International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 1, February 2025, pp. 894~907
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i1.pp894-907

Comparison of machine learning algorithms to identify and prevent low back injury

Christian Ovalle Paulino[1], Jorge Huamani Correa[2]
[1]Department of Engineering, Faculty of Engineering, Universidad Tecnológica del Perú, Lima, Perú
[2]Academic Centre for Materials and Nanotechnology, AGH University of Krakow, Kraków, Poland

Article Info

Article history:

Received Apr 4, 2024
Revised Aug 22, 2024
Accepted Sep 3, 2024

Keywords:

Algorithm comparison
Computational medicine
Lumbar injuries
Machine learning
Predictive models

ABSTRACT

With the advancement of technology, remote work and virtual classes have become increasingly common, leading to prolonged periods in front of computers and, consequently, to discomfort and even lower back pain. This study compares machine learning algorithms to identify and prevent low back pain, a common health problem. A predictive model for early diagnosis and prevention of these injuries was developed using datasets from open data repositories. Six machine learning models were used to train the data. Results showed that logistic regression was the most effective model, with performance curves of 70%, 90%, and 99%. Performance metrics indicated 86% accuracy, 85% recall, and 86% F1-score. Accuracy of 70%, recall of 71%, and F1-score of 63% reflect the robust ability of the model to address the problem. In addition, an intuitive interface was implemented using Gradio Software to improve data visualization.

Corresponding Author:

Christian Ovalle Paulino
Department of Engineering, Faculty of Engineering, Universidad Tecnológica del Perú
Lima, Perú
Email: dovalle@utp.edu.pe

1. INTRODUCTION

Low back disorders considerably strain public and occupational health, representing around 40% of all musculoskeletal disorders related to work globally. It is of concern that 75% of these injuries originate in everyday activities, such as lifting, leading to one in three workers worldwide facing low back problems, making them one of the leading causes of work absenteeism [1]–[3]. Moreover, it is projected that approximately 200 billion dollars are spent each year on treating lower back pain [4]. Consequently, due to the high cost and time required for diagnosis, the need for specialized knowledge in this area becomes evident [5].

In this context, the increasing incidence of malignant spinal abnormalities highlights the urgent need for early detection to preserve the quality of life [6]. Beyond the spinal degeneration associated with aging, which can cause acute or chronic low back pain and functional disability at all ages, there are several additional conditions, such as scoliosis and injuries resulting from improper posture, that exacerbate the problem. Studies show that maintaining poor posture while sitting for long periods can cause various health problems, including upper and lower back and neck discomfort. This results from uneven pressure distribution on the spine, potentially leading to bone injuries, sarcopenia, and impaired circulation. Therefore, it is essential to maintain proper sitting posture, especially for those engaged in long work or study sessions [7]–[13].

To tackle these problems, recent advancements in artificial intelligence (AI) have opened up new possibilities for diagnosing and managing lower back injuries. AI has demonstrated its usefulness in delivering accurate and understandable information to healthcare providers, enhancing its dependability across different contexts [13]–[16]. Specifically, machine learning (ML) has emerged as a promising approach to tackling the difficulties associated with diagnosing lumbar conditions [17], [18].

Additionally, machine learning holds the potential to transform medical practice by offering physicians precise and tailored information, which could help minimize medical errors and surpass the effectiveness of conventional methods [19]. Machine learning techniques are transforming the healthcare sector globally, equipping professionals with innovative tools that improve the quality and efficiency of medical care [20]. This highlights their positive impact on patient care [21].

Despite the significant advances in AI and machine learning, there are still substantial challenges to overcome. For instance, the occurrence of false positives in magnetic resonance imaging (MRI) analysis when using Bayes' theorem is a notable issue. This underscores the ongoing need for the development of methods that can enhance the accuracy and reliability of lumbar diagnostics, thereby improving healthcare [22]–[25].

It is crucial to delve into the pioneering research on the application of AI and ML in the management of lumbar injuries. For instance, a groundbreaking study [26] devised a predictive model using deep learning and ML techniques to forecast recovery outcomes following lumbar disc herniation, thereby aiding clinical decision-making. This research retrospectively examined clinical data from 470 patients and applied a range of algorithms, such as random forest (RF), extreme gradient boosting (XGBoost), support vector machine (SVM), decision tree (DT), K-nearest neighbor (KNN), logistic regression (LR), light gradient boosting machine (LGBM), and multilayer perceptron (MLP). The results revealed a low correlation between the features, as depicted in the correlation matrix heat map. Another study [27] crafted a machine-learning algorithm to evaluate the connection between lumbar disc height on radiographs and the presence of disc bulges or herniations. By analyzing data from 458 patients, they identified crucial factors linked to lumbar disc herniation (LDBH), including L4-5-disc height, age, and L1-2-disc height. A DT-based model was developed for clinical decision-making, achieving F1-score of 0.706, 0.778, 0.569, 0.729, and 0.706 for the least absolute shrinkage and selection operator (LASSO), multivariate adaptive regression splines (MARS), DT, RF, and XGBoost models, respectively, with the MARS model attaining the highest F1-score. In a separate study [28], the effectiveness of a transforaminal epidural steroid injection (TFESI) was evaluated in patients with lumbosacral radicular pain due to lumbar spinal stenosis (LSS), a less commonly studied condition. A convolutional neural network (CNN) was trained with data from 193 patients, achieving an area under the curve (AUC) of 0.920 and an accuracy of 87.2%, demonstrating the model's outstanding predictive capability. Similarly, another study by Haider *et al.* [29] introduced an advanced machine learning technique using bootstrapping and data balancing methods to identify low back pain. They employed a standard dataset containing 310 records and proposed the random forest gradient boosting XGBoost ensemble (RGXE) method, which combined RF, gradient boosting (GB), and XGBoost, surpassing previous methods with a remarkable accuracy of 0.99. A study [30] also created a medical test to help healthcare professionals choose and assign physical treatments for nonspecific low back pain patients. This study assessed several ML algorithms, including LR, DT, SVM, KNN, and GB. The findings revealed that all ML models achieved accuracies exceeding 80%, with SVM being the most precise, reaching an accuracy of over 90%. Considering these studies and their limitations, the current study was designed to address these issues and enhance the diagnosis and treatment of lumbar injuries.

The main objective of this article is to explore and analyze the emerging impact of machine learning in the identification and prevention of lumbar lesions. We will highlight how the proposed advanced methodologies, if implemented, can significantly improve diagnostic accuracy and optimize treatment protocols. This research not only contributes to the medical field but also provides a clear framework for future research and clinical developments. In addition to emphasizing the prevalence and impact of low back injuries, this study will identify specific areas that require improvement, such as reducing false positive diagnoses on MRI scans and optimizing strategies for the management of chronic low back pain. Given that ML has proven to be a highly effective tool in data analysis, it is crucial to thoroughly evaluate the various available algorithms to select the most appropriate one for low back injury diagnosis and treatment. The innovation brought by ML can accelerate the time to diagnosis and thus the initiation of treatment, which reduces the waiting time for patients and prevents their health from deteriorating [6], [15], [23].

The article is organized as follows: section 2 outlines the methodology, describing the approach used. Section 3 presents the results obtained. Section 4 provides a discussion that analyzes and interprets the findings. Finally, section 5 concludes with the study's conclusions.

2. METHOD

The project presented entails applied research at a predictive level, aiming to solve a problem through comprehensive, organized, and systematic application of acquired knowledge to find a solution. In this research, we have developed a comprehensive solution for diagnosing and treating lumbar lesions, opting for a pre-experimental design. This choice is justified by its capacity to identify and address potential technical and operational issues before impacting actual patients [9]. Additionally, the quantitative approach facilitates the iterative development of the solution, minimizing risks and ensuring its safety and reliability as it undergoes refinement. Thus, this design allows for adjustments and improvements based on test results and feedback [6].

