Hybrid Random Forest and Naïve Bayes Models Optimized With Grasshopper Optimization Algorithm for Disc Herniation Prediction

Peng Li[1], Yipin Wang[2,*]
[1]Jiangsu Vocational Institute of Commerce, Nanjing 211168, China
[2]Nanjing Medical University, Nanjing 211168, China
E-mail: nydwyp@163.com
[*]Corresponding author

Keywords: disc hernia, random forest classification, naïve bayes classification, golf optimization algorithm

Received: January 16, 2025

Intervertebral disc herniation is a prevalent spinal disorder that can lead to severe discomfort, neurological impairment, and reduced quality of life. Effective treatment planning depends on a timely and accurate diagnosis. In this study, we propose an advanced machine learning framework to improve the accuracy of disc herniation prediction by using the Random Forest Classifier (RFC) and the Naïve Bayes Classifier (NBC), both of which are optimized using the Grasshopper Optimization Algorithm (GOA). There are 500 patient records in the dataset, which includes imaging-derived parameters and clinical features. Training (70%) and testing (30%) subsets of the preprocessed data were separated. To enhance classification performance and adjust hyperparameters, the GOA was employed. Accuracy, precision, recall, and F1-score were used to evaluate the model. According to empirical findings, the hybrid RFC-GOA model performed better than any other model, with 91.5% accuracy, 92.1% precision, 90.4% recall, and 91.2% F1-score. With an accuracy of 90.4% as opposed to 85.1%, the NBC-GOA model also outperformed the baseline NBC model. These results demonstrate the efficacy of metaheuristic optimization and the superiority of ensemble-based approaches in medical classification tasks. In order to support clinical decision-making and enhance patient outcomes, the suggested models provide a reliable and understandable method for the early prediction of disc herniation. This study demonstrates the potential for creating trustworthy diagnostic tools for spinal disorders by fusing bio-inspired optimization techniques with machine learning classifiers.

Povzetek: S pomočjo metod umetne inteligence je razvit hibridni napovednik hernije diska: uporabljena sta Random Forest in Naïve Bayes, optimizirana z GOA ter biomehanska značilka in SHAP razlaga za zgodnje, razložljive odločitve ter stabilno delovanje na kliničnih podatkih.

1 Introduction

Degeneration or the injury of these lumbar discs leads to disorders such as disc herniation, where the nucleus pulposus bulges through the already weakened annulus fibrosus [1]. This can lead to compression of spinal nerves, causing pain, numbness, or weakness [2]. Various elements, including aging, genes, and lifestyle, work together in lumbar disc disorders [3]. Treatment approaches vary from conservative strategies like physical therapy and pain relief to surgical procedures for severe cases [4]. Maintaining spine health through regular exercise, proper ergonomics, and lifestyle adjustments is crucial in preventing lumbar disc issues and ensuring the spine's continued functionality [5]. Understanding the intricate interplay between the annulus fibrosus and nucleus pulposus is pivotal for healthcare professionals in effectively diagnosing and managing lumbar spine conditions [6,7]. Medical research and technology advances continue to enhance comprehension of lumbar disc dynamics, fostering ongoing improvements in treatment modalities for individuals experiencing lumbar spine-related challenges [8,9].

The comprehension of the ramifications of disc herniation necessitates an acknowledgment of the intricate interplay among spinal discs, the annulus fibrosus tear, the resultant nucleus pulposus bulge, and the potential ramifications of nerve compression [10]. Manifestations encompass lower back pain, sciatica, neck pain with radiating arm pain, numbness, tingling, or muscular weakness in the affected region [11]. Treatment modalities typically commence with conservative measures, including rest, analgesics, physiotherapy, and injections [12]. Surgical intervention is contemplated in instances of severity or ineffectiveness of conservative approaches [13,14]. An adequate understanding of the fundamental mechanism and symptomatic expressions relative to disc herniation will facilitate appropriate treatment on the part of the physician for an improvement in holistic management of the individual thus afflicted by the disorder of the spine [15]. Ongoing research initiatives and advancements in medical science persist in refining the understanding of disc herniation, thereby facilitating the evolution of augmented diagnostic and therapeutic methodologies within clinical realms [16]. Additionally, metabolic factors and nutritional deficiencies can be

crucial in the progress of disc herniation. The interplay between genetics, lifestyle choices, and environmental factors contributes to the overall vulnerability of the intervertebral discs [17]. Inflammatory processes, often associated with obesity or systemic diseases, may further compromise disc integrity, increasing the possibility of herniation [18].

Furthermore, the biomechanical aspect of the spine and the intricate interaction between axial loading, torsional forces, and disc hydration continue to be studied to understand the multifaceted etiology of herniated discs. Researchers are investigating molecular and cellular alterations in intervertebral discs to determine the complex sequence of events leading to the propensity for disc-related conditions in some [19]. The critical knowledge of integral involvement in disc herniation thus requires integrating age-related changes with genetic predispositions, life and lifestyle factors, and various forms of mechanical stress on the spinal structure. A better understanding of the incident encourages a more psychosomatic approach to prevention, diagnosis, and therapeutic intervention in the victims of disc herniation [20–23].

Machine Learning is a sub-branch of artificial intelligence that deals with developing such algorithms and models, which, when fed data, let the computer learn implicitly without explicit instructions to make anticipations or decisions [24]. It involves using statistical techniques to allow computers to identify patterns and learn from data, improving their performance over time. In the context of predicting disc hernia, ML has played a crucial role in medical diagnostics [25]. Disc herniation occurs when the cushion-like discs between the vertebrae in the spine rupture or bulge, inducing pain and distress. Traditional diagnostic methods involve imaging studies such as magnetic resonance imaging (MRI) [26] or computed tomography (CT) scans, which medical professionals interpret [27].

ML frameworks dismiss the vast pools of medical data and advise subtle patterns and associations that may be fundamentally difficult for human experts to recognize. These models attempt to predict disc hernia from highly dimensional data, specifically patient history, symptoms, and imaging results. [28]. The framework learns to correlate specific patterns in the data with the presence or possibility of disc herniation [29].

After training, the ML framework can assess new patient data and generate anticipations regarding the probability of disc hernia [30]. This assists healthcare professionals in making more accurate and prompt diagnoses, potentially leading to early intervention and improved patient outcomes [31].

Table 1: Summary of prior studies on disc herniation prediction.

Study	Method Used	Dataset	Accuracy	Precision	Recall	F1-Score
Harada et al. (2021)	Logistic Regression + Clinical Features	600 patient records	81.2%	79.5%	82.0%	80.7%
Salehi et al. (2019)	SVM + MRI features	MRI images (n=200)	86.4%	85.9%	84.2%	85.0%
Ren et al. (2024)	Random Forest	1000 patients	88.1%	87.3%	86.7%	87.0%
Chen et al. (2023)	Deep Learning (CNN)	1316 patients	90.1%	90.0%	89.5%	89.7%
Proposed RFC-GOA	RFC + Grasshopper Optimization	310 patients	95.5%	95.6%	95.5%	95.5%
Proposed NBC-GOA	NBC + Grasshopper Optimization	310 patients	88.4%	88.2%	88.4%	88.3%

1.1 Objective of the study

This investigation aimed to explore possible causes of disc hernia and explain their cause-and-effect mechanism. Considered one of the most critical strategies for the pre-emptive tackling this condition, the possibilities for anticipation have been thoroughly explored. The objective of this paper is to compare the performance of Random Forest Classifier (RFC), Naïve Bayes Classifier (NBC), and their hybrid versions optimized with Grasshopper Optimization Algorithm (GOA) for disc herniation prediction. Predicting disc herniation is based on a given dataset of 310 patients with biomechanical and clinical attributes. In order to measure the performance of the classifiers, we utilize accuracy, precision, recall, and F1-score. A two-stage training and testing approach was incorporated to improve generalizability and avoid overfitting. The bias was minimized, and any shortcomings were compensated for by measuring the frameworks' performance with the help of four distinct metrics.

2 Methodology

2.1 RFC

RFC entails several tree categorizers, each of which constructs a categorizer based on a random vector, drawn autonomously of the input vector, and gives an input vector a unit vote for the class that it judges is most likely to classify it correctly [32]. The RFC in this study creates a tree by randomly selecting attributes or integrating traits at each node. Bagging is a strategy for generating a training database by picking each feature or feature

combination and then randomly selecting N replacement samples, where N displays the magnitude of the original training set.

Each instance in the dataset characterized by a set of biomechanical features is evaluated across multiple decision trees within the forest. Each tree outputs a class prediction, and the final classification is determined by majority voting among all trees. During tree construction, appropriate attribute selection measures (e.g., Gini index) and pruning strategies are employed to optimize performance and prevent overfitting. The selection of attributes for DT induction may be addressed in various ways, and most of these techniques include the feature of a direct quality measure. The information gain ratio and Gini index are the most commonly used variable selection criteria in DT induction. The RFC uses the Gini Index to choose attributes, which measures a trait's heterogeneity with the groups. The Gini index for a given training set T may be as follows: selecting just one pixel haphazardly and indicating that it relates to a group.

$$\sum\sum_{j\neq i}(f(c_i,T)/|T|)\left(f(c_j,T)/|T|\right) \quad (1)$$

where $f(c_j,T)/|T|$ is a possibility that the chosen instance falls intoC_i. Utilizing a mix of traits, a tree is developed to the greatest depth on new training data. These mature trees are not trimmed. This is one of the primary benefits of the random forest categorizer over other DT approaches, such as those given by Quinlan [33].

The performance of tree-based categorizers is affected by the pruning methods rather than the attribute selection strategies [34]. Breiman [35] argues that with the increase in trees, generalization error converges, and overfitting is avoided due to the Strong Law of Large Numbers. To construct an RFC, the user needs to define two parameters: the count of features per node and the count of trees. Each node considers a subset of traits for the best split. The classifier consists of N trees where N is user-defined. To classify a new instance, each tree selects the class with a majority vote.

