OPEN ACCESS ISSN: 2835-9496
Leelarungrayub J, et al. (2024): 390
DOI: 10.47991/2835-9496/AJCMR-120

Research Article

American Journal of Clinical and Medical Research

Comparison of Algorithms for Classification and Prediction of Inspiratory Muscle Weakness-Based on the Orange Data Mining Tool

Jirakrit Leelarungrayub Ph.D[1,2*], Pongkorn Chantaraj Ph.D[2], Supattanawaree Thipcharoen Ph.D[2]

[1]Department of Physical Therapy, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai 50200, Thailand
[2]Department of Data Science and Digital Innovation, Faculty of Innovation Technology and Creativity, The Far Eastern University, Chiang Mai, 50100 Thailand.

[*]**Corresponding author:** Jirakrit Leelarungrayub, Ph.D, Department of Physical Therapy, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai 50200, Thailand. Tel: (+66)53936072; Fax: (+66)53936042; Email: donrawee.leela@cmu.ac.th

Citation: Leelarungrayub J, Chantaraj P, Thipcharoen S (2024) Comparison of Algorithms for Classification and Prediction of Inspiratory Muscle Weakness-Based on the Orange Data Mining Tool. Ameri J Clin Med Re: AJCMR-120.

Received Date: 10 February, 2024; **Accepted Date:** 19 February, 2024; **Published Date:** 23 February, 2024

Abstract

Background: *This work aims to discover which machine learning algorithms have the highest accuracy and precision in the classification and best prediction of inspiratory muscle weakness (IMW) or non-weakness (non-IMW).*

Methods: *Data from multiple datasets, including anthropometrics such as age, weight, height, body mass index (BMI), waist, and lung function (forced vital capacity, FVC and maximal inspiratory mouth pressure, PImax) from 200 participants were analyzed in data science procedures; data collection, pre-processing, and data analysis on classification and prediction the IMW and non-IMW groups under the Orange Data Mining Tool.*

Results: *After the outlier identification with the Radivz visualization technique, the final 171 data were added to the data analysis. The following information was provided for 62 IMW and 109 non-IMW participants: average age (28.9±8.3 & 29.3±9.5 years), weight (54.3±5.8 & 60.1±8.2 kg), height (1.6±0.9 & 1.7±0.1 m), BMI (20.9±1.4 & 21.3±1.3 kg.m^{-2}), waist (29.1±2.9 & 30.3±3.5 inches), FVC (2.94±0.5 & 3.3±0.4 L), and PImax (66.1±8.8 & 104.0±17.4 cmH_2O). In contrast to K-Nearest Neighbor (KNN), Neural Network, and Naïve Bayes model, the best-classified models with the highest accuracy and precision were Logistic Regression, Random Forest, and Decision Tree. In addition, the confusion matrix showed the highest predicted proportion for both IMW and non-IMW classes was a Decision Tree model. The association result revealed a significant ($p<0.05$) relationship between PImax and the FVC, height, weight, waist, and BMI. Therefore, this study shows that the Decision Tree is the best model for classifying and predicting inspiratory muscle weakness in the clinic.*

Keywords: *Algorithms; Inspiratory muscle weakness; Machine learning; Orange Data Mining.*

Introduction

Classification and Prediction of inspiratory muscle weakness are very important for clinical practice and research [1]. The clinical approach to identifying respiratory dysfunction before presenting the symptoms; for example, sleep apnea, morning headache, and hypersomnolence or disease is very challenging [2]. Eighty centimeters of water (cmH_2O) of maximal inspiratory mouth pressure (PImax) has been reported to be a cut-off point to exclude significant inspiratory muscle weakness (IMW) in healthy individuals between the ages of 18 and 80 years old (ATS/ERS, 2002) [3]. In contrast, patients with chronic obstructive pulmonary disease (COPD) have been suggested to have less than 60 cmH_2O [4]. Thus, the main clinical choice on respiratory weakness or non-weakness can be categorized. Furthermore, there were different reference equations for PImax [5,6] and these were correlated with variables like height, weight, BMI, forced expiratory volume at one second (FEV1), peak expiratory flow (PEF), and forced vital capacity (FVC) [7,8], as well as six-minute walking distance [9]. Regretfully, no evidence has been found to support the association between these variables and the final classification of inspiratory muscle weakness or non-weakness in the data. In general, data mining is helpful for mining and delving deeply into data in many formats to gain patterns and knowledge discovery (KD). It is one of many study strategies for categorizing and predicting some target variables [10]. The prior study demonstrated how different data sets can be presented as having an impact on target categories using classification algorithms including multilayer perception, Naïve Bayes, Sequential Minimal Optimization (SMO), and Decision Tree [11]. For instance, the best-predicted model method for predicting fruit sweetness using artificial intelligence was logistic regression, as demonstrated by research on the Orange tool [12]. Currently, additional research is still needed to determine which algorithms can be used for classification and prediction, particularly when estimating the worth of new databases using data mining models [13].