The project is meticulously structured into four essential phases following the cross-industry standard process for data mining (CRISP-DM) methodology, the most widely used reference model for developing data mining projects [31]. This methodology is renowned for its structured and systematic data analysis and knowledge extraction approach, making it applicable to various projects [32]. The process commences with understanding the business and data, which is crucial for the company as it enhances the likelihood of success for their data mining endeavors [33]. Given the above, the four phases of the current research on a predictive model based on ML algorithms to identify and prevent lumbar injuries are detailed below.

2.1. Data pre-processing

Dataset pre-processing is not just an initial step but a crucial one in our research. At this stage, our main objective is to address missing and null values that could disrupt our predictions. The primary focus is cleaning the data and converting all feature data into numerical values, enabling the algorithm to operate effectively. Your role in this process is significant. To achieve this, we conducted rigorous pre-processing based on datasets obtained from Kaggle, extracting the most relevant values and addressing issues related to the lumbar region. One of the datasets, named "*column_3C_weka.csv*," referred to as "*Column*" henceforth, includes diagnostics related to the spine and is based on six values representing angles of the spine's essential parts. The other dataset, named "*Dataset_spine.csv*," hereafter referred to as "*Spine*" for brevity, contains data related to abnormal or normal spines and 12 variables representing different parts of the spine.

It is essential to note that the quality of both datasets significantly influences the ML model's effectiveness. Access to these online data sources was facilitated through an internet connection, allowing us to obtain relevant medical information. Our comprehensive pre-processing procedure involved meticulously applying Dummy coding and labeling for categorical variables based on meaningful datasets such as Kaggle.

Figures 1 and 2 depict the datasets *Column* and *Spine*, respectively. For the definition of the column code, we assigned the value 1 for a hernia and 0 for no hernia, as well as 1 for a normal condition and 0 for an abnormal condition. Additionally, 1 was assigned to indicate the presence of spondylolisthesis and 0 for its absence. Regarding the spine code, 1 signifies a normal condition, while 0 denotes an abnormal condition.

1	pelvic_incidence	pelvic_tilt	lumbar_lordosis_angle	sacral_slope	pelvic_radius	degree_spondylolisthesis	class
2	63.0278175	22.55258597	39.60911701	40.47523153	98.67291675	-0.254399986	Hernia
3	39.05695098	10.06099147	25.01537822	28.99595951	114.4054254	4.564258645	Hernia
4	68.83202098	22.21848205	50.09219357	46.61353893	105.9851355	-3.530317314	Hernia
5	69.29700807	24.65287791	44.31123813	44.64413017	101.8684951	11.21152344	Hernia
6	49.71285934	9.652074879	28.317406	40.06078446	108.1687249	7.918500615	Hernia
7	40.25019968	13.92190658	25.1249496	26.32829311	130.3278713	2.230651729	Hernia
8	53.43292815	15.86433612	37.16593387	37.56859203	120.5675233	5.988550702	Hernia
9	45.36675362	10.75561143	29.03834896	34.61114218	117.2700675	-10.67587083	Hernia
10	43.79019026	13.5337531	42.69081398	30.25643716	125.0028927	13.28901817	Hernia
11	36.68635286	5.010884121	41.9487509	31.67546874	84.24141517	0.664437117	Hernia
12	49.70660953	13.04097405	31.33450009	36.66563548	108.6482654	-7.825985755	Hernia
13	31.23238734	17.71581923	15.5	13.51656811	120.0553988	0.499751446	Hernia
14	48.91555137	19.96455616	40.26379358	28.95099521	119.321358	8.028894629	Hernia
15	53.5721702	20.46082824	33.1	33.11134196	110.9666978	7.044802938	Hernia
16	57.30022656	24.1888846	46.99999999	33.11134196	116.8065868	5.766946943	Hernia
17	44.31890674	12.53799164	36.098763	31.78091509	124.1158358	5.415825143	Hernia
18	63.83498162	20.36250706	54.55243367	43.47247456	112.3094915	-0.622526643	Hernia
19	31.27601184	3.14466948	32.56299592	28.13134236	129.0114183	3.623020073	Hernia
20	38.69791243	13.44474904	31	25.25316339	123.1592507	1.429185758	Hernia
21	41.72996308	12.25407408	30.12258646	29.475889	116.5857056	-1.244402488	Hernia
22	43.92283983	14.17795853	37.8325467	29.7448813	134.4610156	6.451647637	Hernia
23	54.91944259	21.06233245	42.19999999	33.85711014	125.2127163	2.432561437	Hernia
24	63.07361096	24.41380271	53.99999999	38.65980825	106.4243295	15.77969683	Hernia
25	45.54078988	13.06959759	30.29832059	32.47119229	117.9808303	-4.987129618	Hernia
26	36.12568347	22.75875277	29	13.3669307	115.5771163	-3.237562489	Hernia
27	54.12492019	26.65048856	35.32974693	27.47443163	121.447011	1.571204816	Hernia
28	26.14792141	10.75945357	14	15.38846783	125.2032956	-10.09310817	Hernia
29	43.58096394	16.5088837	46.99999999	27.07208024	109.271634	8.992815727	Hernia
30	44.5510115	21.93114655	26.78591597	22.61986495	111.0729197	2.652320636	Hernia
31	66.87921138	24.89199889	49.27859673	41.9872125	113.4770183	-2.005891748	Hernia
32	50.81926781	15.40221253	42.52893886	35.41705528	112.192804	10.86956554	Hernia

Figure 1. Dataset *Column*: The first six columns represent crucial spine measures, while the last column represents the target for each patient

1	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10	Col11	Col12	Class_att
2	63.0278175	22.552586	39.609117	40.4752315	98.6729168	-0.2544	0.74450346	12.5661	14.5386	15.30468	-28.658501	43.5123	Abnormal
3	39.056951	10.0609915	25.0153782	28.9959595	114.405425	4.56425865	0.41518568	12.8874	17.5323	16.78486	-25.530607	16.1102	Abnormal
4	68.832021	22.2184821	50.0921936	46.6135389	105.985136	-3.5303173	0.47488916	26.8343	17.4861	16.65897	-29.031888	19.2221	Abnormal
5	69.2970081	24.6528779	44.3112381	44.6441302	101.868495	11.2115234	0.36934526	23.5603	12.7074	11.42447	-30.470246	18.8329	Abnormal
6	49.7128593	9.65207488	28.317406	40.0607845	108.168725	7.91850062	0.54336047	35.494	15.9546	8.87237	-16.378376	24.9171	Abnormal
7	40.2501997	13.9219066	25.1249496	26.3282931	130.327871	2.23065173	0.78999286	29.323	12.0036	10.40462	-1.512209	9.6548	Abnormal
8	53.4329282	15.8643361	37.1659339	37.568592	120.567523	5.9885507	0.19891957	13.8514	10.7146	11.37832	-20.510434	25.9477	Abnormal
9	45.3667536	10.7556114	29.038349	34.6111422	117.270068	-10.675871	0.13197256	28.8165	7.7676	7.60961	-25.111459	26.3543	Abnormal
10	43.7901903	13.5337531	42.690814	30.2564372	125.002893	13.2890182	0.19040763	22.7085	11.4234	10.59188	-20.020075	40.0276	Abnormal
11	36.6863529	5.01088412	41.9487509	31.6754687	84.2414152	0.66443712	0.36770014	26.2011	8.738	14.91416	-1.702097	21.432	Abnormal
12	49.7066095	13.0409741	31.3345001	36.6656355	108.648265	-7.8259858	0.6880095	31.3502	16.5097	15.17645	-0.502127	18.3437	Abnormal
13	31.2323873	17.7158192	15.5	13.5165681	120.055399	0.49975145	0.60834276	21.4356	9.2589	14.76412	-21.724559	36.4449	Abnormal
14	48.9155514	19.9645562	40.2637936	28.9509952	119.321358	8.02889463	0.13947817	32.7916	7.2049	8.61882	-1.215542	27.3713	Abnormal
15	53.5721702	20.4608282	33.1	33.111342	110.966698	7.04480294	0.08193099	15.058	12.8127	12.00109	-1.734117	15.6205	Abnormal
16	57.3002266	24.1888846	47	33.111342	116.806587	5.76694694	0.41672151	16.5158	18.6222	8.51898	-33.441303	13.2498	Abnormal
17	44.3189067	12.5379916	36.098763	31.7809151	124.115836	5.41582514	0.66404088	9.5021	19.1756	7.25707	-32.893911	19.5695	Abnormal
18	63.8349816	20.3625071	54.5524337	43.4724746	112.309492	-0.6225266	0.56067537	10.769	16.8116	11.41344	2.676002	17.3859	Abnormal
19	31.2760118	3.14466948	32.5629959	28.1313424	129.011418	3.62302007	0.53448124	31.1641	18.6089	8.4402	4.482424	24.6513	Abnormal
20	38.6979124	13.444749	31	25.2531634	123.159251	1.42918576	0.30658054	28.3015	17.9575	14.75417	-14.252676	24.9361	Abnormal
21	41.7299631	12.2540741	30.1225865	29.475889	116.585706	-1.2444025	0.46852593	28.5598	12.4637	14.1961	-20.392538	33.0265	Abnormal
22	43.9228398	14.1779585	37.8325467	29.7448813	134.461016	6.45164764	0.28044621	12.4719	16.8965	10.32658	-4.986668	22.4667	Abnormal
23	54.9194426	21.0623325	42.2	33.8571101	125.212716	2.43256144	0.17524457	23.0791	14.2195	14.14196	3.780394	24.9278	Abnormal
24	63.073611	24.4138027	54	38.6598083	106.42433	15.7796968	0.66638801	11.9696	17.6891	7.63771	-14.183602	44.2338	Abnormal