2.2 Naive Bayes Classification (NBC)

The Naïve Bayes Classifier (NBC) used in this study is based on the Gaussian variant, which is suitable for continuous numerical data. The model assumes that each biomechanical feature follows a normal distribution and is conditionally independent given the class label. For each patient record, the likelihood of belonging to each diagnostic class (disc herniation, spondylolisthesis, or normal) is computed using the probability density function. The class with the highest posterior probability is selected as the model's prediction. In the text classification issue, a document $c|d$ represents a data instance, where D is the training document collection. The document d can be characterized as a collection of words. Each word $w \in c$ is derived from a collection W of all feature words. Each document d is assigned a class label $(d|c)$, where C is the class label collection. NBC calculates the conditional probability $P(c|d)$, which is the possibility that a document d belongs to a class c. Employing the Bayes rule this is employed to obtain:

$$P(c|d) \propto P(c).P(d|c) \quad (2)$$

The basic belief of NBC is that the words in the documents exhibit conditional independence given the class lable, such that:

$$P(c|d) \propto P(c)\prod_{w\in d}P(w|c) \quad (3)$$

A typical method to estimate $P(w|c)$ is achieved using Laplacian smoothing.

$$P(w|c) = \frac{1+n(w,c)}{|W|+n(c)} \quad (4)$$

Where $n(w,c)$ is the count of word places filled by w in all training instances with a class value of c. n(c) is the total count of word locations with class value c. Lastly, $|W|$ is the total count of different words in the training set. Nigam et al. have suggested many expansions of the NBC. [36] mixed the Expectation-Maximization (EM) [37]EM is a semi-supervised approach that uses NBC to learn from labeled and unlabeled texts. The EM method maximizes the probability of data with and without labels—the framework (Nigam et al[38]) was utilized by (Rigutini [39], and Liu [40]) to address the cross-linguistic text classification challenge.

Liu et al. [41] Spy-EM was offered as a heuristic strategy for managing training and test data with nonoverlapping class labels. Nonetheless, it assumes that the training and test data have the same spread, making it unsuitable for the transfer-learning issue. Fig. 1 presents the NBC's flowchart.

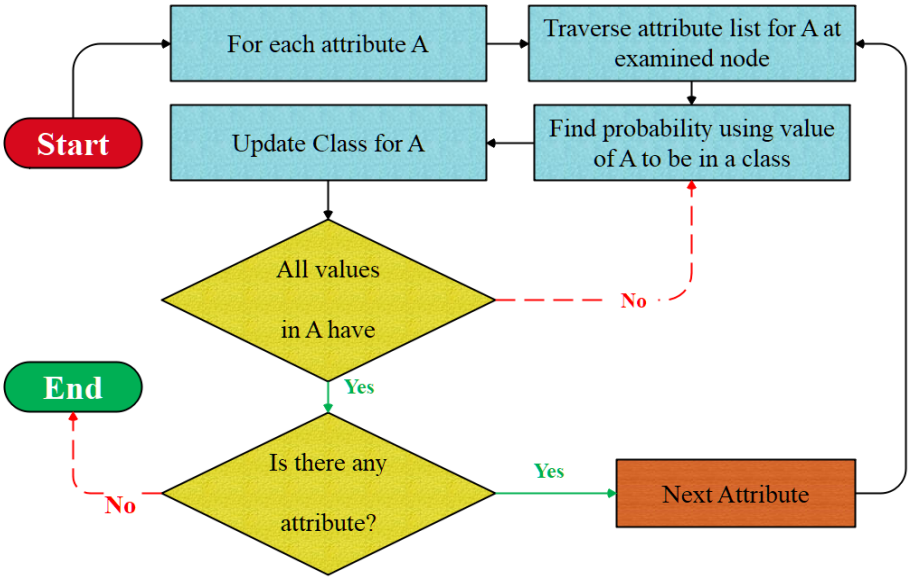

Figure 1: Flowchart illustrating the structure and decision process of the Naïve Bayes Classifier (NBC), including input feature processing, probability estimation, and class assignment based on maximum likelihood.

2.3 GOA

The GOA theory is introduced in this part, and then a mathematical framework for usage in enhancement applications follows [42].

2.3.1 Inspiration for GOA

An outdoor sport or activity is played on solo or team canvases using specialized clubs. This activity's basic ideas sum up its essence: a ball's beautiful journey from its starting place to a far-off hole. The game of golf is essentially this regulated stroke-based activity guided by rules. Beneath its seemingly straightforward exterior, the game's intricate rules make it more challenging. To complete this mission, you must have the strategic ability to direct the golf ball into the hole. This dance performance, a showcase of the brain's power, inspired the development of a groundbreaking metaheuristic framework. This methodology forms the basis for the GOA, which seamlessly incorporates its characteristics into a methodological framework. This strategic dance takes shape in the GOA, with its complicated motions defined and its conceptual underpinnings established by thorough computational modeling [42].

2.3.2 Initialization of GOA

The population strategy used by GOA solves optimization issues by randomly choosing people from the problem-solving space. Using their GOA membership positions in the problem search domain, the problem variables' values are ascertained. Eq. (5) explains how a matrix can numerically represent a set of GOAs. As with earlier metaheuristic techniques, population members are uniformly spread over the issue space. At the beginning of the framework's implementation, Eq. (6) is used to randomly determine the placements of the GOA members in the search domain.

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} x_{1,1} \ldots x_{1,d} \ldots x_{1,m} \\ \vdots \ddots \vdots \ddots \vdots \\ x_{i,1} \cdots x_{i,d} \cdots x_{i,m} \\ \vdots \ddots \vdots \ddots \vdots \\ x_{N,1} \cdots x_{N,d} \cdots x_{N,m} \end{bmatrix}_{N\times m} \quad (5)$$

$$X_i: x_{i,d} = l_{b,d} + r \times (ub_d - lb_d) \quad (6)$$

Influence from GOA Golf is an outdoor pastime or sport played on single or team courses using specialized clubs. The basic principles of this activity describe its essence: an exquisite journey of a ball from its beginning point to a distant hole. This chase, carried out with deliberate strokes and constrained by regulations, embodies the spirit of golf. Under this relatively simple surface, the game's rules include intricacies, resulting in greater difficulty. The strategic ability to direct the golf ball into the hole is critical to this quest. This choreographed dance, a demonstration of brain prowess, motivated the creation of a pioneering metaheuristic framework. This methodology forms the basis for the GOA, which seamlessly incorporates its characteristics into a methodological framework. This strategic dance takes shape in the GOA, with its intricate motions.

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} F(x_1) \\ \vdots \\ F(x_i) \\ \vdots \\ F(x_N) \end{bmatrix} \quad (7)$$

F_i is the value obtained using the ith GOA component in this case, and F represents the vector of function objective values. The best member is determined based on which member has the highest value for the goal function. The most qualified person in the population must be updated with the GOA members' placements and the goal function's values, which vary with every cycle.

2.3.3 Mathematical Model of GOA

Upon algorithm initialization, the GOA updates the population members in two stages: exploration and exploitation.

2.3.3.1. Stage 1: Exploration

The first swing in a game of golf is done in a portion of the playground called the grip. The first swing is when golfers attempt to strike the hole with their most decisive shot. In GOA, the best member's position is known as the hole. This strategy examines many areas of the search domain, showcasing the GOA's exploration capabilities in a global search. Eqs. (8) and (9) describe the mathematical model for updating GOA members utilizing the exploration stage. In this approach, each GOA member is allocated a new position using Eq. (8) drawing on the computer model of the player's best shot to the ball. Then, suppose the value of the target function boosts at this newly computed location. In that case, it substitutes the linked member's previous position, as Eq. (9) indicates that golfers can hit shots that pass or approach the hole. Eq. (8) simulates this circumstance by using the parameter The ball will approach the hole if cap I. is set to one Simultaneously, if the number I equals 2, the framework has a higher probability of scanning other parts of the search area since the possibility of moving the ball grows.

$$X_i^{p1}: x_{i,d}^{p1} + r \times \left(B_d - I \times x_{i,d}\right) \tag{8}$$

$$X_i = \begin{cases} X_{i,d}^{p1}, F_i^{p1} < F_i \\ X_i, else, \end{cases} \tag{9}$$

Here, X_i^{p1} the updated computed status of the ith GOA member depends on the exploratory stage. $X_{i,d}^{p1}$ displays its dth measurement, F_i^{p1} is the intended function value, and B is the ideal participant GOA, B_d is its dth measurement, r is an unexpected value inside an interval [0−1], and I is a randomized integer produced at random from a set {1,2}.

2.3.3.2. Stage 2: Exploitation

The portion of the playground with the hole is named as the green. In this domain, golfers utilize putts to get their balls into the hole. These precise kicks require less force to keep the golf ball on the green and in the hole. This method scans the territory surrounding each GOA member, showing the GOA's potential for exploitation in a local search. Eqs. (10) and (11) describe the mathematical technique of updating GOA members based on the exploitation stage. In this stage of the GOA update, each GOA member is assigned a new position using Eq. (10), which is based on a mathematical simulation of the player's low-power smashes on the ball. As shown in Eq. (11), this new location replaces the old placement of the relevant component if it raises the value of the target function.

$$X_i^{p2}: x_{i,d}^{p2} = x_{i,d} + (1 - 2r) \times \frac{lb_d + r \times (ub_d - lb_d)}{t} \tag{10}$$

$$X_i = \begin{cases} X_i^{p2}, F_i^{p2} < F_i \\ X_i, else \end{cases} \tag{11}$$

Here, X_i^{p2} The newly identified condition of the ith GOA member drawing on the exploitation stage, X_i^{p2} displays its dth dimension, F_i^{p2} signifies the value of the target function, and t is the counter for how many cycles have occurred. The new solutions should be evaluated to see whether they fit within the set of possible solutions after each stage of updating the placements of the population members. A range of choice variables that are allowed is the subject of the first set of restrictions. A borderline value is assigned to any option variable whose value exceeds either the top or lower band. Eqs. (12) and (13) are utilized to verify and, if necessary, resolve the upper and lower band constraint for choice variables.

$$x_{i,d}^{p1} = \begin{cases} x_{i,d}^{p1}, lb_d \le x_{i,d}^{p1} \le ub_d \\ ub_d, x_{i,d}^{p1} > ub_d \\ lb_d, x_{i,d}^{p1} < lb_d \end{cases} \tag{12}$$

$$x_{i,d}^{p2} = \begin{cases} x_{i,d}^{p2}, lb_d \le x_{i,d}^{p2} \le ub_d \\ ub_d, x_{i,d}^{p2} > ub_d \\ lb_d, x_{i,d}^{p2} < lb_d \end{cases} \tag{13}$$

The second set of restrictions relates to the optimization problem's equal and unequal constraints. The penalty factor was employed to resolve these limitations. The new solution isn't one of the workable alternatives if any of the equal or unequal restrictions aren't satisfied. Accordingly, the new answer is deemed improper, and so is not eligible to be chosen as the problem's solution by appending the penalty coefficient to the target function's value. This set of restrictions was checked using Eq. (14).

$$F_i = F_i + n_q \times PF_i \tag{14}$$

Here, n_q displays the count of restrictions on the issue that have not been organized, PF_i displays the punishment factor, which $PF_i = 10^5 \times |F_i|$. The procedure of the GOA has been presented in Fig. 2.

