results are found. For example, in a model with an accuracy of 70%, the model is correct at 70 predictions for every 100 predictions made.
Clustering	A form of prediction that combines results from a collection of models. Increase model accuracy.
Algorithms	Are sets of steps, rules and/or processes that a computer can follow to obtain a solution to a specific type of problem.
Machine Learning	It is a data analysis method that allows systems, after being trained, to create algorithms capable of modifying and improving with experience
Deep Learning	It is often defined as a subset of machine learning. It refers to the use of neural networks in several layers that repeat a task, learn progressively, and seek the gradual improvement of results (58,60)
Baseline	Is the model used as a reference point to compare the performance of another model, usually more complex.
Test base	This is data presented to the model that simulates real forecasts to verify the real performance.
Training base	Set of data used by a training algorithm to create a model.
Class	Category of a set of enumerated target values for a label. For example, in a chronic pain binary classification model, there are two classes, with pain and without pain.
Dataset	A dataset or "dataset" is a collection of data that is usually tabulated.
Ensemble	Is a machine learning technique that combines the result of multiple models in order to produce a better predictive model.
Random Forest	Estimate based on decision tree models.
Framework	Is a library that unites codes with several functions ready to be imported and used. They have many different resources and algorithms already optimized.
Big data	Refers to large sets of data.
Artificial Intelligence	Technological advancement that allows computer systems to perform tasks that would be performed by humans.
Instance	Records in the dataset about which you want to make a prediction. For example, each instance could be a record containing patient information such as age, weight.
Threshold	A threshold value criterion, higher and lower, used as a parameter.
Model	A statistical representation of a forecasting task. Train a model and then use the model to make predictions.
Overfitting	Occurs when the model learns the details in the training data. In machine learning, the aim is to create a model that learns about the data and then makes predictions with the input of new data.
Precision	It is the ratio between the True Positives and all Positives. It would be the measure of patients that were identified among all patients.

Features	It transforms raw data into data processed as features ready to be used by any model. For example, an instance with patient data would have a resource indicating gender.
Label	A response (outcome) to a prediction task. For example, the instance label defined in this project would be a pain, indicating whether or not the patient has improved.
Validation	A process used, as part of training, to assess the quality of a machine learning model. Because the validation set is separate from the training set, validation helps ensure that the model's performance generalizes beyond the training set.

SUMMARY

INTRODUCTION

1 INTRODUCTION

Pain is the most prevalent symptom in humans (1). More than 45% of Brazilians have chronic pain, and in the state of São Paulo, pain is present in 31% of the population (2). Its prevalence is increasing due to the aging of the population, which is due to the improvement of sanitary conditions, nutrition, quality of life, and therapeutic interventions aimed at the treatment of naturally disabling or fatal conditions (3); it is an important social, economic and assistance burden (4).

The International Association for the Study of Pain (IASP) defines pain as “an unpleasant sensory and emotional experience associated, or similar to that associated with actual or potential tissue damage”; and it is characterized as chronic when the pain lasts longer than three months (5).

Patients with chronic pain use health services five times more often than those without pain and require specialized care, generally multidisciplinary, involving medication, physical medicine and rehabilitation procedures, mental health, and, often, neuroanesthetic procedures and functional neurosurgical devices (3). In addition, associated morbidities are common in patients with chronic pain, among them anxiety, depression, sleep disorders, fatigue, and impairment of physical and mental performance (6).

Pain is often a warning of a medical condition or injury. Treatment of the underlying medical condition is crucial and can improve the pain, but the pain may persist despite treatment of the initial condition that caused it. Therefore, the epidemiological issue is aggravated due to incurability, that is, many patients who suffer from it have their suffering refractory to the available treatments and, consequently, the resulting disability becomes progressive (5, 7). According to the Global Burden of Disease Study, low back pain and neck pain are among the three leading global causes of years lived with disability (8).

Predicting the evolution of chronic pain and the possibility of its relief is difficult because it has different etiologies, such as trauma, neoplasms, and functional disorders, among others that are often difficult to detect in the first consultations (9). Chronic pain is usually refractory to treatment because its diagnosis is usually difficult, its treatment is prolonged, the possibility of improvement is low and when it occurs it is slow even with the use of well-structured therapies and with high costs (10). Many

therapeutic schemes have been developed using analgesics and other pharmacological treatments, physical therapies, rehabilitation methods, psychotherapies, and functional anesthetic and neurosurgical procedures to treat and prevent pain (11, 12). As a consequence of population aging, the social burden of chronic pain management must increase considerably and new innovative procedures aimed at its management must be implemented (13). In addition to the primary costs, there are those related to the displacement of patients to treatment centers and abstention from work, school, family, and social activities, which can compromise the follow-up and the result of interventions (14).

Treating patients with pain is increasingly recognized as a mandatory component of curricula related to health areas and government policies (15). The scarcity of knowledge about prevention, application of treatments, rehabilitation, and monitoring of patients with pain can result in ineffective treatments, iatrogenesis, and idiosyncrasy (16). Unsatisfactory pain treatment is a serious public health problem and further compromises patients' quality of life (17) and worsens their isolation, depression, anxiety, frustration, sleep disorders, and fatigue (3).

According to the guidelines proposed by the IASP, treatment should focus on pain mechanisms and the medication mechanism of action (18). Pharmacological interventions are generally the first-line therapies in pain control (19). Although chronic pain is a global problem, there are still gaps.

With the increasing application of information technologies (IT) in health policies and their relevance in patient care, there is a need to develop studies that make it possible to analyze the real scope of these modalities of access to information and understand the themes applied in the medicine to improve the activity and attitude of health professionals, based on the characteristics of patients and doctor-patient relationships (20). In addition, computational science methods can incorporate clinical and experimental data, even complex ones, to better understand the complexity of pain management (21).

Scientific computing, which is the science that studies data processing techniques and methods with a focus on the development of algorithms, made significant progress with the popularization of computer and internet access, the use of databases and their mining to generate knowledge, remote monitoring techniques, the use of virtual reality and computational simulation (22). In medicine, the use of this reality originated in the 20th century with the term “telemedicine” (23).

The use of Artificial Intelligence (AI) in medicine makes it possible to learn (train, build, formulate, or induce a knowledge model) with a data set and look for patterns following algorithms defined by specialists who are capable of proposing solutions to medical problems (24), such as the Machine Learning (ML) technique, which uses algorithms based on mathematics and statistics to perform learning. Among data science techniques, ML stands out as a set of methods that make it possible to detect patterns in data and, based on discovered patterns, predict or classify future data to analyze structures such as subgroups of data or extract information from data. Adequate to generate new knowledge (21). The use of ML is based on the need to process and obtain useful information from large volumes of data in situations where it is impracticable to perform processing and analysis manually. There are expectations that in health, AI will make it possible to improve the prevention, detection, diagnosis, and treatment of diseases and transform healthcare delivery models (25) .

There is great interest that the application of ML methods corroborates to identify predictive variables of improvement in treatments for making clinical decisions with a greater probability of effectiveness (14). Understanding what induces a prediction is essential to determine the appropriately targeted interventions in clinical settings, for this reason, ML methods employed in clinical applications move away from complex models, although more accurate, and mold themselves into more simple interpretable models (e.g., linear) (26).

Although many studies are based on ML to diagnose chronic pain, little attention has been given to its treatment and management. There have been many attempts to carry out traditional studies seeking answers about the possibilities of pain improvement or not with the use of specific characteristics of chronic pain, many of them with a psychological nature. Therefore, research on AI and its methods, such as ML, in treating and rehabilitating patients with chronic pain has become relevant. Some works evaluate the use of AI in the self-management of pain, with the creation of mechanisms that the patient uses to manage his pain, such as applications. Virtual reality enables a wide range of health data, both in clinical and non-clinical settings, and can potentially be applied with ML algorithms to support chronic pain research. These analyses can help predict, identify, and treat diseases (27).

To carry out the present study, structured and unstructured data from patients with chronic pain of different causes were analyzed to highlight the most important attributes of pain characteristics with the objective of mapping the patient's information

collected in their first consultation for each outcome used. The result indicates whether or not the patient will experience significant relief in pain sensation. Furthermore, the analysis of these variables may enable the development of procedures aimed at improving pain and use this technology to detect patients with a high probability of not benefiting from the usual treatments and enable the development of new individualized therapies, medication, physical medicine and rehabilitation, mental health or based on the execution of invasive procedures.

OBJECTIVES

2 OBJECTIVES

2.1 Primary

To identify predictive characteristics of pain improvement in patients who do not respond to pain treatments.

2.2 Secondary

To develop a questionnaire with these predictive characteristics.

LITERATURE REVIEW

3 LITERATURE REVIEW

3.1 Chronic pain

Pain is a biological phenomenon with a subjective psychological dimension that emerges from brain activity; various nuances of behavior and cognition influence it. It is a complex condition dependent on the interaction between biological, psychological, ethical, age, and social factors (28).

According to the IASP, pain was defined in 1979 as “an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage”. In 2020, this concept was revised and aims to expand the scope of the definition of pain, which is now conceptualized as an “unpleasant sensory and emotional experience associated with or similar to that resulting from actual or potential tissue injury” (5, 29). Complementing the new definition, the authors included six explanatory notes about pain:

1. Pain is always a personal experience and is influenced by biological, psychological, and social factors (5, 29).
2. Pain and nociception are not the same phenomenon and cannot be exclusively determined by the activity of sensitive neurons (5, 29).
3. Each person's experience interferes with the concept of pain (5, 29).
4. Self-report of pain should be considered (5, 29).
5. Although the individual generally adapts to pain, it can have negative effects on social and psychological well-being (5, 29).
6. The inability to communicate may not reveal that a human being or a non-human animal is experiencing pain; that is, the verbal description is just one of several behaviors or instruments to express pain (5, 29).

Pain is generally classified as acute or chronic according to duration (30). Chronic pain differs from acute pain in numerous ways as a specific disease or injury usually causes acute pain due to tissue damage inflammation or a disease process that lasts for days or a few weeks and resolves as the cause or the injured tissue heals (31). Pain is considered chronic when it persists or recurs for more than three months, persists for a long time after the cure of the disease or causal injury, and can manifest itself even in the absence of injury, that is, more than what is observed about acute

pain, there is not a direct relation between the magnitude of the tissue injury and the intensity of the pain (32), it may persist for days, weeks, months, years or permanently after the injury, or it may not have a defined etiology, that is, chronic pathological pain becomes a disease (33). Additionally, chronic pain causes suffering, disability, and negative biopsychosocial impact (3). It often becomes some patients' only clinical or predominant problem (34). Chronic pain incorporates several physical, psychological, and social factors (35). It is a complex biopsychosocial experience, clearly influenced by neurological and psychosocial processes (36).

Chronic pain is a serious public health problem. More than 45% of the Brazilian population has chronic pain (2). In developed countries, the prevalence is also high; up to 50% of the general population can be affected by chronic pain (35). These patients use health services five times more than patients without pain and require specialized care (37, 38), generally multidisciplinary, involving medication, physical and mental health rehabilitation, and often physical medicine, neuroanesthetic, and functional neurosurgical devices. It is one of the most common diseases that affect humans and the most common cause of years lived with disability in the world (39, 40). According to the last study published in 2019 by the Global Burden of Disease Study (GBD), chronic pain is among the ten diseases that most affect years lived with disability (AVIs), a metric composed not only of the prevalence of diseases but also of the morbidities, mortalities, and age of occurrence experienced by young people, adults and the elderly. Among the ten diseases most commonly affect humanity, two are chronic pain: musculoskeletal disorders, and headaches, which most burden the health system. Over the past 30 years, global health has improved, but the number of years lived with a disability has remained stable. Among the main public health problems, the inappropriate use of analgesics in therapies that are not efficient in pain management stands out (8).

The reduced work capacity of patients with chronic pain results in enormous costs with sick leave and benefits arising from disability, the high consumption of health resources, and expenses related to issues arising from the labor market (10). The chance of an unemployed individual with chronic pain is twice as high as the general population (41). In addition, mortality rates are 2.5 times higher in individuals affected by chronic pain (42).

Pain is classified, according to its pathophysiology, as nociceptive pain when there is continuous inflammation and demonstrable tissue damage (osteoarthritis,

rheumatoid arthritis, fracture, injury, trauma, etc.) (5); neuropathic pain, when there is damage to structures of the peripheral or central somatosensory system, and can be felt in areas without tissue damage, far from the damage or disease of the nervous system (43); nociplastic pain when there is no identifiable causal lesion (44); or mixed pain when combining nociceptive, neuroplastic and nociplastic profiles. To be considered neuropathic pain, some questions must be accounted for, that is, characteristic verbal descriptors (burning, stinging, numbness) (45), plausible anatomical distribution for the pain complaint and presence of causal health condition, such as diabetes mellitus, neurotoxicity, trauma, among others (43, 46). Nociplastic pain results from abnormal processing or modulation of nociception or other perceptions without clear evidence of tissue damage activating nociceptors and/or injury or disease of the somatosensory system. Pain is mixed when it results from damage to somatosensory nervous system structures and other organs or tissues, as it occurs in cancer patients, for example (47-49).

Despite recent advances in pain research, there are still no effective measures to treat most patients with chronic pain. Frequently, the cause of chronic pain is unknown; its occurrence is unaccompanied by objective findings, and effective methods to detect it are unknown, which makes treatment and management complex and generally ineffective. In addition, due to the high variability of presentation and patient characteristics, appropriate care for patients in chronic pain management requires personalized attitudes regarding pain intensity and duration, disease status, tolerability and tolerance of possible adverse events, and risk of drug abuse (50).

As pain is an individualized experience, it has a multifactorial etiology, and understanding the biological, social, physical, and psychological contexts is essential for the treatment to be satisfactory; it is not directly related to the intensity and nature of nociceptive stimuli and biomedical mechanisms but of the interaction of various psychological and social dimensions. Therefore, it is more accurately qualified and quantified when reported (30, 32). Nevertheless, universal standards still must be for preventing and treating chronic pain (51). Therefore, self-report instruments and standardized questionnaires to assess pain intensity, functional abilities, beliefs and expectations, and emotional distress have been developed in several languages and help with treatment planning (52).

3.2 Pain management

The therapeutic protocol for chronic pain is based on medication, lifestyle changes, and rehabilitation (53).

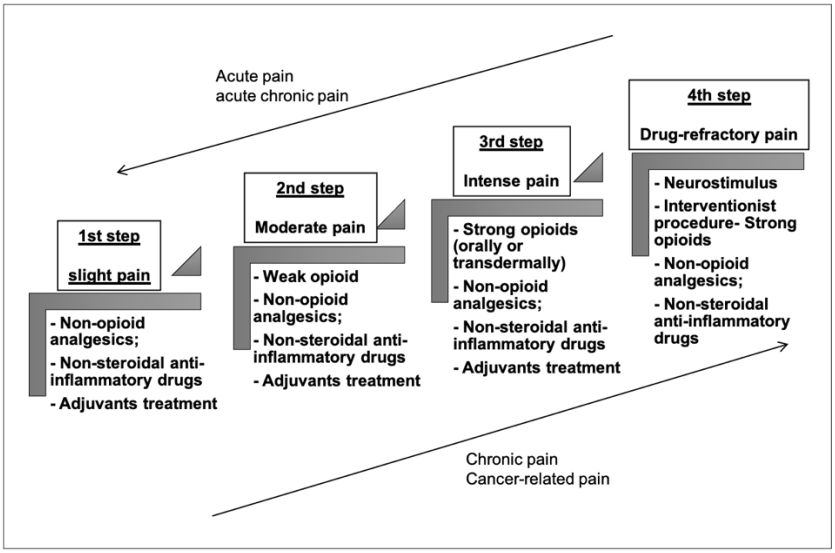
The treatment of chronic pain must be interdisciplinary, as it is a multidimensional experience, i.e., with physical, cognitive, psychological, and behavioral aspects that involve follow-up and clinical assistance, with pharmacological, physical, and psychological methods (32) and must have a multimodal treatment approach (54). The treatment is evaluated by the type of pain, the cause, and personal characteristics (53). It aims to relieve pain, minimize discomfort, and improve the quality of life, functions, and performance of carrying out activities. Therefore, all professionals in the multidisciplinary team must know epidemiology, anatomy, physiology, biochemistry, clinical psychological aspects, and pharmacological, rehabilitation, and reintegration interventions related to pain (3).

One of the main barriers to pain management is the lack of knowledge about diagnoses, therapies, rehabilitation, and reintegration of patients with pain (55). Inadequate management affects approximately 80% of the global population (56). Chronic pain is rarely cured; physicians and patients should focus on improving activities, functionality, and quality of life, even when it persists. Many therapeutic strategies incorporate trial and error in which medications and procedures are tested until the most effective ones are identified, at the expense of medical visits, reassessments, and frequent adjustments of intervention modalities and magnitudes (56, 57). Despite the publication of several guidelines and recommendations for treatment, up to 40% of patients with chronic pain remain symptomatic due to the heterogeneity of chronic pain mechanisms and individual variables that are not linearly related to the etiology of pain but to the interaction of its pathophysiology, individual variables, and social contexts. One of the recommendable precepts for treating chronic pain is the observation of clinical evaluations in sequential appointments (58).

The treatment of chronic pain is based on the World Health Organization (WHO) Analgesic Ladder (Figure 1), which was released in 1986 with the original proposal to treat cancer pain (12, 59, 60) and which was expanded to treat also patients with pain not caused by the oncological disease. According to the WHO Analgesic Ladder (11), non-steroidal anti-inflammatory drugs (NSAIDs) are recommended on the first rung to treat mild pain, weak opioids to treat moderate pain on the second, and potent opioids

to treat intense pain on the third step. Adjuvant drugs can be incorporated at all steps. The Analgesic Ladder is part of the WHO Program entitled “Pain in Cancer and Palliative Care” and has undergone several modifications over the years (61). Despite criticism, it is still a reference for controlling pain in cancer patients and non-oncological acute and chronic pain conditions. It was recommended to use the WHO Analgesic Ladder only in a unidirectional way; that is, if there is a need for more advanced interventions due to the worsening of the pain, the treatment included drugs from higher steps. Recently, the bidirectional application of the Analgesic Ladder has been proposed; that is, the treatment must adapt to the intensity of the pain in ascending steps. When improvement occurs, it must be used in the descending direction. A step was also added to the ladder. The fourth step includes non-pharmacological procedures, that is, interventional and minimally invasive, recommended to treat persistent pain, combined with other methods. Among these procedures, epidural analgesia, prolonged administration of analgesics and anesthetics with or without implantable pumps or not, and functional neurosurgical analgesic procedures (for example, lumbar percutaneous adhesiolysis, rhizotomy, cordotomy) stand out; electrical neuromodulation of the nervous system (for example, brain stimulation of the spinal cord or peripheral nervous system), anesthetic blocks of peripheral nerves with physical or chemical means, and among other procedures (12, 61).

Figure 1 - Adapted WHO Analgesic Ladder



Fonte: Anekar, Cascella (2022). WHO Analgesic Ladder is still a reference in pain management in oncological patients and is being extended for pain settings in acute and non-oncological diseases (11).

Non-biologically based treatments are essential to improve therapeutic outcomes in chronic pain. Among them are rehabilitation, physiotherapy, occupational therapy, nutrition, and psychological therapies, including cognitive behavioral therapy and acupuncture (62). Drug treatment uses analgesics, tricyclic or dual antidepressants, anticonvulsants, and weak or strong opioids (63). A balance must be found between effective treatment and acceptable side effects (54). Unfortunately, the pharmacological, interventional, behavioral, and surgical therapies used to treat chronic pain have limited effectiveness in reducing pain, functional recovery, and returning to work. In addition, there is a high occurrence of continuous dependence on opioid analgesics (64).

The critical challenge in pain management arises because pain cannot be measured directly. Although several methods for multidimensional pain assessments have been proposed, the gold standard of data collection is self-report (52). The current definition of pain by the IASP evidences verbal self-report compared to non-verbal assessments, as it makes it possible to register valuable information, even in animals and humans with compromised cognition or linguistic inability (5).

Self-reports and non-standard or low-precision questionnaires are the most used assessment tools in patients with pain and one of the most expressive causes of assessment errors compared to other areas of medicine; i.e., the importance of the appropriate collection is emphasized, of real-life data to develop strategies and improve the care of patients with pain using AI (14).

Improving the treatment and management of chronic pain and improving patients' quality of life also reduce social costs (27). A comprehensive assessment of each patient's specific biological, psychosocial, and psychobehavioral etiologies will likely improve the quality of care. The first consultation is a special moment in the patient's clinical history, as it can improve adherence to treatment and engagement with long-term care (65). Treatments with virtual tools have become fundamental in reaching patients regardless of location (53).

3.3 Health records

The health record, also called medical records, have common characteristics that must be adopted in accordance with the Resolution of the Federal Council of Medicine.

The medical record is a single document about the patient's health and the care provided to him; it has a legal, confidential, and scientific character (66).

The health record is essential in clinical practice, as health professionals use the clinical and evolutionary history of the patient during their follow-up to make clinical and managerial decisions to support research and professional training; therefore, the quality of records is a portrait of the quality of care provided (67).

All records must be completed objectively, allowing for their reading, identification, and location. However, some studies point to the low quality of the health record in terms of illegibility, inconsistency of records, and lack of basic information on patient follow-up (68-70).

The international literature advocates structuring the health record electronically, with standardization of patient assessment methods (71, 72) to improve the quality of care. However, in Brazil, the digitization of medical records is a late novelty; there are still specific medical records for each health unit, which are often on paper or digitalized, which makes it difficult to update and share patient information and places a logistical burden on health systems (73).

Some studies demonstrate that even a simple platform, but one that meets clinical and documental needs related to medical evaluation, can be effective for recording patient treatment. Wilsey et al. demonstrated the importance of using a simple documentation system developed with the Microsoft Access program to help prescribe opioids and their use in 1,400 patients (61).

Data availability and quality are essential to establish treatment, rehabilitation, and reintegration strategies for patients with chronic pain. In the 1970s and 1980s, few hospitals recorded structured data in computer systems. Hospitals that collected standardized data used their nomenclatures and definitions, a methodology that made attempts to algorithmically model pain difficult (51). Over time, vast amounts of data became available, and advances were made in the standardization of medical nomenclatures (74). Based on data from standardized records, it is possible to analyze this information and direct it toward adequate decision-making in management and care (75, 76).

The use of IT in health began to be incorporated into electronic records at various stages of patient care, from initial care to diagnostic configuration, decision-making, and patient follow-up throughout treatment. Using standardized systems to assess

care can improve care providers' quality and efficiency and reduce interventions' costs (74).

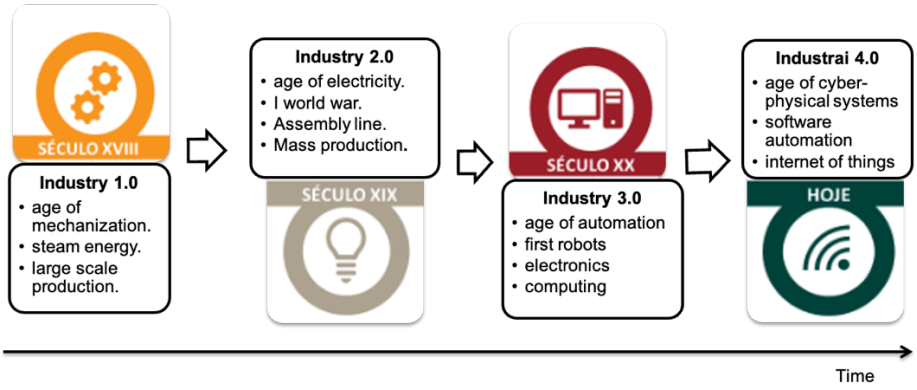
With the advancement of science, there is a demand for new methods of analyzing complex and unstructured data, that is, the analysis of large databases (big data). In the big data revolution, precision medicine and electronic medical records have great prominence, despite being at the beginning, as there is a large amount of quality data and the need for sampling as a tool to work with this variety of data and, thus, reduce the bias (77, 78). Scientific knowledge from current research points to results based on large averages. Personalized medicine or precision medicine aims to improve these results by customizing the treatment according to the biological characteristics of the patients. However, due to the multicausality of the diseases, it takes work to achieve 100% accuracy (79). Therefore, using standardized electronic medical records and digitizing all patient data by health services to encourage these new big data analyses is critical (73, 80).

New techniques to ensure data confidentiality using encryption techniques will be increasingly incorporated into scientific research (81).

3.4 Artificial Intelligence (AI)

For many years scientists have been studying the possibilities of making computers learn similarly to human beings. The first industrial revolutions presented alternatives to automate processes to supply human labor efficiently and at a lower cost using machines. The Industrial Revolution 4.0 enabled processes and machines to communicate using computers and networks dedicated to these operations (Figure 2). All industrial revolutions focused on the independence of machines in production processes. The data analysis area followed this process by developing algorithms that respond autonomously and adapt to the data automatically without human intervention (73, 82-85).

Figure 2 – Stages of Industrial Revolution



Source: Sakurai, Zuchi (2018) Graphic representation of the Industrial Revolution timeline depicting the historic milestones to the present day (82).

In 1950 the British mathematician Alan Turing, considered the “father of computational science and AI”, published a book entitled “Computing Machinery and Intelligence” with the subtitle “Can machines think?”, the objective was to show if computers were capable of learning and convince a human being that they were talking to someone else and not a machine; this experiment was called the “imitation game” and later the “Turing test” (86).