Materials and Methods

The data in this study was sampled for the initial trial performance in a big data experimental research project that was integrated with the data analysis process and ethically approved by the Faculty of Associated Medical Sciences Ethic Committee, Chiang Mai University, Chiang Mai, Thailand (Study Code: AMSEC-61EX-096). The Orange data mining tool has been used for prediction accuracy checks as well as classification and analysis of those forecasts. In the field of data science research, this tool-data mining and structural algorithms-is highly beneficial for machine learning algorithmic applications.

Sample Size and Lung Function Evaluation

The sample size in this study was calculated by the G*Power program (version 3.0.10) and was used to analyze F-test (multiple regression; Omnibus (R2 deviation from zero). The

effect size of f^2 at 0.15, α error prob of 0.05, Power (1-β err prob) at 0.80, and six variables (FVC, age, weight, height, BMI, and waist) were applied to calculate and predict the PImax. Consequently, it is necessary to gather a minimum of 180 healthy volunteers for the sample size. Furthermore, to avoid having an inadequate sample size, an additional 10-15% was included, and 200 people in total were recruited. The target participants were recruited from the Sansai area and Center in Chiang Mai province. They did not have a history of respiratory diseases such as asthma, chronic lung disease, or bronchitis, nor of other diseases, neurological disorders, or thoracic deformities. They were also non-athletic healthy (more than 10 hours of exercise training per week) and had quit smoking or had quit at least a year ago [14]. Additionally, individuals were excluded if they had a history of influenza or severe acute respiratory syndrome coronavirus (SARS-CoV) 30 days before data collection [15]. Following their approval of the study procedure, each participant signed a permission form.

Anthropometric and Lung Function Evaluation

At the Sansai Hospital in Chiang Mai, Thailand, all data were gathered. All tests were placed in a closed setting at a controlled temperature of 26 degrees Celsius. A digital scale (TANITA Corporation, Tokyo, Japan) was used to measure the anthropometric measurements of body weight, while a stadiometer (Physician, AD, Medical, Inc., USA) measuring millimeters was used to determine the height. Weight divided by height squared ($Kg.m^{-2}$) was used to compute the Body Mass Index (BMI). Using a non-stretch tap horizontally positioned at the navel level during expiration, the waist circumference (in inches) was manually measured [16]. Spirometry (Easy on-PC Spirometry, and Medical Technologies, Zurich, Switzerland) was used to assess the FVC by the accepted guidelines [17]. The respiratory pressure meter (MicroRPM) (CareFusion, UK 232 Ltd, United Kingdom) was used to evaluate the PImax under the established guidelines of the ATS/ERS Statement of Respiratory Muscle Testing (ATS/ERS, 2022) [18]. The simulation process depicted in **Figure 1** consists of three steps: (1) data cleaning for outliers, (2) data processing for accuracy and prediction analysis, correlation, and (3) statistical analysis**.**