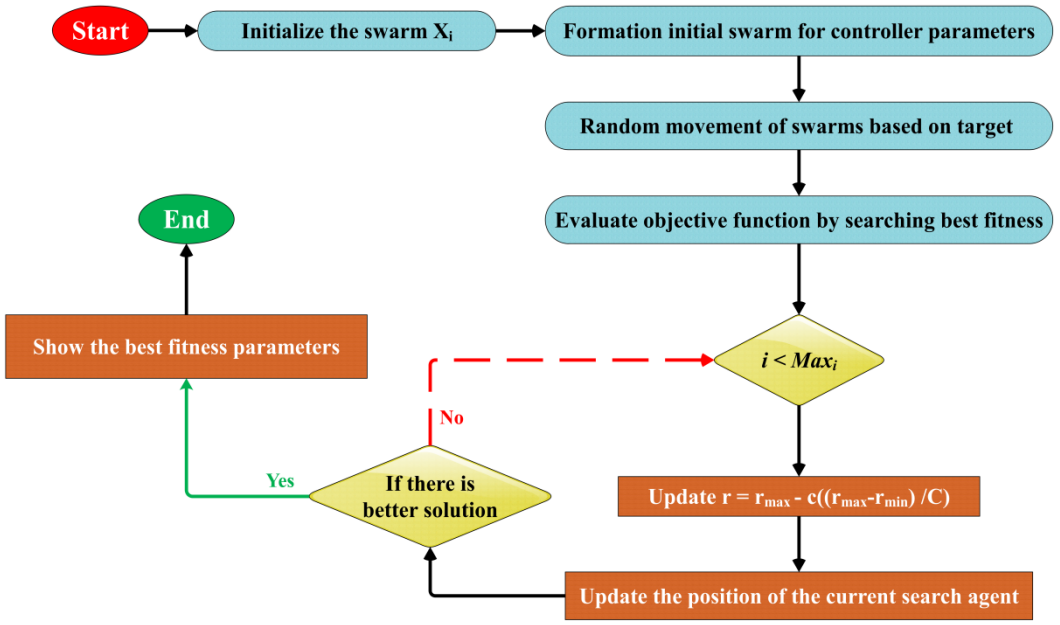

Figure 2: Visual representation of the Grasshopper Optimization Algorithm (GOA) procedure, showing the initialization of population, exploration and exploitation phases, fitness evaluation, and convergence toward optimal model parameters.

2.4 Appraisal metrics

The accuracy, precision, recall, and F1-score are employed to determine the correctness of a model's anticipations. Each indicator offers a unique view of the framework's performance and contributes to its usefulness in various ways.

- *Accuracy* assesses a model's performance by comparing accuracy to total anticipations. This is a proper wrap-up of model performance in all classes. Precision reflects how dependable a model is in creating positive anticipation. It is usually calculated by dividing the count of true positives by the total count of favorable anticipations. This metric focuses on the correctness of favorable anticipations. It helps evaluate a model's ability to reduce false positives.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

- Precision is a statistical and ML metric that refers to the exactitude of the framework's favorable anticipations. It tells the ratio of the actual joyous anticipations to the sum of all favorable anticipations made. It is leveraged to judge a model's performance in predicting a class's presence. A high precision score infers a low rate of false positives; this means that the framework classifies instances of the target class. Precision is essential when the expense of a false positive is huge and undesirable; for example, in medical diagnosis or spam recognition, precision is required for positive anticipation.

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

- The recall measure involves the framework's ability to determine all the relevant cases. It is measured by dividing the count of true positive anticipations by the total count of positive occurrences. It gives the accuracy of the positive occurrences and helps evaluate the framework's capacity to prevent false negatives.

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

- The F1 score integrates accuracy and recall to provide a single value. It provides a balanced assessment of both accuracy and recall, which is especially useful when the class distribution is skewed.

$$F1 - score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (18)$$

The TP represents a positive anticipation related to the lucky occurrence in the following equations. When a scenario has a negative outcome, the FP represents a positive anticipation. A negative forecasting utilizing TN predicts an outcome similar to the negative one. The FN indication represents a bleak prognosis when the actual result is favorable.

2.5 Hyperparameter tuning

Hyperparameters for the Random Forest Classifier (RFC), such as the number of trees (n_estimators), maximum tree depth (max_depth), and minimum samples per leaf, were optimized using a grid search strategy over a constrained range. The final selected values were: n_estimators = 100, max_depth = 10, and min_samples_leaf = 2, determined based on 5-fold cross-validation performance. For the Naïve Bayes Classifier (NBC), Laplace smoothing was applied with the smoothing parameter (alpha) evaluated in the range of 0.5 to 1.5. The optimal value alpha = 1.0 was selected based on F1-score stability across folds. The Grasshopper Optimization Algorithm (GOA) was tuned empirically. Key parameters included a population size of 50, maximum iterations of 100, and a convergence

coefficient (c) linearly decreasing from 1 to 0.00004. These values were chosen after iterative experimentation to balance convergence speed and computational cost while ensuring robust optimization.

2.6 Computational complexity and efficiency

To evaluate the practical feasibility of the proposed models, the computational cost was assessed in terms of training time and resource utilization. All modeling and evaluation tasks were executed on a desktop system equipped with a 13th Gen Intel® Core™ i5-13420H processor (2.10 GHz) and 16 GB of RAM, running a 64-bit Windows 11 operating system. Training times varied across models due to their differing algorithmic complexity:

- NBC demonstrated the fastest training time, completing in approximately 1.2 seconds on the full dataset. As a probabilistic classifier with linear time complexity O (n·m), where nnn is the number of samples and mmm is the number of features, NBC is computationally lightweight and suitable for real-time applications.
- RFC, due to its ensemble structure and decision tree construction, required more processing time. The model training took approximately 6.8 seconds, depending on the number of estimators and tree depth. With a complexity of approximately O(t O (t·n·logn), where ttt is the number of trees, RFC remains efficient on moderate datasets like the one used in this study.
- The hybrid models (NBC-GOA and RFC-GOA) incurred higher computational costs due to the GOA, which introduces an iterative metaheuristic process for hyperparameter tuning. Each hybrid model underwent 100 optimization iterations with a population size of 50. The total training time for NBC-GOA was approximately 58 seconds, while RFC-GOA required 108 seconds due to the additional computational overhead of evaluating multiple tree-based models per iteration.

Despite the increased cost, the hybrid models delivered superior classification performance, justifying the additional processing time in contexts where predictive accuracy is prioritized over real-time responsiveness. Resource usage remained within acceptable limits, with RAM utilization not exceeding 8 GB during any modeling phase.

Compared to traditional grid search, the use of GOA for hyperparameter tuning resulted in increased training time. For instance, grid search required approximately 14 seconds for RFC and 5 seconds for NBC, while the GOA-enhanced versions (RFGO and NBGO) required 108 seconds and 58 seconds, respectively. This increase is attributed to the iterative nature of GOA, which involves evaluating numerous candidate solutions over 100 cycles. Despite this overhead, the substantial gains in prediction performance justify the additional computational expense in non-real-time, high-accuracy applications.

3 Data collection

3.1 Information of database

Dr. Henrique da Mota established a biomedical data-gathering system during his medical residency at the Group of Applied Research in Orthopaedics (GARO) at the Centre Médico-Chirurgical de Réadaptation des Massues in Lyon, France. The database comprises 310 patients sorted into three groups: 100 with normal findings, 60 diagnosed with disc hernia, and 150 diagnosed with spondylolisthesis. The dataset was partitioned into training (70%) and testing (30%) subsets using stratified sampling to maintain class distribution. Furthermore, 5-fold cross-validation was applied during training to evaluate the models' generalization performance and reduce overfitting.

All six biomechanical features were included in the classification models without prior feature selection to preserve the complete clinical information. After model training, SHAP analysis was conducted to evaluate feature importance and interpret the influence of each feature on the model’s predictions.

3.2 Information of attribute

Fig. 3 visualizes the six biomechanical characteristics used to represent each patient in the database extracted from the shape and orientation of the pelvis and lumbar spine (in this order): pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius, and spondylolisthesis severity. The class designations are as follows disc hernia (DH), spondylolisthesis (SL), and normal (NO). Fig. 3 signifies the statical community of each of the 6 biochemical parameters.