In 1956, the term AI was introduced during a Dartmouth College conference by John McCarthy et al., who defined it as “the science and engineering of making intelligent machines”. The conference spawned this new area of interdisciplinary research that provided the intellectual framework for subsequent computer research and development efforts (25). As it is a new area, it was also discussed which field the AI area was related to, whether computer science, engineering, or biomedicine. AI is a subfield of Computer Science and is a term used for the ability of a computer to model intelligent behavior with minimal human intervention (87).

In the 1980s, a social revolution was expected due to increased access to computers and AI, with health areas being the main focus (88). However, some issues limited the growth in the use of computers, such as storing large amounts of data in the computer's main memory, which justified the introduction of the term "big data", which describes the set of large numbers of data that, despite the challenge of analyzing a large volume of information would make it possible to develop new knowledge, which would not be possible with usual data analysis (89).

The most extensive relation between statistics and data science is forecasting and regression. Basic statistics focuses on performing inference; it uses a small

sample to draw conclusions about large populations. There has been growing development in the discipline of statistics in recent years related to data modeling. Exploratory data analysis (EDA) emerged in 1977 with John W. Turkey through the book Exploratory Data Analysis. It allowed quantitative analysis, greater access to extensive data, and the development of new technologies (77).

With the increase in data and its complexity, the choice of traditional statistical methodology to analyze data presents limitations. Traditional statistics aims to find and interpret the relationship between data (confidence intervals and results). Compared to typical statistical analysis, AI works with high-dimensional data and many structures that cannot be represented in a simple statistical model, such as a linear regression graph. Furthermore, running conventional analyzes needs many assumptions about data to work correctly; the larger the volume of data, the harder it is to find patterns. AI methods, on the other hand, require considerably fewer assumptions about the data. One of the few assumptions is that the random variables are independent and identically distributed. ML methods are based on high accuracy and prediction but less on interpreting the generated model (90). In ML techniques, statistical methods assess relations between data characteristics and the proposed result. However, the use of the MA technique in a real clinical environment is limited due to the difficulty in interpreting its predictions, which limits the optimized use of methods, such as deep learning and models, to support medical decisions (26).

In April 2021, GM Ordinance No. 4,617 was approved in Brazil, which instituted the Brazilian Strategy for Artificial Intelligence and its thematic axes to "guide the actions of the State in favor of fostering research, development, and innovation in artificial intelligence and guaranteeing innovation in the productive and social environment of the area" (91).

The data comes from several sources: text, video, questionnaires, etc. The biggest challenge in data science is to work with raw data, that is, unstructured data, such as free texts. Unstructured data must be transformed into structured data, i.e.is, numerical or categorical data, being processed and manipulated. Converting numerical data into categorical data is often used in data analysis. The use of AI relies on deploying algorithmic software routines designed to analyze data and built upon large pre-filtered structured and unstructured datasets (88). Data is sorted by type to determine which programming language will best process it, for example, R or Python (77).

Before starting any analysis, it is necessary to standardize the programming language. Many languages and software are available for data analysis, for example, Stata, SPSS, R, Python, etc. The most commonly used languages are R, which is based on data analysis, and Python, a general programming language both have open and free libraries. Python is easily assimilated and very close to “natural language” and is considered the best language for programmers in other areas of expertise. It is commonly used due to its versatility and scope of applicability (92).

AI is divided, according to the Association for the Advancement of Artificial Intelligence (AAAI), into the following sub-areas: AI Applications; Data Mining and Big Data; ML; Natural Language Processing; Automated Planning; Reasoning and Probabilistic Reasoning; Knowledge Representation; Robotics and Perception; Agent-Based and Multiple-Agent Systems and Research (or Pathfinding) (92).

AI is classified into virtual and physical. The physical part is a machine's hardware, and the virtual part is the software. The virtual component is divided into several branches, including MA, which uses mathematical algorithms to learn through experience based on historical data (89, 90). It differs from classical statistical methods because they are data-driven and do not impose a linear structure to the data (77).

The term Machine Learning was first used in 1959 by Arthur Scherbius to describe algorithms that give "computers the ability to learn without being programmed". One of the basic requirements for any intelligent behavior is learning (93). Learning is the ability to adapt, modify and improve behaviors and responses. The term “learning”, used in ML analyses, can also be described as “training”, “building”, “formulating,” or “inducing” (89). In ML, systems learn patterns from the analysis of millions of data. Develops self-learning and self-improvement algorithms. The evolution of ML techniques aimed at increasing the storage capacity of current computers. It is defined as using algorithms that aim to extract information from data and represent them through a mathematical model used to make predictions or inferences by entering new data sets. It aims to discover patterns that explain the relations of the data among themselves, which would be impossible to be carried out by human beings (88). When making an estimate (test or prediction) for a data set of unknown values, it is referred to that the model is being “applied” (21). The focus of this technology is to be able to make decisions in isolation or with minimal human intervention (88).

ML may seem complex, but it combines simple ideas to produce highly accurate models that can “learn” from previous data. The ML application process can be divided into a few steps: defining the study problem; knowing the data that will be used; preparing these data in the pre-processing, transformation, and selection of variables; choosing the best algorithm; stipulating the method that will evaluate the performance of the generated models; finding better results; establishing the method for presenting the data (26).

Commonly used types of ML algorithms are (88):

- Unsupervised learning: These are classification and prediction algorithms based on previous examples. There is no label information; the estimated desired outputs are not informed, so the data are “unlabeled”. During training, the algorithm does not receive the expected outcomes and possibly discovers their relations through data exploration. The learning process, in this case, intends to identify similar data and group or organize them.
- Supervised learning: It is the ability to find patterns. The data is labeled; it is known which output (outcome) is expected for each data input. The model is built from the input data (or dataset) in ordered pairs (input and desired output). Learning (or training) consists of presenting the algorithm with sufficient examples (records or instances) of desired inputs and outputs. The algorithm learns the general rule, correctly maps the inputs and outputs, and generates the final model.

An ML model is a set of variables (or resources), where each variable has a “weight” that the machine learns through data inclusion. The final result, or equation, is generated from an ML algorithm. The development of this equation determines the model's performance, and its success depends on the algorithm used. Several different algorithms can be tested with the test data, varying the parameters of these algorithms in search of what performs best by the data scientist. To measure the result of the model, you can use cross-validation tests, separate the data into a training and test set, and others. For the final equation, the model relies on input data feeds to define the response of the output values. Interpretable models are simpler models, such as tree and linear regression models, in which, due to their simplicity, little effort is required to elaborate an explanation from the learned model (27).

To evaluate model performance, i.e., which model produces the most accurate predictions, the Cross Validation (CV) technique, consists of dividing the data into sets. One set is used for training, and the other for testing and evaluating the model's performance, basically alternating between training and testing and vice versa. Cross-validation with five folds (or parts) is built into the XGBoost algorithm; it is unnecessary to program and specify the number of iterations required in the run. In this method, the available models are randomly partitioned into equal subsets of size N, obtaining the mean of N different estimates. Each estimate is obtained by keeping one of the N subsets for validation and the remaining N - 1 subsets for training. In the first iteration, the data were divided into five parts, four parts for training and 1 part for testing. In the second iteration, 1 part, different from the previous one, is used for testing while the rest is for training. This process is repeated five times so that every database goes through training and testing (94).

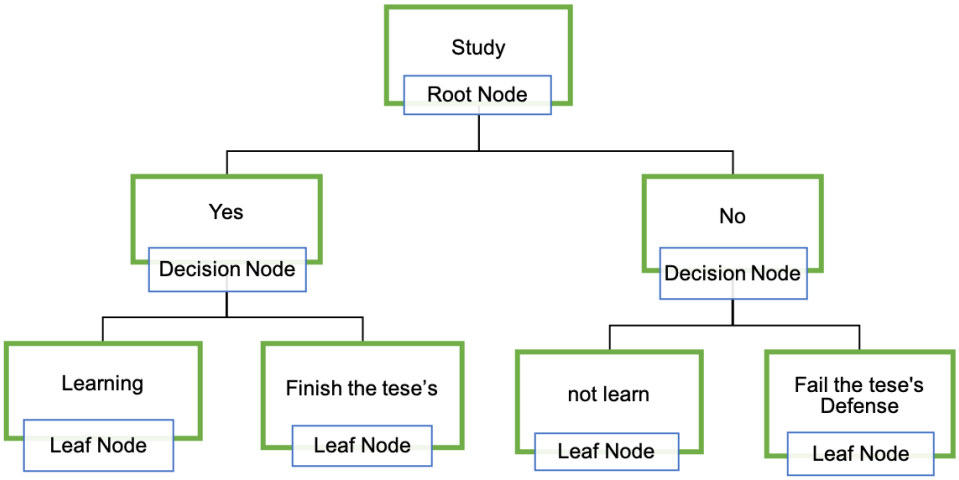
Initially, before training the model, the data are randomly separated on a training and test basis, which means learning the weights to minimize the prediction error. In the training and testing technique, during the development of the MA application, a database, known as the training base, that is, the data are “taught” as training datasets, will be processed by the chosen algorithm, which will apply several rules, calculations, parameters, and comparisons, in search of a final equation that best fits the parameters used in this training. Algorithms learn from unprogrammed data and can be used to gain insights and predict decisions from large and complex data sets. Creating reliable ML models requires large volumes of training data. Therefore, you can find different models that perform similarly in training. These analyses can help better predict, identify, and treat diseases in healthcare. After the model was trained, the test base was used to evaluate performance with data input different from those already used. Test data, which is data, was not used in the development of the algorithm. In this step, the algorithms use parameters based on training data to confirm their performance and identify the one that presents the best result with new input data. Next, it is necessary to define the metric to evaluate the model, for example, the ROC curve, the area under the curve (AUC), and the confusion matrix. After these steps, the final model will be defined, and it can be used or not, in predicting new values, depending on its performance (27, 58).

There are several algorithms for ML analysis, the most used are:

a. Explanatory Algorithms (95) (Linear Regression, Logistic Regression, SHAP, LIME): Creates explanatory models to understand the relations of model variables and identifies variables that have a statistical relations with the result.
b. Clustering algorithms (95) (K-Means, Hierarchical Clustering): are used to perform cluster analysis.
c. Dimensionality Reduction Algorithms (95) (PCA, LDA): reduces the number of input variables.
d. Similarity Algorithm (77, 95) (KNN, Euclidean Distance, Cosine, Levenshtein, Jaro-Winkler, SVD, etc.): compare the similarity of the data.
e. K-Nearest Neighbors (KNN): a method for classification and regression. Classify a record by assigning it to the class that has its counterpart.
f. Ensemble Learning Algorithm (77, 96) (Random Forest, XGBoost, LightGBM, CatBoost): it is based on combining several simple prediction models (weak learner), training them for the same “task” and producing a grouped model more complex (strong learner), from the sum of the parts of the simpler models, aiming to reduce the biases and make them more robust, reducing the individual disadvantages of the final model. For a prediction of a single model, many algorithms can be used. The most used ones are:
    - Random Forest (97), which is a machine learning algorithm based on “decision trees” (or classification). This algorithm allows combining numerous decision trees, which were trained separately, in a single model, which is their set. Individually, the predictions made by decision trees may not be accurate, but when combined and analyzed together, they allow for more robust and accurate overall predictions and the algorithm
    - Extreme Gradient Boosting (XGBoost) (96) is also based on decision trees. XGBoost is based on training several simple models to generate, in the end, a robust model. It combines software and hardware optimization techniques to produce superior results and, in this way, uses fewer computing resources in a reduced time. The technique is superior and complementary to Random Forest, as it optimizes the available space on the disk while the big data analysis takes place. This algorithm already has the cross-validation method integrated into each interaction, thus optimizing the need to program the analyzes and specify the exact number of necessary reinforcement interactions in a single execution.

The Tree Model is a classification and regression method. Currently, decision tree-based algorithms are considered best in class. Decision trees (Figure 3) are question-and-answer flowcharts that graphically represent the decision-making process. They are techniques for learning relations rules between variables. The process starts with several questions with an initial estimate based on knowledge about the analyzed problem and refines them as more information is obtained; the process ends with the analysis of results and decision-making. The decision tree lists all possible alternatives for each question in the “True” (value 1) and “False” (value 0) modes. Decision trees are nodes, edges, and terminal nodes generated through descending processes. First, nodes are split based on the underlying distribution of a given input variable, resulting in multiple subnodes. Then, this process is repeated, resulting in a predicted output. Simple tree models are easily interpretable (78, 95).

Figure 3 – Illustration of a decision tree.

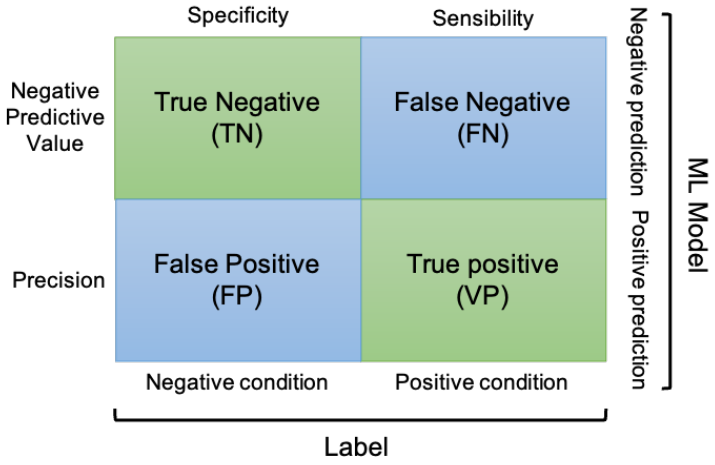


Source: author themselves

The metrics to evaluate the generated models can be carried out in several ways. The results of the models can be represented, for example, through a confusion matrix (Figure 4) or by the ROC curve and AUC, among the most used metrics in MA. The confusion matrix details model performance and illustrates the number of correct and incorrect classifications categorized by response types by the model. It is used to calculate “False Positive” values, i.e., when the expected result is negative, but the model assumes it to be positive; “False Negatives”, when the expected result is

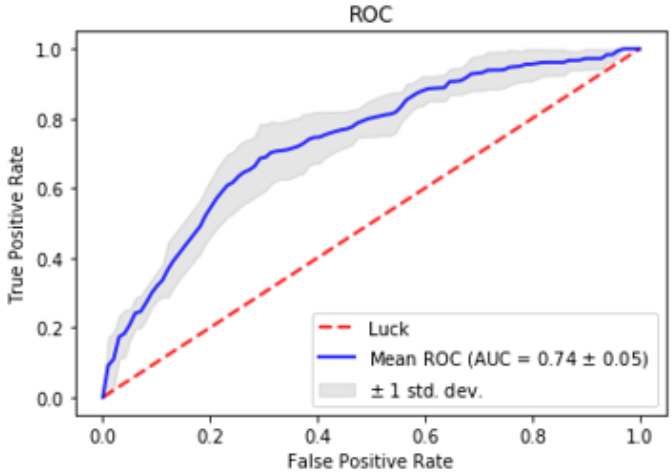
positive, but the model results in negative; “True Positives”, when the expected result is positive and the model is positive; and “True Negatives”, when the expected result is negative, and the model results in negative (77). The AUC curve is derived from the ROC curve. The Area Under the Curve (AUC) is the total area under the Receiver Operating Characteristic (ROC) curve. It evaluates each model and demonstrates how well it can distinguish between positive and negative. The AUC value is associated with greater efficiency of the classifier with a lower false positive rate; the higher the AUC, the better (Figure 5) (78, 80).

Figure 4 – Example of a confusion matrix for a binary response.



Source: author themselves

Figure 5 – Exemplification of the ROC and AUC curve



Source: author themselves

To explain the output of the ML models, the Shapley Additive Planation (SHAP) technique is used, which is used in game theory to determine how much each player contributes to the gain in a collaborative game. This method assigns an average contribution value among all possible characteristic combinations. SHAP is a library that uses Shapley Values; that is, it assigns importance to each variable to justify the result of any machine learning model. SHAP values consider that the result of each combination of factors should be used to rank the importance of a single factor, having as input an instance and a trained model that assigns a contribution to each variable; the sum of all contributions corresponds to the output predicted by the model (98-100).

SHAP is more challenging to interpret than a linear or logistic regression model, but it is a visual, intuitive, and simple form of representation (99, 100). In some cases, even if models contribute to each feature, the interpretability is limited; the features must be interpretable. Therefore, SHAP can be described as a method based on assigning a value to each resource in the cooperation based on its contribution to the model's decisions. There are other feature assignment methods, but SHAP is the only one with the three desirable properties: 1- The model is explained truthfully, and it has precision; 2 - Missing features do not impact model decisions; 3 - If a model changes so that the contribution of some resource increases or remains the same, the allocation of that resource must not decrease, regardless of the other resources, maintain consistency (101).

Some SHAP graphics options are used as explanatory methods for models, such as: “KernelSHAP”, which is an agnostic explainer; “LinearSHAP”, for linear models; “DeepSHAP”, for deep models; and “TreeSHAP” more geared towards tree-based models (85). The most widely used is TreeSHAP, which only authorizes allowed paths within the “tree”, which means that it does not include unrealistic combinations of features, as in other methods based on permutation (98).

The characteristics of a SHAP chart can be evaluated as follows (Figure 6):

- The X axis is the SHAP values. Positive values contribute to the model responding positively to the desired outcome; Negative values represent that the model responds negatively to the expected outcome.
- The Y axis is the model's variables (or features). They are arranged in order of importance, from most important to least.
- Each colored dot represents a sample (or dataset); that is, each dot is a sample, a patient. The further to the right, the more positive the contribution of the

variable to that sample. The more comprehensive the range of SHAP values, the better the variable for the model. Points around the value 0 have weak contributions.

- The colors represent the increase or decrease of the variable value. The red color is high values, and the blue color is lower. Variables with a transparent color split, i.e., red and blue from opposite sides, demonstrate good prediction as you can visualize their contribution to the model.

Figure 6 – Chart representation SHAP Summary plot.

Source: author themselves

Shapley Additive Explanation (SHAP) framework is an ML machine explainability technique (99, 102). The base value is the standard output predicted by the model if there is no variable information, and the model output value represents the prediction considering the provided input. A contribution in this scale is associated with each variable, with a length and a color. The length indicates the impact of this variable on the prediction, and the color indicates whether the impact is positive or negative. For example, red indicates a positive impact on the prediction, while blue indicates a negative impact. It is considered a classification threshold; if the prediction output value is more significant, it will be represented by the red color and positively contribute to the model. As a value lower than the threshold is considered a negative prediction when evaluating the same model with the base value, variables with a lower value will have a negative impact and are represented in blue.

The ability to interpret a single prediction is useful, but the model learns relations that are often complex. For example, variables with the same value can have positive and negative impacts for different instances (or even no impact). The summary plot in the SHAP framework was used to generate an overview of the importance of each variable and its impact.

The models differ significantly in terms of their explanatory factors. These separate models are not good predictors, but together (average), they can predict better than any individual model. Each generated SHAP chart represents the individual description of each model and how the variables impact this model. For each model, a SHAP was generated; for example, if the algorithm generates ten models, 10 SHAPs will be generated. The algorithm's number of models must be fixed and known in advance. However, this difference does not impact performance since the configuration found performs best.

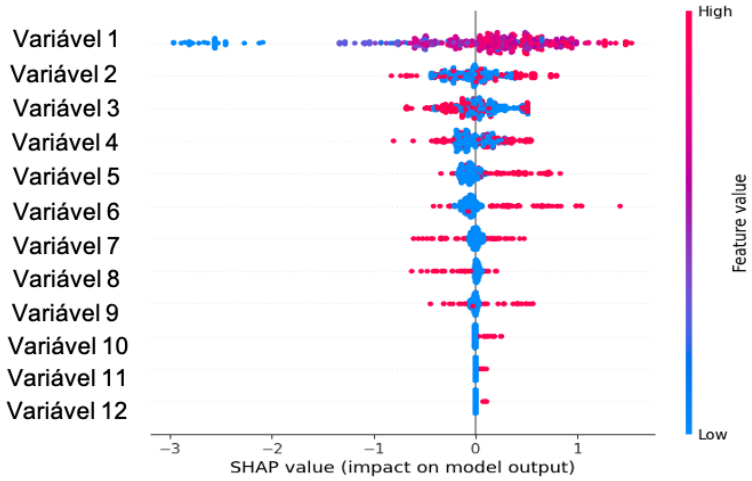
SHAP charts make it possible to understand the “role” of each variable (feature). It can be interpreted according to length and color. Color has a positive (red color) or negative (blue color) impact on pain improvement; that is, the blue color meant that the patient generally did not improve; variables represented with red meant that the patient probably improved. The length of the variable had an impact on the prediction; i.e., the longer the length, the greater the impact on the model.

Figure 7 is the representation of the model presented by the summary plot. The summary plot aims to present an overview of the contribution of the variables. In this graph on the left, there is a list of variables that are ordered from the highest to the lowest impact on the model's decisions; that is, the variables “1”, “2,” and “3” are the ones that most impact the model's prediction. In the center of the graph (value 0) is a set of points, each representing an input instance (dataset) of the model. On the “x” axis of the image, it is possible to observe a scale from -3 to approximately 2. The “x” coordinate of the point quantitatively measures the impact (as can be seen on the scale), while the “y” coordinate indicates the variable to which that point refers. The red and blue colors mean that, in that instance, the variable's value is high (red) or low (blue). The interpretation of the meaning of color in the models must be individual for each variable; for example, for a medicine, the color blue (0) indicates that the patient did not use it, while the color red (1) indicates that the patient used the medicine; for a disease, for example, the variable “6”, blue (0) indicates that the patient did not have the disease and red (1) indicates that the patient had it. They were considering the

variable “10”. Among the 12 variables that comprise the model, the “10” is the 10th in importance; it can assume two values, 0, which indicates that the patient does not use this medication, and 1, which indicates that they do. As can be seen, there is a large number of points in blue positioned on the 0 axes in the graph; this indicates, based on what was learned by the model, that the information that the patient does not use this medication has no impact on the prediction for the patient's improvement.

On the other hand, the information that the patient uses this medication (red, indicating the value 1) has a small positive impact on the prediction (points positioned on the right side of the value 0 of the scale). Other variables present a more complex behavior. For example, the variable “7”, with value 1, can negatively and positively impact the prediction. It is also possible to observe that the positive or negative impacts are quite similar in terms of the impact and the number of impacted instances. Unlike variable “6”, which, although the indication of the occurrence of the disease by the patient (value 1, high) can have a positive and negative impact, it is possible to observe that the impact is most often positive in the prediction. In addition, the dimension of the impact is more significant. On the other hand, the absence of disease has a small and mostly negative impact on the prediction.

Figure 7 – Presentation of the overview of the behavior of the variables in a model with the graph.



Source: author themselves. Explanation factors (as seen as SHAP summary graphs) associated to prototypical models. Model built with XGBoost.

3.5 The use of Artificial Intelligence in pain

Scientific literature shows that the use of AI can contribute to improving the performance and quality of health care (79). In medicine, it makes it possible to propose solutions to clinical problems based on the analysis of large volumes of data thanks to the use of computers and algorithms defined by specialists in the field (24). It provides several clinical applications, including neuronavigation, image processing, three-dimensional modeling and printing, prosthetic fabrication, stereotactic radiosurgery, and clinical trial management (89). Patient data can be collected using electronic medical records by entering information on anamnesis, clinical and complementary exams, disease and patient evolution, use of prescribed medications, and use of defined algorithms that can be updated. The analysis of these data and proposal of differential diagnoses, with the respective probabilities of occurrence, support the decision as a way to reduce the possibility of mistakes by the team. Studies indicate that data from the anamnesis, the quality of the physical examination, and the volume of complementary tests requested by physicians to solve cases vary significantly, emphasizing the importance of professional experience in solving possible diagnoses. Electronic medical or health records are essential tools for personalized medicine, early detection, and targeted prevention to increase clinical value and reduce healthcare costs (88).

AI plays a significant role in healthcare, and to keep up with this progress, it is essential that professionals can evaluate AI-based studies (103).

The use of AI in healthcare to assist professionals in clinical diagnostic decisions and therapy management is increasing (104). For example, works in the scientific literature use ML to predict the evolution of chronic pain with algorithms to calculate dynamic changes in chronic pain risk scores based on various aspects of health behavior. However, they usually use only three sources of information: depression, nutrition, and physical activities (51).

AI methods or ML algorithms have been used to obtain physiological information from individuals with chronic pain, including respiratory rate, blood oxygen concentrations, heart rate, body temperature, and blood pressure (53).

Some AI studies evaluate the improvement of chronic pain treatment via the prediction of therapeutic responses with trained, supervised learning algorithms instead of using traditional statistics that divide individuals into binary response categories. Studies usually classify patients as “responders” and “non-responders,” and two “arms”, active and placebo, are used. However, one cannot legitimately label

an individual treated with a given active treatment as a 'responder' (or not) because it is unknown what would have happened had they been allocated to the comparator arm (53).

AI can help fine-tune the evaluation of patients and minimize the evaluator's biases. It makes it possible to assess patients' risk, analyze the disease (for example, decoding ECG and imaging findings), select the best therapy based on the patient's clinical history and clinical trial results, track diseases, and detect early signs of worsening alerts. Currently, several therapies are used to control chronic pain, based on drug treatments, mind-body therapies, cognitive behavioral therapy, rehabilitation methods, digital therapies, telerehabilitation, chatbots, remote patient monitoring, AI, and immersive medicine technology (51).