Figure 2. Dataset *Spine*: The first eleven columns represent variables representing different parts of the spine, while the last column classifies the spine as abnormal or normal

We meticulously focused on crucial spine measures, including pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral tilt, pelvic radius, degree of spondylolisthesis, pelvic slope, direct tilt, thoracic slope, cervical tilt, sacral angle, and scoliotic slope. This meticulous approach to data preparation instills confidence in the thoroughness of our research. It lays the foundation for a detailed and accurate analysis of low back problems in the subsequent steps.

2.2. Classification methods

Six machine learning models were selected for consideration in our research:

- Logistic regression (LR): Evaluates the connection between the categorical dependent variable (to be predicted) and one or more independent variables (which influence the former) by estimating probabilities using a logistic function [34].
- Support vector machine (SVM): It is an algorithm suitable for anticipating urban logistics demand. It offers specific benefits in solving problems with limited datasets and nonlinear functions and identifying patterns in multidimensional spaces [35].
- K-nearest neighbor (KNN): It is a classification method that determines a data point's class by examining its closest neighbors' classes. These neighbors are identified based on their distance from the data point, commonly calculated using Euclidean distance [36].
- Convolutional neural network (CNN): It is a specialized deep neural network comprising input, hidden, and output layers. The hidden layers are particularly notable for incorporating convolutional layers specifically designed to perform convolution operations [37].
- Decision tree (DT): It is a simple and easily understandable approach in machine learning that is applied across various disciplines. It is practical, requires less data, and provides interpretability. Decision tree generates models organized in the structure of a tree and can be used for both regression and classification problems [38].
- Extreme gradient boosting (XGBoost): It is based on DT and employs a serial training process with datasets to combine weaker predictors into stronger predictors. It iteratively optimizes the objective function until reaching its lowest value, at which point the training process stops [36].

This choice was guided by prior research, creating a solid and evidence-backed strategy [4], [8]. Furthermore, we incorporated 12 critical variables into the model development, including pelvic incidence, pelvic tilt, and the lumbar lordosis angle. We followed the standard practice of dividing the data into 80% for training and 20% for testing, which refined our methodology and facilitated a more thorough analysis of lumbar issues.

Each algorithm was tailored with specific settings. For instance, logistic regression was configured with default settings, support vector machine with a linear kernel, K-nearest neighbor with Euclidean distance and 10 neighbors in the *Column* dataset and 20 neighbors in the *Spine* dataset, convolutional neural network with 3 layers of 32, 64, and 64 neurons per layer, and 100 epochs. Decision tree and XGBoost were configured using default settings.

2.3. Evaluation metrics

In this section, we define the appropriate evaluation metrics for assessing the performance of each algorithm in diagnosing and treating low back problems. The selected metrics provide a comprehensive view of each model's effectiveness, ensuring it meets the client's standards. We evaluate the relevant metrics using a confusion matrix [39]. The metrics chosen for evaluation are as follows:

– Precision (PR)

Precision measures the performance of an ML algorithm by determining the ratio of correctly predicted positive cases to the total predicted positive cases. Where TP represents actual positive values, and FP represents false positive values. It is calculated as (1):

$$Precision\ (PR) = \frac{TP}{TP + FP} \tag{1}$$

– Recall (RC)

Recall indicates how well the model identified all confirmed cases in a given class. This is particularly relevant in the medical context where accurate detection of all lumbar lesions is critical for effective treatment. It is calculated as (2):

$$Recall\ (RC) = \frac{TP}{FN+TP} \tag{2}$$

– F1-Score

The F1-score is a key metric for assessing lumbar injury diagnostic models. It represents the harmonic mean of precision and recall, providing a more balanced performance measure [40]. It is computed as (3):

$$F1 - Score = 2\frac{RC.PR}{RC + PR} \tag{3}$$

Furthermore, the area under the curve (AUC) is utilized to evaluate the discrimination ability of a binary classification model. A higher AUC value indicates a higher accuracy of the ML algorithm. AUC is commonly used in the context of the ROC curve, which represents the classification prediction results in a two-dimensional plane [41].

These metrics are typically analyzed using statistical approaches. Although creating rules for predicting results may involve rough-set technology. When information about granules is provided in the prediction process, an interpretation of the “fuzziness” based on rough sets can be achieved [42].

2.4. Web environment integration

At this project stage, we focus on integrating the most effective model into a web environment to ensure the development of a solution conducive to adopting these technologies in online medical practice [23]. We have implemented an interface with Gradio, an open-source Python library, to achieve this goal. Gradio plays a crucial role in implementing a model in an accessible web environment, significantly enhancing the speed and effectiveness of decision-making, particularly in health for disease control and prevention plans [43].

Gradio is built upon a web interface that facilitates interaction with an ML system using models and algorithms. It allows for inputting extensive data for early disease diagnosis [44]. Hence, Gradio is an excellent choice for integrating the selected model into a web environment, enabling the rapid creation of user interfaces for learning models.

This detailed approach, marked by thorough and precise evaluation, has been implemented in the project. It has offered a complete and accurate assessment of the model's performance in the multiclass classification task, confirming its effectiveness for diagnosing and treating lumbar lesions. These procedures are illustrated in the schematic diagram presented in Figure 3.

Thus, a process flow diagram illustrating the sequence of tasks and activities was considered, as depicted in Figure 4. The process initiates when the user enters the system and inputs the patient's data for evaluation. Subsequently, the system validates the accuracy of the entered data. In case of any inaccuracies, the system will flag an error. Following validation, the system processes the entered data through the trained model, providing the user with the results.