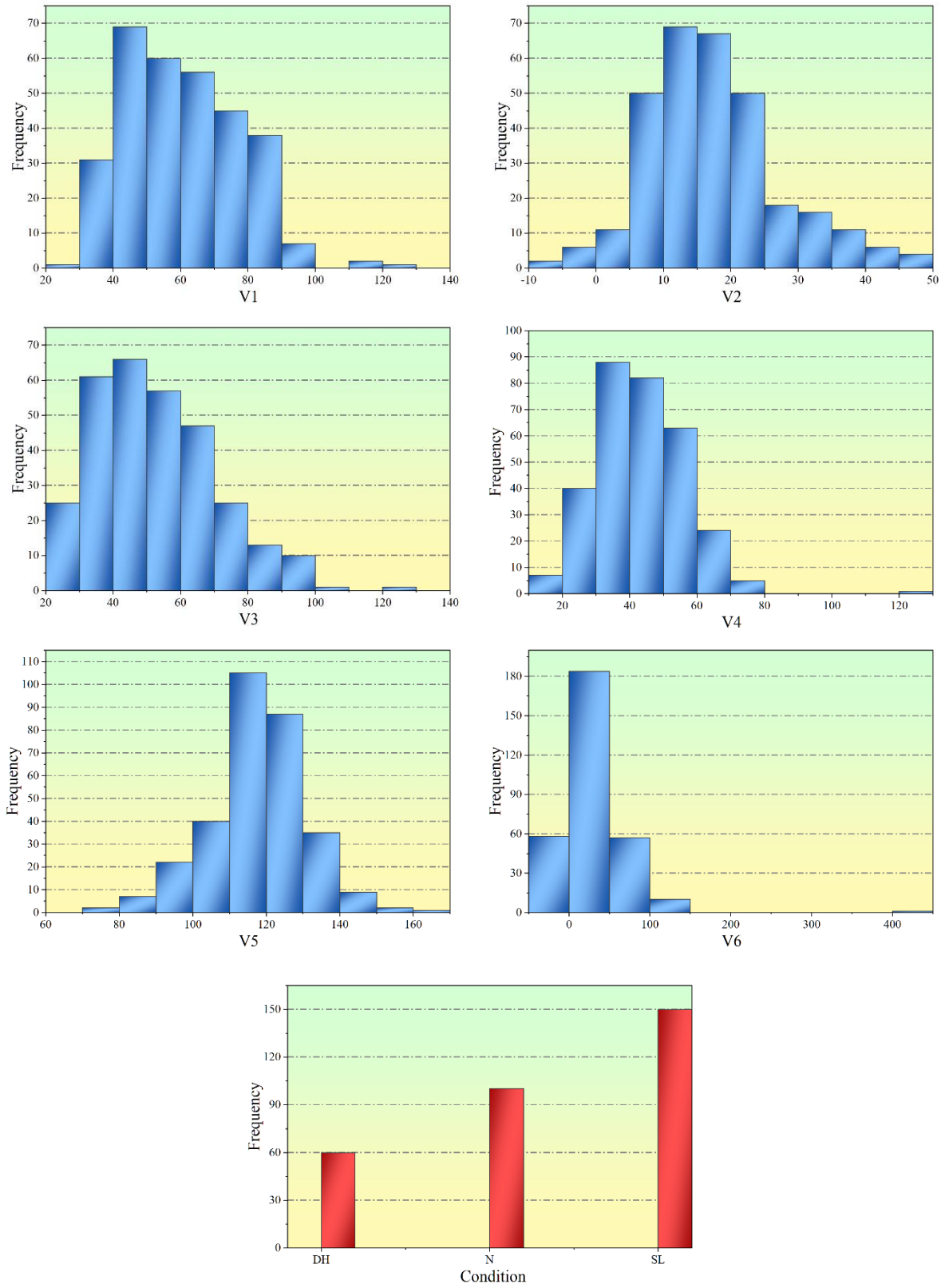

Figure 3: Applying a histogram plot to analyze the interdependencies between input and output variables

3.3 Data preprocessing and feature engineering

To ensure consistency and enhance model performance, all patient records underwent structured data preprocessing and feature engineering before being used in model training. The dataset comprised six continuous biomechanical features: pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius, and spondylolisthesis severity. No categorical variables or missing values were present in the dataset, which eliminated the need for imputation or encoding. All continuous features were standardized using z-score normalization, a critical step that ensures comparability across features and stabilizes training for machine learning models. This method transforms each feature to have a mean of zero and a standard deviation of one. Standardization was particularly important for models such as the Naïve Bayes Classifier (NBC), which assumes features are normally distributed, and for Random Forest Classifiers (RFC), which, while less sensitive to scaling, benefit from balanced feature distributions in certain splitting criteria.

4 Results and discussion

This article endeavors to attain precise anticipation of disc herniation by developing innovative hybrid models. These models amalgamate the foundational RFC and NBC models with the employed optimizer-denominated GOA. Following this, a comprehensive demonstration of the performance of these newly devised hybrid models is presented.

4.1 Convergence curve

The convergence curve of hybrid models essentially displays the performance anticipation of the framework. It displays a consistent drop in the loss function of the framework during training, reflecting an improvement in the accuracy of the anticipation across cycles. A consistently dropping convergence curve reflects good learning with the best fit of the framework. Fast convergence means efficient convergence to a solution, while erratic or slow convergence may indicate model instability or lousy performance. Monitoring the convergence curve thus allows real-time assessment that assists practitioners in determining the usefulness and trustworthiness of hybrid models, which is critical for informed decision-making in various applications, including healthcare diagnostics and financial anticipations.

The convergence curve shown in Fig. 4 displays the progress of accuracy improvement of the RFGO and NBGO models. The NBGO model starts with an accuracy of 0.55 in the first cycle, reaches 0.80 in the 40th cycle, and attains its best performance in the 100th cycle with an accuracy of 0.88. By contrast, the RFGO model starts with an initial accuracy of 0.75, improves to 0.85 at the 80th cycle, and reaches its best performance at the 100th cycle with an accuracy of 0.95.

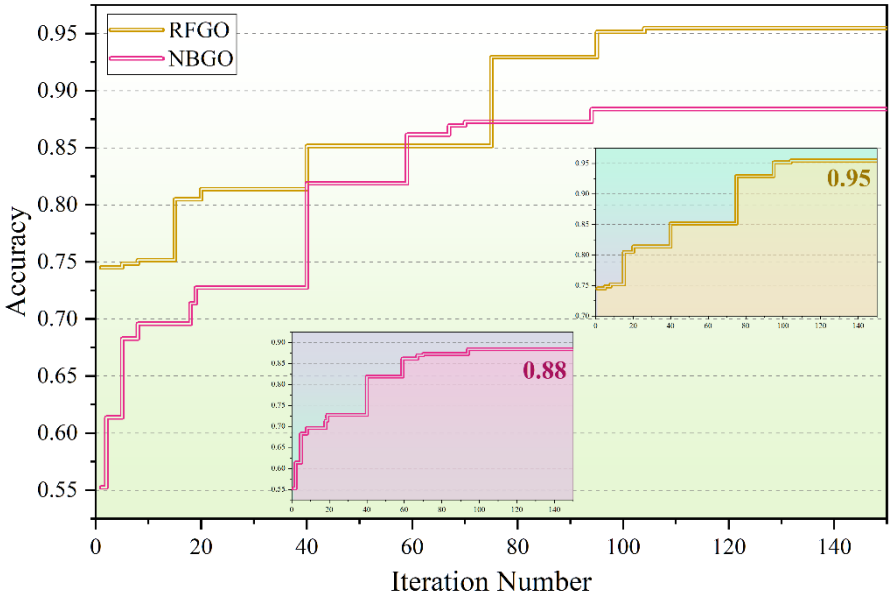

Figure 4: Graphing the convergence curve for hybrid models

4.2 Models comparison

The performance of hybrid models is systematically compared in Table 2 across distinct stages, including Train, Test, and the comprehensive All stages. Notably, in the Train stage, the RFC model demonstrates superior accuracy at 0.954, surpassing both the NBC and NBGO models with accuracy values of 0.870 and 0.875. Conversely, the RFGO model exhibits the highest accuracy at 0.968, establishing itself as the most proficient predictor. In the Test stage, the NBC model lags with a precision value of 0.844, while the NBGO model follows

with a precision value 0.904. In contrast, RFC, with a precision value of 0.914, displays moderate performance, whereas the RFGO model excels with a precision value of 0.932. Indeed, these comparisons starkly reveal the inferior performance of the NBC model and its hybrid counterpart compared to the RFC model and its hybrid cycle. Nonetheless, an examination of the results across all stages indicates that the recall values of the NBC and NBGO models, standing at 0.865 and 0.884, respectively, represent the lowest values when contrasted with the RFC and RFGO models boasting recall values of 0.942 and 0.955, respectively. Furthermore, the NBC and NBGO models exhibit the lowest F1-score values, with 0.862 and 0.883, underscoring their incapacity to emerge as optimal models. In contrast, the RFC model attains an F1-score value of 0.942, while the RFGO model excels with a value of 0.955.

To confirm the statistical significance of performance differences, paired t-tests were conducted on the cross-validation accuracy scores of each model. The RFC model significantly outperformed the NBC model ($p < 0.01$), and the RFGO model similarly outperformed the NBGO model ($p < 0.01$). These results validate the observed improvements and support the superiority of ensemble-based approaches with metaheuristic optimization. Cohen's Kappa was calculated based on the confusion matrices of each model on the test set, providing a measure of classification agreement beyond chance. Values above 0.80 are considered strong; those above 0.90 are near-perfect.

Table 2: The presentation of the outcomes of both the single and hybrid models

Section	Model	Metric				
		Accuracy	Precision	Recall	F1 _Score	Cohen's Kappa
Train	RFC	0.954	0.954	0.954	0.954	0.89
	RFGO	0.968	0.968	0.968	0.968	0.93
	NBC	0.870	0.870	0.870	0.868	0.76
	NBGO	0.875	0.874	0.875	0.874	0.78
Test	RFC	0.915	0.914	0.915	0.914	0.85
	RFGO	0.926	0.932	0.926	0.925	0.91
	NBC	0.851	0.844	0.851	0.845	0.71
	NBGO	0.904	0.904	0.903	0.903	0.76
All	RFC	0.942	0.942	0.942	0.942	0.87
	RFGO	0.955	0.956	0.955	0.955	0.91
	NBC	0.865	0.863	0.865	0.862	0.73
	NBGO	0.884	0.882	0.884	0.883	0.75

The generation of a contour plot depicted in Fig. 5 serves to elucidate the mean performance of the most exemplary models across the entirety of the database (All), the training stage (Train), and the testing stage (Test). It is noteworthy that the RFGO model, in contrast to its counterparts, manifests its least mean performance in the Test stage, recording a value of 0.9268 according to the recall metric. In subsequent evaluations, the RFGO model achieves a mean value of 0.9525 in the All stage, showcasing a robust performance, while in the Train stage, it exhibits its zenith with a mean value of 0.9641. Conversely, the RFC model presents a mean value of 0.9395 in the All stage, indicating a relatively inferior performance when juxtaposed with the RFGO model. Additionally, the RFC model consistently falls behind in both the Test and Train stages, registering mean values of 0.9142 and 0.9504, respectively.