A meaningful follow-up interpretation of the patient's journey can be provided to assess national guidelines and use ML to chart treatment adherence, making problem-solving and decision-making more informed and improving communication between patients and professionals of health. Ultimately, it makes it possible to create the best possible scenarios to improve the management and self-management of chronic pain. Therefore, in health sectors, the interpretability of an ML model is a desired feature, which should help the team make decisions (13).

METHODS

4 METHODS

An open observational study was carried out with an exploratory nature on pain in chronic patients. Information on pain screening, pain assessment, medications used, individual and hereditary antecedents, life habits, impressions of change in relation to pain, rehabilitation methods performed, and other interventions the patient had access to were accessed.

The existence of response predictors of pain improvement or not was evaluated with the analysis of standardized data extracted from the information of the first consultation. After the analysis, the main response predictors were identified.

4.1 Study location and period

Data collection took place from March 2017 to March 2018 at the Ambulatory of the Pain Center of the Neurological Clinic (PCNC) of the Hospital das Clínicas of the Faculty of Medicine of the University of São Paulo (HCFMUSP), using the electronic platform entitled “Interface Eletrônica de Atendimento Ambulatorial” (IEAA).

4.2 Ethical and legal aspects

The project was submitted to the Research Ethics Committee of HCFMUSP and approved (CAAE nº 10854912.2.0000.0068 opinion issued on 20/12/2013). Furthermore, the “Electronic Interface for Ambulatory Care” system was patented via the University of São Paulo (BR1020140115030 – deposit made on 13/05/2014) and has been routinely used for over ten years in patient care at the Neurosurgery Division of HCFMUSP/SP.

4.3 Study population

Data from patients treated at chronic pain outpatient clinics for one year were collected based on information from medical consultations recorded on the “Electronic Interface” service platform, which consists of a standardized script for consultations and records the care information.

4.3.1 The inclusion criteria were:

1. Age equal to or greater than 18 years.
2. Patients who attended at least one follow-up consultation at the PCNC of HCFMUSP in the 12 months following the first consultation.
3. Patients with data entered into the Electronic Interface during outpatient care.
4. Patients with information in the mandatory questionnaires were completed entirely.

4.3.2 The exclusion criteria were:

1. Age less than 18 years old.
2. Inability to answer the questions during the consultation performed by the pain group.
3. Patients who did not attend more than one consultation within the 12-month follow-up period.
4. Patients who did not respond to the mandatory questionnaires or who responded with missing information.

4.4 Data collection

The evaluation and follow-up of the patients were carried out using the tool “Electronic Interface for Ambulatory Care” (IEAA), developed in the Discipline of Neurosurgery of the Department of Neurology at FMUSP, to constitute a standardized online service platform, providing an essential path of clinical information about patients with pain, allowing for the addition of new scales, clinical inventories, and specific research according to demand. This tool enables a uniform and systematic assessment of patients and provides medical and legal documentation of the clinical and sociodemographic conditions of patients for the multidisciplinary team. The use of standardized support tools during care can be considered costly, but their application during care can induce physicians to avoid “undertreatment” and “undercare” of the patient.

The IEAA offers the care team standardized follow-up to evaluate patients. It is composed of validated, detailed and self-administered instruments. It also makes it possible to create personalized doctor-patient treatment agreements, provides consistent, detailed, and available documentation for the medical record at any time, even outside the scheduled consultation day or at another institute, and aggregates information and data for analysis. The IEAA makes it possible to monitor patients regardless of the institution where they are treated. It was formalized as a care instrument and received support from the Research Program for the SUS (PPSUS) during 2016-2018.

In the present project, the data computerization tool was used to track the degree of pain improvement or worsening and markers of good and bad prognosis. Analogously, the information can be used to readjust the prescription of medications or interventions with higher costs and, potentially, less clinical effectiveness.

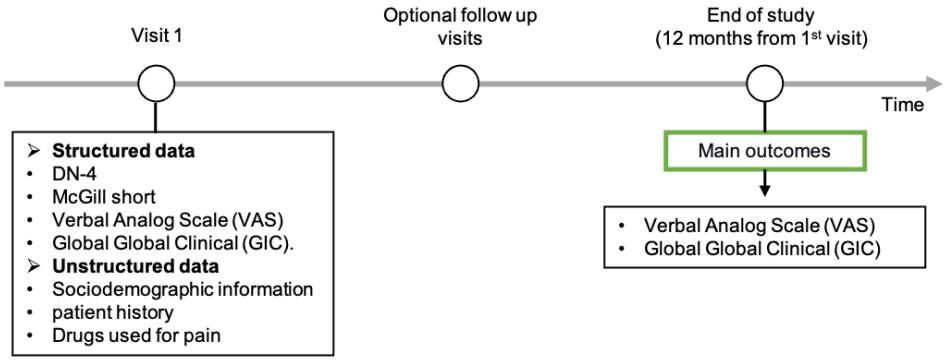
The IEAA digital mask provides an extensive database with patient information throughout their care history. Data registered at the IEAA are anonymized in several layers; the patients are characterized according to code numbers and can only be identified through physical access to a file protected by the institution.

4.4.1 Clinical evaluation and treatment of the patient with pain

The assessment process consisted of self-reports and patient-directed reports during the first visit. From the collected data, a big data study was carried out.

Figure 8 represents the timeline of the evaluations carried out in each service.

Figure 8 – Design of the study.



The outcomes are the improvement of 30% of pain according to the Verbal Analog Scale (VAS); improvement of the pain according to the Global Impression Change Scale (GIC).

4.4.1.2 Assessment instruments

Clinical assessment included patients' pain characteristics, symptoms, functional status, and clinical history. The evaluations were carried out with tools that aimed to locate and quantify the subjective experience of pain.

The patient's clinical history, personal and family history, life habits, pain characteristics, intensity, duration, mode of onset, duration of disease and pain, temporal relations, location, the associated somatic symptoms, the drugs used, and the respective follow-up dates.

4.4.1.3 Unstructured data Sociodemographic assessment:

1. The sociodemographic assessment carried out was based on gender, age, location of residence, skin color, and education.

2. Medications used: The medication treatments for pain relief that the patient had access to at the hospital pharmacy were recorded.

4.4.1.4 Structured data

1. Background assessment: A questionnaire was used to assess the associated morbidities, present or past, life habits, and hereditary history of the patients. (Appendix 1)

2. Interrogation about the different systems: The different systems were evaluated (cephalic segment, neck, respiratory system, circulatory system, digestive system, locomotor system, nervous system, genitourinary system, and endocrine system). (Appendix 2)

3. Assessment of pain: The assessment is divided into nine axes: Installation of pain; Duration of pain; duration of illness; Verbal pain intensity; Plan of pain location; Character of the pain; Frequency of pain episodes; Temporal predominance of pain; Pain evolution. (Appendix 3)
4. McGill Pain Questionnaire (105, 106) – short version: It is a multidimensional scale to characterize pain. It uses quantitative measures that make it possible to assess the essential qualities of the painful condition. It consists of 15 descriptors, grouped into three groups and classified as absent or present: Sensitive-discriminative dimension of pain; Affective-motivational dimension of pain, Cognitive-evaluative dimension of pain. It also makes it possible to record the location of the pain. A figure with the human body parts divided and numbered from 1 to 53 is presented to the patient. (Appendix 4)
5. Verbal Analog Scale (VAS) (107, 108): scale used to assess pain intensity. Scores range from 0 to 10, with 0 representing no pain, five moderate pain, and 10 “worst possible pain”. (Appendix 5)
6. Douleur Neuropathique Pain 4 Questions – DN4 (45, 109): scale to assess the occurrence or not of neuropathic pain, incorporating both symptoms and signs on physical examination. One point is awarded for each positive item and 0 for each negative item. The total score consisted of the sum of 10 items, and the diagnosis of neuropathic pain was established when the sum of “positive” values was more significant than or equal to 4. (Appendix 6)
7. Global Impression Clinical (GIC) (110, 111): this is a self-evaluation of the patient and an evaluation of the physician on the perception of the global evolution of the disease about the treatment used. (Appendix 7)

4.5 Application of Machine Learning

4.5.1 Knowing the data

After collecting the data, they were arranged in an Excel spreadsheet and later in SPSS to evaluate and understand each variable and its relationship with the others. Based on the analysis of the raw data, three parameters were selected as an outcome to show the reduction or not of pain:

1. 30% reduction in pain intensity according to the Analog Verbal Scale (VAS); the difference between pain intensities reported in the first and last consultation was calculated.
2. 50% reduction in pain intensity, according to the Analog Verbal Scale (VAS). The difference between the reported pain intensities between the first and the last consultation was calculated.
3. Global Impressions Clinical (GIC) according to the degree of improvement or worsening of the pain sensation according to the patient and the evaluator. An improvement was considered when the value was at least two on the final visit.

4.5.2 Preparing the data

The raw data were processed to avoid errors in the dataset. First, however, the data underwent a selection and adequacy that defined which ones would be excluded or kept.

After this stage, the data was pre-processed, where the missing data were formatted, organized, and treated. Data were simplified to “Boolean” variables (true or false), with some exceptions, such as age. This processing facilitated the development of the algorithm and was chosen because it worked best with discrete and binary values. For example, the variable 'verbal intensity' admitted the values 'strong', 'moderate', and 'weak' as an answer. In the pre-processing, three variables were created: 'strong verbal intensity', 'moderate verbal intensity', and 'weak verbal intensity'. The values 1 for “true” and 0 for “false” were assigned to each variable. Data pre-processing has a positive impact on a model generation.

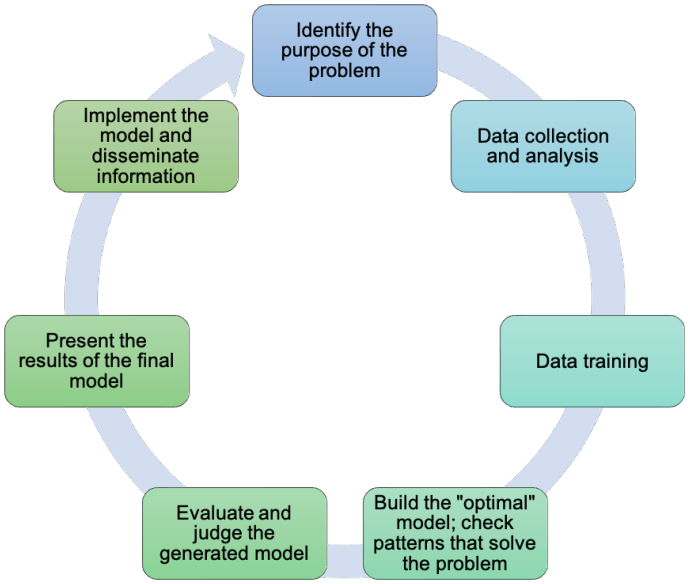
The last phase of this step was the selection of variables (or features), where it was evaluated, which variables would be most important for the study problem, which is to map the patient information collected in their first consultation to the correct outcome and select the outcome (or label); outliers were identified; and the variables in the database were organized to be presented to the algorithm. Finally, data were labeled, i.e., data pairs (patient information - result). Labeled data is used for training prediction models. The algorithm carried out in this phase was developed and used to evaluate the problem proposed in this project.

In this study, the outcome was pain improvement; as output values, the labels VAS < 30%, VAS < 50%, and GIC “better” or “much better” were used according to the perception of the doctor and the patient.

4.5.3 Machine Learning method

After the selection and pre-processing stages, the stages in which the data were treated, the testing phase began, and the technique of choice for analyzing these data was with the ML algorithms (Figure 9). The same tests were repeated in some algorithms using the cross-validation technique, with training and test data variations.

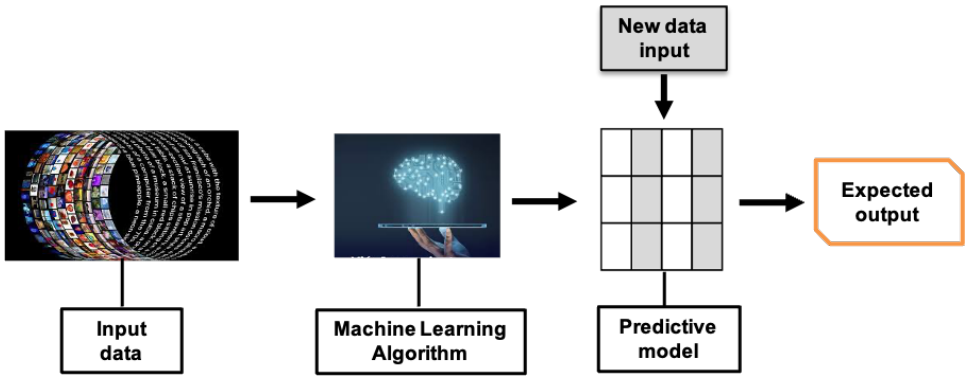
Figure 9 – Flowchart of a project with Machine Learning.



Source: Author themselves. Illustrating the planning step by step to use a machine learning technique.

Ultimately, the techniques were supervised ML (Figure 10) with Ensemble Learning methods. The algorithms used for data analysis were XGBoost and Random Forests, two well-known state-of-the-art algorithms for tabulated data, which are also robust in high-dimensional problems. The existing implementations in the Scikit-learn (https://scikit-learn.org/) (78) and https://xgboost.readthedocs.io/en/stable/ libraries were used. The performances of these algorithms were superior compared to routine analyses with regressions. The choice was based on the study problem that would be better analyzed through the Decision Tree in order to find the best rules to be later used for forecasting.

Figure 10 – Study outline with Supervised Machine Learning



Source: Author themselves. A great volume of data is collected and input in the system; a predictive model is generated with the input data during training; the expected output is evaluated; following each new data entry in the predictive model developed, the expected outcome is the same.

The programming language used was Python. (Access: https://www.python.org/).

The final model's performance was the ROC/AUC metrics (ROC curve and area under the curve). The ROC curve graph was generated only in the comparison phase of the proposed MA approach with other approaches in the literature. Given the number of experiments performed, only AUC values were generated in other cases. In the models generated from the final model, the performance was evaluated using the confusion matrix and AUC values.

Cross-validation was used to obtain a more accurate estimate of how the final model would behave for real data. It was performed in 5 folds, that is, a total of 5 variations of the model are estimated, each with four folds for training and one-fold for testing.

4.5.3 Model development

The AI algorithm used was ML, with supervised learning algorithms trained to predict the probability of response to treatment. The initial model was built from input data (or dataset) that were arranged in ordered pairs with the desired “input” and “output”. These data were labeled; for each input, the output was already expected.

A prediction model is a function f(X) that receives information from patient X as input and returns the expected result as output so that the result can be anticipated for future patients. It is a weighted combination of features. The features included were

the evaluated attributes. A prediction model is obtained after a well-defined error-minimization strategy. There are many possible minimization strategies, but trees were used in this study.

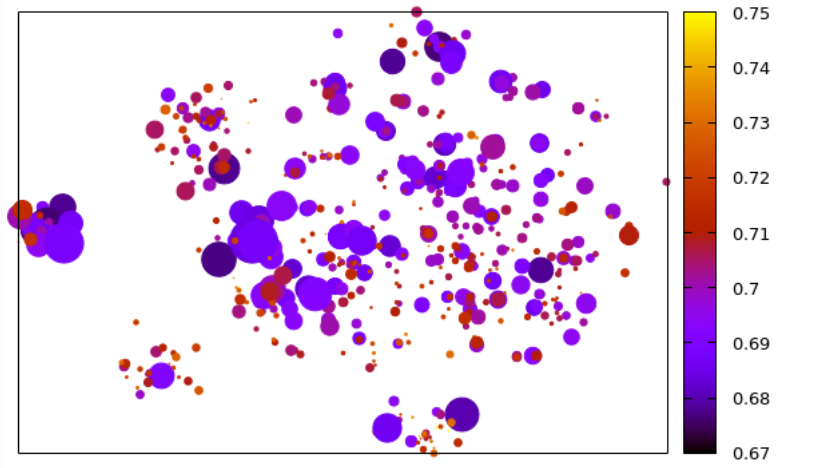
The models were built using base algorithms available in the SciKit-Learn libraries: XGBoost or Random Forests. A model from the XGBoost library and Random Forest was fed separately, with each label's resources. Each variable has a “weight” that the machine learns by adding the data. The model learned from data from subpopulations in which each subpopulation is associated with a specific subset of features; e.g., the characteristics of the original set were decomposed into several subsets. The input data were divided into training and test sets. The training set was used to build the model. Data learning, or data training, consisted of introducing several examples (records or instances) of desired inputs and outputs, using samples of cases for which the true classification was known to induce a set of training examples. During training and testing, some variations may occur due to parameter adjustments to find the combination that presents the best result. The models underwent a fine adjustment in the parameters of each algorithm, with many repetitions of processes, and this was possible due to the computational power used.

Then, the cluster learning method was used to promote diversity when learning the set. When we combine these different models, we have an ensemble, the predictive power being more significant than the models alone. Each generated ensemble model was evaluated according to its attributes. The models were grouped, separated by labels, according to their explanatory factors, and sampled in space. The points (or models) were grouped so that models that included the same attributes belonged to the same group. For this method, the set of characteristics representative of the characteristics to build the model was selected. The best models (best AUC) were selected to create the ensemble. The higher the AUC, the more clustered the models were.

Model performance was evaluated using appropriate statistics. The standard measure for quantifying forecasting performance is based on model sensitivity and specificity. The trade-off between these two measures produces a characteristic curve, summarized by its area (AUC). Model reliability was evaluated by cross-validation; that is, the same model is trained and evaluated using different subsets of data. Measures of dispersion about the mean (or average), especially the mean absolute deviation (MAD), provide how well the model can be expected to work for future patients. Figure

11 exemplifies a model space where each point is a model, their colors indicate AUC and their sizes indicate MAD (the more significant, the more reliable the model). A model's coordinates correspond to its predicted results. Specifically, each model receives patients and returns the corresponding results. The final vector of results associated with a model is projected into two-dimensional space for easy visualization.

Figure 11 – Representation of a model space



Each point represents a model x'. Models are arranged according to the probabilities attributes of patients, thus, models that specify similar probabilities to the same patients are placed next to each other in space.

After this process, the best learning models were selected, evaluated, and re-evaluated until a satisfactory model was found for the presented problem. A final predictive model was then created to map the evolution of pain in patients with chronic pain.

After defining the model, or final algorithm, all new data entries were performed through this programmed algorithm for the output of each label. Thus, several models for each outcome were generated.

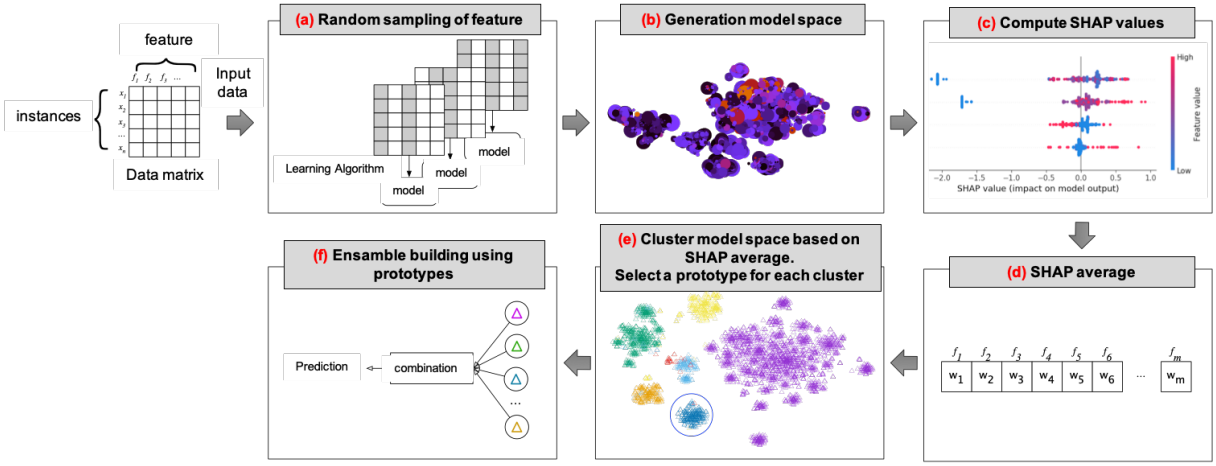
The generated models were arranged in a "model space" and graphically represented by SHAP. Models with a good threshold performed better than the average for each label. For each algorithm in each label, a space of unique and distinct models was generated in order to build simple and interpretable models. A prediction model's performance strongly depends on the characteristics that make up the model. A model composed of many features is likely to suffer from poor generalization, i.e., the model will not work well for future patients. Selecting the optimal set of features to compose a model requires exhaustive research on all possible feature sets.

The alternative was the sampling of forecasting models. Sampling involves randomly selecting a set of features and then training a prediction model composed of those features. This process was repeated many times, resulting in different (i.e., thousands) sets of features and models to characterize the model space. For example, you can filter out models that perform well and are reliable.

When presenting the model space, the maximum number of resources that make up each model was defined; each model should have at most 15 features as a good compromise between interpretability and performance. Using this threshold, the validity and feasibility of the work can be tested. For the elaboration of this project, “TreeSHAP” was used to calculate the explanation model, since the learning algorithms of this model are based on trees driven by gradients or random “forests”.

Figure 12 represents a view of the methodology (framework) used in this work. The algorithm was developed by UFMG, validated in thesis format (112) and article publication (113), and used as a case study for the problem presented in this project.

Figure 12 – Illustrated description of the methodology used.



Source: Costa, Moreira (2021) (112). Artistic representation of the framework overview. The framework began with an input table matrix. (n x m) being “n” the number of instances and “m” the number of resources. Figures represent a) A random sample of resources sets. The learning algorithm is used to induce a model for each sample set; b) The set of all generated models integrate a space in the model; c) and d) The calculation of the mean values Shapley Additive exPlanations (SHAP) for each model in the space of the model; e) The clustering of the model space based on the mean of SHAP.

The time limit for model optimization was defined as 24 hours; this is the approximate time in the worst-case scenario to execute the proposed approach.

4.5.3 Interpretation of the model

Objectively, the stages of the algorithm used can be described as follows: After generating the model space, the Shapley Additive Planation (SHAP) method was used to intuitively interpret the models graphically. With the models defined, each model's average of the SHAP values was calculated. Then, it was grouped in the model space using the average of the SHAP values as a criterion. Models with similar explanations were in the same grouping, while models with considerably divergent explanations were far apart and thus in different groupings. A prototype was then selected to represent each grouping. The method for choosing the prototype was to select the model with the best performance using the AUC of the cluster. Results were always given in pairs of AUC Curve and SHAP explanations. The selection of a prototype for each grouping made it possible to build a combination of models that, in addition to presenting a better performance, were more interpretable since they used only a fraction of the original variables. It is important to emphasize that the algorithm used carried out these steps (113, 114).

4.5.4 Development of the questionnaire

Each label spawned some models that were considered great. All generated models were grouped for each label, and among the generated models, each one was evaluated to verify which ones could generate a 12-question questionnaire with the best performance. The approach itself performed feature selection. Only some generated models met the criterion of few features and satisfactory performance.

After performing the ensemble, a unique model was generated for each label. Considering the AUC values, the best ensemble was evaluated, and a questionnaire was created with the variables of this ensemble model. It was necessary to restrict the number of resources to limit the number of questions. While the initial experiment allowed up to 15 features per model, in this case, it was limited to 7 features, with satisfactory performance.

The features that generated the 12 questions came from a single ensemble which, in turn, consisted of a combination of base models (XGBoost and Random Forest).

4.6 Statistical analysis

Data analyses were carried out in partnership with the Federal University of Minas Gerais (UFMG) and with the company Kunumi, a pure tech producing AI knowledge and tools.

The data were analyzed and prepared correctly. The data consisted of 338 attributes involving pain characteristics, socioeconomic status, and prescribed treatments and were not limited to continuous types.

The sample was characterized with descriptive statistics using means, standard deviations, and extensive data analysis. The software used for the descriptive statistical analysis was SPSS (Statistical Package for the Social Sciences) version 24.0 from IBM (International Business Machines).

The specifications related to the hardware on which the works were performed were CPU Intel® CoreTM i3-6100 @ 3.70 GHz, 16 GB DDR3 1,600 MT/s, and 256 GB SSD.

RESULTS

5 RESULTS

The results of this project were divided into two groups:

1. Descriptive analysis: descriptive statistical analysis to characterize the sample, using means and standard deviations;
2. Model analysis: analysis of models generated with the aid of AI.

5.1 Descriptive analysis

Of the 934 patients initially evaluated for inclusion in the study, 506 were included (Figure 13). The pain clinic evaluated all patients to establish the diagnosis of chronic pain. Women represented more than 57.7% of patients. The mean age of patients treated was 55.66±14.20 years. Patients from the southeast region of Brazil predominated (96.3%), with a low level of education (35%) and white (74.7%) (Table 1).

Figure 13 - Selection of research participants. The value in parentheses refers to the number (n) of patients at each selection stage.