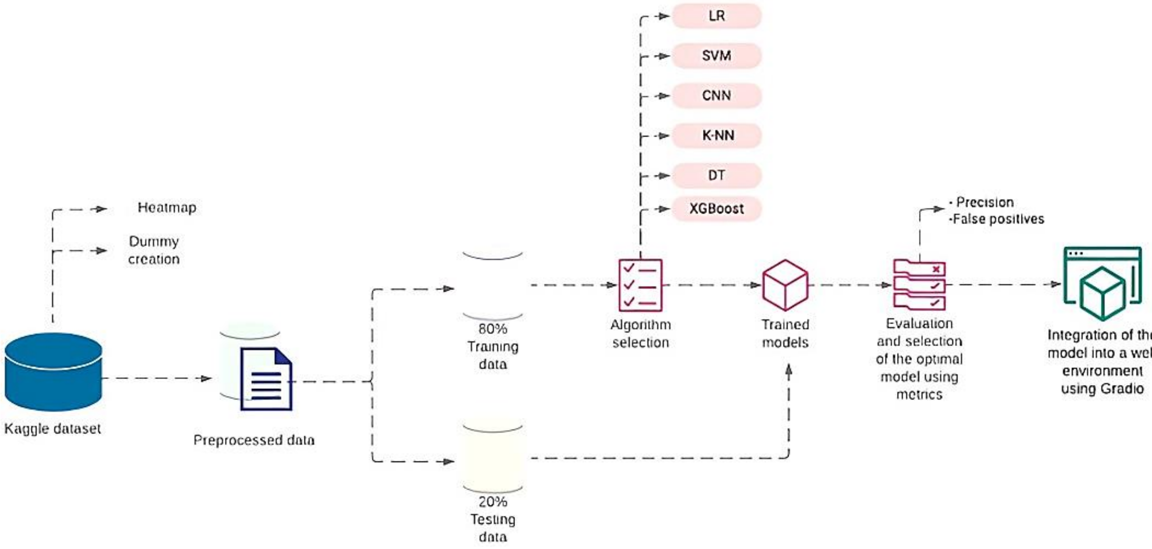


Figure 3. Schematic diagram of predictive model using in the training models

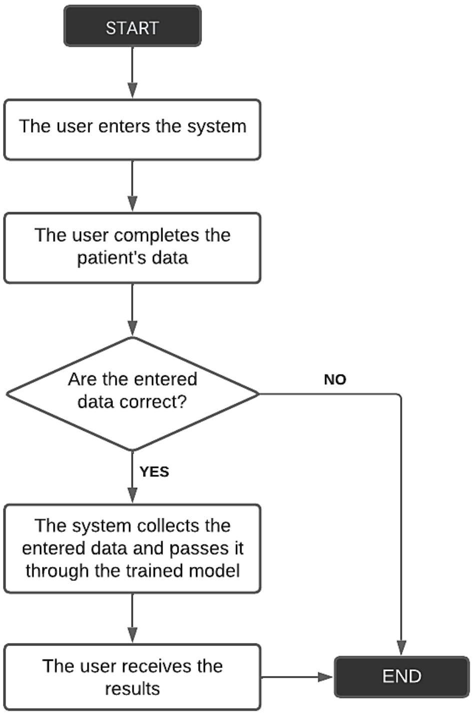


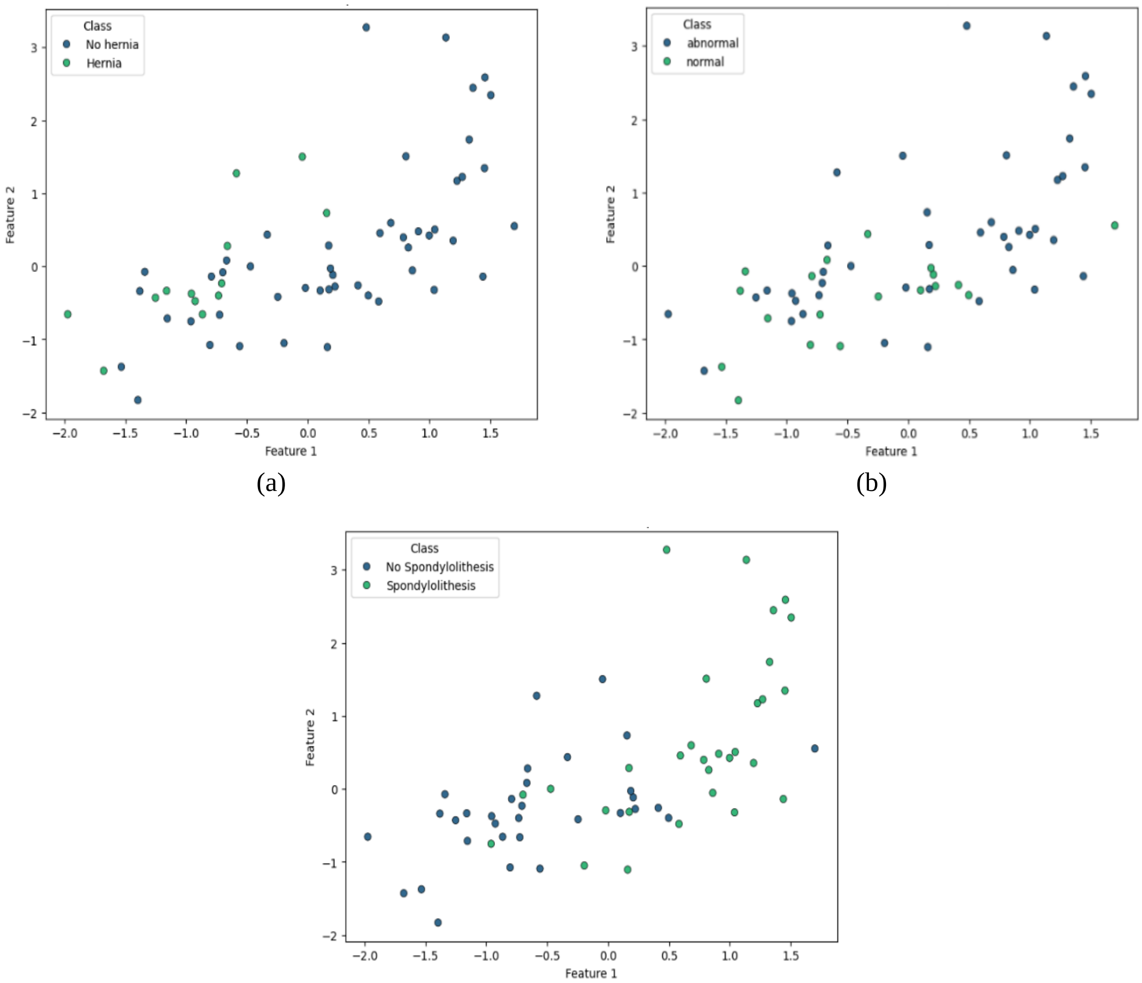
Figure 4. Process flowchart

3. RESULTS AND DISCUSSION

In this section, we present the results obtained to further explore the effectiveness of these algorithms in healthcare settings. These tools not only improve accuracy in identifying and preventing low back injuries but also have the potential to offer new perspectives in medical care. This inspiring potential provides a solid foundation for future research and clinical development.

3.1. Dispersion with data separation

As stated in section 2.2, we adopted the standard 80-20 ratio to partition the data. Subsequently, to obtain detailed insights into our variables and classes, Figures 5(a) to 5(c) depict the plotted *Column* datasets. It is observed that the *Column* datasets exhibit an ascending trend. Similarly, Figure 6 illustrates the data division of the *Spine* dataset, revealing a low correlation among its variables, a testament to the model's ability to handle complex data with confidence.