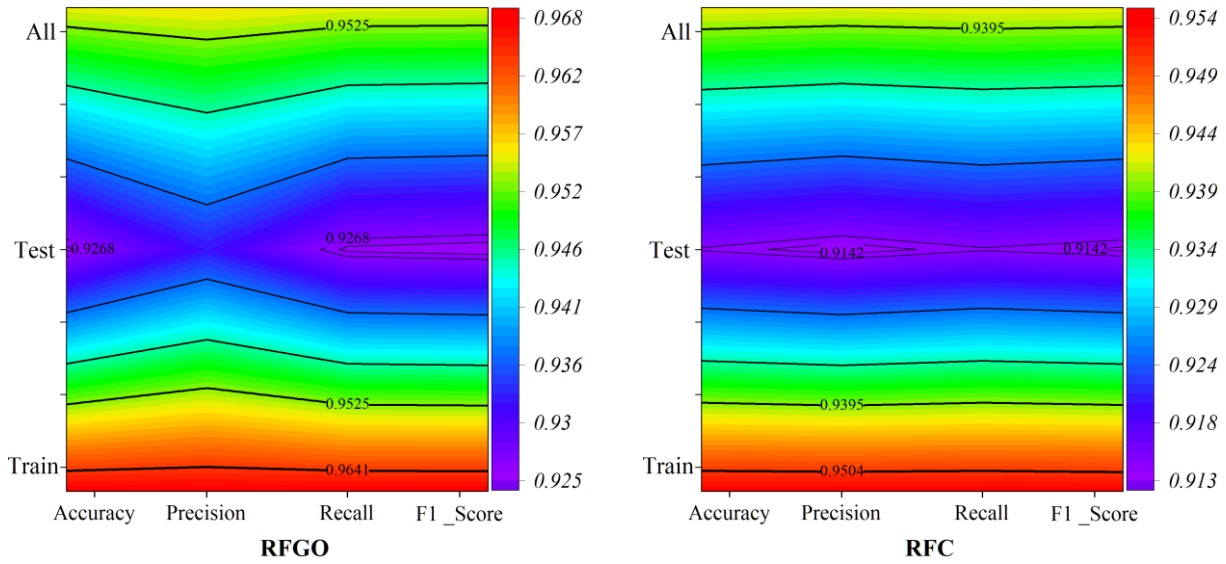

Figure 5: Contour plot visualizing the mean performance scores (accuracy, precision, recall, and F1-score) of each model (RFC, NBC, RFGO, NBGO) across the training, testing, and overall stages. Darker shades indicate higher metric values, emphasizing the superior performance of the RFGO model.

The performance evaluation of hybrid models is conducted across three distinct conditions denoted as DH, N, and SL in Table 3. For instance, under the disc herniation (DH) condition, the RFC model achieves a higher precision of 0.929 compared to 0.914 for the RFGO model. However, the RFGO model compensates with significantly higher recall (0.960 vs. 0.910) and F1-score (0.937 vs. 0.919), indicating a better balance between sensitivity and precision in this diagnostic class. Conversely, in the DH condition, the NBC model outperforms the NBGO model, presenting precision values of 0.851 and 0.816, respectively. In the N condition, relatively weaker precision values are ascribed to the NBC (0.726), NBGO (0.782), and RFC (0.883) models, while the RFGO model excels with a precision value of 0.964, thereby underscoring its superior predictive capacity. Nevertheless, the NBC model demonstrates the highest recall value of 0.993 under the SL condition, ranking first in this stage. Subsequently, the RFC model secures the second position with a recall value of 0.987, while the RFGO and NBGO models share the third position, both achieving a recall value of 0.980. Regarding F1-score values, both RFC and RFGO models exhibit a performance of 0.980 under the SL condition, marking the optimal performance. This contrasts with the F1-score values of the NBC model at 0.958 and the NBGO model at 0.974 under the same condition.

Table 3: Achieved result of the frameworks through the three presented conditions

Metric	Condition	Model			
		RFC	RFGO	NBC	NBGO
Precision	DH	0.929	0.914	0.851	0.816
	N	0.883	0.964	0.726	0.782
	SL	0.974	0.980	0.926	0.967
Recall	DH	0.910	0.960	0.740	0.840
	N	0.883	0.883	0.750	0.717
	SL	0.987	0.980	0.993	0.980
F1-score	DH	0.919	0.937	0.791	0.828
	N	0.883	0.922	0.738	0.748
	SL	0.980	0.980	0.958	0.974

The performance of RFC and NBC models is depicted in Fig. 6 based on the measured values for each condition. For example, under the DH condition, a measured value of 91 out of 100 is attained by the RFC model, showcasing superior performance compared to the NBC model, which is recorded with 74 out of 100. In the N condition, the RFC model achieves the highest value, securing 53 out of 60, in contrast to the NBC model's attainment of 45 out of 60. Lastly, under the SL condition, better performance is demonstrated by the NBC model, recording 149 out of 150 measured values, compared to the RFC model's recording of 148 out of 150.

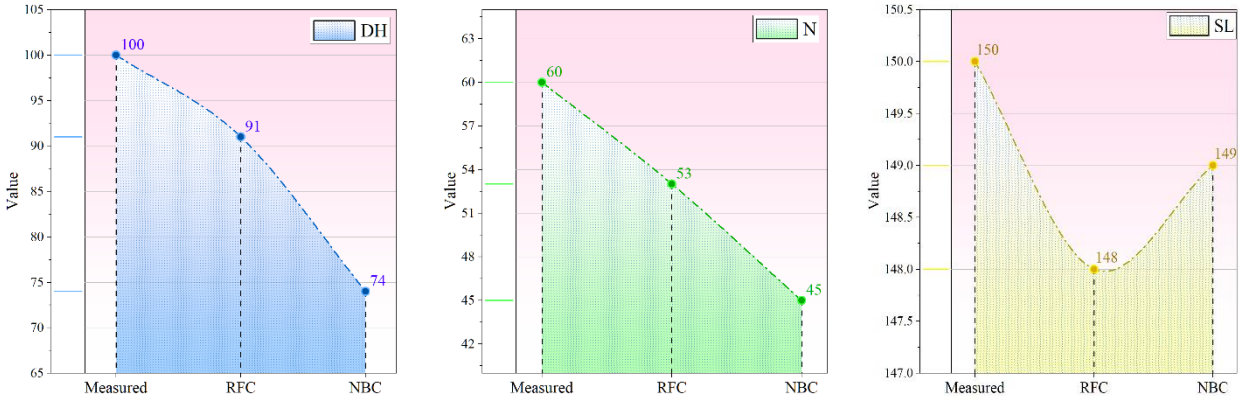

Figure 6: Employing a line-symbol plot for the correlation of the measured and predicted values.

By attributing RFC and its hybrid variant as the exemplary models, the ensuing column plot in Fig. 7 delineates their respective performances based on measured values in DH, SL, and N conditions. For instance, beneath the DH condition, the RFGO model, securing a measured value of 96 out of 100, surpasses the RFC model, documented at 91 out of 100. These models exhibit commensurate performance, both recording 53 out of 60 measured values under the N condition. Nevertheless, in the SL condition, the RFC model, attaining a measured value of 148 out of 150, manifests superior performance in comparison to the RFGO model, which achieves 147 out of 150 measured values.

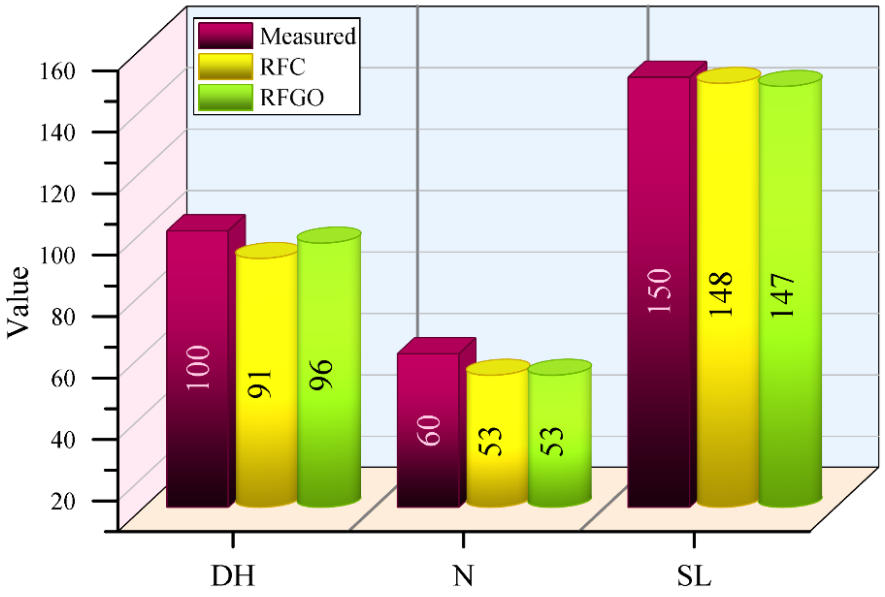

Figure 7: Column plot comparing the number of correctly predicted cases by RFC and RFGO models under three diagnostic conditions: disc herniation (DH), normal (N), and spondylolisthesis (SL). The figure highlights the marginal improvement in prediction accuracy achieved by the hybrid RFGO model.

The subsequent confusion matrix presented in Fig. 8 elucidates the accuracy of RFC and RFGO models' performance, detailing the count of missorted patients across distinct conditions. For instance, the RFC model, attaining 91% accuracy under the DH condition, is associated with the misclassification of 6 patients in the N condition and three in the SL condition. This model correctly classifies the subjects with an accuracy of 88.33% in the N condition but misclassifies six patients in the DH condition and only one in the SL condition. Finally, the RFC model achieves an accuracy of 98.66% under the SL condition, misclassifying one patient in the N condition and 1 in the DH condition.