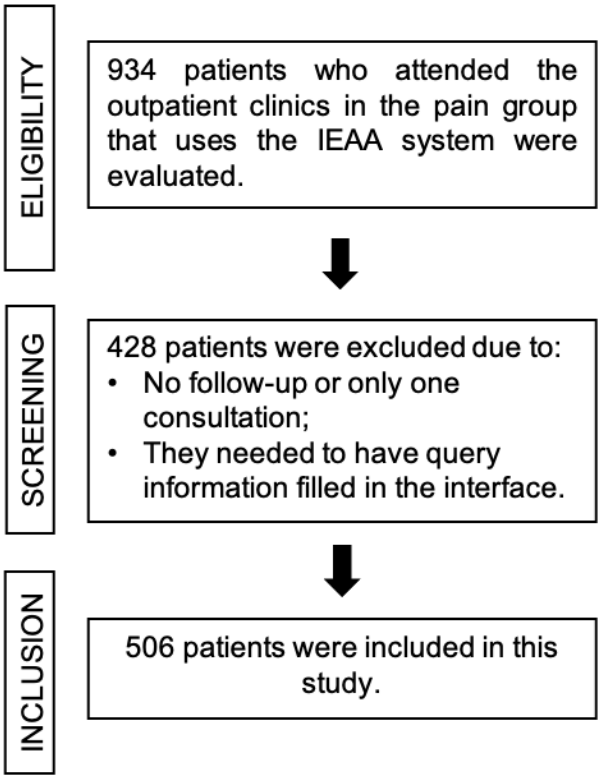


Table 1 – Distribution of patients according to sex, the state where they lived, level of education, and skin color in absolute numbers (n) and percentages (%).

	N = 506
Age (years)	55,66±14,20
Sex	
Male	214 (42,3%)
Female	292 (57,7%)
State	
Bahia	1 (2,0%)
Distrito Federal	1 (2,0%)
Espirito Santo	2 (4,0%)
Minas Gerais	1 (2,0%)
Pernambuco	1 (2,0%)
São Paulo	471 (93,3%)
Education level	
Primary education completes	177 (35%)
Primary education incomplete	61 (12,1%)
Secondary education complete	109 (21,5%)
Secondary education incomplete	12 (2,4%)
Tertiary education complete	37 (7,3%)
Tertiary education incomplete	14 (2,8%)
Postgraduation	3 (0,6%)
Literate	17 (3,4%)
Illiterate	5 (1,0%)
Not informed	42 (8,3%)
Ethnicity	
White	378 (74,7%)
Brown	48 (3,8%)
Black	26 (5,1%)
Asian	9 (1,8%)
Not informed	45 (8,9%)

All patients underwent standardized neurological examination (Table 2), and the main pain intensity, measured by VAS, was 6.36 (± 2.66). The mean value of the positive DN-4 questionnaire for neuropathic pain was 4.43. The patients had pain predominantly located in the lower limbs (right and left), mainly in the feet (right and left) (Table 3). Among the 506 individuals with chronic pain, it can be noted that the regions most affected by pain were located in the right hemibody. (Figure 14).

Table 2 – Distribution of absolute numbers (n) presented in means and standard deviations of the reduced McGill questionnaire, according to sensitive, affective, and evaluative dimensions. Pain intensities according to VAS. And neuropathic pain screening number with DN4.

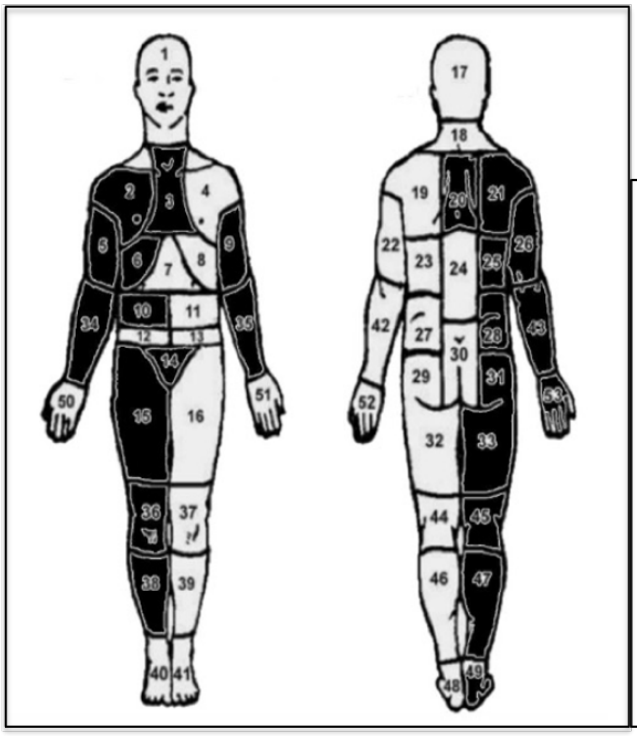
	N = 506
McGill Questionnaire short	
Sensitive	3,16±2,35
Affective	2,62±1,51
Evaluative	1,26±0,50
Analog Verbal Scale	
0	26 (5,1%)
1	7 (1,4%)
2	19 (3,8%)
3	24 (4,7%)
4	26 (5,1%)
5	69 (13,6%)
6	59 (11,7%)
7	77 (15,2%)
8	93 (18,4%)
9	43 (8,5%)
10	63 (12,5%)
Neuropathic pain Inventory (DN-4)	4,43±2,38

Table 3 – Distribution of affected body segments in absolute numbers (n) and percentages (%).

	N = 506 (%)
Location of the pain according to body segment	
Head	99 (19,6%)
Right shoulder	82 (16,2%)
Lower thorax	30 (5,9%)
Left shoulder	101 (20,0%)
Right arm	59 (11,7%)
Right hypochondrium	58 (11,5%)
Epigastric region	27 (5,3%)
Left hypochondrium	63 (12,5%)
Left arm	76 (15,0%)
Right flank	56 (11,1%)
Left flank	60 (11,9%)
Right periumbilical region	56 (11,1%)
Left periumbilical region	69 (13,6%)
Genital region	28 (5,5%)
Right thigh	95 (18,8%)
Left thigh	93 (18,4%)
Posterior region of the head	26 (5,1%)
Cervical Posterior region	68 (13,4%)
Left scapula	103 (20,4%)
Posterior thorax	28 (5,0%)
Right scapula	88 (17,4%)
Left arm posterior side	63 (12,5%)
Left posterior hemithorax	67 (13,2 %)
Lumbar	54 (10,7%)
Right posterior hemithorax	59 (11,7%)
Right arm posterior side	53 (10,5%)
Left lumbar	111 (21,9%)
Right lumbar	103 (20,4%)
Left glute region	109 (21,5%)

Coccygeal region	92 (18,2%)
Right glute region	101 (20,0%)
Posterior side of the left thigh	101 (20,0%)
Posterior side of the right thigh	84 (16,6%)
Right forearm	75 (14,8%)
Left forearm	81 (16,0%)
Right knee	107 (21,1%)
Left Knee	111 (21,9%)
Right leg	103 (20,4%)
Left leg	113 (22,3%)
Right foot	115 (22,7%)
Left foot	116 (22,9%)
Posterior side of the left forearm	50 (9,9%)
Posterior side of the right forearm	41 (8,1%)
Left popliteal region	75 (14,8%)
Right popliteal region	72 (14,2%)
Posterior side of the left leg	103 (20,4%)
Posterior side of the right leg	87 (17,2%)
Left heel	107 (21,1%)
Right heel	94(18,6%)
Internal region of the right hand	101 (20,0%)
Internal region of the left hand	101 (20,0%)
External region of the left hand	91 (18,0%)
External region of the right hand	76 (15,0%)

Figure 14 – Artistic representation of body regions affected by pain.



1. Head, 2. Right shoulder, 3. Anterior thorax, 4. Left shoulder, 5. Right arm, 6. Right hypochondrium, 7. Epigastric region, 8. Left hypochondrium, 9. Left arm, 10. Right flank, 11. Left flank, 12. Right periumbilical region, 13. Left periumbilical region, 14. Genital, 15. Right thigh, 16. Left thigh, 17. Head (posterior region), 18. Cervical (posterior region), 19. Left scapula, 20. Posterior thorax, 21. Right scapula, 22. Left arm (posterior), 23. Left thorax, posterior), 24. Lumbar, 25. Right thorax (posterior), 26. Right arm (posterior), 27. Left lumbar region, 28. Right lumbar region, 29. Left gluteus, 30. Coccygeal region, 32. Left thigh (posterior), 33. Right thigh (posterior), 34. Right forearm, 35. Left forearm, 36 Right knee, 37. Left knee, 38. Right leg, 39. Left leg, 40. Right foot, 41. Left foot, 42. Left forearm (posterior), 43. Right forearm to (posterior), 44. Left popliteal region, 45. Right popliteal region, 46. Left (posterior) leg, 47. Right (posterior) leg, 48. Left calcaneal region, 49. Right calcaneal region, 50. Right hand (inner region), 51. Left hand (inner region), 52. Left hand (outer region), 53. Right hand (outer region)

Source: Costa, Moreira (2021) (113). Areas in black are the most often reported by patients.

The “Global Impression of Change” questionnaire showed that more than 30% of the patients reported that the current treatment did not improve their pain, proportional to the opinion of the evaluating physician (Table 4).

Table 4 – Distribution of absolute numbers (n) and percentages (%) of responses to the Global Impression of Change questionnaires, according to the doctor's and patient's versions.

Global Impression Clinical	N = 506
Patient's version	
Slightly better	97 (19,2%)
Slightly worse	63 (12,5%)
Better	74 (14,6%)
Much better	30 (5,9%)
Much worse	17 (3,4%)
worse	53 (10,5%)
No changes	168 (33,2%)
Doctor's version	
Slightly better	90 (17,8%)
Slightly worse	60 (11,9%)
Better	91 (18,0%)
Much better	28 (5,5%)
Much worse	9 (1,8%)
Worse	48 (9,5%)
No changes	176 (34,8%)

Patients with chronic pain were subdivided according to pain diagnoses and pain syndrome (Tables 5 and 6). Neuropathic pain was the most prevalent main pain syndrome observed in 59.3% of patients. Among those with non-neuropathic pain as their main pain, fibromyalgia was the most common, affecting 8.1% of patients. As a secondary pain syndrome, fibromyalgia affected 5.9% of patients, followed by nociceptive pain, which affected 4%. Finally, as tertiary pain, migraine affected 4.9% of patients. In addition, other pain syndromes were reported, with 33% having myofascial pain syndrome.

Table 5 – Distribution of diagnoses classified as primary, secondary, and tertiary pain in absolute numbers (n) and percentages (%).

	N = 506
Primary pain	
Neuropathic pain	300 (59,3%)
Nociceptive pain	9 (1,8%)
Nociplastic pain	
Fibromyalgia	41 (8,1%)
Migraine	16 (3,2%)
Not informed	140 (27,7%)
Secondary pain	
Neuropathic pain	0
Nociceptive pain	20 (4.0%)
Nociplastic pain	
Fibromyalgia	30 (5.9%)
Migraine	12 (2.4%)
Tertiary pain	
Neuropathic pain	0
Nociceptive pain	0

Nociplastic pain	
Fibromyalgia	2 (0.4%)
Migraine	25 (4.9%)

Table 6 – Main pain syndromes in absolute numbers (n) and percentages (%).

	N = 506
Myofascial syndrome	167 (33.0%)
Core pain	71 (14.0%)
Postherpetic neuralgia	35 (6.9%)
Polyneuropathies	82 (16.2%)
Pain in phantom limb	11 (2.2%)
Trigeminal neuralgia	15 (3.0%)
Carpal tunnel syndrome	15 (3.0%)

As expected, patients with chronic pain had other baseline diseases (Table 7). Among the 506 individuals with chronic pain, 24.9% reported systemic arterial hypertension, and 70% had already suffered a cerebrovascular accident (CVA). The pain can be derived from some trauma. More than 16% of the patients reported pain installed after the trauma, with 5.9% motorcycle accidents and 4% firearm accidents.

Table 7 - Distribution of diagnoses of the most prevalent conditions reported in medical records in absolute numbers (n) and percentages (%).

Ailments reported in patient's records at the first consultation	N = 506
Hansen's disease	13 (2,6%)
Spinal cord injury	21 (4,2%)
Cancer	42 (8,3%)
Systemic Arterial Hypertension	126 (24,9%)
Diabetes mellitus	88 (17,4%)
Dyslipidemia	48 (9,5%)
Hypothyroidism	44 (8,7%)
Cerebrovascular accident	70 (13,8%)
Depression	45 (8,9%)
Traumas	82 (16,2%)
Car accidents	8 (1,6%)
Motorcycle accidents	30 (5,9%)
Being run over	8 (1,6%)
Firearm injury	20 (4,0%)
Fall	10 (2,0%)
Trauma	6 (1,2%)

Most patients had already been treated with medication before the first appointment by the ACDNC at HCFMUSP. Patients arrived at the pain clinic with treatment for chronic pain previously prescribed by other specialties. 47.5% of patients were using amitriptyline, and 41.5% were using gabapentin. (Table 8).

Table 8 – Distribution of drugs for pain registered in the first consultation in absolute numbers (n) and percentages (%).

Drugs being used	n = 506
Amitriptyline 25 Mg Pill.	241 (47,5%)
Amitriptyline + Lindocaine Gel (4% + 2%) Gel	46 (9,0%)
Baclofen 10 Mg Pill.	68 (13,5%)
Carbamazepine 200 Mg Pill.	58 (11,5%)
Celecoxib 200 Mg Caps.	7 (1,4%)
Ketoprofen 100 Mg Pill. Ent.	6 (1,2%)
Cyclobenzaprine 5 Mg Pill.	34 (6,7%)
Codeine 30 Mg Pill.	37 (7,3%)
Diclofenac Sodium 50 Mg Pill.	3 (0,6%)
Dipyrone 500 Mg Pill.	146 (28,9%)
Dipyrone 500 Mg / Ml Sol. Oral	58 (11,4%)
Duloxetine 30 Mg Caps.	16 (3,2%)
Phenytoin 100 Mg Pill.	6 (1,2%)
Fentanyl 25 Mcg / H patch	4 (0,8%)
Fentanyl 50 Mcg / H patch	2 (0,4%)
Gabapentin (Manipulated) 100 Mg Caps.	6 (1,2%)
Gabapentin 300 Mg Caps.	210 (41,5%)
Gabapentin 400 Mg Caps.	99 (19,5%)
Carboximetilcelulose Gel (manipulated) 3% Gel	3 (0,6%)
Carboximetilcelulose Gel 3% + Lidocaine 2% Gel	1 (0,2%)
Ibuprofen 300 Mg Pill.	5 (1,0%)
Lamotrigine 100 Mg Pill.	79 (15,6%)
Lamotrigine 25 Mg Pill.	52 (10,2%)
Levomepromazine (1mg / Drop) 4% Sol. Oral	7 (1,4%)
Lidocaine 2% Gel	8 (1,5%)
Metadona (Manipulated) 1Mg Caps.	1 (0,2%)
Metadona 10 Mg Pill.	95 (18,7%)
Morphine 10 Mg Pill.	12 (2,3%)
Morphine 30 Mg Caps. Lib.	1 (0,2%)
Morphine 30 Mg Pill.	6 (1,2%)
Naproxen 250 Mg Pill.	13 (2,5%)
Oxcarbazepine 300 Mg Pill.	6 (1,2%)
Oxycodone 10 Mg Pill.	14 (2,7%)
Oxycodone 20 Mg Pill.	2 (0,4%)
Paracetamol 500 Mg Pill.	59 (11,6%)
Piroxicam 20 Mg Pill. Subl.	2 (0,4%)
Pregabalin 75 Mg Caps.	22 (4,3%)
Muscle Relaxant HC Pill.	100 (19,7%)
Topiramate 100 Mg Pill.	5 (1,0%)
Topiramate 25 Mg Pill.	6 (1,2%)
Topiramate 50 Mg Pill.	7 (1,4%)
Tramadol 50 Mg Caps.	168 (33,2%)
Venlafaxine 75 Mg Caps. Lib.	101 (19,9%)
No pain medication	102 (20,1%)
Not informed	18 (3,5%)

A comparative analysis was conducted regarding the number of consultations to predict with a high degree of confidence whether the patient would respond positively to conventional treatment for chronic pain with a minimum number of follow-up consultations (Table 9). Two hundred and sixty-five patients were seen in at least

three consultations; 30% showed improvement with treatment, 34% were men aged 54 years on average, and the mean VAS value was 4.93. The increase in the number of consultations followed the reduction in the number of patients treated; fewer patients had more than three consultations. Table 9 shows patient information broken down by the number of appointments. As the number of appointments increased, so did the percentage of patients for whom the treatment was effective.

Accumulating resources from previous queries leads to better results. In general, there is a gain in prediction performance as more features are added to the model (i.e., more queries). The gain provided is quite significant, especially with the addition of the second appointment. However, forecast performance tends to stabilize. A significant disadvantage of systematic resource accumulation is that it markedly increases the total number of resources. For example, the data resulting from accumulating features from the first five queries has 1660 features. Algorithms that are not robust to handle so much functionality have limitations.

Table 9 – Distribution of the absolute number (n) and percentages (%) of consultations about the effect or not of the treatment, age (years), the average of the Brief Pain Questionnaire (McGill), and pain intensity (VAS).

	3 consultations	4 consultations	5 consultations
Effective treatment (n = 506)	152 (30,03%)	82 (16,20%)	57 (11,26%)
Sex – Male	52 (34,21%)	27 (32.93%)	17 (29,82%)
Age (years)	54,64 ± 9,2	54,11 ± 9,3	52,81 ± 10,2
Intensity of pain (0-10)	4,93 (2,0−7,25)	5,54 (3,25−8,0)	4,51 (2,0−7,0)
Not effective treatment (n = 506)	113 (22,33%)	51 (10,07%)	24 (4,74%)
Sex – Male	56 (49,56%)	21 (41,18%)	10 (41,67%)
Age (years)	57,21 ± 9,2	55,55 ± 9,5	50.21 ± 12,04
EVA (0−10) Intensity of pain	6,09 (5,0−8,0)	6,02 (4,5−8,0)	5,79 (5,0−8,0)

5.2 Model analysis

Data consisted of attributes extracted from self-reports of patients registered at the first consultation. The models' results predicted a significant reduction in pain at the end of treatment.

The models generated from the XGBoost, and Random Forest algorithms were based on the interaction of many data. The trained data model consisted of a mixture of subpopulations in which each subpopulation was associated with a specific subset of features. Sampling consisted of randomly selecting a subset of variables. As a

result, 1,000 random subsets were generated for each size (n), where 1 ≤n ≤15; the model could not have more than 15 variables, adding up to an approximate total of 15,000 subsets. The sample generated a total of 150,000 models using the XGBoost algorithm and another 150,000 models using the Random Forests algorithm. Each of them provided a different explanation for the phenomenon. Therefore, many contrasting interpretations or competing explanations exist for the same phenomenon.

As a comparison parameter, a single model was trained and fed with all the variables to obtain the average performance (average of XGBoost and the Random Forest). In the end, the following mean AUCs were found for each label: AUC of 0.65 for EVA 30, 0.615 for VAS 50, and 0.569 for GIC (physician and patient) (Table 10). Models with a good threshold performed better than the average for each label.

Table 10 – Distribution of mean AUC values when training a single model with data separated by labels.

	XGBoost	Random Forests	
Label	AUC	AUC	Mean
VAS 30	0,648	0,652	0,650
VAS 50	0,634	0,597	0,615
GIC	0,564	0,575	0,569

The mean was obtained by training a single model and feeding it with all the variables. To evaluate the performance of the models, a standard AUC (area under ROC) measure was used. A crossed validation of the five folds was used, hence, data were organized in five folds, and, in every run, four folds were used as training sets and the remaining fold was used as a test set. A separate validation set was also employed, which was used to select the best models. The mean value of AUC was reported in the five runs. All of this process was carried out separately for each label, namely, VAS 30, VAS 50, and GIC .

The experimental results revealed that the method reached an AUC of 0.84 based only on data from the first consultation, which followed the selection procedure of characteristics and explanation diversities. Using data from the second query, the AUC value increased to 0.896, with the model explanations from query 1 as meta-features and combining them with the new data from query 2. As a result, prediction performance increased to 0.975 in 5 queries, an increase of only 16% compared to consultation 1, reducing the treatment planning period; test results showed that it was optional to compare several consultations to obtain a good performance of the predictive model.

The performance threshold of 0.650 for the VAS label 30 resulted in one sampled model space for XGBoost and another for Random Forests. Although this performance

threshold appeared low, it exceeded the physician's estimated first-visit performance. As a result, the XGBoost model space comprises 2,830 models out of the original 150,000 models. In contrast, the Random Forests model space comprised 2,507 models.

The generated models were projected on a two-dimensional plane, according to the probabilities assigned by the model, using the t-sne technique, an adequate tool for better visualization of the perception of high-dimensional datasets (114).

The paired correlation was performed between the GIC patient version, GIC evaluator version, VAS 30, and VAS 50 labels. It was observed that the VAS 30 and VAS 50 labels were highly correlated and reached a correlation value of 0.85. However, the GIC label was not highly correlated with VAS 30 and VAS 50 and presented correlation values of 0.1 and 0.097, respectively, meaning that patients' and physicians' self-assessments could be discrepant. It has been shown that when patients experience a 30% reduction in pain intensity, most of the time, they also experience a 50% reduction.

Although the type of pain is important for diagnosis and treatment, the response to treatment was more based on the personal characteristics of patients and less related to the type of pain itself. Patients were usually treated according to their main pain complaint, with pain etiology not being the main factor directly considered in the models.

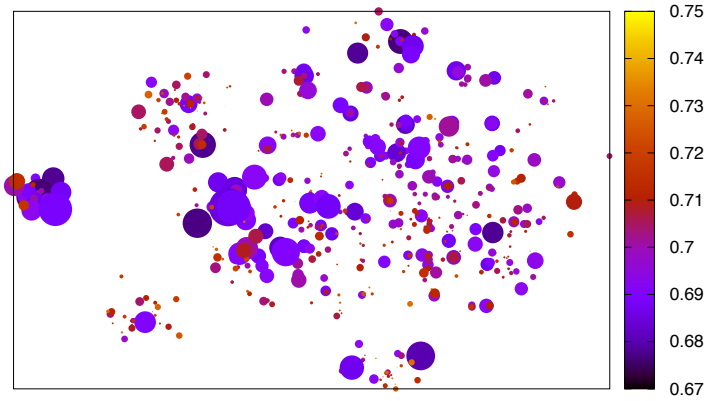
The best models for each label will be presented below and explained briefly for observation.

5.2.2 Templates for the label VAS 30

The XGBoost model space generated for VAS 30 comprised 2,830 models out of the original 150,000 models, while the Random Forests model space comprised 2,507 models. Figure 15 represents the model space, where each point corresponded to a model, and the point size indicated the variation of the validation error; the color scale was associated with the performance achieved by the model, with light colors representing the best performance. It demonstrates predictive models that had a prediction effectiveness greater than 0.65 of AUC. Each dot has color, size, and position (x and y coordinates). The color shows the average AUC value; the lighter, the more influential the model. Each model was evaluated several times, and an

average AUC was obtained. The size of the points shows the degree of reliability in the model, that is, the degree of variation of the AUC. In this case, the higher the point, the more reliable the model was. Light and large dots were best. The coordinates were assigned based on the probabilities of improvement for each patient. Then, each model (or point) was applied to all patients, generating a probability vector (each dimension of the vector was the probability that the patient in question improved). So, if one point is located close to the other, these two points correspond to predictive models and were right for the same patients. If two points are positioned far from each other, they correspond to the correct models for different patients. The spaces of the generated models were filtered, and only the models considered good were selected, with superior performance than the average. Depending on the space, this filtering removed 90% to 98% of the models.

Figure 15 – Graphical representation of the model space generated for the VAS 30 label.



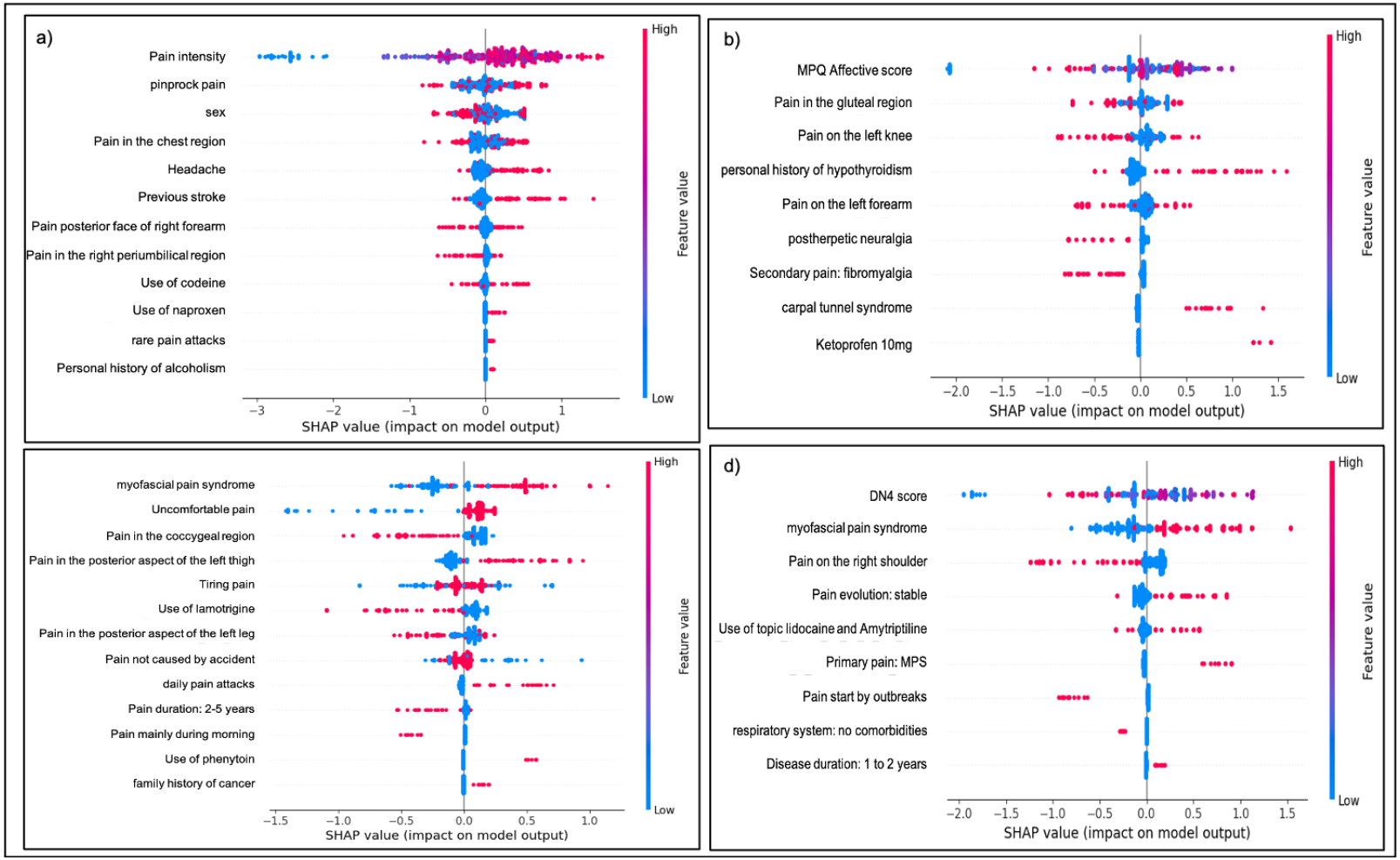
Visualization t SNE of the space of models generated where each point represents a model *X*'. The models are displayed according to the attribute probabilities by model. The colors indicate the mean (cross-validation) of the AUC values and the size is relative to the variance with smaller points indicating that the corresponding model has a lower variance.