(a) (b)

(c)

Figure 5. Scatter plots of *Column* dataset of (a) hernia and no hernia, (b) normal and abnormal, and (c) spondylolisthesis and no spondylolisthesis

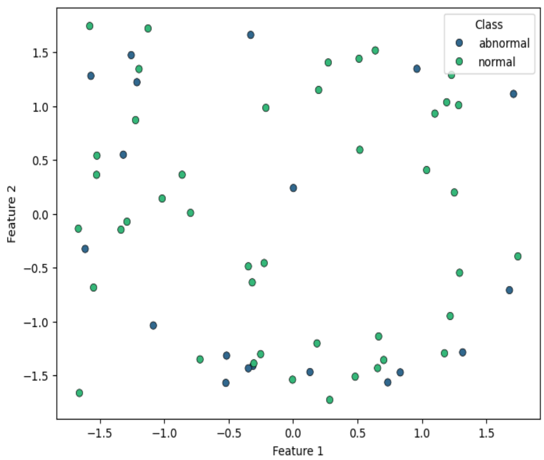
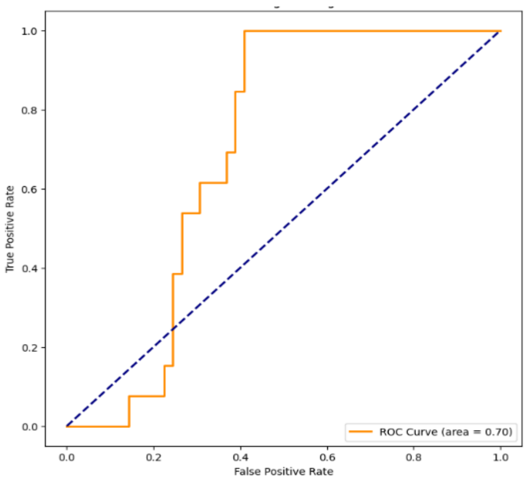


Figure 6. Scatter plot of the *Spine* dataset of normal and abnormal

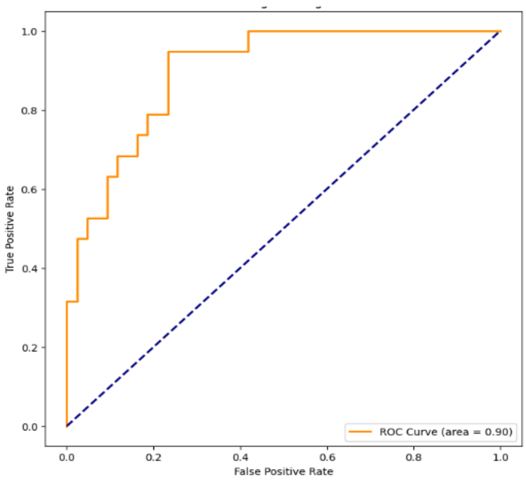
3.2. Model training

For the training of models on the *Column* dataset, the logistic regression model, trained with variables such as *pelvic_incidence*, *pelvic_tilt*, *lumbar_lordosis_angle*, *sacral_slope*, *pelvic_radius*, and *degree_spondylolisthesis*, provided a comprehensive diagnosis for each case, classifying the conditions as hernia, spondylolisthesis, or standard back. Figures 7(a) to 7(c) is presented below to visualize the training specifics of each model in the *Column* dataset. Conversely, within the *Spine* dataset, the LR model was trained with variables such as *pelvic_slope*, *Direct_tilt*, *thoracic_slope*, *cervical_tilt*, *sacrum_angle*, and *scoliosis_slope*, yielding a more generalized outcome by classifying the spine as "abnormal" or "normal." Figure 8 is provided below to offer a detailed insight into the input of each model within the *Spine* dataset.

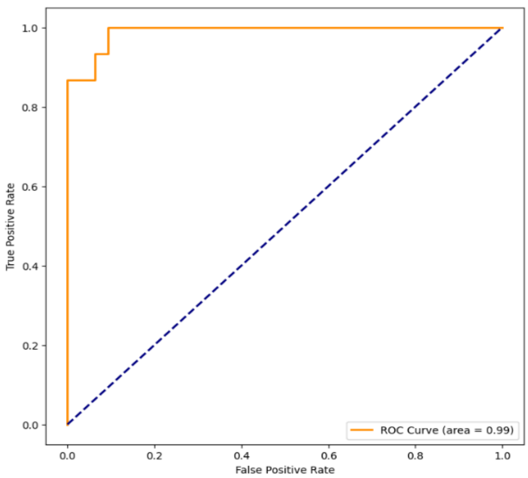
It is essential to highlight that the ROC curve was employed for training measurement, exhibiting satisfactory outcomes, particularly for LR in the *Column* dataset, with curves of 0.70 in Figure 7(a), 0.90 in curve Figure 7(b), and 0.99 in Figure 7(c). In contrast, a lower training score of 0.48 is observed for the *Spine* dataset, as depicted in Figure 8. However, other metrics will be utilized to assess its predictive capability adequately.



(a)



(b)



(c)

Figure 7. Training plots of *Column* dataset for (a) hernia and no hernia, (b) normal and abnormal, and (c) spondylolisthesis and no spondylolisthesis

Regarding the performance metrics, a comparative analysis of the performance of the 6 selected ML models is presented in Table 1, where they were evaluated using the selected metrics. Firstly, the metrics values for the *Column* dataset are presented, followed by those for the *Spine* dataset for each model; it is noteworthy that the order of the datasets remains consistent (refer to Table 1). Additionally, it can be observed that the LR model exhibited exceptional performance in both datasets. For the *Spine* dataset, it attained a precision of 86%, recall of 85%, and F1-score of 86%, indicative of its high accuracy in identifying positive and negative cases, providing a detailed diagnosis by categorizing the spine into specific conditions such as hernia, spondylolisthesis, or normal back. Furthermore, on the *Spine* dataset, the LR model achieved a precision of 70%, a recall of 71%, and an F1-score of 63%. Though still commendable, the model binarily classifies the spine as “abnormal” or “normal,” offering a more generalized diagnosis in this dataset. These findings underscore the LR model's adaptability to different datasets and the significant role of selecting pertinent variables for each diagnostic assessment context, highlighting the importance of your work in this field.

Figure 9 displays a heatmap illustrating the correlation between the variables in the dataset. This visualization aids in identifying patterns and trends within the data by showcasing their relationships. In this heatmap, red denotes a positive correlation, with more intense shades indicating a stronger positive correlation; blue signifies a negative correlation, with darker hues representing a stronger negative correlation. White indicates the absence of correlation.

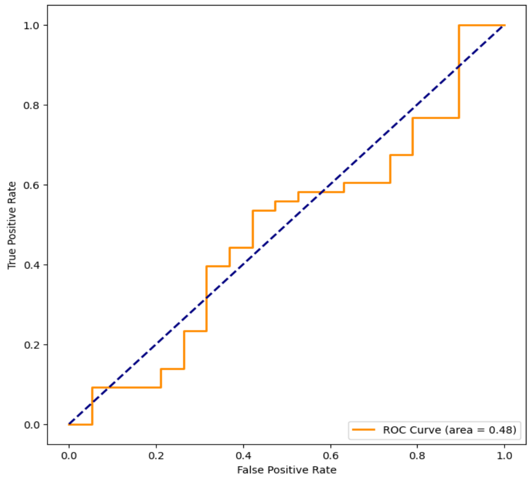


Figure 8. Training plot of the *Spine* dataset for normal and abnormal

Table 1. Training results of ML models

A	B	C	D
Models		Metrics	
	Precision	Recall	Fl-score
LR	0.86	0.85	0.86
	0.70	0.71	0.63
SVM	0.85	0.85	0.85
	0.79	0.69	0.57
CNN	0.80	0.81	0.80
	0.62	0.53	0.56
K-NN	0.72	0.71	0.71
	0.60	0.65	0.53
DT	0.66	0.66	0.66
	0.56	0.45	0.49
XGBoost	0.80	0.81	0.80
	0.67	0.69	0.68

3.3. Implementation in Gradio software

The system's graphical interface offers a user-friendly experience, featuring clearly defined data entry sections for inputting relevant information about low back injuries. Users can provide details such as pain location, intensity, associated symptoms, and other essential factors. Additionally, there is a dedicated section for visualizing diagnostic results, including the type of low back injury, treatment recommendations, and rehabilitation strategies. This tool is designed to cater to both experienced and inexperienced users, ensuring ease of use and interaction. It facilitates fluid navigation between different sections and functionalities, presenting accurate and personalized results and recommendations regarding lower back injuries.

Implemented in a robust and easily accessible web environment, the solution ensures that users, including those lacking technical expertise, can effectively utilize it to diagnose and treat low back injuries. Leveraging Gradio further enhances the tool's utility, enabling professionals from diverse disciplines and clinical settings to benefit from this innovative solution in managing spine health-related cases. Figure 10 depicts the system's graphical user interface, illustrating its accessibility to a broad audience. This interface empowers healthcare professionals and other users with intuitive access to diagnose and treat low back injuries.