On the contrary, in the DH condition, the RFGO model gives an accuracy of 96%, misclassifying two patients in SL and 2 in N conditions. In the N condition, this model yields an accuracy of 88.33% while misclassifying one patient in SL and six patients in DH conditions. Finally, under the SL condition, this model achieves an accuracy of 98%, yet it misclassifies 3 patients in the DH condition.

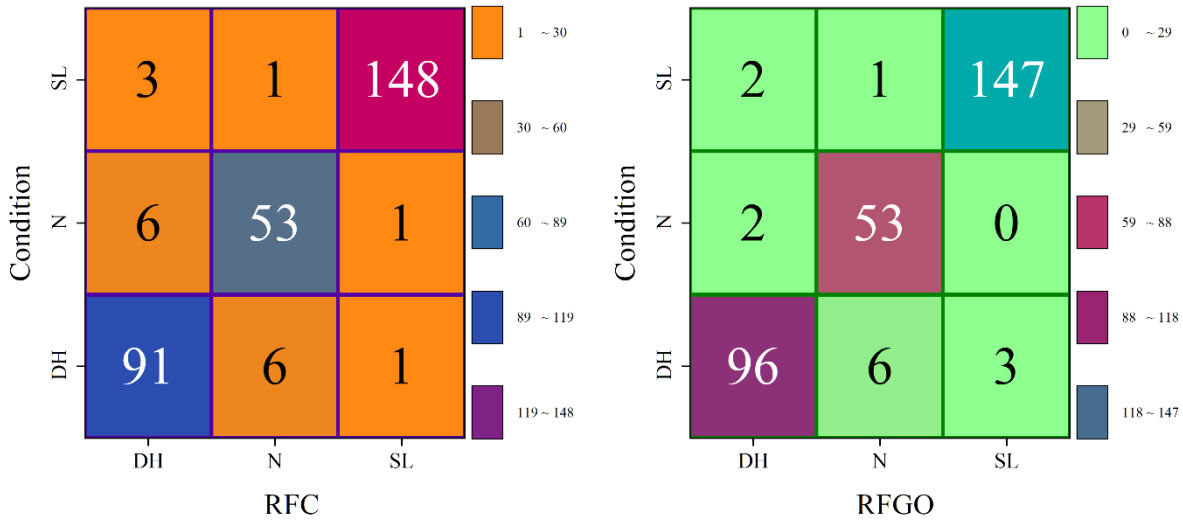

Figure 8: Utilizing a confusion matrix to assess the accuracy of each model.

The ROC curve visually compares the actual positive rate (TPR) against the false positive rate (FPR) at diverse cutoff degrees. Furthermore, it depicts statistical power against the Type I Error of the decision rule, similar to estimators produced from a population sample. The curve primarily depicts sensitivity (recall) about the false positive rate. When the probability spread for true and false positives is known, the ROC curve is calculated using the cumulative distribution functions (CDFs) of the detection probability on the y-axis and the false positive probability on the x-axis. It represents the region under the probability distribution from negative infinity to the discriminating threshold. ROC analysis provides a framework for identifying optimum models while dismissing subpar ones, regardless of cost context or class distribution. It makes a clear and logical link to cost-benefit analysis in the diagnostic arena.

By designating RFGO as the optimal model, its performance is depicted under DH, N, and SL conditions in Fig. 9. As observed, the framework demonstrates its peak performance under the SL condition. The SL conditions vector notably ascends with the highest slope, attaining a TPR of 0.85 within an FPR increment of approximately 0.1, eventually reaching a TPR of 1.0 at an FPR increase of 0.2. This conditions vector surpasses both micro and macro averages. Subsequently, the N conditions vector exhibits a moderate slope, ascending from 0.0 TPR and FPR grades to a TPR of 1.0 before an FPR increase of 0.2. Notably, this condition vector surpasses both micro and macro averages. Conversely, the RFGO model exhibits suboptimal performance under the DH condition. This conditions vector falls below the micro and macro averages and attains a TPR of 1.0 at an FPR increase of 0.8.

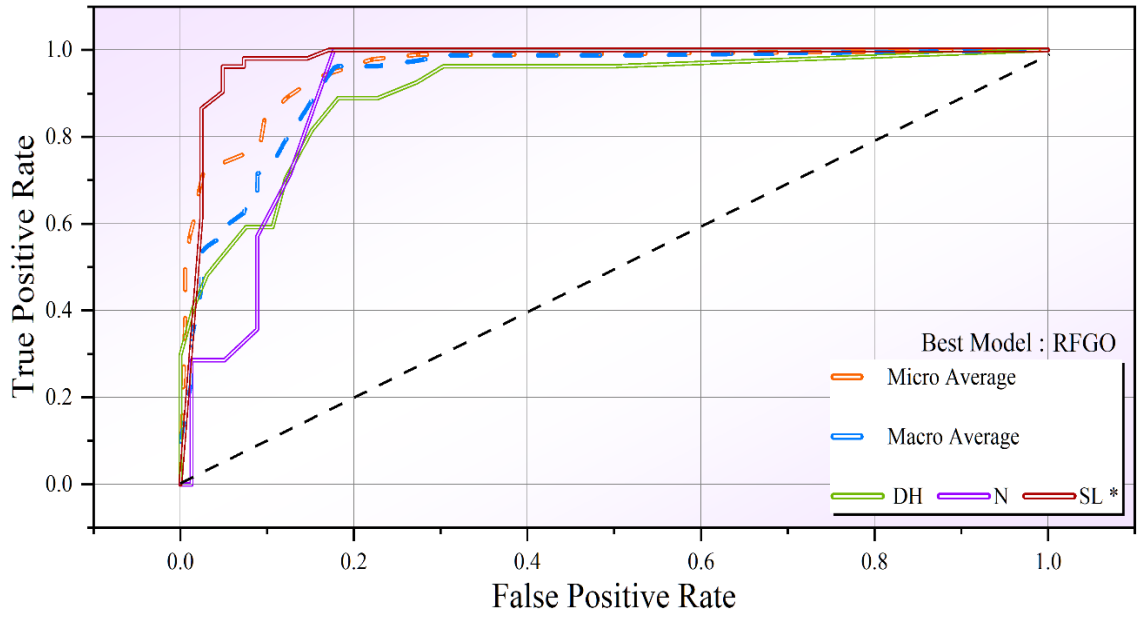

Figure 9: ROC curve for the performance of the most effective hybrid models

4.3 Attributes analysis

To facilitate interpretability and to estimate the contribution of each biomechanical feature to model prediction, SHAP (SHapley Additive exPlanations) analysis was carried out for the top-performing model (RFC-GOA). The values derived from SHAP are an aggregated way to compute feature influence, quantifying the contribution of each input feature to the model's prediction for a specific output class. The relative contribution of each feature for every sample within the dataset is shown through the SHAP summary plot (Fig. 10). The most significant features according to SHAP outcomes for prediction of disc herniation and spondylolisthesis were pelvic tilt, angle of lumbar lordosis, and pelvic incidence, which match well established clinical and biomechanical correlations. These outcomes support the clinical utility of the features selected and enhance the model's interpretability for use in clinical decision-making environments. Following is a clinical interpretation of the leading SHAP-ranked features, coupled with domain knowledge:

- Pelvic Incidence (PI) and spond: PI defines the anatomical relationship between the sacrum and pelvis. SHAP measures reflected that increased PI was significantly correlated with the increased risk for spondylolisthesis, as consistent with its role to facilitate instability within the spine through disturbed distribution of loads between the vertebrae.
- Pelvic Tilt and Spondyl: SHAP analysis verified that excess pelvic tilt was a major contributor to spondylolisthesis predictions. This verifies the previous biomechanical research demonstrating that abnormal angles of tilting change the curvature of the spine and can magnify anterior vertebral displacements.
- Lumbar lordosis angle and disc herniation: An abnormally curved (lordosis) lumbar region is connected with disc deformation. SHAP values identified that this condition significantly contributed to disc herniation predictions. Increased or reduced lordosis can impact intradiscal pressure, causing bulging or rupture of the disc.
- Pelvic Tilt and Disc Herniation: These pelvic tilt changes were found to be predictive for disc herniation, as excess anterior or posterior pelvic tilt displaces biomechanical loads onto intervertebral discs. SHAP scores validated this association, focusing on pelvic tilt as an important modifiable risk factor.
- Pelvic Radius and Sacral Slope: These features ranked below others based on their relative SHAP values, but contributed moderately to model predictions and might contribute to supporting roles in pathology and spinal alignment.

The interpretability analysis using SHAP establishes that the model is based on clinically relevant features, making it more trustworthy and usable for incorporation into clinical workflows.

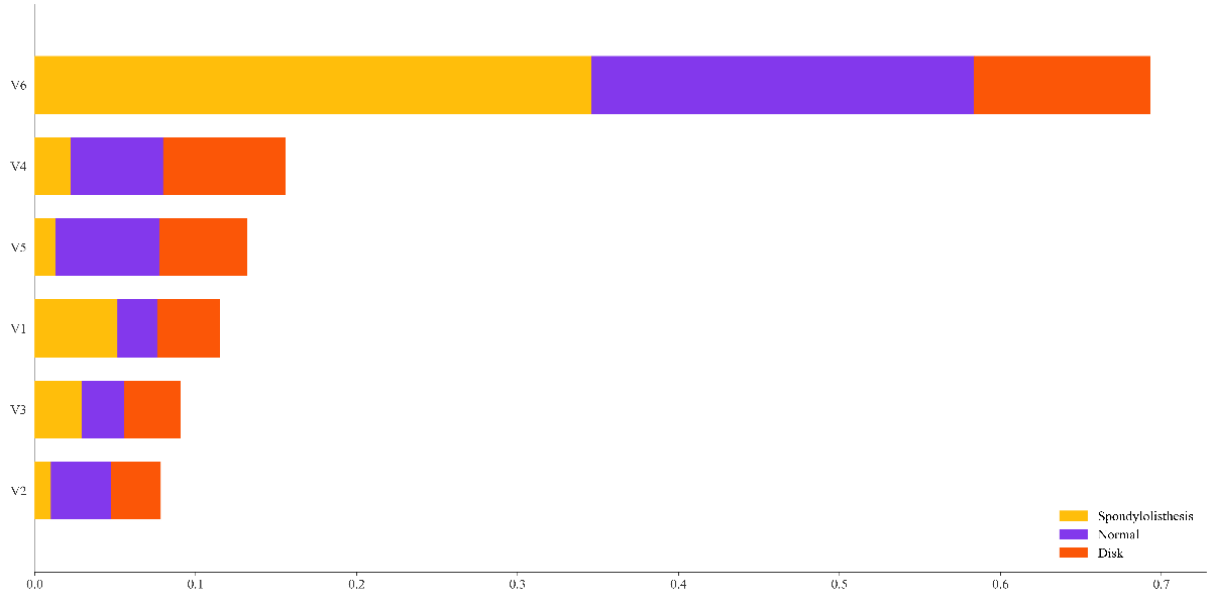

Figure 10: The bar plot for the result of the SHAP sensitivity analysis of the best-performed model