After that, the attributes present in each model were evaluated. Models which included the same attribute were grouped together. Finally, to explain the model, the used methods was Shapley Additive Planation (SHAP) to intuitively interpret the models graphically. Each SHAP graph generated represented an individual description of each model and the how the variables individually impacted this model. A SHAP was created for each model.

In Figure 16 is the representation of the best models, selected according to the distribution of the variables evaluated according to the level of interference on the occurrence of pain, based on the coding expressed in the ordinate column presented to the right of the SHAP graphs. The figure depicts an overview of the models which had an improvement of 30% as the outcome, selected as “great” from the analyses of the AUCs’ means. There was great diversity in these prototype models, evidencing 64 distinct resources, of which 8 resources were present in 7 models and just 1 resource in three models.

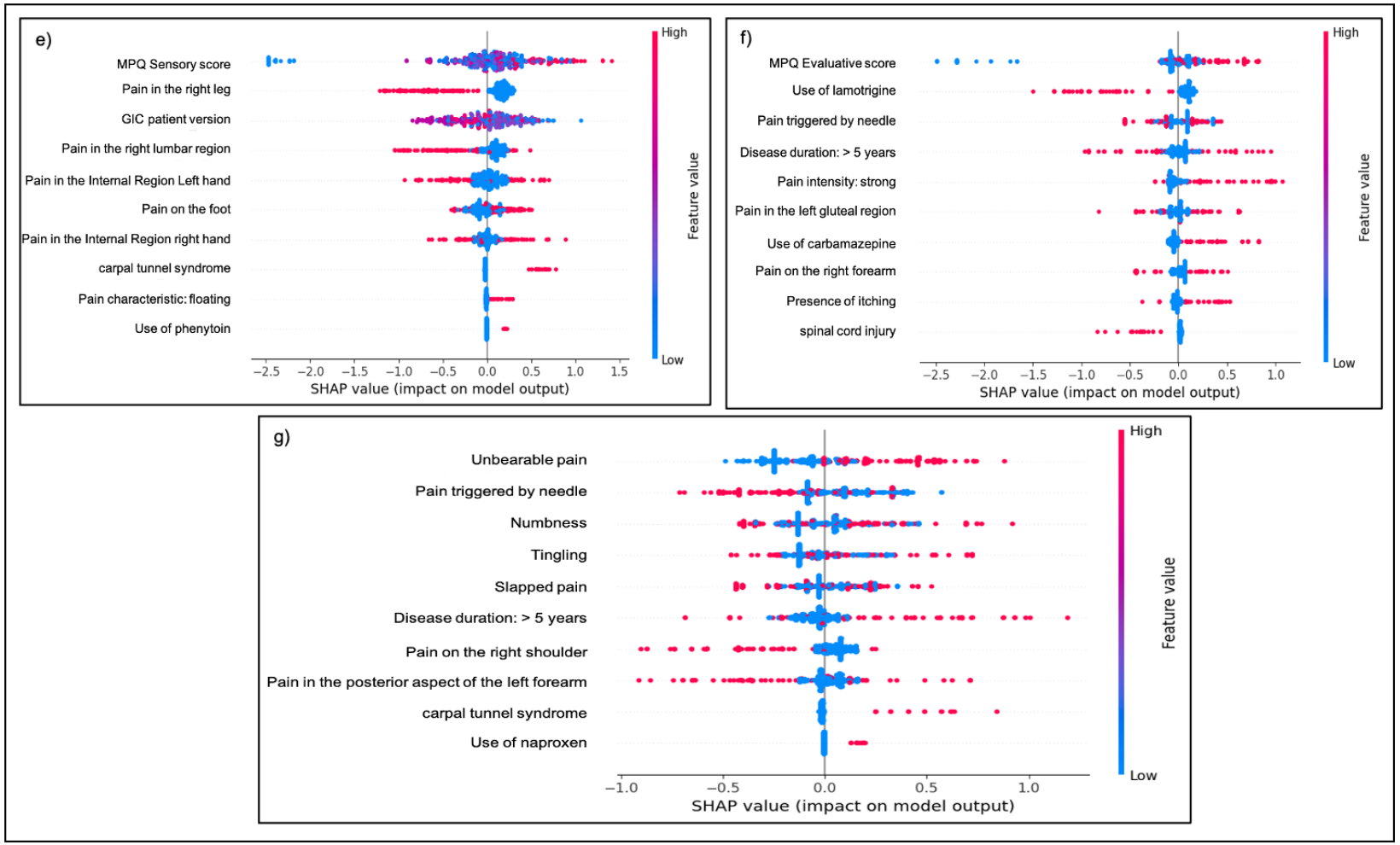
The represented model in the graph in Figure 16 “a” aimed at predicting the improvement of the patient based on the result of VAS 30 using 12 features. The most important feature for the decision making of the model is located on the upper part of the SHAP graph. In this case the ”intensity of pain” was the most relevant feature, followed by the “pin like”, and “sex”. The “personal history of alcoholism” was the least important in this case. The models differed significantly in terms of their explanatory factors. The diversity becomes clear as one inspects the prototype models, since each model employed a very different set of resources from the other prototype models. As the models were selected maximizing the diversity of explanations, it was expected that the number of shared resources would be few. The most relevant resources for each prototype could be extracted directly from their SHAP values. The most relevant characteristics of the final model was the combination of the most relevant characteristics in their prototype model. The most relevant characteristics of each model of this prototype were: “pain intensity”, McGill score – affective dimension”, “Myofascial Pain Syndrome", “DN-4 score total”, McGill score – sensory dimension”, “McGill score – evolving dimension”, and “unbearable pain”.

Figure 16 – Explanation factors presented as summary SHAP graphs with the label VAS 30.



A set of graphs of summaries showing only the most important resources. They were extracted from the Random Forest and XGBoost models. Each line represents a variable. The datasets were constructed from the concatenation of the first visit, combined with the SHAP obtained from a trained machine learning model with all the data from the first visit. Explanatory factors (seen as SHAP summary graphs) associated to prototype models with the label VAS 30. It continues.

Figure 16 - Explanatory factors are presented as SHAP summary charts for the VAS 30 label .



A set of graphs of summaries showing only the most important resources. They were extracted from the Random Forest and XGBoost models. Each line represents a variable. The datasets were constructed from the concatenation of the first visit, combined with the SHAP obtained from a trained machine learning model with all the data from the first visit. Explanatory factors (seen as SHAP summary graphs) associated to prototype models with the label VAS 30. Conclusion.

Considering the 30% reduction in pain, of the 506 patients, there was an improvement in 210, while the remaining 296 did not improve (Table 11).

Table 11 – Distribution of patients with 30% or no improvement in pain according to VAS, in absolute numbers (n) and percentages (%).

VAS 30	n = 506
Improved	210 (41,5%)
Did not improve	296 (58,5%)

Among the 210 patients who improved, the model could correctly predict 171 but was wrong in 39. Within the 296 who did not improve, the model could correctly predict “no improvement” of pain in 165 patients but was wrong in the prediction in 131 cases (Table 12).

Table 12 - Confusion matrix for evaluating the models generated based on the label VAS 30.

		Predicted value	
	Real	No	Yes
VAS 30	No	165	131
	Yes	39	171

5.2.1 Models for the label VAS 50

The XGBoost model space generated for VAS 50 comprised 1,408 models out of the original 150,000 models, while the Random Forests model space comprised 11,829 models.

The results of the generated prototype models were displayed in representative SHAP graphs associated with prototype models generated in the VAS 50 model space to provide an overview of the most important features.

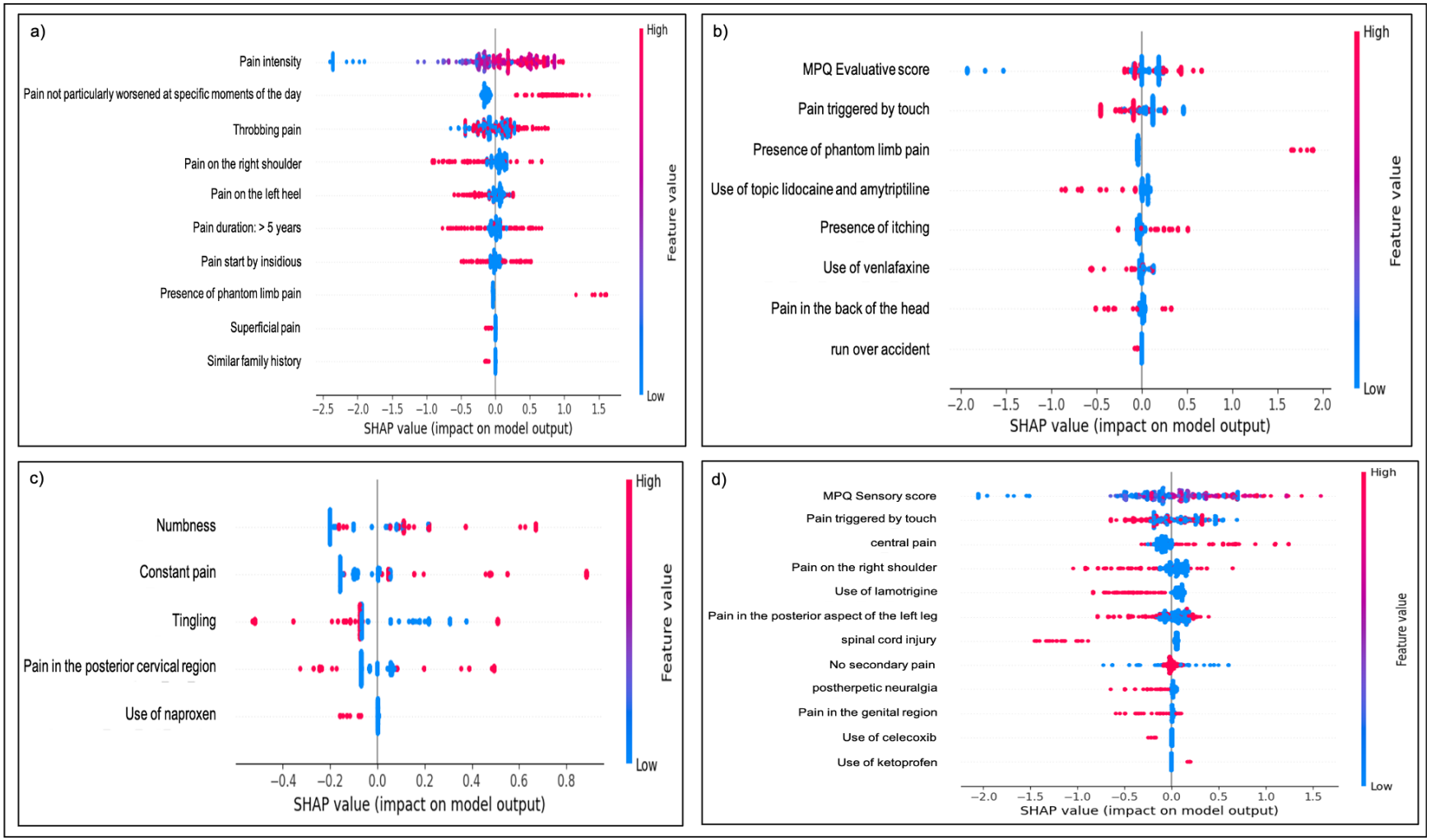
Figure 17 represents the best models selected according to the distribution of the evaluated variables, referring to the degree of interference in the occurrence of pain, based on the coding expressed in the ordinate column presented to the right of the SHAP graphs. The figure provides an overview of which ensembles were most important in the models, with a 50% improvement in pain as the outcome, selected as “optimal” from the analysis of the mean AUCs. The models differed greatly depending

on the extent, location, and duration of pain and prescribed medications. There was great diversity in these prototype models, showing 79 distinct features within the 11 models, of which 14 features were present in 11 models and only one feature in four models.

The model represented in Figure 17 by graph “a” aimed to predict the patient's improvement based on the result of the VAS 50 using ten features. The most important feature for model decision-making is located at the top of the SHAP chart. In this case, “pain intensity” was the most relevant feature, followed by “pain predominance” and “throbbing pain”. “Similar family history” was the least important in this case. Therefore, according to the values of the SHAP graph for the “pain intensity” feature of model “a”, it was considered that high values of pain intensity increased the possibilities of a significant reduction in the magnitude of pain at the end of the treatment. Very low values (blue dots) are concentrated on the left side of the graph; low pain intensity did not contribute significantly to the patient's improvement. The points in red for the second variable, which refer to the predominance of pain, are predominantly positioned to the right of the value “0”, that is, well delimited, demonstrating that the lack of a moment in the day when pain predominates had an positive impact in predicting improvement; patients who had a specific time of pain predominance had little impact on prediction. The sensation of throbbing pain, of the pain-sensitive dimension variable, had a positive but small impact on pain improvement; the absence of throbbing pain had almost no impact on the improvement or worsening of pain in this model. The model suggested that the feature “pain located in the right shoulder” was revealed to have a negative impact on this model and less probability of improvement if the patient had it. “Phantom limb pain” was related to a great positive impact on the prediction of pain improvement.

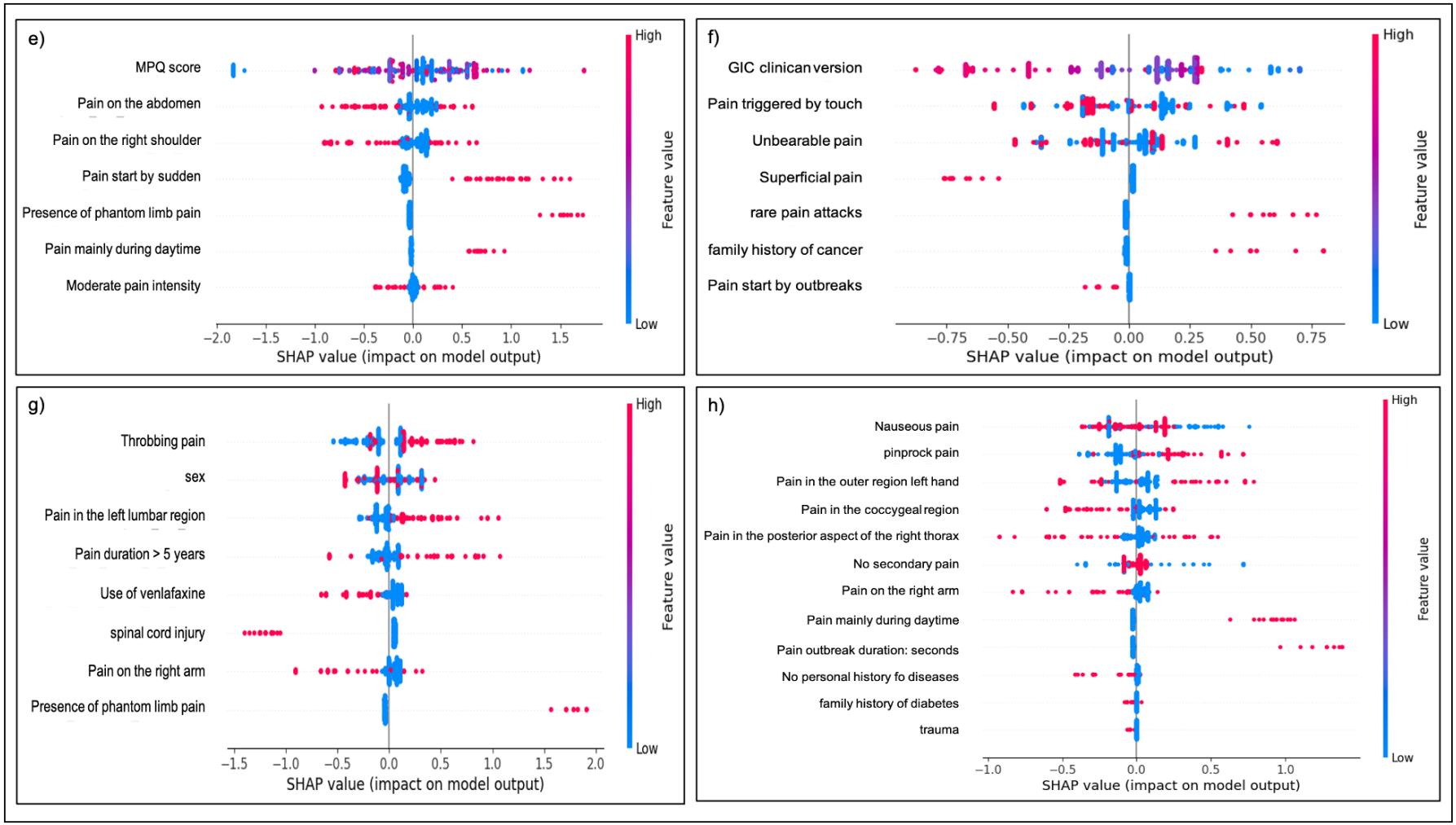
The most relevant characteristics of each prototype model were: “pain intensity”; “McGill score - evolutionary dimension”; “numbness”; “McGill score - sensitive dimension”; “McGill total score”; “GIC Doctor's Version”; “throbbing pain”; “nauseating pain”; “DN-4 score”; “severe pain intensity” and “McGill score - affective dimension”.

Figure 17 - Explanatory factors presented as SHAP summary charts for VAS 50 label.



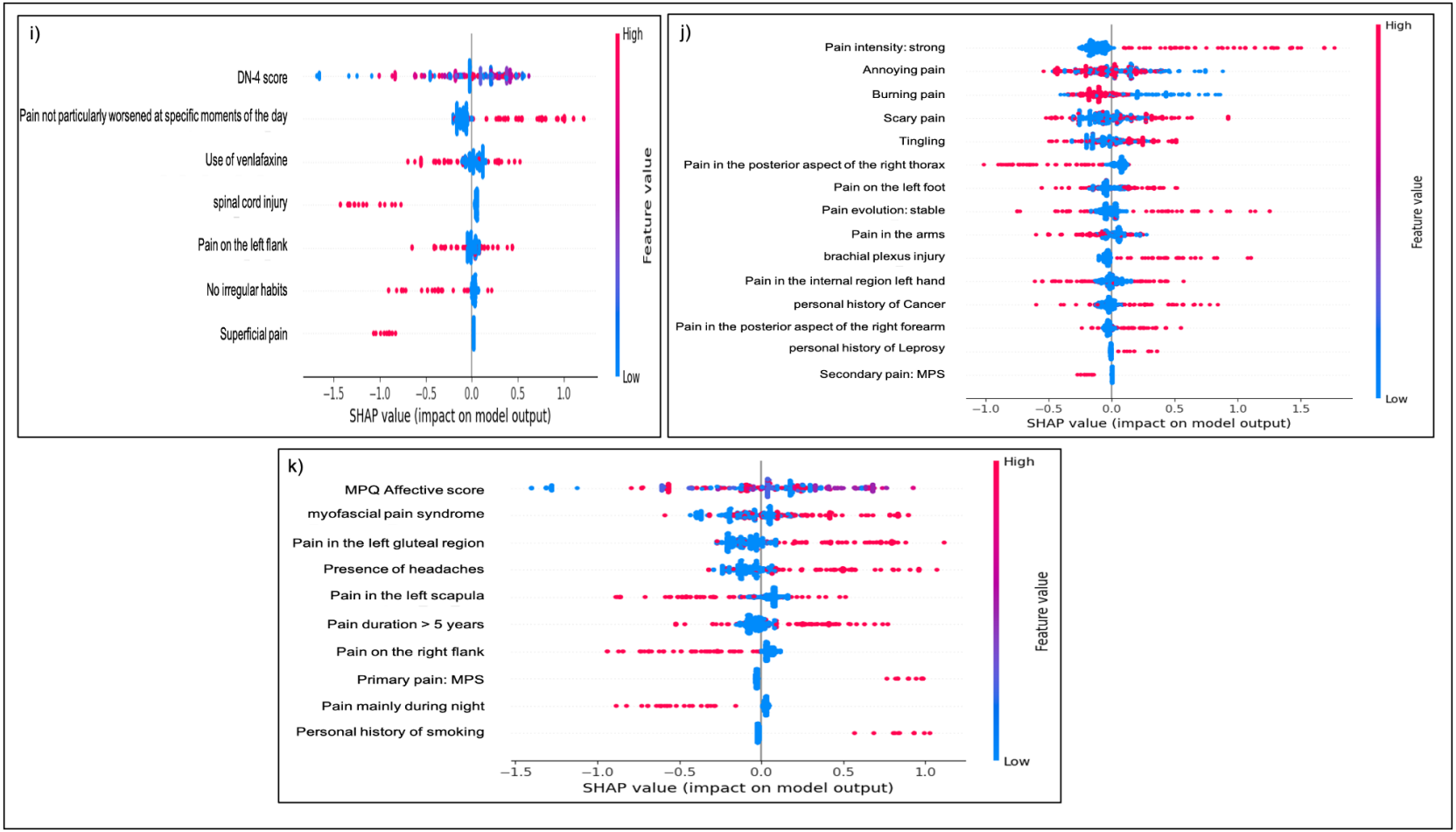
A set of graphs of summaries showing only the most important resources. They were extracted from the Random Forest and XGBoost models. Each line represents a variable. The datasets were constructed from the concatenation of the first visit, combined with the SHAP obtained from a trained machine learning model with all the data from the first visit. Explanatory factors (seen as SHAP summary graphs) associated with prototype models for EVA 50 label. Four of the eleven models of prototypes built with XGBoost. (It continues below).

Figure 17 – Explanatory factors presented as SHAP summary charts for VAS 50 label .



A set of graphs of summaries showing only the most important resources. They were extracted from the Random Forest and XGBoost models. Each line represents a variable. The datasets were constructed from the concatenation of the first visit, combined with the SHAP obtained from a trained machine learning model with all the data from the first visit. Explanatory factors (seen as SHAP summary graphs) associated with prototype models for VAS 50 label. Four of the eleven models of prototypes built with XGBoost. (It continues below).

Figure 17 – Explanatory factors presented as SHAP summary charts for VAS 50 label .



A set of graphs of summaries showing only the most important resources. They were extracted from the Random Forest and XGBoost models. Each line represents a variable. The datasets were constructed from the concatenation of the first visit, combined with the SHAP obtained from a trained machine learning model with all the data from the first visit. Explanatory factors (seen as SHAP summary graphs) associated with prototype models for VAS 50 label. Four of the eleven models of prototypes built with XGBoost. Conclusion.

Of the 506 patients included in the present study, 160 patients reported at least 50% pain relief at the end of follow-up, while 346 did not. Although this number is not precise, it was considered the worst scenario, since the traditional management to treat chronic pain represents, at least, confidence in the patient's improvement. (Table 13).

Table 13 – Distribution of records of pain improvement of 50% or not according to the VAS assessment of pain intensity (0 to 10) in absolute numbers (n) and percentages (%).

VAS 50	n = 506
Improved	160 (31,62%)
Did not improve	346 (68,38%)

The matrix revealed that of the 160 patients who improved, the model could correctly predict 105 and was wrong in 55. On the other hand, of the 346 who did not improve, the model could correctly predict non-improvement in 246 cases, but it needed to be corrected in 100 cases. Table 14).

Table 14 - Confusion matrix for evaluating the models generated based on the VAS 50 label.

		Predicted Value	
	Real	No	Yes
VAS 50	No	246	100
	Yes	55	105

5.2.3 Global Impression Clinical assessment

Data referring to the report of pain improvement were evaluated according to the versions of the patient and the evaluator regarding the Clinical Global Impression (GIC) scale, classified as much worse, worse, slightly worse, unchanged, slightly better, better, much better. The outcome values were “better” or “much better”.