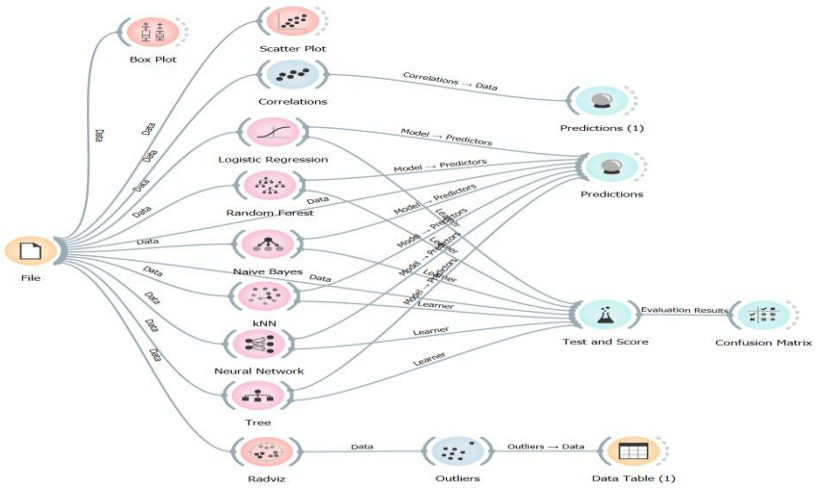

Figure 1: Simulation workflow of the model in the Orange Data Mining tool.

Data Cleansing Processing

The data from 200 participants were analyzed using the radial coordinate visualization (*Radviz*) technique [19, 20] to separate the participants' clinical and to clean outliers using the Euclidean metric. The participants' characteristics included age, weight, height, body mass index, waist, and lung function (FVC and PImax). Prior research revealed that the Euclidean matrix is one of the several options for computer techniques used in outlier detection, which can be broadly classified into five categories: distance-based, density-based, tree-based, clustering-based, and neural network-based techniques [21].

Classification Algorithm Model and Prediction Evaluation

The prediction accuracy of inspiratory muscular strength weakness (IMW) (PImax < 80 cmH_2O) or non-IMW (PImax ≥ 80 cmH_2O) was assessed using several algorithms. Various tools for classification models were chosen, including Decision Tree, Random Forest, K-Nearest Neighbor (KNN), Neural Network, Naïve Bayes, and logistic regression. The test and score on classification accuracy (CA) and prediction were analyzed on the Confusion matrix. For a detailed explanation of each algorithm, consider the following: Naïve Bayers is a straightforward probabilistic classifier based on applying the Bayes theorem with strong independence assumptions [22]; KNN is frequently used in pattern recognition and data mining for classification purposes due to its simplicity and low error rate [23]; Logistic Regression is used to generate observations about a set of categories and changes the output by using a sigmoid logistic function to return the possible value; Neural networks function similarly to networks of neurons that interpret sensory data using machine perception, tagging, or input methods [24]; and Decision trees are represented by a structure resembling a tree, in which each branch indicates a potential test result and each inner mode represents a test for a specific feature [25].

Statistical Analysis

The classification, prediction, correlation, and statistical analysis were performed in the Orange Data Mining tool. The area under the curve (AUC), classification accuracy (CA), weight average of precision and recall (F1), Precision, and the Matthews correlation coefficient (MCC) between models were observed. The percentages of prediction of the actual and predicted respiratory muscle weakness and non-weakness on the confusion matrix from each model were also compared. Finally, the correlation between the dataset and PImax was analyzed by a Pearson correlation test, as same as the student t-test in the Boxplot visualization was evaluated with a significant level of 0.05.

Results

A cut-off value of 80 cmH_2O was used to classify the 200 participants in the study into two groups: clinical IMW (n = 80) and non-IMW (n = 120). The participants' average age was 29.7±9.6 years, weight was 58.1±8.7 Kg, height was 1.6±0.1 m, BMI was 21.2±1.4 $Kg.m^{-2}$, waist was 20.0±3.6 inches, FVC was 3.4±0.7 litter, and PImax was 92.9±27.2 cmH_2O. Following the data set collection, Radiviz visualization was used to clean the data and identify any outliers, as seen in **Figure 2**. Data processing was completed with the final dataset of 171 individuals (**Table 1**). The box-plot presentation with a statistical analysis comparing IMW and non-IMW revealed that, except for age and height ($p>0.05$), there were significant differences in weight, waist, body mass index, FVC, and PImax ($p<0.05$).