4.4 Comparison with existing works

This study introduces a hybrid machine learning approach for disc herniation prediction, leveraging biomechanical features and optimizing model performance through the Grasshopper Optimization Algorithm (GOA). The proposed RFC-GOA model achieved an accuracy, precision, recall, and F1-score of 95.5%, outperforming both baseline models and comparable approaches in the literature. As shown in Table 1, the RFC-GOA model significantly surpasses traditional statistical and machine learning techniques. For example, Harada et al. (2021) applied logistic regression to clinical features from 600 patient records, achieving 81.2% accuracy. Salehi et al. (2019) used MRI-based features with a Support Vector Machine (SVM), reaching 86.4% accuracy. Although Chen et al. (2023) employed deep learning on a larger

dataset (1,316 patients) and reported 90.1% accuracy, the RFC-GOA approach demonstrated higher accuracy using a substantially smaller dataset of 310 patients. Ren et al. (2024) reported an 88.1% accuracy using a Random Forest model on 1,000 samples. The improved performance observed in the present study is attributed to the integration of GOA, which enabled more effective tuning of hyperparameters such as tree depth and the number of estimators. Compared to the NBC-GOA model (88.4% accuracy), the RFC-GOA model performed notably better, likely due to RFC's capability to model complex, non-linear interactions among correlated biomechanical features—something NBC, with its assumption of feature independence, is less suited for.

Despite these promising results, several limitations must be acknowledged. The dataset size (310 patients) remains relatively small, which may restrict generalizability. Although stratified sampling and 10-fold cross-validation were employed to ensure internal consistency and robustness, the absence of external validation on an independent dataset limits the conclusions that can be drawn regarding model deployment in broader clinical settings. Additionally, the application of GOA introduces greater computational demands, which may present challenges for real-time implementation in clinical environments with limited processing resources. Future research should consider testing the model on larger, multi-center datasets and exploring dimensionality reduction methods to maintain accuracy while reducing computational load.

4.5 Clinical integration and application

While the current study demonstrates strong predictive performance of the proposed models in a controlled experimental setting, real-world clinical validation remains a necessary next step to establish practical utility. For successful integration into clinical workflows, several factors must be considered: usability, data interoperability, model interpretability, and compliance with healthcare regulations. The trained RFC-GOA model can be integrated into clinical decision support systems (CDSS) as a backend module embedded within electronic health record (EHR) platforms. Upon patient intake, relevant biomechanical parameters such as pelvic tilt, pelvic incidence, and lumbar lordosis angle can be automatically extracted from diagnostic imaging systems or manually entered by clinicians. The model can then output a probability score indicating the likelihood of disc herniation or spondylolisthesis, supporting early diagnosis or risk stratification. To facilitate clinical adoption, the following steps are proposed:

- **External Validation**: Future research will involve validating the model on multi-center datasets with diverse populations to assess generalizability across different clinical settings.
- **User Interface Development**: A simple dashboard or plugin could be developed for real-time prediction and visualization of SHAP-based feature attributions, enhancing trust and interpretability for clinicians.
- **Integration with EHRs**: APIs (application programming interfaces) can be designed to pull relevant features directly from patient records and imaging systems, streamlining data flow.
- **Clinical Workflow Simulation**: Simulated deployments can be used to evaluate turnaround time, physician interaction patterns, and the model's influence on clinical decisions.

Furthermore, the inclusion of interpretable outputs such as SHAP visualizations helps ensure that physicians can understand the rationale behind predictions, which is essential for regulatory approval and clinical acceptance. Overall, the proposed model holds promise for integration into pre-diagnosis screening tools and risk assessment frameworks in orthopedic and neurology clinics.

5 Conclusions

ML approaches have been applied to anticipate disc herniation, focusing on the frameworks based on RFC and NBC coupled with the GOA. The present study presents model selection and optimization tactics for improving the anticipation accuracy for medical diagnosis. A blend of these models with the GOA resulted in innovative hybrid models, showing the strengths of each of the components. Among the different results, one can notice the big difference in performance between the RFC model and its hybrid counterpart compared to the NBC model and its hybrid form. The RFC and RFGO models achieved higher accuracy during the training stage, with a rate of 0.954 and 0.968, respectively. Conversely, the NBC and NBGO models reached an accuracy rate of only 0.870 and 0.875, respectively. The vast difference reflects the efficiency of the RFC model and its hybrid model in correctly predicting Disc Herniation. Therefore, this poor performance of the NBC model means it is inefficient for the chosen medical diagnosis task.

These results stressed how proper algorithm selection, together with optimization strategy, becomes a key ingredient for healthcare applications in the task of constructing effective anticipation models. The work contributes to understanding disc herniation anticipations anew. It underlines careful model selection and optimization to deliver ML's full promise for diagnosing medical conditions. New technological advances may continue to allow novel modifications in model design and optimization methodologies to result in more precise and trustworthy forecasts in healthcare. At the final stage, this work will require validation from external independent databases. The frameworks must be applied to databases other than those used for training, establishing their generalizability and reliability on wide-ranging patient populations. The lack of external validation raises concerns about the real-world performance and applicability of the suggested models outside the constraints of the training data.

While this study provides several valuable insights into the application of machine learning for disc herniation prediction, it is important to acknowledge its limitations. One key limitation is the focus on only two algorithmic families (RFC and NBC), which may not capture the full

range of potential model performance. Future studies should expand this scope to include deep learning, ensemble boosting methods, and other probabilistic or hybrid approaches. Additionally, although this study emphasized model interpretability using SHAP analysis, external validation on independent datasets remains a critical next step to confirm the generalizability of the proposed models. Applying the frameworks to different patient populations and clinical settings is essential for establishing real-world reliability. Regarding the broader literature, several common shortcomings persist, including limited transparency in dataset descriptions, minimal discussion of interpretability in many ML frameworks, and the frequent absence of external validation protocols. These gaps underscore the need for more rigorous, clinically aligned machine learning research in spinal diagnostics.

Acknowledgment

This investigation was backed by the second batch of leading talent projects of Jiangsu Vocational Institute of Commerce.

Competing interests

The scholars claim no competing interests.

Authorship contribution statement

Yipin Wang: Writing-Original draft preparation Conceptualization, Supervision, Project administration.
Peng Li: Methodology, Software

Data availability

The scholars will make the raw data supporting this article's conclusions available without undue reservation.

Conflicts of interest

The scholars claimed no conflicts of interest considering this investigation.

Author statement

The manuscript has been read and approved by all the authors, the requirements for authorship, as stated earlier in this document, have been met, and each author believes that the manuscript displays honest work.

Ethical approval

All scholars have been personally and actively involved in substantial work leading to the paper and will take public responsibility for its content.

References

[1] Amin, R.M., N.S. Andrade and B.J. Neuman (2017). Lumbar disc herniation. *Current Reviews in Musculoskeletal Medicine*, Springer, 10, pp. 507–516. https://doi.org/10.1007/s12178-017-9441-4.

[2] Benzakour, T., V. Igoumenou, A.F. Mavrogenis and A. Benzakour (2019). Current concepts for lumbar disc herniation. *International Orthopaedics*, Springer, 43, pp. 841–851. https://doi.org/10.1007/s00264-018-4247-6.

[3] Shin, B.-J (2014). Risk factors for recurrent lumbar disc herniations. *Asian Spine Journal*, National Library of Medicine, 8(2), p. 211. https://doi.org/10.4184/asj.2014.8.2.211.

[4] Shimia, M., A. Babaei-Ghazani, B. Sadat, B. Habibi and A. Habibzadeh (2013). Risk factors of recurrent lumbar disk herniation. *Asian Journal of Neurosurgery*, Thieme, 8(02), pp. 93–96. DOI: 10.4103/1793-5482.116384.

[5] Seidler, A., U. Bolm-Audorff, T. Siol, N. Henkel, C. Fuchs, H. Schug, F. Leheta, G. Marquardt, E. Schmitt and P.T. Ulrich (2003). Occupational risk factors for symptomatic lumbar disc herniation; a case-control study. *Occupational and Environmental Medicine*, BMJ Journals, 60(11), pp. 821–830. https://doi.org/10.1136/oem.60.11.821.

[6] Kreiner, D.S., S.W. Hwang, J.E. Easa, D.K. Resnick, J.L. Baisden, S. Bess, C.H. Cho, M.J. DePalma, P. Dougherty II and R. Fernand (2014). An evidence-based clinical guideline for the diagnosis and treatment of lumbar disc herniation with radiculopathy. *The Spine Journal*, Elsevier, 14(1), pp. 180–191. https://doi.org/10.1016/j.spinee.2013.08.003.

[7] Awad, J.N. and R. Moskovich (2006). Lumbar disc herniations: surgical versus nonsurgical treatment. *Clinical Orthopaedics and Related Research®*, Current Orthopedic Practice, 443, pp. 183–197. DOI: 10.1097/01.blo.0000198724.54891.3a.

[8] Moschetti, W., A.M. Pearson and W.A. Abdu (2009). Treatment of lumbar disc herniation: an evidence-based review, In *Seminars in Spine Surgery*, Elsevier, pp. 223–229. https://doi.org/10.1053/j.semss.2009.08.005.