Regarding GIC, 18,575 models were generated for XGBoost and 10,035 models for Random Forests.

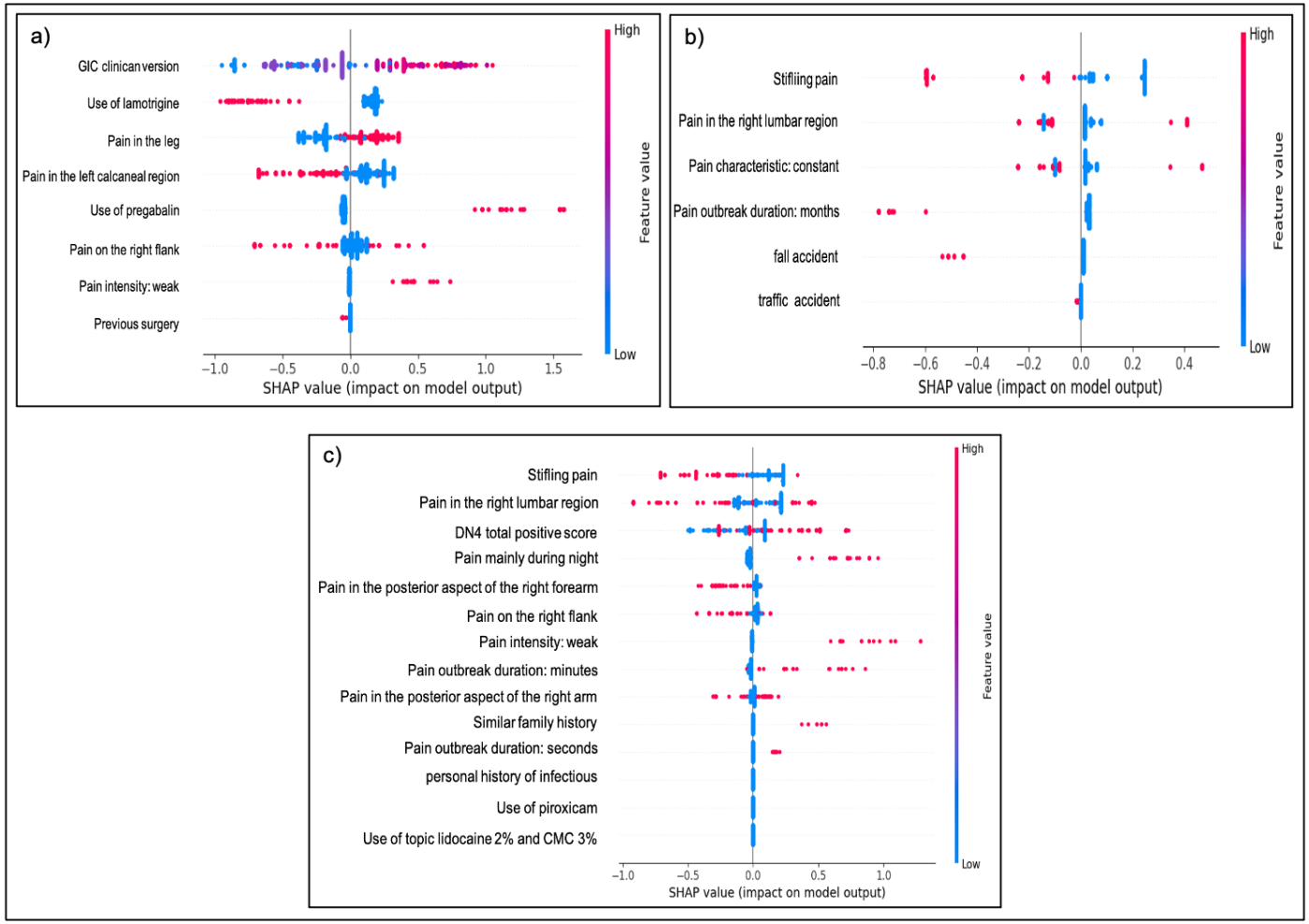
5.2.3.1 Patient version

Figure 18 shows the best models selected according to the distribution of the evaluated variables, referring to the degree of interference in the occurrence of pain, based on the coding expressed in the ordinate column shown to the right of the SHAP graphs. The figure reveals an overview of which models were selected as “optimal” based on the analysis of the average AUCs, which had the patient’s version of pain as the outcome, with “better” or “much better” as positive parameters (value 1)., according to the GIC scale. The models differed greatly depending on the extent, location, and duration of pain and prescribed medications. There was great diversity in these prototype models, showing 49 distinct ensembles within the five prototype models, of which seven features were present in 4 models and four features were present in only 1 model.

The model represented in Figure 18 by graph “a” uses eight features. It can be seen in this model that the variable “GIC medical version” was the most relevant, followed by the variable “lamotrigine 100mg” and “leg pain”; that is, they are the three variables that most impacted the prediction of this model; following the same philosophy, knowing whether “the patient underwent surgery” was the least important feature.

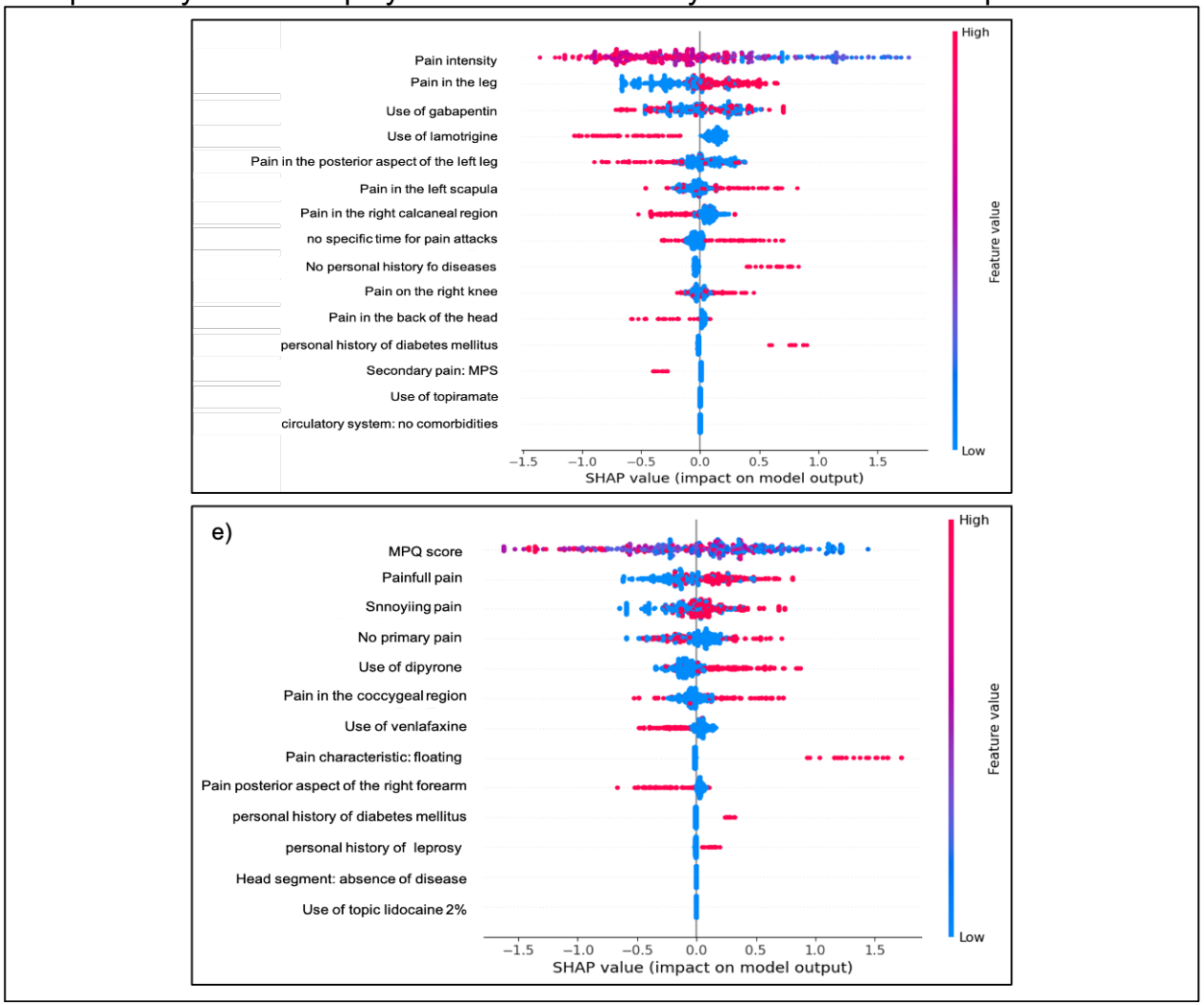
The most relevant features of the final model would be the combination of the most relevant features in its prototype models. For example, using the GIC scale by patient version as the label, XGBoost as the learning algorithm, and DBScan as the clustering algorithm, the most relevant characteristics of each prototype model were: “GIC physician version”; “suffocating pain”; “pain intensity”; “McGill total score”.

Figure 18 - Explanatory factors displayed as SHAP summary charts for GIC label patient's version.



A set of graphs of summaries showing only the most important resources. They were extracted from the Random Forest and XGBoost models. Each line represents a variable. The datasets were constructed from the concatenation of the first visit, combined with the SHAP obtained from a trained machine learning model with all the data from the first visit. Explanatory factors (seen as SHAP summary graphs) associated with prototype models for GIC patient's version. Four of the eleven models of prototypes built with XGBoost.. Continuation

Figure 18 - Explanatory factors displayed as SHAP summary charts for GIC label patient's version.



A set of graphs of summaries showing only the most important resources. They were extracted from the Random Forest and XGBoost models. Each line represents a variable. The datasets were constructed from the concatenation of the first visit, combined with the SHAP obtained from a trained machine learning model with all the data from the first visit. Explanatory factors (seen as SHAP summary graphs) associated with prototype models for GIC patient's version... Conclusion.

According to the patient version GIC scale results, out of the 506 patients, 130 patients improved, and the remaining 376 did not. (Table 15).

Table 15 - Distribution of patients with or without pain improvement according to the patient's version of the Clinical Global Impression (GIC) scale, in absolute numbers (n) and percentages (%).

GIC patient	n = 506
Improved	130 (25,7%)
Did not improve	376 (74,3%)

The model correctly predicted the outcome of 73 patients out of the 130 who improved, but it was wrong in 57. Of the 376 who did not improve, the model was able to predict the absence of improvement in 288 cases correctly, but it needed to be corrected in the prediction in 88 cases (Table 16).

Table 16 - Confusion matrix for evaluating the models generated based on the patient's version GIC label.

		Predicted Value	
	Real	No	Yes
GIC patient	No	288	88
	Yes	57	73

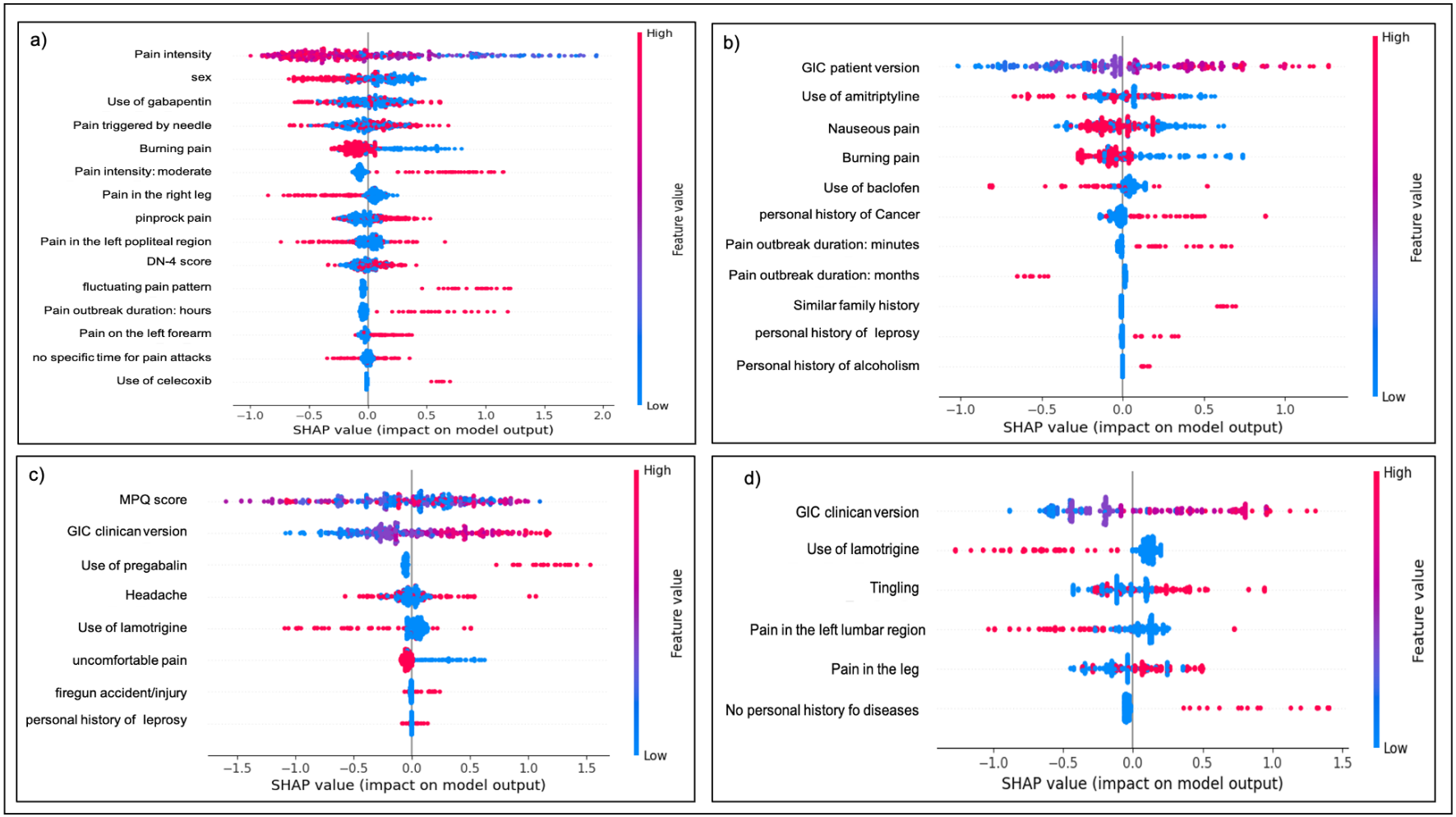
5.2.3.2 Doctor’s version

Figure 19 represents the best models selected according to the distribution of the evaluated variables, referring to the degree of interference in the occurrence of pain based on the coding expressed in the ordinate column presented to the right of the SHAP graphs. The figure provides an overview of which ensembles were most important in the models that had the doctor's version of the patient's pain sensation as the outcome, having as positive parameters (value 1) “better” or “much better”, according to the scale of GIC. The models differed greatly depending on the extent, location, and duration of pain and prescribed medications. There was great diversity in these prototype models, showing 36 distinct ensembles within the 4 prototype models, of which 4 features were present in 3 models and only 1 feature in three models.

The model represented in Figure 19 by graph “a” used 15 features. It can be observed in this model that the variable “pain intensity” was the most relevant, followed by the variable “gender (f/m)” and “gabapentin use”; that is, they were the three variables that most impacted the prediction of this model; following the same philosophy, knowing whether “the patient used celecoxib” was the least important feature.

The most relevant features of the final model would be the combination of the most relevant features in its prototype models. Using the GIC scale by the doctor's version as a label, the most relevant characteristics of each prototype model were: “pain intensity”; “GIC patient version”; “McGill total score”; "GIC Doctor's Version”.

Figure 19 – Explanation factor presented as SHAP summary graphs for the GIC label Doctor's version.



A set of graphs of summaries showing only the most important resources. They were extracted from the Random Forest and XGBoost models. Each line represents a variable. The datasets were constructed from the concatenation of the first visit, combined with the SHAP obtained from a trained machine learning model with all the data from the first visit. Explanatory factors (seen as SHAP summary graphs) associated with prototype models for GIC doctor's version.

According to the results of the physician's version of the Clinical Global Impression scale, of the 506 patients, 137 patients improved, and the remaining 369 did not (Table 17).

Table 17 – Distribution of patients in terms of pain improvement or not, according to the Clinical Global Impression (GIC) scale according to the doctor's version in absolute numbers (n) and percentages (%).

GIC Doctor	n = 506
Improved	137 (27,1%)
Did not improve	369 (72,9%)

The model correctly predicted the outcome of 94 patients out of the 137 who improved, but it was wrong in 43. Of the 369 who did not improve, the model was able to predict the absence of improvement in 235 cases correctly, but it needed to be corrected in the prediction in 134. (Table 18).

Table 18 - Confusion matrix for evaluating the models generated based on the GIC label evaluator's version.

		Predicted Value	
	Real	No	Yes
GIC doctor	No	235	134
	Yes	43	94

5.2.5 Final questionnaire

The final questionnaire was obtained after generating the ensemble for each label and evaluating the corresponding AUCs.

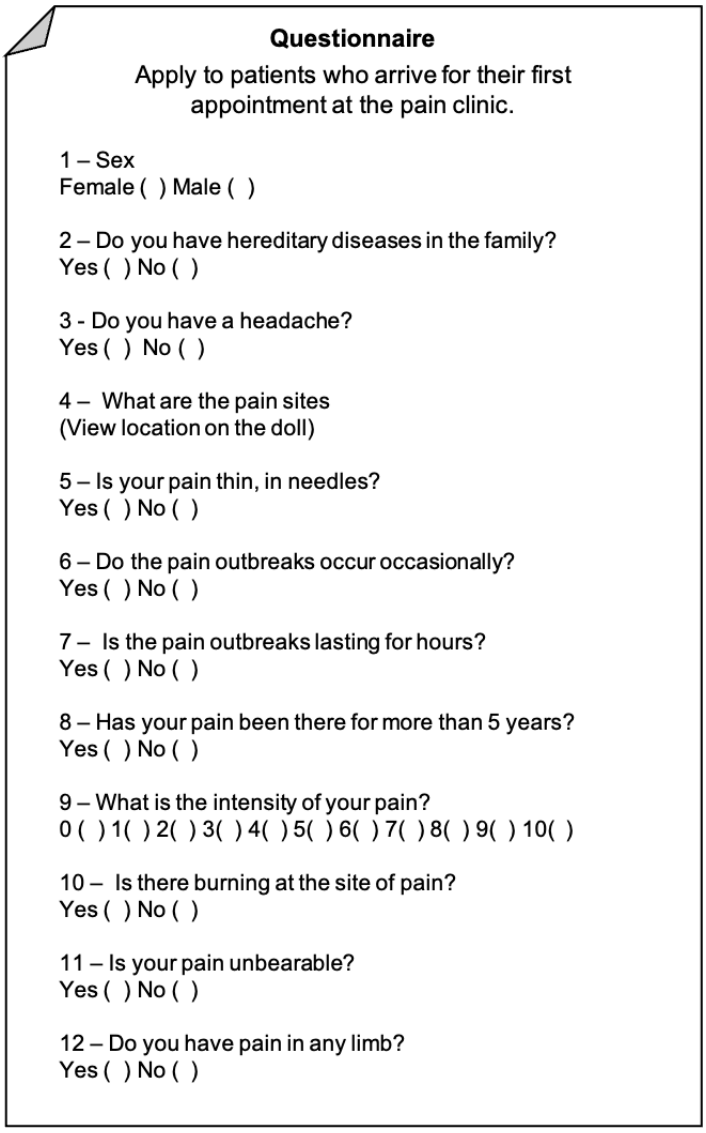
An AUC > 0.70 was used as a filter to develop the questionnaire experiment. In addition, the following conditions were added to the system: 1) the model was composed of a reduced number of features, and 2) ensembles were generated with a maximum of 38 features. To elaborate on the questionnaire, the objective was to reach maximum performance with the information from the first consultation.

After analyzing the generated models, the patient version GIC label reached the best AUC, with an ensemble model with 14 unique features converted into 12 questions.

The algorithm developed from the analysis of the data of the current project was practical not only in identifying, in an isolated way, the variables that most contributed to the outcome but also in the combination of these variables that achieved an excellent performance.

Figure 20 shows the questionnaire developed in this work.

Figure 20 – Questionnaire presented with the best features of the models generated from the GIC label.



Questionnaire

Apply to patients who arrive for their first appointment at the pain clinic.

1 – Sex
Female ( ) Male ( )

2 – Do you have hereditary diseases in the family?
Yes ( ) No ( )

3 - Do you have a headache?
Yes ( ) No ( )

4 – What are the pain sites
(View location on the doll)

5 – Is your pain thin, in needles?
Yes ( ) No ( )

6 – Do the pain outbreaks occur occasionally?
Yes ( ) No ( )

7 – Is the pain outbreaks lasting for hours?
Yes ( ) No ( )

8 – Has your pain been there for more than 5 years?
Yes ( ) No ( )

9 – What is the intensity of your pain?
0 ( ) 1( ) 2( ) 3( ) 4( ) 5( ) 6( ) 7( ) 8( ) 9( ) 10( )

10 – Is there burning at the site of pain?
Yes ( ) No ( )

11 – Is your pain unbearable?
Yes ( ) No ( )

12 – Do you have pain in any limb?
Yes ( ) No ( )

Questionnaire developed from the best AUCs of the variables with the best labels, namely, the patient's version of the GIC scale.

The questionnaire was entered into the IEAA system and will be completed during the first consultation of patients treated at the PCNC at HCFMUSP. In addition, the internal validation of the questionnaire will be carried out with a prospective study of these data.

DISCUSSION

6 DISCUSSION

This study aimed to apply an artificial intelligence approach to identify predictive parameters of the evolution of pain relief in patients with chronic pain and to obtain a questionnaire containing these parameters to help in the clinical management of these patients. Despite the existence of several guidelines and recommendations for its treatment, up to 40% of patients with chronic pain remain symptomatic despite the best medical treatment (58), and precisely defining the best therapy for a patient is still a challenge.

Chronic pain is considered a complex disease, resulting from associated morbidities and other symptoms resulting from it or therapeutic interventions. This explains why it is misdiagnosed in more than 75% of patients, and treatments are often ineffective. In addition, how pain is described, interpreted, and understood varies considerably from the perspectives of health professionals and sufferers, especially when the pain is chronic (13, 115).

The descriptive analysis carried out in patients with pain makes it possible to observe several factors that determine its occurrence and perpetuation, such as sex, age, life habits, individual and hereditary antecedents, and types of pain, among other clinical determinants. The management of these risk factors can make it possible to guide prophylactic attitudes towards chronic pain, aiming at reducing its duration and severity, rehabilitation and integration of patients. In addition, some of these factors are relevant for predictions, assessments, management, and prognosis of chronic pain conditions, and others are potentially important for identifying new models of therapeutic interventions (115).

The complexity of chronic pain is challenging because it stems from several mechanisms that can cause or aggravate it, which justifies the various strategies for its management. Furthermore, the journey to defining the cause of the pain can take many years. Therefore, patient-centered multidisciplinary assessment is extremely important and is the basis for appropriately managing chronic pain (116).

6.1 Demographic characteristics of the study cohort

Most (96%) of the patients included in the present study came from the southeastern region of Brazil; however, it must be considered that most of the patients

treated at HCFMUSP live in the state of São Paulo. According to studies, approximately 31% of adult residents of the city of São Paulo have chronic pain, which is considered a public health problem (2, 58, 117).

The prevalence of chronic pain in women in Brazil is approximately 70%; in this series, it was observed that pain was more prevalent in women (57.7%); however, it is important to emphasize that more women than men are treated in outpatient service environments , that is, there is greater demand for assistance to treat their pain (2, 114, 118, 119).

Pain prevalence is highest in adults aged 45 to 65, and the mean age of patients seen in the pain clinic was 55 years (58, 114).

74% of the patients treated at the outpatient clinic declared themselves white. However, there are poorly understood ethnic variations related to pain, and it is questionable whether white people experience less pain than black people (120, 121).

Although most patients had a certain level of education, 35% had completed primary education, 21% had completed secondary education, and only 7.3% had completed higher education. Other studies also indicate a correlation between chronic pain and socioeconomic deprivation; individuals with low educational levels and low social status are more likely to have chronic pain (121-123).

The patients included in this series had other comorbidities such as systemic arterial hypertension (SAH) (25%), diabetes mellitus (17%), and cerebrovascular accident (CVA) (14%). Previous studies report the presence of additional chronic conditions in up to 88% of individuals with chronic pain, suggesting that those with chronic diseases are more prone to chronic pain. If, on the one hand, chronic pain increases the risk for SAH, on the other hand, the predisposition to chronic pain increases in those with SAH. The presence of comorbidities is related to the successful therapeutic management of chronic pain as it limits the applicability of disease-specific clinical guidelines and reduces options for adequate analgesic therapy (124-126). The predisposition to the occurrence of chronic painful conditions may occur in patients with a hereditary history, such as diabetes mellitus, hypertension, hypothyroidism, migraine, degenerative neuropathies, and other neurological diseases, Hansen's disease and other infectious diseases, fibromyalgia or other functional diseases (35, 127).

Some studies emphasize that improving pain can improve subjective health, that is, the patient's perception of pain and well-being (128).

Based on this criterion, it was decided to use the VAS instruments as an outcome, which evaluates the intensity of pain, and GIC, which evaluates the doctor's and the patient's perception regarding the improvement in the treatment.