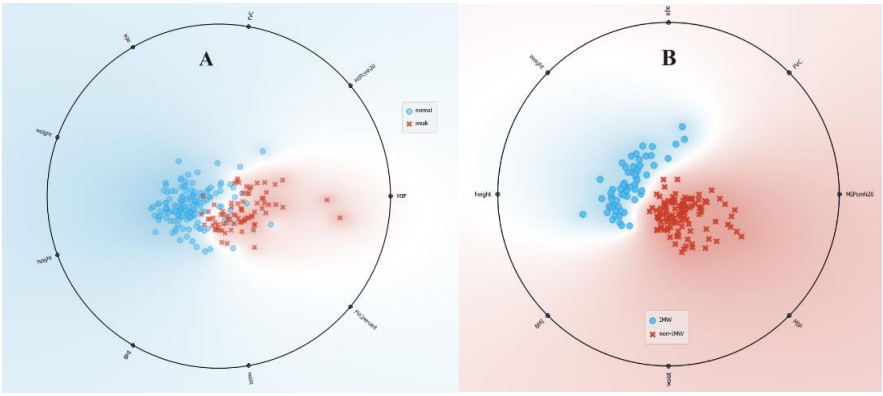

Figure 2. Pre-processing of original data (A) and after cleansing data (B) with Radviz visualization and outlier cleansing with Euclidean metric.

Table 1: Characteristics and lung function between participants in two groups.

	Inspiratory muscle weakness (IMW) (n=62)	Non-Inspiratory muscle weakness (non-IMW) (n =109)	p
Characteristics			
Age (years)	28.9±8.3	29.3±9.5	0.79
Height(m)	1.6±0.9	1.7±0.1	0.06
Weight (kg)	54.3±5.8	60.1±8.2	0.00
Waist (inch)	29.1±2.9	30.3±3.5	0.02
Body mass index (BMI)	20.9±1.4	21.3±1.3	0.04
Lung function			
FVC (L)	2.9±0.5	3.3±0.4	0.00
FVC (% predicted)	82.5±8.5	92.2±9.2	0.00
PImax (cmH_2O)	66.1±8.8	104.0±17.4	0.00
Note: FVC = forced vital capacity, PImax = maximal inspiratory mouth pressure.			

Table 2 compares the six algorithms evaluated on all datasets to the categorized target of IMW and non-IMW. The results showed that after comparison work outperformed six techniques in classifiers, the area under the curve (AUC), classification accuracy CA), weighted average of precision and recall (F1), Precision, Recall, and the Matthews correlation coefficient (MCC) were shown. The highest accuracy of 100% and precision of 100% were provided by Logistic Regression, Random Forest, and Decision Tree. On the other hand, the accuracy scores of 99.4%, 97.7%, and 84.8% for KNN, Neural Network, and Naïve Bayes were fewer than those of the precision scores, which were 99.4%, 97.7%, and 86.8%, respectively.

Table 2: Performance comparison for classification algorithm models.

Model	AUC	CA	F1	Precision	Recall	MCC
Logistic Regression	1.000	1.000	1.000	1.000	1.000	1.000
Random Forest	1.000	1.000	1.000	1.000	1.000	1.000
Decision Tree	1.000	1.000	1.000	1.000	1.000	1.000
KNN	1.000	0.994	0.994	0.994	0.994	0.987
Neural Network	0.998	0.977	0.977	0.977	0.977	0.949
Naïve Bayes	0.952	0.848	0.851	0.868	0.848	0.701
Note: AUC = Area under the curve, CA = classification accuracy, F1 =weighted average of precision and recall, MCC = The Matthews correlation coefficient. KNN = K-Nearest Neighbor						

Following the interpretation of the classification model, the confusion matrix was used to examine the percentage of prediction on each model for both the actual and predicted classes of IMW and non-IMW. In a matrix, the prediction error is determined by off-diagonal entries, whereas the accurate prediction is determined by diagonal elements. IMW and non-IMW class precision are displayed in **Table 3** based on the results. When compared to other models, such as Random Forest (98.4% & 100%), Logistic Regression (95.2% & 97.2%), KNN (92.3% & 98.1%), Neural Network (89.2% & 96.2%), and Naïve Bayes (72.2% & 94.6%). The Decision Tree model demonstrated the best prediction at 100% in both IMW and non-IMW between predicted and actual data. The correlation analysis revealed a significant ($p<0.05$) relationship between PImax and the FVC (+0.47) (**Figure 3**), height (+0.45), weight (+0.44), waist (+0.34), and BMI (+0.21). In addition, the statistical analysis on FVC and PImax showed a moderate positive correlation ($r = +0.45$). Furthermore, statistical analysis revealed a significant difference ($p<0.01$) in FVC and PImax between the IMW and non-IMW groups (**Figure 4**).