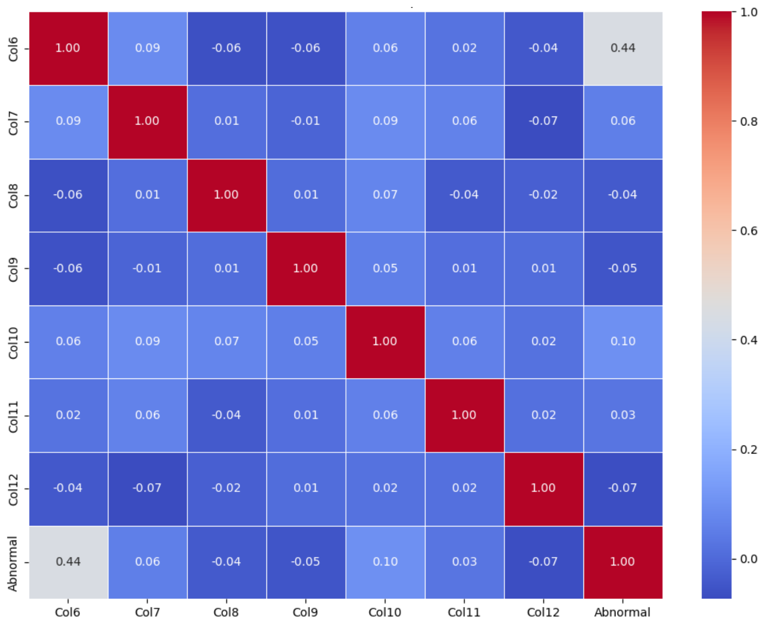


Figure 9. Heat correlation map of the variables taking into account in the dataset *Spine*

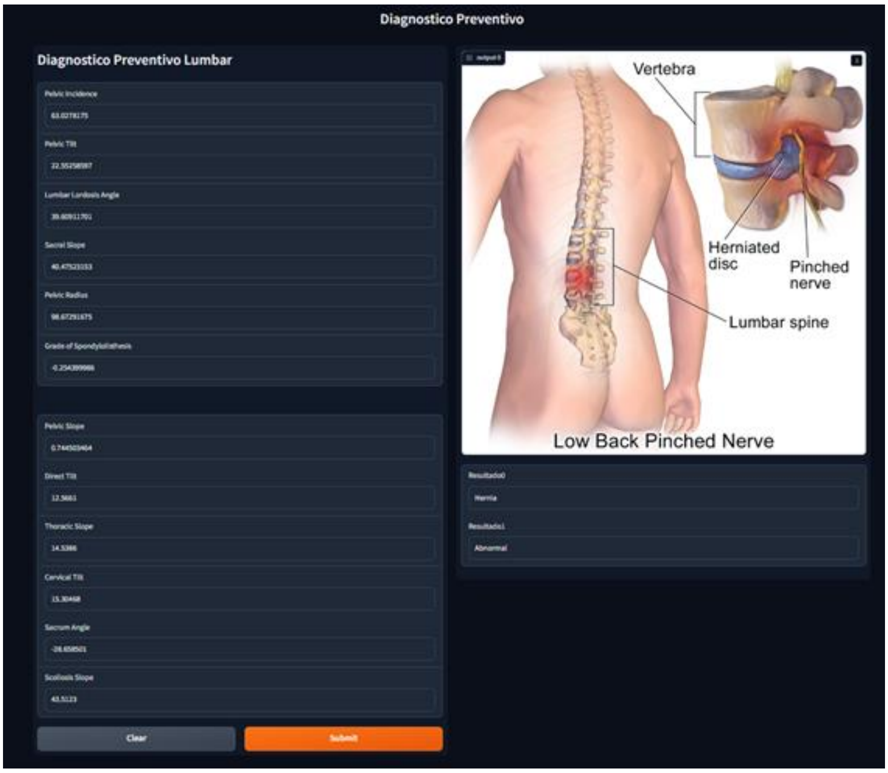


Figure 10. System graphical user interface

4. DISCUSSION

In this section, we review the results of our research and compare them with findings from related studies. Various ML algorithms have been employed to identify and prevent low back injuries, focusing on model comparison and their integration into an accessible web environment. In the present study, it was found that in the *Column* dataset, the LR model showed an accuracy of 86%, recall of 85%, and an F1-score of 86%, and in the *Spine* dataset, an accuracy of 70%, recall of 71%, and an F1-score of 63% was observed.

These results align with earlier research that applied machine learning algorithms to lumbar issues. For example, the case study [26] developed a predictive model for recovery after lumbar disc herniation using deep learning and machine learning techniques. This study employed algorithms such as random forest, LR, SVM, KNN, and XGBoost. Among these, the logistic regression algorithm achieved an accuracy of 0.735, a recall of 0.755, and an F1-score of 0.745. A previous study [27] designed a machine learning algorithm to establish the relationship between lumbar disc height on radiographs and the presence of disc bulges or herniated discs, achieving F1-scores of 0.706, 0.778, 0.569, 0.729, and 0.706 for the LASSO, MARS, DT, RF, and XGBoost models, respectively, with the MARS model showing the highest F1-score another study [28] investigated the outcomes of TFESI for patients with lumbosacral radicular pain caused by LSS, utilizing the CNN algorithm. This study achieved an accuracy of 0.733, a recall of 0.917, and an F1-score of 0.815. A previous study [29] proposed an advanced machine learning method using data balancing and bootstrapping techniques to detect low back pain. They used algorithms such as RF, DT, SVM, KNN, LR, and XGBoost, highlighting this with an accuracy of 0.910, a recall of 1. 000, and an F1-score of 0.950. A previous study [30] created a medical test to assist healthcare professionals in selecting and assigning physical treatments for patients with non-specific low back pain (LBP). This study employed algorithms including LR, DT, SVM, KNN, and XGBoost. Among these, the SVM algorithm was the most accurate, achieving an accuracy of 0.957 and a recall of 0.953. For further details are shown in Table 2.

Table 2. Comparison of the results obtained with similar research

Reference	Algorithm	Precision	Recall	Fl-score
[26]	MLP	0.750	0.755	0.753
	RF	0.783	0.629	0.698
	LR	0.735	0.755	0.745
	SVM	0.809	0.916	0.859
	KNN	0.869	0.510	0.643
	XGBoost	0.787	0.958	0.864
[27]	LASSO regression	0.600	0.857	0.706
	MARS	0.676	0.924	0.778
	DT	0.547	0.592	0.569
	RF	0.675	0.794	0.729
	XGBoost	0.600	0.857	0.706
[28]	CNN	0.733	0.917	0.815
	Stochastic gradient descent (SGD)	0.958	0.852	0.902
[29]	RF	0.880	0.950	0.910
	DT	0.820	0.820	0.820
	SVM	0.830	0.840	0.860
	KNN	0.760	0.970	0.860
	LR	0.810	0.900	0.850
	XGBoost	0.910	1.000	0.950
	RGXE	0.910	1.000	0.950
[30]	LR	0.867	0.867	-
	DT	0.814	0.813	-
	SVM	0.957	0.953	-
	KNN	0.813	0.813	-
	XGBoost	0.820	0.820	-

Comparing the results obtained in this research with previous findings underscores the relevance and efficacy of ML models in diagnosing and treating low back problems. The variability in performance metrics, primarily observed across different datasets, emphasizes the importance of tailoring the models to the specificities of each medical context. The versatility of the LR model underscores its potential to address diverse challenges in the field of low back health.

5. CONCLUSION

The use of ML algorithms to diagnose and treat low back injuries has yielded valuable and promising insights. From the introduction, in which the significant burden of low back disorders in occupational and public health was emphasized, to the detailed methodology addressing challenges through developing a comprehensive solution, each phase of this study has contributed to advancing the understanding and application of AI in healthcare. The findings of this study underscore the potential of ML for enhancing diagnostic accuracy, optimizing treatment strategies, and ultimately improving patient outcomes in managing musculoskeletal conditions.