6.2 Characteristics of the accessed health record

This study showed the importance of a systematic and standardized medical record to assess and treat chronic pain and the possibility of using these complex and multivariate data in future research. Furthermore, only an organized record of multiple variables allows designing research strategies to assess possible interactions between variables and their impact on the mechanism and management of the disease.

Currently, most electronic records of pain-related data are not standardized. For example, diagnostic codes need to be clearly identified, and instruments for self-description and self-assessment of pain need to be more accurate in identifying pain-related issues (129, 130). To overcome the challenges related to the communication of chronic pain, a digital tool called "Electronic Interface for Ambulatory Care" (IEAA) was created by the Discipline of Neurosurgery of the Department of Neurology at FMUSP, which provides an extensive, consistent database of patients throughout their clinical care history. This tool was created in 2014 and formalized as a care instrument in 2016. The use of the IEAA allowed the experimental design of the present study with the analysis of multiple variables involved in pain with the outcome of its evolution and the search for predictors of its evolution.

6.3 Characteristics of pain

The analysis of data from this series emphasized the importance of self-reporting in the treatment and monitoring of the evolution of pain. These previous corroborating studies considered self-report as the gold pain assessment standard (131, 132).

Pain location: In the present case series, the pain was more frequently located on the right side of the body and manifested in multiple locations. The number of pain areas was the most accurate variable. The occurrence of acute or chronic pain elsewhere in the body was the most important clinical risk factor for developing chronic pain. It is described that the presence of more than one cause of chronic pain and pain of longer duration is associated with worse quality of life (133-136).

Type of pain: Neuropathic pain, detected using the DN-4 questionnaire, was present in 59.3%, according to the IASP criteria (7), as the main pain syndrome. Studies endorse these data, pointing out that neuropathic pain affects approximately 14% of the world population (137, 138). In this sample, it was also observed that nociceptive pain was present as the primary pain in approximately 2% of the evaluated patients. A systematic review study analyzed that the prevalence of nociceptive pain in Brazil is approximately 16% and is Brazil's most reported pain mechanism. Nociplastic pain was present in approximately 11% of patients, but it is important to emphasize that patients from the specific outpatient clinic for fibromyalgia were not evaluated. The prevalence of nociplastic pain in Brazil is around 12%, with 5% for fibromyalgia and 2% for chronic migraine (2, 139).

Pain intensity: Chronic pain intensity ranged from moderate to severe, with pain reported by more than 45% of patients with pain in Brazil (140, 141). However, studies indicate a lack of uniformity in measurement instruments and the use of “weak”, “moderate” and “strong” terminologies to classify pain intensity (140, 141).

Pain management: Most patients (47.5%) evaluated in the present series used the tricyclic antidepressant amitriptyline as an analgesic, which acts in the blockade and reuptake of serotonin and norepinephrine neurotransmitters in the pain suppressor system (142). Another widely used drug was gabapentin (61%), an anticonvulsant drug indicated for the pain of postherpetic neuralgia and painful peripheral diabetic neuropathy, but with limited evidence for treating other types of pain (143).

Chronic pain management focuses on rehabilitation and improving quality of life, not necessarily on the cure. The economic burden of treating chronic pain is significant. In addition to drug treatment, the costs of rehabilitation and reintegration of patients are added. Delayed diagnosis and poor pain management can result in more suffering, increased costs, and worsening impairments and disabilities (13, 144). Satisfactory pain management depends on a comprehensive assessment of the biological etiology of pain in conjunction with the patient's specific psychosocial and behavioral dimensions.

6.4 Application of Artificial Intelligence (AI) and Machine Learning (ML) in data analysis

The growing availability of “big data” enables new lines of research in chronic pain while at the same time implying the incorporation of new techniques for data mining and improving knowledge (145).

Interpreting a large amount of data and its relationship with the disease is a great challenge when predicting outcomes that can make the most effective therapy for the patient (146).

The analysis of longitudinal data, which are measures observed over time in many subjects, are generally multivariate and quite unbalanced; that is, they are unevenly arranged over the period studied. Longitudinal data analysis is traditionally performed using simple statistical methods. However, longitudinal data analysis derived from patient records may violate this assumption, as the observations correlate for the same patient but are independent. In addition, you often see different results tracked repeatedly at various intervals and/or with varying frequencies. Therefore, classical statistical models may not be applicable to analyze longitudinal data of this magnitude (147).

This research aimed to overcome the limits of basic statistics, which is a science that evaluates the behavior of data to identify patterns or correlations, that is, averages, standard deviations, variance, covariance, a correlation between data points, univariate and multivariate (148), which is limited in identifying “responders” and “non-responders” considering only individuals in the intervention group, that is, labeling an individual who received a certain active treatment as a “responder” (or not) because it is not known what would happen to that person if he belonged to the comparator (or placebo) group. To correctly infer whether or not a given participant responded to a given treatment, it is necessary to know what would happen if a key event (treatment) occurred and it did not, which is not possible in the real world (14).

The idea of using AI was precisely to find the best predictors, which often need to be more readily identifiable during the query. A supervised ML algorithm was used to analyze the data collected to identify these response predictors. It is a predictive and non-intuitive analysis technique that evaluates many data instead of dividing individuals into binary response categories, as usually used in basic statistics (14). A similar study uses this technique to assess pain predictors (149).

The development and popularization of machine learning techniques and the improvement of diagnostic instruments based on these technologies have revolutionized decision-making by health professionals (13). However, although many

studies have relied on ML to diagnose chronic pain, little has been attributed to its treatment (150-152).

Research on ML approaches to treat, rehabilitate and self-manage chronic pain is needed. Many articles use AI in self-care, aiming to create mechanisms so that the patient manages his pain using, for example, applications and virtual reality. Using these tools in patient care generates a wide range of health data, both within and outside of clinical settings, that can potentially be used in future BF studies to support research on chronic pain (153-156).

ML techniques require fewer assumptions about the data compared to simple statistics. ML focuses on mathematics and algorithms, uses large volumes and varieties of data to make predictions, and is concerned with solving problems, which implies predictions and pattern recognition. It is an excellent attribute for conducting longitudinal studies with extended follow-up and multiple measures (147).

Based on understanding data, ML is increasingly used to help health professionals and patients. Renowned for its ability to find complex relationships and identify critical patterns in datasets, ML offers new opportunities to use health and healthcare-related information effectively (13). This approach is very attractive as the potential clinical gains are significant. For example, even a small increase in the probability of response to treatment in a particular individual can be dramatic. However, attempts to achieve clinical gains through personalized medicine often must catch up to the original expectation (14).

The approach used was ensemble learning, a little explored technique because it evaluates explanatory modeling and predictive modeling and uses a combination of models to generate predictions in search of the best performance (157, 158). The SHAP (102) has recently emerged as an explanatory method of models, which is used to obtain the prediction explanation from the generated model. Model interpretability is always a challenge, especially when using the ensemble method, one of the techniques chosen in this study, due to the complexity of the data (159). The model explanation is more concerned with providing insight into the contribution of each feature from the model output.

There are several ML algorithms; it is necessary to try them in each presented problem to identify which obtains better performance because there is no algorithm that is superior to the other. The XGBoost (160) and Random Forest (97) algorithms

are not explanatory. The model was tested to avoid malfunction and validate explainability (149).

We evaluated the joint learning approach to predict the evolution of pain relief in patients with unknown chronic pain conditions because, despite the existence of several guidelines and recommendations for its treatment, up to 40% of patients with chronic pain may remain symptomatic despite the best medical treatment. Precisely defining the best therapy for a patient is still a challenge.

Research using AI tools has revealed that improved data increases accuracy performance. As the number of consultations considered increased, an increase in the AUC was observed with our proposed approach. On the other hand, the performance of the model reached a threshold of improvement; that is, when accumulating raw data from several sequential queries, it did not reflect significant changes, but there was a decrease in the performance of the model due to the large number of features added in each additional visit. The study showed that the first consultation was an essential event in the history of pain, influencing adherence to treatment and involvement with long-term self-management, corroborating previous studies (65).

The accuracy performance can be justified as traditional machine learning approaches are generally not adapted to handle longitudinal data.

139 variables were identified in the generated models, including factors related to and associated with the pain phenotype, such as pain etiology, morbidities, and factors related to treatment. The variables were listed in order of importance in the figures (Figure 16, Figure 17, Figure 18, and Figure 19). Many variables that were used did not show significant influence on the model. Hierarchical classifiers allowed for classifying features in the top-down direction of the decision tree. In all labels, the characteristics that appeared as the most important were: pain intensity, McGill questionnaire score, and GIC. The classifiers trained to make these assignments were complex, so the analysis of the models was better together than the evaluation of each separately. Interestingly, the model selected the best predictors without bias (or prior knowledge of predictors that were by default considered most important) (149). Detailed discussions about the strengths and limitations of these methods will be the scope of future studies.

The improved performance has generated a significantly reduced set of features. This remarkably reduced subset produced side benefits such as improving the explainability of predictions. The explanations of the models were used as meta-

features that functioned as a memory of previous queries. This method strongly indicated that the resource weights contain more decisive information than the resources in the previous queries. The approach that used the model explanations meta-feature with a diversification set of explanations achieved an AUC of 0.975.

Finally, the study demonstrated that models generated by the GIC label patient version had the best AUC when crossed with each other. To group these predictive variables, a questionnaire with 12 questions was generated, with the objective of being applied in the first consultation of the patient with pain at the HCFMUSP pain clinic, for a better prediction of pain evolution.

The acceptance of the use of AI in the clinical environment is slow, mainly due to the fact that the introduction of the use of new technologies is related to the increase in available tools that will be used during the consultation, which can be seen as difficulty and complexity in work by the doctor. However, implementing these technologies in the long term will facilitate and simplify some costly activities. In addition, using these tools will be just one more source of information to help improve diagnostic accuracy; the final judgment will remain with the physician (149).

It is expected that new analyses will be carried out using AI on query data due to the enormous amount of information produced and stored by modern technology. Current ML algorithms provide tools to help the multidisciplinary team make decisions based on these data relationships.

There are limitations to this study. First, this study is a single-center study. The development of the project presented a computational limitation. The proposal demanded a considerable time to learn a model, but with some optimizations, it was possible to reduce it. There is an obstacle to calculating the SHAP value, as not all supervised learning algorithms allow this calculation efficiently. There is an independent version of the algorithm, but its computational complexity is high, and it was impossible to apply it in our present study.

CONCLUSION

7 CONCLUSION

- Consistent systematized recording of evolutionary data related to individuals with pain, especially with chronic pain, obtained by the IEAA, was essential for the application of AI tools in ML algorithms.
- The ensemble algorithm was the most efficient in the ensemble analysis of the models.
- The application of Random Forest and XGBoost algorithms were the ones with the best performance in dealing with many variables.
- A good prediction was obtained using data from the first consultation.
- The use of model explanations meta-features allowed to reach an AUC of 0.975.
- The experimental design with ML allowed the generation of a 12-question questionnaire with the best predictors of pain evolution.
- ML analyzes are not intuitive; therefore, the selected variables can be considered unimportant in relation to evaluations in other care settings, for example, in the office, but this factor is considered positive.
- Machine learning has a high level of complexity, so you should not focus on understanding what is behind it but on the result.
- Chronic pain is a public health problem, and new strategies are needed to facilitate and streamline safe and effective care.

REFERENCES

REFERENCES

1. Ciampi de Andrade D. Introduction for special issue on pain in developing countries. Pain Rep. 2019;4(6):e800.
2. Aguiar DP, Souza CPDQ, Barbosa WJM, Santos-Júnior FFU, Oliveira ASD. Prevalence of chronic pain in Brazil: systematic review. Brazilian Journal Of Pain. 2021.
3. Teixeira MJ, Figueirão JB, Yeng LTY, de Andrade DC. DOR MANUAL PARA O CLÍNICO. 2 º ed. Brasil: EDITORA ATHENEU LTDA; 2018. 582 p.
4. Sá KN, Moreira L, Baptista AF, Yeng LT, Teixeira MJ, Galhardoni R, et al. Prevalence of chronic pain in developing countries: systematic review and meta-analysis. Pain Rep. 2019;4(6):e779.
5. Raja SN, Carr DB, Cohen M, Finnerup NB, Flor H, Gibson S, et al. The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises. Pain. 2020;161(9):1976-82.
6. Mills S, Torrance N, Smith BH. Identification and Management of Chronic Pain in Primary Care: a Review. Current Psychiatry Reports. 2016;18(2).
7. Treede RD, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, 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(1):19-27.
8. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020;396(10258):1204-22.
9. Pombo N, Araújo P, Viana J. Knowledge discovery in clinical decision support systems for pain management: a systematic review. Artif Intell Med. 2014;60(1):1-11.
10. Gaskin DJ, Richard P. The economic costs of pain in the United States. J Pain. 2012;13(8):715-24.
11. Anekar AA, Cascella M. Who Analgesic Ladder. Itália: StatPearls; 2022.
12. Maltoni M, Scarpi E, Modonesi C, Passardi A, Calpona S, Turriziani A, et al. A validation study of the WHO analgesic ladder: a two-step vs three-step strategy. Support Care Cancer. 2005;13(11):888-94.
13. Hepburn L-A, Jaatun EAA, editors. Making Sense of Data : Translating complexity for improved chronic pain management. 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom); 2018 2018-09-01: IEEE.
14. Norbury A, Seymour B. Response heterogeneity: Challenges for personalised medicine and big data approaches in psychiatry and chronic pain. F1000Res. 2018;7:55.
15. Tucker KL. The debate on elder abuse for undertreated pain. Pain Med. 2004;5(2):214-7; discussion 8-28.
16. Glajchen M. Chronic pain: treatment barriers and strategies for clinical practice. J Am Board Fam Pract. 2001;14(3):211-8.
17. Fishman SM, Gallagher RM, Carr DB, Sullivan LW. The case for pain medicine. Pain Med. 2004;5(3):281-6.
18. Lippe PM, Brock C, David J, Crossno R, Gitlow S. The First National Pain Medicine Summit--final summary report. Pain Med. 2010;11(10):1447-68.
19. Peiró AM. Pharmacogenetics in Pain Treatment. Adv Pharmacol. 2018;83:247-73.

20. Fernandez JM, Cenador MBG, Manuel López Millan J, Méndez JAJ, Ledesma MJS. Use of Information and Communication Technologies in Clinical Practice Related to the Treatment of Pain. Influence on the Professional Activity and the Doctor-Patient Relationship. J Med Syst. 2017;41(5):77.
21. Lötsch J, Ultsch A. Machine learning in pain research. Pain. 2018;159(4):623-30.
22. Cafezeiro I, Costa LCD, Kubrusly RDS. Ciência da Computação, Ciência da Informação, Sistemas de Informação:uma reflexão sobre o papel da informação e da interdisciplinaridade na configuração das tecnologias e das ciências. Perspectivas em Ciência da Informação. 2016;21(3):111-33.
23. Maldonado JMSDV, Marques AB, Cruz A. Telemedicine: challenges to dissemination in Brazil. Cadernos de Saúde Pública. 2016;32(suppl 2).
24. Lobo LC. Inteligência Artificial e Medicina. Revista Brasileira de Educação Médica. 2017;41(2):185-93.
25. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. NPJ Digit Med. 2018;1:5.
26. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine learning predictions to help anesthesiologists prevent hypoxemia during surgery. The preprint server for biology; 2017. Contract No.: doi.org/10.1101/206540.
27. Jenssen MDK, Bakkevoll PA, Ngo PD, Budrionis A, Fagerlund AJ, Tayefi M, et al. Machine Learning in Chronic Pain Research: A Scoping Review. Applied Sciences. 2021;11(7):3205.
28. Bevers K, Watts L, Kishino ND, Gatchel RJ. The Biopsychosocial model of the assessment, prevention, and treatment of chronic pain. US Neurology. 2016;12(2):7.
29. Cohen M, Quintner J, van Rysewyk S. Reconsidering the International Association for the Study of Pain definition of pain. Pain Rep. 2018;3(2):e634.
30. Walk D, Poliak-Tunis M. Chronic Pain Management: An Overview of Taxonomy, Conditions Commonly Encountered, and Assessment. Med Clin North Am. 2016;100(1):1-16.
31. MA A, PS S. Management of chronic pain. Lancet. 1999.
32. Hylands-White N, Duarte RV, Raphael JH. An overview of treatment approaches for chronic pain management. Rheumatol Int. 2017;37(1):29-42.
33. Ana MP. Chapter Ten - Pharmacogenetics in Pain Treatment. In: Kim Brøsen and Per D, editor. Pharmacogenetics. 83: Academic Press; 2018. p. 247-73.
34. Albe-Fessard D, Berkley KJ, Kruger L, Ralston HJ, 3rd, Willis WD, Jr. Diencephalic mechanisms of pain sensation. Brain Res. 1985;356(3):217-96.
35. Mills SEE, Nicolson KP, Smith BH. Chronic pain: a review of its epidemiology and associated factors in population-based studies. British Journal of Anaesthesia. 2019;123(2):e273-e83.
36. Agoston AM, Sieberg CB. Nonpharmacologic Treatment of Pain. Semin Pediatr Neurol. 2016;23(3):220-3.
37. Mann EG, Johnson A, Vandenkerkhof EG. Frequency and characteristics of healthcare visits associated with chronic pain: results from a population-based Canadian study. Canadian Journal of Anesthesia/Journal canadien d'anesthésie. 2016;63(4):411-41.
38. Kay C, Wozniak E, Bernstein J. Utilization of Health Care Services and Ambulatory Resources Associated with Chronic Noncancer Pain. Pain Med. 2017;18(7):1236-46.

39. Williams ACC, Craig KD. Updating the definition of pain. Pain. 2016;157(11):2420-3.
40. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain. 2008;9(2):105-21.
41. Institute of Medicine (US) Committee on Advancing Pain Research Cr, and, Education. Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research. 2011.
42. Macfarlane GJ, Barnish MS, Jones GT. Persons with chronic widespread pain experience excess mortality: longitudinal results from UK Biobank and meta-analysis. Ann Rheum Dis. 2017;76(11):1815-22.
43. Bouhassira D. Neuropathic pain: Definition, assessment and epidemiology. Rev Neurol (Paris). 2019;175(1-2):16-25.
44. Fitzcharles MA, Cohen SP, Clauw DJ, Littlejohn G, Usui C, Häuser W. Nociplastic pain: towards an understanding of prevalent pain conditions . *The Lancet*. 2021;397:2098-110.
45. Santos JG, Brito JO, de Andrade DC, Kaziyama VM, Ferreira KA, Souza I, et al. Translation to Portuguese and validation of the Douleur Neuropathique 4 questionnaire. J Pain. 2010;11(5):484-90.
46. Treede RD, Jensen TS, Campbell JN, Cruccu G, Dostrovsky JO, Griffin JW, et al. Neuropathic pain: redefinition and a grading system for clinical and research purposes. Neurology. 2008;70(18):1630-5.
47. Bennett MI, Smith BH, Torrance N, Lee AJ. Can pain can be more or less neuropathic? Comparison of symptom assessment tools with ratings of certainty by clinicians. Pain. 2006;122(3):289-94.
48. Kosek E, Cohen M, Baron R, Gebhart GF, Mico JA, Rice ASC, et al. Do we need a third mechanistic descriptor for chronic pain states? Pain. 2016;157(7):1382-6.
49. Jensen TS, Gebhart GF. New pain terminology: a work in progress. Pain. 2008;140(3):399-400.
50. Zorba Paster R. Chronic pain management issues in the primary care setting and the utility of long-acting opioids. Expert Opin Pharmacother. 2010;11(11):1823-33.
51. Navani A, Li G. Chronic pain challenge: A statistical machine-learning method for chronic pain assessment. Journal on Recent Advances in Pain. 2016;2(3):82-6.
52. Tracey I, Woolf CJ, Andrews NA. Composite Pain Biomarker Signatures for Objective Assessment and Effective Treatment. Neuron. 2019;101(5):783-800.
53. Rejula V, Anitha J, Belfin RV, Peter JD. Chronic Pain Treatment and Digital Health Era-An Opinion. Front Public Health. 2021;9:779328.
54. Dale R, Stacey B. Multimodal Treatment of Chronic Pain. Med Clin North Am. 2016;100(1):55-64.
55. JE M, NL B. Assessment of knowledge about cancer pain management by physicians in training. J Pain Symptom Manage. 1997:21-8.
56. Mędrzycka-Dąbrowska W, Dąbrowski S, Basiński A. Problems and Barriers in Ensuring Effective Acute and Post-Operative Pain Management--an International Perspective. Adv Clin Exp Med. 2015;24(5):905-10.
57. Knab JH, Wallace MS, Wagner RL, Tsoukatos J, Weinger MB. The use of a computer-based decision support system facilitates primary care physicians' management of chronic pain. Anesth Analg. 2001;93(3):712-20.
58. Leão Ferreira KA, Bastos TR, Andrade DC, Silva AM, Appolinario JC, Teixeira MJ, et al. Prevalence of chronic pain in a metropolitan area of a developing country: a population-based study. Arq Neuropsiquiatr. 2016;74(12):990-8.

59. Ventafridda V, Saita L, Ripamonti C, De Conno F. WHO guidelines for the use of analgesics in cancer pain. Int J Tissue React. 1985;7(1):93-6.
60. Karp JF, Shega JW, Morone NE, Weiner DK. Advances in understanding the mechanisms and management of persistent pain in older adults. Br J Anaesth. 2008;101(1):111-20.
61. Yang J, Bauer BA, Wahner-Roedler DL, Chon TY, Xiao L. The Modified WHO Analgesic Ladder: Is It Appropriate for Chronic Non-Cancer Pain? . Journal of Pain Research. 2020;Volume 13:411-7.
62. Mathews M. Multimodal treatment of pain. Neurosurg Clin N Am. 2014;25(4):803-8.
63. Protocolo Clínico e Diretrizes Terapêuticas da Dor Crônica . PORTARIA Nº 1083. , (02 DE OUTUBRO DE 2012. R etificada em 27 de novembro de 2015. E revoga a Portaria n°859/SAS/MS de 04 de novembro de 2000.).
64. Davis KD, Aghaeepour N, Ahn AH, Angst MS, Borsook D, Brenton A, et al. Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: challenges and opportunities. Nature Reviews Neurology. 2020;16(7):381-400.
65. White K, Lee J, Williams AC. Are patients' and doctors' accounts of the first specialist consultation for chronic back pain in agreement? Journal of Pain Research. 2016;Volume 9:1109-20.
66. RESOLUÇÃO CFM nº 1.638/2002. Define prontuário médico e torna obrigatória a criação da Comissão de Revisão de Prontuários nas instituições de saúde., (09 de agosto de 2002.).
67. Donabedian A. The quality of care. How can it be assessed? JAMA. 1988;260(12):1743-8.
68. Scochi MJ. Indicadores da qualidade dos registros e da assistência ambulatorial em Maringá, (estado do Paraná, Brasil), 1991: um exercício de avaliação. Cadernos de Saúde Pública. 1994;10(3):356-67.
69. Modesto MG, Moreira EC, Almeida-Filho ND. Reforma sanitária e informação em saúde: avaliação dos registros médicos em um distrito sanitário de Salvador, Bahia. Cadernos de Saúde Pública. 1992;8(1):62-8.
70. Jannati A, Sadeghi V, Imani A, Saadati M. Effective coverage as a new approach to health system performance assessment: a scoping review. BMC Health Services Research. 2018;18(1).
71. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. BMJ. 2003;326(7398):1070.
72. Heymann AD, Hoch I, Valinsky L, Shalev V, Silber H, Kokia E. Mandatory computer field for blood pressure measurement improves screening. Fam Pract. 2005;22(2):168-9.
73. Chiavegatto Filho ADP. Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. Epidemiologia e Serviços de Saúde. 2015;24(2):325-32.
74. Clauw DJ, D'Arcy Y, Gebke K, Semel D, Pauer L, Jones KD. Normalizing fibromyalgia as a chronic illness. Postgrad Med. 2018;130(1):9-18.
75. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. 2017;4(4):CD001431.
76. Brennan S, Joanne EWhitty, PaulaBuchan, HeatherGreen, SallyBosch, MarijeFiander, MichelleStrobel, Natalie ADonoghue, EmmaPage, Matthew JYost, JenniferGonçalves-Bradley, Daniela C. Continuous quality improvement: effects on

professional practice and healthcare outcomes in primary care. Cochrane Database of Systematic Reviews Journal. 2017.
77. Bruce P, Bruce A. Practical Statistics for Data Scientists: O'Reilly Media, Inc; 2017.
78. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J. 2019;6(2):94-8.
79. Iriart JAB. Precision medicine/personalized medicine: a critical analysis of movements in the transformation of biomedicine in the early 21st century. Cad Saude Publica. 2019;35(3):e00153118.
80. Ho D, Teo G. Digital Medicine – The New Frontier for AI in Healthcare. Advanced Therapeutics. 2020;3(4).
81. Alberdi A, Aztiria A, Basarab A. On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey. Artificial Intelligence in Medicine. 2016;71:1-29.
82. de Lima AG, Pinto GS. Indústria 4.0: um novo paradigma para a indústria. Revista Interface Tecnológica. 2019;16(2):299-311.
83. Sakurai R, Zuchi JD. AS REVOLUÇÕES INDUSTRIAIS ATÉ A INDUSTRIA 4.0. *Revista Interface Tecnológica*. 2018;15:480-91.
84. Mohd Javaid and Abid H. Industry 4.0 applications in medical field: A brief review. Current Medicine Research and Practice. 2019;9(3):102-9.
85. Tang CS, Veelenturf LP. The strategic role of logistics in the industry 4.0 era. Transportation Research Part E: Logistics and Transportation Review. 2019;129(C):1-11.
86. Turing AM. COMPUTING MACHINERY AND INTELLIGENCE. Computing machinery and intelligence. 1950;59:433-60.
87. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Family Med Prim Care. 2019;8(7):2328-31.
88. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism. 2017;69S:S36-S40.
89. Dagi TF, Barker FG, Glass J. Machine Learning and Artificial Intelligence in Neurosurgery: Status, Prospects, and Challenges. Neurosurgery. 2021;89(2):133-42.
90. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nature Methods. 2018;15(4):233-4.
91. Portaria GM Portaria MCTI n o 4.617, Instituto a Estratégia Brasileira de Inteligência Artificial e seus eixos temáticos., (6 de abril de 2021).
92. Grus J. Data Science do Zero: Mídia O'Reilly; 2016. 336 p.
93. Ismail Fawaz H, Forestier G, Weber J, Idoumghar L, Muller P-A. Deep learning for time series classification: a review. Data Mining and Knowledge Discovery. 2019;33(4):917-63.
94. Simeone O. A Brief Introduction to Machine Learning for Engineers. Foundations and Trends® in Signal Processing. 2018;12(3-4):200-431.
95. Mahesh B. Machine Learning Algorithms—A Review. International Journal of Science and Research. 2020;9:381-6.
96. Wengang Zhang and Chongzhi Wu and Haiyi Zhong and Yongqin Li and Lin W. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. Geoscience Frontiers. 2021;12(1):469-77.
97. BREIMAN L. Random Forests. Machine Learning [Internet]. 2001; 45:[5–32 pp.].
98. Kjersti Aas and Martin Jullum and Anders L. Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. Artificial Intelligence. 2021;298:103502.