Table 3: Prediction analysis with Confusion matrix.

(1) Decision Tree

		Predicted		
		IMW	Non-IMW	∑
Actual	IMW	100%	0.00%	62
	Non-IMW	0.0%	100%	109
	∑	62	109	171

(2) Random Forest

		Predicted		
		IMW	Non-IMW	∑
Actual	IMW	98.4%	0.0%	62
	Non-IMW	1.6%	100.0%	109
	∑	62	109	171

(3) Logistic Regression

		Predicted		
		IMW	Non-IMW	∑
Actual	IMW	95.2%	2.8%	62
	Non-IMW	4.8%	97.2%	109
	∑	62	109	171

(4) K-Nearest Neighbor (KNN)

		Predicted		
		IMW	Non-IMW	∑
Actual	IMW	92.3%	1.9%	62
	Non-IMW	7.7%	98.1%	109
	∑	62	109	171

(5) Neural Network

		Predicted		
		IMW	Non-IMW	∑
Actual	IMW	89.2%	3.8%	62
	Non-IMW	10.8%	96.2%	109
	∑	62	109	171

(6) Naïve Bayes

		Predicted		
		IMW	Non-IMW	∑
Actual	IMW	72.2%	5.4%	62
	Non-IMW	27.8%	94.6%	109
	∑	62	109	171

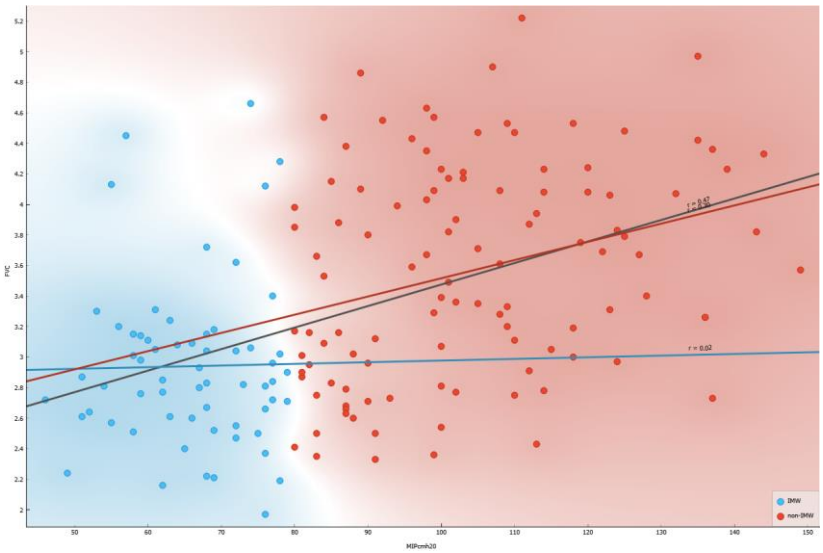

Figure 3: A scatter plot of the correlation between FVC and PImax (r= + 0.47), with a clinical cut-point off value at 80 cmH2O, separates the inspiratory muscle weak (IMW) (blue color) and non-IMW (red color) groups.