In the solution generation phase, six algorithms, LR, SVM, CNN, K-NN, DT, and XGBoost, were implemented and evaluated, showcasing the LR model's versatility in addressing various data contexts. The results were presented at descriptive and inferential levels, considering metrics such as precision, recall, and F1-score and employing hypothesis tests to support the statistical validity of conclusions. LR demonstrated effectiveness by providing a detailed diagnosis in the *Column* dataset and a more generalized categorization in the *Spine* dataset. The ROC curve analysis further validated the model's accuracy, reaching 86% in the *Column* dataset and 70% in the *Spine* dataset. The model's effectiveness and adaptability to different contexts were underscored, emphasizing the importance of selecting relevant variables for each diagnostic evaluation, especially in identifying and preventing lumbar injuries. Moreover, the successful implementation of the Gradio software, offering an intuitive graphical interface for healthcare professionals, highlights the practical potential and accessibility of the developed solution.

While significant progress has been made, challenges and opportunities for improvement remain. Enriching the dataset and exploring more complex variables are identified as future directions. Furthermore, implementation in clinical settings and ongoing evaluation of the model's effectiveness are crucial for ensuring its long-term utility. Future work could expand research towards integrating additional data and exploring advanced ML techniques to enhance the model's accuracy and robustness, thereby continuously improving medical care, particularly in low back injury.

In conclusion, this study comprehensively evaluated the machine learning algorithms used to diagnose lumbar injuries and demonstrated their effectiveness and practical potential. While acknowledging areas for improvement, such as improving accuracy and reducing false positives, it is hoped that this research will serve as a solid foundation for future advances in leveraging artificial intelligence to identify and prevent low back injuries, thereby improving patient care and musculoskeletal health outcomes.

ACKNOWLEDGEMENTS

We are deeply grateful to the Universidad Tecnológica of Peru for its commitment to providing us with the essential tools and management support, which has significantly enriched the quality of our research.

REFERENCES

[1] E. S. Matijevich, P. Volgyesi, and K. E. Zelik, "A promising wearable solution for the practical and accurate monitoring of low back loading in manual material handling," *Sensors*, vol. 21, no. 2, Jan. 2021, doi: 10.3390/s21020340.

[2] F. Pistolesi and B. Lazzerini, "Assessing the risk of low back pain and injury via inertial and barometric sensors," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 11, pp. 7199–7208, Nov. 2020, doi: 10.1109/TII.2020.2992984.

[3] C. A. Nurse, L. J. Elstub, P. Volgyesi, and K. E. Zelik, "How accurately can wearable sensors assess low back disorder risks during material handling? exploring the fundamental capabilities and limitations of different sensor signals," *Sensors*, vol. 23, no. 4, Feb. 2023, doi: 10.3390/s23042064.

[4] N. Aggarwal, "Prediction of low back pain using artificial intelligence modeling," *Journal of Medical Artificial Intelligence*, vol. 4, Mar. 2021, doi: 10.21037/jmai-20-55.

[5] A. Al-Kubaisi and N. N. Khamiss, "A transfer learning approach for lumbar spine disc state classification," *Electronics*, vol. 11, no. 1, Dec. 2022, doi: 10.3390/electronics11010085.

[6] R. V. S. Lalitha, P. E. S. N. Krishna Prasad, T. Rama Reddy, K. Kavitha, R. Srinivas, and B. Ravi Kiran, "Efficient adaptive enhanced adaboost based detection of spinal abnormalities by Machine learning approaches," *Biomedical Signal Processing and Control*, vol. 80, Feb. 2023, doi: 10.1016/j.bspc.2022.104367.

[7] J. V. Shinde, Y. V. Joshi, and R. R. Manthalkar, "Multidomain feature level fusion for classification of lumbar intervertebral disc using spine MR images," *IETE Journal of Research*, vol. 68, no. 6, pp. 4346–4359, Jul. 2022, doi: 10.1080/03772063.2020.1792358.

[8] M. Tajdari *et al.*, "Image-based modelling for adolescent idiopathic scoliosis: mechanistic machine learning analysis and prediction," *Computer Methods in Applied Mechanics and Engineering*, vol. 374, Feb. 2021, doi: 10.1016/j.cma.2020.113590.

[9] J. Rim, S. Ryu, H. Jang, H. Zhang, and Y. Cho, "Machine learning-based pain severity classification of lumbosacral radiculopathy using infrared thermal imaging," *Applied Sciences (Switzerland)*, vol. 13, no. 6, Mar. 2023, doi: 10.3390/app13063541.

[10] H. Y. Tang, S. H. Tan, T. Y. Su, C. J. Chiang, and H. H. Chen, "Upper body posture recognition using inertial sensors and recurrent neural networks," *Applied Sciences*, vol. 11, no. 24, Dec. 2021, doi: 10.3390/app112412101.

[11] M. Han, L. Liu, M. Hu, G. Liu, and P. Li, "Medical expert and machine learning analysis of lumbar disc herniation based on magnetic resonance imaging," *Computer Methods and Programs in Biomedicine*, vol. 213, Jan. 2022, doi: 10.1016/j.cmpb.2021.106498.

[12] A. B. Oktay, N. B. Albayrak, and Y. S. Akgul, "Computer aided diagnosis of degenerative intervertebral disc diseases from lumbar MR images," *Computerized Medical Imaging and Graphics*, vol. 38, no. 7, pp. 613–619, Oct. 2014, doi: 10.1016/j.compmedimag.2014.04.006.

[13] M. C. Tsai, E. T. H. Chu, and C. R. Lee, "An automated sitting posture recognition system utilizing pressure sensors," *Sensors*, vol. 23, no. 13, Jun. 2023, doi: 10.3390/s23135894.

[14] A. Bottrighi *et al.*, "A machine learning approach for predicting high risk hospitalized patients with COVID-19 SARS-Cov-2," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, Dec. 2022, doi: 10.1186/s12911-022-02076-1.

[15] E. Jamshidi *et al.*, "Using machine learning to predict mortality for COVID-19 patients on day 0 in the ICU," *Frontiers in Digital Health*, vol. 3, Jan. 2022, doi: 10.3389/fdgth.2021.681608.

[16] C. Panigutti, A. Perotti, A. Panisson, P. Bajardi, and D. Pedreschi, "FairLens: Auditing black-box clinical decision support systems," *Information Processing and Management*, vol. 58, no. 5, Sep. 2021, doi: 10.1016/j.ipm.2021.102657.

[17] Z. Yu, P. Thakolkaran, K. Shea, and T. Stanković, "Artificial neural network supported design of a lattice-based artificial spinal disc for restoring patient-specific anisotropic behaviors," *Computers in Biology and Medicine*, vol. 152, Jan. 2023, doi: 10.1016/j.compbiomed.2022.106475.

[18] L. Lin *et al.*, "Quantifying axial spine images using object-specific bi-path network," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 8, pp. 2978–2987, Aug. 2021, doi: 10.1109/JBHI.2021.3070235.

[19] F. Rui *et al.*, "Development of a machine learning-based model to predict hepatic inflammation in chronic hepatitis B patients with concurrent hepatic steatosis: a cohort study," *eClinicalMedicine*, vol. 68, Feb. 2024, doi: 10.1016/j.eclinm.2023.102419.

[20] C. H. Lai, P. K. L. Mok, W. W. Chau, and S. W. Law, "Application of machine learning models on predicting the length of hospital stay in fragility fracture patients," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, Jan. 2024, doi: 10.1186/s12911-024-02417-2.

[21] P. M. E. L. van Dam *et al.*, "Machine learning for risk stratification in the emergency department (MARS-ED) study protocol for a randomized controlled pilot trial on the implementation of a prediction model based on machine learning technology predicting 31-day mortality in the emergency department," *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, vol. 32, no. 1, Jan. 2024, doi: 10.1186/s13049-024-01177-2.