99. Lundberg SMaLS-I. A Unified Approach to Interpreting Model Predictions. In: Garnett IGaUVLaSBaHWaRFaSVaR, editor.2017.
100. Lundberg S, Lee S-I. An unexpected unity among methods for interpreting model predictions. 29th Conference on Neural Information Processing Systems (NIPS 2016); Barcelona, Spain2016.
101. Ribeiro MT, Singh S, Guestrin C. Anchors: High-Precision Model-Agnostic Explanations. Proceedings of the AAAI Conference on Artificial Intelligence. 2018;32(1).
102. Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions. Annual Conf on Neural Information Processing Systems2017. p. 4768–77.
103. Faes L, Liu X, Wagner SK, Fu DJ, Balaskas K, Sim DA, et al. A clinician's guide to artificial intelligence: how to critically appraise machine learning studies. Translational vision science & technology. 2020;9(2):7-.
104. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. npj Digital Medicine. 2020;3(1).
105. Ferreira KASL, de Andrade DC, Teixeira MJ. Development and validation of a Brazilian version of the short-form McGill pain questionnaire (SF-MPQ). Pain Manag Nurs. 2013;14(4):210-9.
106. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. Pain. 1975;1(3):277-99.
107. Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain. 2001;94(2):149-58.
108. Price DD, Patel R, Robinson ME, Staud R. Characteristics of electronic visual analogue and numerical scales for ratings of experimental pain in healthy subjects and fibromyalgia patients. Pain. 2008;140(1):158-66.
109. Bouhassira D, Attal N, Alchaar H, Boureau F, Brochet B, Bruxelle J, et al. Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic questionnaire (DN4). Pain. 2005;114(1-2):29-36.
110. Lima MSD, Soares BGDO, Paoliello G, Vieira RM, Martins CM, Mota Neto JID, et al. The Portuguese version of the Clinical Global Impression - Schizophrenia Scale: validation study. Revista Brasileira de Psiquiatria. 2007;29(3):246-9.
111. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont). 2007;4(7):28-37.
112. COSTA ABd. Ensemble learning diversificando as explicações: prevendo a evolução do alívio da dor. Belo Horizonte, MG: Universidade Federal de Minas Gerais; 2021.
113. Costa ABD, Moreira LI, Andrade DCD, Veloso A, Ziviani N. Predicting the Evolution of Pain Relief: Ensemble Learning by Diversifying Model Explanations. ACM Trans Comput Healthcare. 2021;2(4).
114. Hayar MASP, Salimene ACDM, Karsch UM, Imamura M. Aging and chronic pain: a study of women with fibromyalgia. Acta Fisiátrica. 2014;21(3).
115. Santana AN, Cifre I, de Santana CN, Montoya P. Using Deep Learning and Resting-State fMRI to Classify Chronic Pain Conditions. Front Neurosci. 2019;13:1313.
116. Dorner TE. Pain and chronic pain epidemiology. Wiener klinische Wochenschrift. 2018;130(1-2):1-3.
117. Goldberg DS, Mcgee SJ. Pain as a global public health priority. BMC Public Health. 2011;11(1):770.

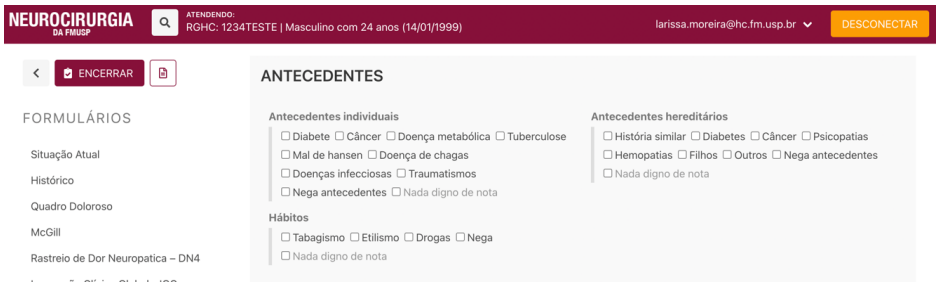
118. Ferreira KDS, Speciali JG. Epidemiology of chronic pain in the office of a pain specialist neurologist. Arquivos de Neuro-Psiquiatria. 2015;73(7):582-5.
119. Greenspan JD, Craft RM, Leresche L, Arendt-Nielsen L, Berkley KJ, Fillingim RB, et al. Studying sex and gender differences in pain and analgesia: A consensus report. Pain. 2007;132(Supplement 1):S26-S45.
120. Campbell CM, Edwards RR. Ethnic differences in pain and pain management. Pain Management. 2012;2(3):219-30.
121. Janevic MR, Mclaughlin SJ, Heapy AA, Thacker C, Piette JD. Racial and Socioeconomic Disparities in Disabling Chronic Pain: Findings From the Health and Retirement Study. The Journal of Pain. 2017;18(12):1459-67.
122. Poleshuck EL, Green CR. Socioeconomic disadvantage and pain. Pain. 2008;136(3):235-8.
123. Maly A, Vallerand AH. Neighborhood, Socioeconomic, and Racial Influence on Chronic Pain. Pain Manag Nurs. 2018;19(1):14-22.
124. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. The Lancet. 2012;380(9836):37-43.
125. Guthrie B, Payne K, Alderson P, Mcmurdo MET, Mercer SW. Adapting clinical guidelines to take account of multimorbidity. BMJ. 2012;345(oct04 1):e6341-e.
126. Bruehl S, Olsen RB, Tronstad C, Sevre K, Burns JW, Schirmer H, et al. Chronic pain-related changes in cardiovascular regulation and impact on comorbid hypertension in a general population: the Tromsø study. Pain. 2018;159(1):119-27.
127. Jackson T, Wang Y, Fan H. Self-efficacy and chronic pain outcomes: a meta-analytic review. J Pain. 2014;15(8):800-14.
128. Reyes-Gibby CC, Aday L, Cleeland C. Impact of pain on self-rated health in the community-dwelling older adults. Pain. 2002;95(1-2):75-82.
129. Fodeh SJ, Finch D, Bouayad L, Luther SL, Ling H, Kerns RD, et al. Classifying clinical notes with pain assessment using machine learning. Medical and Biological Engineering and Computing. 2018;56(7):1285-92.
130. Dulhare UN, Ahmad K, Ahmad KAB. Machine Learning e Big Data: Conceitos, Algoritmos, Ferramentas e Aplicações. 1ª edição ed: Wiley-Scrivener; 2020.
131. Apkarian VA, Hashmi JA, Baliki MN. Pain and the brain: Specificity and plasticity of the brain in clinical chronic pain. Pain. 2011;152(3):S49-S64.
132. Mäntyselkä PT. Chronic Pain and Poor Self-rated Health. JAMA. 2003;290(18):2435.
133. Carvalho RCD, Maglioni CB, Machado GB, Araújo JED, Silva JRTD, Silva MLD. Prevalence and characteristics of chronic pain in Brazil: a national internet-based survey study. Brazilian Journal Of Pain. 2018;1(4).
134. Elliott AM, Smith BH, Hannaford PC, Smith WC, Chambers WA. The course of chronic pain in the community: results of a 4-year follow-up study. Pain. 2002;99(1-2):299-307.
135. Pagé MG, Fortier M, Ware M, Choinière M. As if one pain problem was not enough: prevalence and patterns of coexisting chronic pain conditions and their impact on treatment outcomes. Journal of Pain Research. 2018;Volume 11:237-54.
136. Cruz-Almeida Y, King CD, Goodin BR, Sibille KT, Glover TL, Riley JL, et al. Psychological Profiles and Pain Characteristics of Older Adults With Knee Osteoarthritis. Arthritis Care & Research. 2013;65(11):1786-94.
137. Colloca L, Ludman T, Bouhassira D, Baron R, Dickenson AH, Yarnitsky D, et al. Neuropathic pain. Nat Rev Dis Primers. 2017;3:17002.

138. Finnerup NB, Haroutounian S, Kamerman P, Baron R, Bennett DLH, Bouhassira D, et al. Neuropathic pain: an updated grading system for research and clinical practice. Pain. 2016;157(8):1599-606.
139. Szok D, Csáti A, Vécsei L, Tajti J. Chronic Migraine as a Primary Chronic Pain Syndrome and Recommended Prophylactic Therapeutic Options: A Literature Review. Life. 2023;13(3):665.
140. Dellaroza MSG, Pimenta CADM, Duarte YA, Lebrão ML. Dor crônica em idosos residentes em São Paulo, Brasil: prevalência, características e associação com capacidade funcional e mobilidade (Estudo SABE). Cadernos de Saúde Pública. 2013;29(2):325-34.
141. Watkins EA, Wollan PC, Melton LJ, 3rd, Yawn BP. A population in pain: report from the Olmsted County health study. Pain Med. 2008;9(2):166-74.
142. Gillman PK. Tricyclic antidepressant pharmacology and therapeutic drug interactions updated. British Journal of Pharmacology. 2007;151(6):737-48.
143. Wiffen PJ, Derry S, Bell RF, Rice AS, Tölle TR, Phillips T, et al. Gabapentin for chronic neuropathic pain in adults. Cochrane Database of Systematic Reviews. 2017;2020(2).
144. Mäntyselkä PT, Kumpusalo EA, Ahonen RS, Takala JK. Direct and indirect costs of managing patients with musculoskeletal pain-challenge for health care. Eur J Pain. 2002;6(2):141-8.
145. Ultsch A, Kringel D, Kalso E, Mogil JS, Lötsch J. A data science approach to candidate gene selection of pain regarded as a process of learning and neural plasticity. Pain. 2016;157(12):2747-57.
146. Demchak B, Kreisberg JF, Fuxman Bass JI. Theory and Application of Network Biology Toward Precision Medicine. Journal of Molecular Biology. 2018;430(18):2873-4.
147. Mikael Lopes J, Barbosa O.G. Guedes M, Limeira Cavalcanti R, Gabriel De Souza C. Dados Longitudinais e o Viés da Correlação de Medidas: A Alternativa dos Modelos Mistos. Arquivos Brasileiros de Cardiologia. 2019.
148. Windish DM, Diener-West M. A clinician-educator's roadmap to choosing and interpreting statistical tests. J Gen Intern Med. 2006;21(6):656-60.
149. Miettinen T, Mäntyselkä P, Hagelberg N, Mustola S, Kalso E, Lötsch J. Machine learning suggests sleep as a core factor in chronic pain. Pain. 2021;162(1):109-23.
150. Boissoneault J, Sevel L, Letzen J, Robinson M, Staud R. Biomarkers for Musculoskeletal Pain Conditions: Use of Brain Imaging and Machine Learning. Current Rheumatology Reports. 2017;19(1).
151. Lannon E, Sanchez-Saez F, Bailey B, Hellman N, Kinney K, Williams A, et al. Predicting pain among female survivors of recent interpersonal violence: A proof-of-concept machine-learning approach. PLOS ONE. 2021;16(7):e0255277.
152. Robinson ME, O'Shea AM, Craggs JG, Price DD, Letzen JE, Staud R. Comparison of Machine Classification Algorithms for Fibromyalgia: Neuroimages Versus Self-Report. The Journal of Pain. 2015;16(5):472-7.
153. Le C, Deleat-Besson R, Turkestani NA, Cevidanes L, Bianchi J, Zhang W, et al. TMJOAI: An Artificial Web-Based Intelligence Tool for Early Diagnosis of the Temporomandibular Joint Osteoarthritis. Clin Image Based Proced Distrib Collab Learn Artif Intell Combat COVID 19 Secur Priv Preserv Mach Learn (2021). 2021;12969:78-87.
154. Jamison RN, Xu X, Wan L, Edwards RR, Ross EL. Determining Pain Catastrophizing From Daily Pain App Assessment Data: Role of Computer-Based Classification. J Pain. 2019;20(3):278-87.

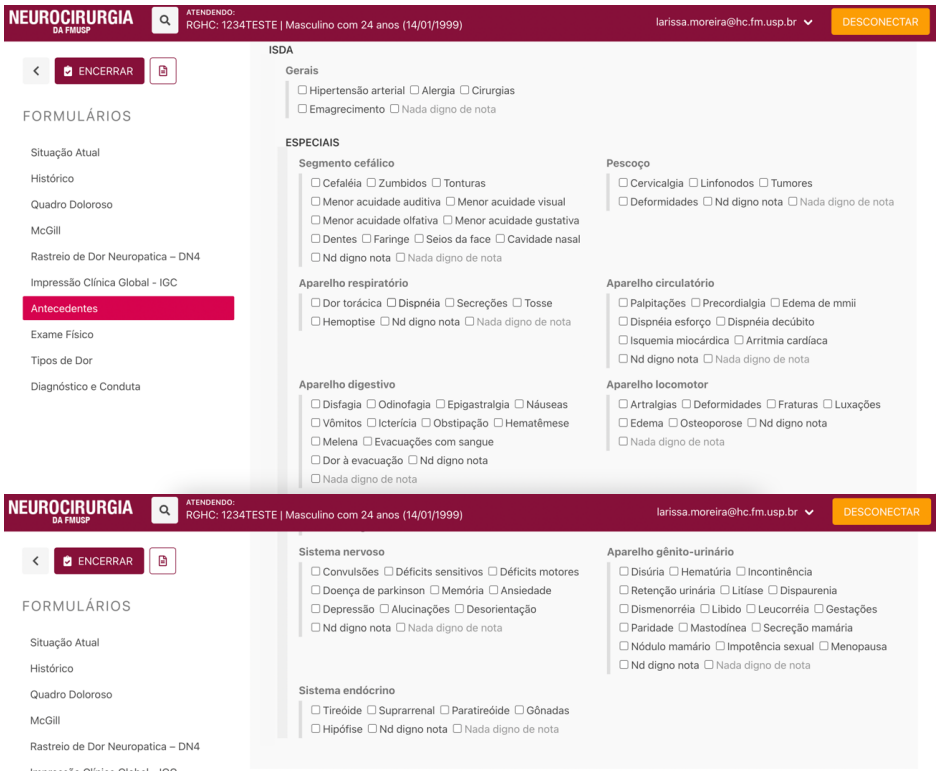
155. Zhang M, Zhu L, Lin SY, Herr K, Chi CL, Demir I, et al. Using artificial intelligence to improve pain assessment and pain management: a scoping review. J Am Med Inform Assoc. 2023;30(3):570-87.
156. Fritz RL, Wilson M, Dermody G, Schmitter-Edgecombe M, Cook DJ. Automated smart home assessment to support pain management: multiple methods analysis. Journal of Medical Internet Research. 2020;22(11).
157. Kuncheva LI, Whitaker CJ. Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. 2003;51(2):181-207.
158. Opitz DW. Feature Selection for Ensembles. Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh Innovative Applications of Artificial Intelligence Conference Innovative Applications of Artificial Intelligence. {USA}: American Association for Artificial Intelligence; 1999. p. 379–84 , numpages = 6.
159. Mamoshina P, Vieira A, Putin E, Zhavoronkov A. Applications of Deep Learning in Biomedicine. Mol Pharm. 2016;13(5):1445-54.
160. Chen TG, Carlos. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*2016. p. 785–94.

Appendix

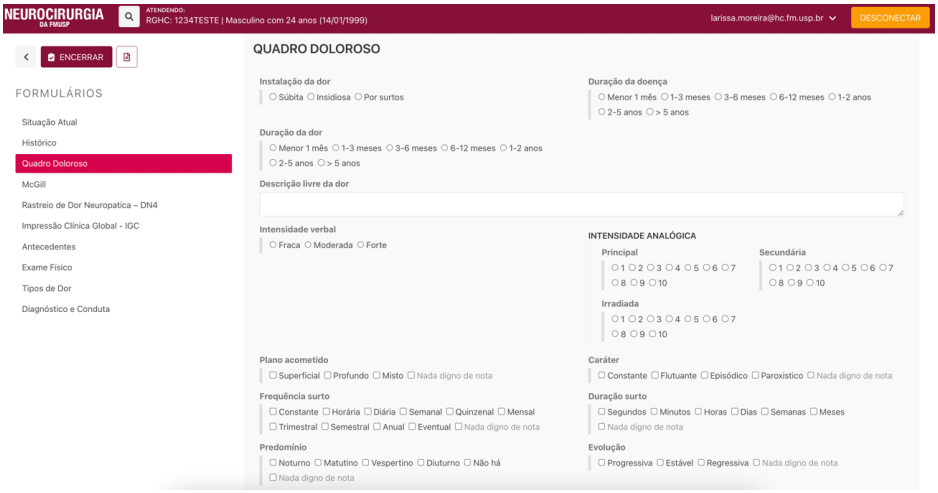
Appendix 1 - Patient history



Appendix 2 – ISDA Scale

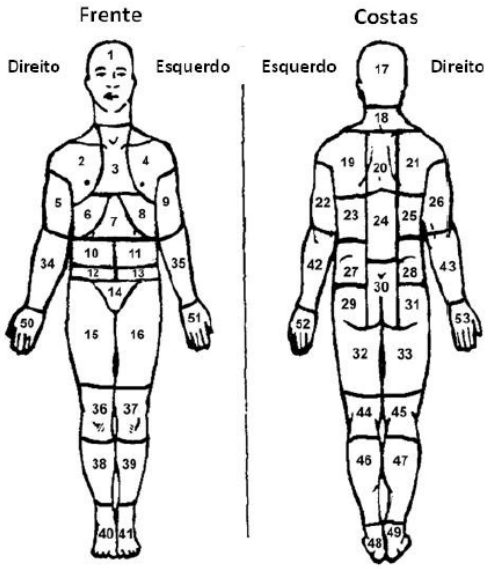


Appendix 3 - Pain assessment



Appendix 4 - The short-form McGill Pain Questionnaire

Data:			
Nº	**Dimensões da dor:**		
1	Dimensão Sensorial:		
		Ausente	Presente
	Latejante	( )	( )
	Pontada	( )	( )
	Choque	( )	( )
	Fina/agulhada	( )	( )
	Fisgada	( )	( )
	Queimação	( )	( )
	Espalha	( )	( )
	Dolorida	( )	( )
2	Dimensão Afetiva		
		Ausente	Presente
	Cansativa	( )	( )
	Enjoada	( )	( )
	Sufocante	( )	( )
	Apavorante	( )	( )
	Aborrecida	( )	( )
3	Dimensão Avaliativa		
		Ausente	Presente
	Que incomoda	( )	( )
	Insuportável	( )	( )
4	Questão 2: Marque sobre o diagrama, com um X, a (s) área (s) onde você sente dor e onde a dor é mais intensa:		



1- cabeça	17- cabeça (região	28- região lombar D	41- pé E
2- ombro D	posterior)	29- glúteo E	42- antebraço (posterior) E
3- tórax anterior	18- cervical (região	30- região coccígea	43- antebraço (posterior) D
4- ombro E	posterior)	31- glúteo D	44- região poplítea E
5- braço D	19- escapula E	32- coxa (posterior) E	45- região poplítea D
6- hipocôndrio D	20- tórax posterior	33- coxa posterior D	46- perna (posterior) E
7- região epigástrica	21- escapula D	34- antebraço D	47- perna (posterior) D
8- hipocôndrio E	22- braço (posterior) E	35- antebraço E	48- região calcânea E

Appendix 5 – Douleur Neuropathique 4 Questions (DN4)

Nº	Questões
1	Questão 1: A sua dor tem uma ou mais das seguintes características?
	Queimação: ( ) Sim ( ) Não
	Sensação de frio doloroso: ( ) Sim ( ) Não
	Choque elétrico: ( ) Sim ( ) Não
2	Questão 2: Há presença de um ou mais dos seguintes sintomas na mesma área da sua dor?
	Formigamento: ( ) Sim ( ) Não
	Alfinetada e agulhada: ( ) Sim ( ) Não
	Adormecimento: ( ) Sim ( ) Não
	Coceira: ( ) Sim ( ) Não
	Exame do paciente:
3	Questão 3: A dor está localizada numa área onde o exame físico pode revelar uma ou mais das seguintes características?
	Hipoestesia ao toque: ( ) Sim ( ) Não
	Hipoestesia a picada de agulha: ( ) Sim ( ) Não
4	Questão 4: Na área dolorosa a dor pode ser causada ou aumentada por:
	Escovação: ( ) Sim ( ) Não
	Escore total:

Appendix 6 – Global Impression Clinical scale (GIC)

Appendix 7

Predicting the Evolution of Pain Relief: Ensemble Learning by Diversifying Model Explanations

ANDERSON BESSA DA COSTA, Computer Science Department, Universidade Federal de Minas Gerais
LARISSA MOREIRA, Pain Center, Department of Neurology, Universidade de São Paulo
DANIEL CIAMPI DE ANDRADE, Pain Center, Instituto do Câncer do Estado de São Paulo
ADRIANO VELOSO, Computer Science Department, Universidade Federal de Minas Gerais
NIVIO ZIVIANI, Computer Science Department, Universidade Federal de Minas Gerais, and Kunumi

Modeling from data usually has two distinct facets: building sound explanatory models or creating powerful predictive models for a system or phenomenon. Most of recent literature does not exploit the relationship between explanation and prediction while learning models from data. Recent algorithms are not taking advantage of the fact that many phenomena are actually defined by diverse sub-populations and local structures, and thus there are many possible predictive models providing contrasting interpretations or competing explanations for the same phenomenon. In this article, we propose to explore a complementary link between explanation and prediction. Our main intuition is that models having their decisions explained by the same factors are likely to perform better predictions for data points within the same local structures. We evaluate our methodology to model the evolution of pain relief in patients suffering from chronic pain under usual guideline-based treatment. The ensembles generated using our framework are compared with all-in-one approaches of robust algorithms to high-dimensional data, such as Random Forests and XGBoost. Chronic pain can be primary or secondary to diseases. Its symptomatology can be classified as nociceptive, nociplastic, or neuropathic, and is generally associated with many different causal structures, challenging typical modeling methodologies. Our data includes 631 patients receiving pain treatment. We considered 338 features providing information about pain sensation, socioeconomic status, and prescribed treatments. Our goal is to predict, using data from the first consultation only, if the patient will be successful in treatment for chronic pain relief. As a result of this work, we were able to build ensembles that are able to consistently improve performance by up to 33% when compared to models trained using all the available features. We also obtained relevant gains in interpretability, with resulting ensembles using only 15% of the total number of features. We show we can effectively generate ensembles from competing explanations, promoting diversity in ensemble learning and leading to significant gains in accuracy by enforcing a stable scenario in which models that are dissimilar in terms of their predictions are also dissimilar in terms of their explanation factors.

CCS Concepts: • **Computing methodologies** → **Ensemble methods**; • **Applied computing** → *Health informatics;*

This work was partially funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and by projects ATMOSPHERE (Horizon 2020 grant No 777154), RNP-MCTIC (grant No 51119), and MASWeb (grant FAPEMIG/PRONEX APQ-01400-14), and by the authors individual grants from CNPq, FAPEMIG, and Kunumi.
Authors' addresses: A. B. da Costa and A. Veloso, Computer Science Department, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; emails: {anderson.bessa, adrianov}@dcc.ufmg.br; L. Moreira, Pain Center, Department of Neurology, Universidade de São Paulo, São Paulo, São Paulo, Brazil; email: larissaiulle@usp.br; D. C. de Andrade, Pain Center, Instituto do Câncer do Estado de São Paulo, São Paulo, São Paulo, Brazil; email: ciampi@usp.br; N. Ziviani, Computer Science Department, Universidade Federal de Minas Gerais, and Kunumi, Belo Horizonte, Minas Gerais, Brazil; email: nivio@dcc.ufmg.br.

2637-8051/2021/08-ART36 $15.00
https://doi.org/10.1145/3466781