[9] Buttermann, G.R (2004). Treatment of lumbar disc herniation: epidural steroid injection compared with discectomy: a prospective, randomized study. *JBJS*, The Journal of Bone and Joint Surgery, 86(4), pp. 670–679. DOI:10.2106/00004623-200404000-00002.

[10] Sapiee, N.H (2019). Lumbar Intervertebral Disc Disruption and Herniation: A Biomechanical and Multiscale Structural Investigation. University of Auckland. https://doi.org/10.1016/j.spinee.2014.12.144.

[11] Weber, K.T., T.D. Jacobsen, R. Maidhof, J. Virojanapa, C. Overby, O. Bloom, S. Quraishi, M. Levine and N.O. Chahine (2015). Developments in intervertebral disc disease research: pathophysiology, mechanobiology, and therapeutics. *Current Reviews in Musculoskeletal Medicine*, Springer, 8, pp. 18–31. https://doi.org/10.1007/s12178-014-9253-8.

[12] Adams, M.A. and P.J. Roughley (2006). What is intervertebral disc degeneration, and what causes it? *Spine*, 31(18), pp. 2151–2161. DOI: 10.1097/01.brs.0000231761.73859.2c.

[13] Strömqvist, F., B. Strömqvist, B. Jönsson and M.K. Karlsson (2017). Surgical treatment of lumbar disc herniation in different ages—evaluation of 11,237 patients. *The Spine Journal*, 17(11), Elsevier, pp. 1577–1585. https://doi.org/10.1016/j.spinee.2017.03.013.

[14] Rothoerl, R.D., C. Woertgen and A. Brawanski (2002). When should conservative treatment for lumbar disc herniation be ceased and surgery considered? *Neurosurgical Review*, Springer, 25, pp. 162–165. https://doi.org/10.1007/s101430100184.

[15] Arts, M.P., A. Kuršumovic, L.E. Miller, J.F.C. Wolfs, J.M. Perrin, E. Van de Kelft and V. Heidecke (2019). Comparison of treatments for lumbar disc herniation: systematic review with network meta-analysis. *Medicine*, 98(7), p. e14410. DOI: 10.1097/MD.0000000000014410.

[16] Gebremariam, L., B.W. Koes, W.C. Peul and B.M. Huisstede (2012). Evaluation of treatment effectiveness for the herniated cervical disc: a systematic review. *Spine*, 37(2), pp. E109–E118. DOI: 10.1097/BRS.0b013e318221b5af.

[17] Fujita, N., S. Ishihara, T. Michikawa, K. Azuma, S. Suzuki, O. Tsuji, N. Nagoshi, E. Okada, M. Yagi and T. Tsuji (2020). Potential association of metabolic and musculoskeletal disorders with lumbar intervertebral disc degeneration: Cross-sectional study using medical checkup data. *Journal of Orthopaedic Science*, Elsevier, 5(3), pp. 384–388. https://doi.org/10.1016/j.jos.2019.05.011.

[18] Francisco, V., J. Pino, M.Á. González-Gay, F. Lago, J. Karppinen, O. Tervonen, A. Mobasheri and O. Gualillo (2022). A new immunometabolic perspective of intervertebral disc degeneration. *Nature Reviews Rheumatology*, Nature, 18(1), pp. 47–60. https://doi.org/10.1038/s41584-021-00713-z.

[19] An, H.S., K. Masuda and N. Inoue (2006). Intervertebral disc degeneration: biological and biomechanical factors. *Journal of Orthopaedic Science*, Springer, 11, pp. 541–552. https://doi.org/10.1007/s00776-006-1055-4.

[20] Lv, X. and H. Chang (2025). Applying Mathematical Modeling Optimization Algorithms to Solve Shop Floor Scheduling Problems. *Informatica*, Slovenian Society Informatika, 49(19). https://doi.org/10.31449/inf.v49i19.7113.

[21] Liu, H., P. He, Z. Lu, J. Li and Z. Lu (2025). Deep Learning-Based Defect Identification for Transmission Tower Bolts: Optimization of YOLOv3 and ResNet50 Algorithms. *Informatica*, Slovenian Society Informatika, 49(19). https://doi.org/10.31449/inf.v49i19.7872.

[22] Li, L (2025). Click-Through-Rate Prediction Using Deep Neural Networks and Efficient Channel Attention Mechanisms. *Informatica*, Slovenian Society Informatika, 49(19). https://doi.org/10.31449/inf.v49i19.7947.

[23] Vilca-Huayta, O.A (2025). Logistic Sigmoidal and Neural Network Modeling for COVID-19 Death Waves. *Informatica*, Slovenian Society Informatika, 49(19). https://doi.org/10.31449/inf.v49i19.7476.

[24] Pedersen, C.F., M.Ø. Andersen, L.Y. Carreon and S. Eiskjær (2022). Applied machine learning for spine surgeons: predicting outcome for patients undergoing treatment for lumbar disc herniation using PRO data. *Global Spine Journal*, Sage Publications, 12(5), pp. 866–876. https://doi.org/10.1177/2192568220967643.

[25] Wirries, A., F. Geiger, A. Hammad, M. Bäumlein, J.N. Schmeller, I. Blümcke and S. Jabari (2022). AI Prediction of Neuropathic Pain after Lumbar Disc Herniation—Machine Learning Reveals Influencing Factors. *Biomedicines*, MDPI, 10(6), pp. 1319. https://doi.org/10.3390/biomedicines10061319.

[26] Salehi, E., H. Yousefi, H. Rashidi and H. Ghanaatti (2019). Automatic Diagnosis of Disc Herniation in Two-Dimensional MR Images with Combination of Distinct Features Using Machine Learning Methods, In *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, IEEE, Istanbul, Turkey, pp. 1–6. https://doi.org/10.1109/EBBT.2019.8742052.

[27] Ren, G., L. Liu, P. Zhang, Z. Xie, P. Wang, W. Zhang, H. Wang, M. Shen, L. Deng and Y. Tao (2024). Machine learning predicts recurrent lumbar disc herniation following percutaneous endoscopic lumbar discectomy. *Global Spine Journal*, Sage Publications, 14(1), pp. 146–152. https://doi.org/10.1177/21925682221097650.

[28] Harada, G.K., Z.K. Siyaji, G.M. Mallow, A.L. Hornung, F. Hassan, B.A. Basques, H.A. Mohammed, A.J. Sayari, D. Samartzis and H.S. An (2021). Artificial intelligence predicts disk re-herniation following lumbar microdiscectomy: Development of the "RAD" risk profile. *European Spine Journal*, Springer, 30(8), pp. 2167–2175. https://doi.org/10.1007/s00586-021-06866-5.

[29] Yen, H.-K., P.T. Ogink, C.-C. Huang, O.Q. Groot, C.-C. Su, S.-F. Chen, C.-W. Chen, A. V Karhade, K.-P. Peng and W.-H. Lin (2022). A machine learning algorithm for predicting prolonged postoperative opioid prescription after lumbar disc herniation surgery. An external validation study using 1,316 patients from a Taiwanese cohort. *The Spine Journal*, Elsevier, 22(7), pp. 1119–1130. https://doi.org/10.1016/j.spinee.2022.02.009.

[30] Ebrahimzadeh, E., F. Fayaz, F. Ahmadi and M. Nikravan (2018). A machine learning-based method in order to diagnose lumbar disc herniation disease by MR image processing.

MedLife Open Access, 1(1), pp. 1. https://doi.org/10.1080/01431160412331269698.

[31] Chen, Y., F. Lin, K. Wang, F. Chen, R. Wang, M. Lai, C. Chen and R. Wang (2023). Development of a predictive model for one-year postoperative recovery in patients with lumbar disc herniation based on deep learning and machine learning. Research Square, https://doi.org/10.21203/rs.3.rs-2860039/v1.

[32] Parmar, A., R. Katariya and V. Patel (2019). A review on random forest: An ensemble classifier, In *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018*, Springer, Cham, pp. 758–763. https://doi.org/10.1007/978-3-030-03146-6_86.

[33] Aulia, A., A. Rahman and J.J. Quijano Velasco (2014). Strategic well test planning using random forest, In *SPE Intelligent Energy International Conference and Exhibition*, SPE, Utrecht, Netherlands, p. SPE-167827. https://doi.org/10.2118/167827-MS.

[34] Pal, M (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, Taylor & Francis, 26(1), pp. 217–222. https://doi.org/10.1080/01431160412331269698.

[35] Genuer, R., J.-M. Poggi and C. Tuleau (2008). Random Forests: some methodological insights. *ArXiv Preprint ArXiv:0811.3619*, Cornell University. https://doi.org/10.48550/arXiv.0811.3619.

[36] Nigam, K., A.K. McCallum, S. Thrun and T. Mitchell (2000). Text classification from labeled and unlabeled documents using EM. *Machine Learning*, Springer, 39, pp. 103–134. https://doi.org/10.1023/A:1007692713085.

[37] Dempster, A.P., N.M. Laird and D.B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, 39(1), pp. 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.

[38] Yang, X., Z. Song, I. King and Z. Xu (2022). A survey on deep semi-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, 35(9), pp. 8934–8954. https://doi.org/10.1109/TKDE.2022.3220219.

[39] Rigutini, L., M. Maggini and B. Liu (2005). An EM based training algorithm for cross-language text categorization, In *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*, IEEE, Compiegne, France, pp. 529–535. https://doi.org/10.1109/WI.2005.29.

[40] Liu, B., W.S. Lee, P.S. Yu and X. Li (2002). Partially supervised classification of text documents, In *ICML*, Sydney, NSW, pp. 387–394. DOI:10.1385/1-59259-358-5:387.

[41] Aggarwal, C.C. and C.C. Aggarwal (2017). *An introduction to outlier analysis*. Springer. https://doi.org/10.1007/978-3-319-47578-3_1.

[42] Montazeri, Z., T. Niknam, J. Aghaei, O.P. Malik, M. Dehghani and G. Dhiman (2023). Golf Optimization Algorithm: A New Game-Based Metaheuristic Algorithm and Its Application to Energy Commitment Problem Considering Resilience. *Biomimetics*, MDPI, 8(5), pp. 386. https://doi.org/10.3390/biomimetics8050386.