[22] F. Galbusera, T. Bassani, F. Costa, M. Brayda-Bruno, A. Zerbi, and H. J. Wilke, "Artificial neural networks for the recognition of vertebral landmarks in the lumbar spine," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, vol. 6, no. 4, pp. 447–452, Dec. 2018, doi: 10.1080/21681163.2016.1261370.

[23] Z. Han, B. Wei, X. Xi, B. Chen, Y. Yin, and S. Li, "Unifying neural learning and symbolic reasoning for spinal medical report generation," *Medical Image Analysis*, vol. 67, Jan. 2021, doi: 10.1016/j.media.2020.101872.

[24] T. Sustersic, V. Rankovic, M. Peulic, and A. Peulic, "An early disc herniation identification system for advancement in the standard medical screening procedure based on Bayes theorem," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 151–159, Jan. 2020, doi: 10.1109/JBHI.2019.2899665.

[25] O. K. Oyedotun, E. O. Olaniyi, and A. Khashman, "Disk hernia and spondylolisthesis diagnosis using biomechanical features and neural network," *Technology and Health Care*, vol. 24, no. 2, pp. 267–279, Mar. 2016, doi: 10.3233/THC-151126.

[26] Y. Chen *et al.*, "Development of a predictive model for 1-year postoperative recovery in patients with lumbar disk herniation based on deep learning and machine learning," *Frontiers in Neurology*, vol. 15, Jun. 2024, doi: 10.3389/fneur.2024.1255780.

[27] P. C. Lin *et al.*, "Development of a machine learning algorithm to correlate lumbar disc height on X-rays with disc bulging or herniation," *Diagnostics*, vol. 14, no. 2, Jan. 2024, doi: 10.3390/diagnostics14020134.

[28] J. K. Kim and M. C. Chang, "Convolutional neural network algorithm trained on lumbar spine radiographs to predict outcomes of transforaminal epidural steroid injection for lumbosacral radicular pain from spinal stenosis," *Scientific Reports*, vol. 14, no. 1, Apr. 2024, doi: 10.1038/s41598-024-59288-w.

[29] M. Haider *et al.*, "Novel ensemble learning algorithm for early detection of lower back pain using spinal anomalies," *Mathematics*, vol. 12, no. 13, Jun. 2024, doi: 10.3390/math12131955.

[30] F. Villalba-Meneses *et al.*, "Classification of the pathological range of motion in low back pain using wearable sensors and machine learning," *Sensors*, vol. 24, no. 3, Jan. 2024, doi: 10.3390/s24030831.

[31] S. Tripathi, D. Muhr, M. Brunner, H. Jodlbauer, M. Dehmer, and F. Emmert-Streib, "Ensuring the robustness and reliability of data-driven knowledge discovery models in production and manufacturing," *Frontiers in Artificial Intelligence*, vol. 4, Jun. 2021, doi: 10.3389/frai.2021.576892.

[32] K. Okoye, J. T. Nganji, J. Escamilla, and S. Hosseini, "Machine learning model (RG-DMML) and ensemble algorithm for prediction of students' retention and graduation in education," *Computers and Education: Artificial Intelligence*, vol. 6, Jun. 2024, doi: 10.1016/j.caeai.2024.100205.

[33] J. Gopsill, L. Giunta, M. Goudswaard, C. Snider, and B. Hicks, "Data mining prototyping knowledge graphs for design process insights," *Journal of Engineering Design*, pp. 1–27, Jan. 2024, doi: 10.1080/09544828.2024.2302746.

[34] J. G. Yu, B. Selby, N. Vlahos, V. Yadav, and J. Lemp, "A feature-oriented vehicle trajectory data processing scheme for data mining: a case study for statewide truck parking behaviors," *Transportation Research Interdisciplinary Perspectives*, vol. 11, Sep. 2021, doi: 10.1016/j.trip.2021.100401.

[35] N. Yu, W. Xu, and K. L. Yu, "Research on regional logistics demand forecast based on improved support vector machine: a case study of Qingdao city under the new free trade zone strategy," *IEEE Access*, vol. 8, pp. 9551–9564, 2020, doi: 10.1109/ACCESS.2019.2963540.

[36] S. Roopashree, J. Anitha, S. Challa, T. R. Mahesh, V. K. Venkatesan, and S. Guluwadi, "Mapping of soil suitability for medicinal plants using machine learning methods," *Scientific Reports*, vol. 14, no. 1, Feb. 2024, doi: 10.1038/s41598-024-54465-3.

[37] L. Zheng, M. Mueller, C. Luo, and X. Yan, "Predicting whole-life carbon emissions for buildings using different machine learning algorithms: A case study on typical residential properties in Cornwall, UK," *Applied Energy*, vol. 357, Mar. 2024, doi: 10.1016/j.apenergy.2023.122472.

[38] M. Asfand-e-yar, Q. Hashir, A. A. Shah, H. A. M. Malik, A. Alourani, and W. Khalil, "Multimodal CNN-DDI: using multimodal CNN for drug to drug interaction associated events," *Scientific Reports*, vol. 14, no. 1, Feb. 2024, doi: 10.1038/s41598-024-54409-x.

[39] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, vol. 91, pp. 216–231, Jul. 2019, doi: 10.1016/j.patcog.2019.02.023.

[40] I. Emmanuel, Y. Sun, and Z. Wang, "A machine learning-based credit risk prediction engine system using a stacked classifier and a filter-based feature selection method," *Journal of Big Data*, vol. 11, no. 1, Feb. 2024, doi: 10.1186/s40537-024-00882-0.

[41] C. Gui, "Link prediction based on spectral analysis," *PLoS ONE*, vol. 19, no. 1 January, Jan. 2024, doi: 10.1371/journal.pone.0287385.

[42] I. Düntsch and G. Gediga, "Indices for rough set approximation and the application to confusion matrices," *International Journal of Approximate Reasoning*, vol. 118, pp. 155–172, Mar. 2020, doi: 10.1016/j.ijar.2019.12.008.

[43] S. R. Shah, S. Qadri, H. Bibi, S. M. W. Shah, M. I. Sharif, and F. Marinello, "Comparing inception V3, VGG 16, VGG 19, CNN, and ResNet 50: A case study on early detection of a rice disease," *Agronomy*, vol. 13, no. 6, Jun. 2023, doi: 10.3390/agronomy13061633.

[44] N. A. Hroub, A. N. Alsannaa, M. Alowaifeer, M. Alfarraj, and E. Okafor, "Explainable deep learning diagnostic system for prediction of lung disease from medical images," *Computers in Biology and Medicine*, vol. 170, Mar. 2024, doi: 10.1016/j.compbiomed.2024.108012.

BIOGRAPHIES OF AUTHORS

Christian Ovalle Paulino is an associate professor at the Faculty of Engineering of the Universidad Tecnologica del Perú, Lima-Peru. He has a Ph.D. in systems engineering with a specialization in artificial intelligence. His research areas are process mining, business data analysis, and pattern recognition. He is the CEO of the 7D consultancy dedicated to the Investigation of intelligent solutions. He has participated in different research projects, receiving awards from the Peruvian Ministry of Defense and the Armed Forces Army for the best general researcher, which is a technology-based company and his innovative products received national and international recognition. Dr. Ovalle has filed a number of patents and industrial designs on his innovative ideas. His research interests include data mining, artificial intelligence, image/signal processing, bibliometrics, and pattern recognition. He can be contacted by email: dovalle@utp.edu.pe.

Jorge Huamani Correa Ph.D. in physics from Universidade Federal Fluminense (Brazil), M.Sc. in physics from Universidade de Brasilia (Brazil) and B.Sc. from Universidad Nacional del Callao. He has experience in the area of condensed matter physics, acting mainly in electronic properties and quantum transport in graphene and two-dimensional materials, as well as magnetic properties in nanowires, strain effects and spin-orbit coupling. He has knowledge of programming languages such as Fortran and Python, focusing in the latter on data science and machine learning. He can be contacted by e-mail: c22273@utp.edu.pe.