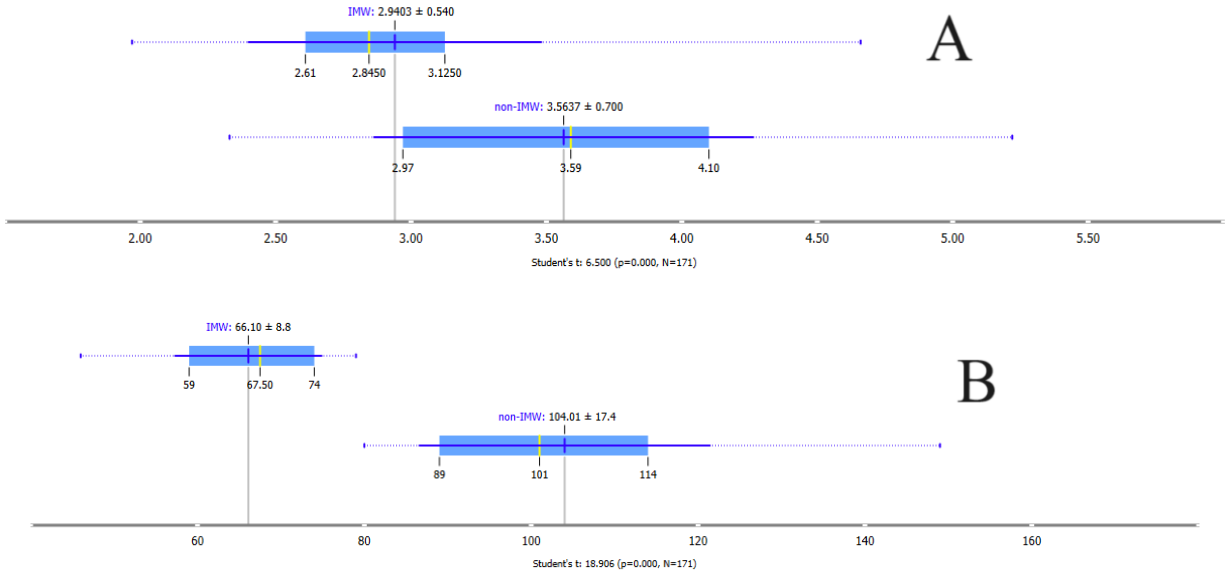

Figure 4: Statistics comparing the two groups of IMW (inspiratory muscle weakness) and non-IMW (non-inspiratory muscle weakness) for forced vital capacity (FVC) (A) and PImax (B).

Discussion

This work describes a unique statistical computer analysis method called data-mining machine learning (ML) that can be used for both medical research and rehabilitation. The basic step of dataset analysis, the first step of data cleansing, is very important. Data outliers have an impact on the analysis's findings. To efficiently assess highly dimensional data sets, the radial coordinate visualization (*Radviz*) technique has been frequently employed [19,20] and outlier cleansing with Euclidean metric. This metric can be used to evaluate the outlier roughly categorized into distance-bases, density-based, tree-bases, clustering bases, and neural network-based methods [21]. Building a model that correctly predicts the class labels of new instances based on their features is the aim of classification. Therefore, two types of classification; binary and multi-classes are performed. Before developing a classification, model using different techniques like correlation analysis, information gain, and principal component analysis, preprocessing to ensure the quality of the data, such as handling outliers, missing values, or data transformation, should be taken into consideration after data collection [26]. The primary tool utilized in this work was Orange, however, other tools like RapidMiner, Spark, or Weka were also can be used [27]. In a prior study, the orange tool was employed to analyze data and demonstrated how the environment affects students' lives [28]. According to ATS/ERS (2002) standard criteria, the study's cut-off point for respiratory muscle weakness and non-weakness was established. The PImax value transformation data at 80 cmH_2O for healthy individuals aged 18–80 years, classified as either IMW or non-IMW following the ATS/ERS guideline [18].

This procedure, which divided the target categories into IMW and non-IMW groups, displayed the straightforward outlier protocol using a Radviz technique visualization. Multi-dimensional data was applied to the 2D design through the analysis of complicated datasets with outliers and all variables, including age, weight, height, waist, BMI, FVC, and PImax [29]. As a result, after cleansing (**Figure 1. B**) as opposed to before cleansing (**Figure 1. A**), this procedure produced the clearest results. The present procedure aligns with prior research indicating that impure data may occasionally impact a model's classification accuracy in German credit data, BUPA liver illnesses, Johns Hopkins Ionosphere, and Pima Indians Diabetes [30].

Six algorithms, such as Decision Tree, Random Forest, Neural Networks, K-Nearest Neighbor, Naïve Bayes, or Logistic Regression, were chosen as classifiers in the classification model analysis of this study, as shown in **Table 2**. Every classifier displayed variations in its application; for instance, a set of tree classifiers that combined to use a random vector sampled separately from the input vector in Random Forest, which contained both observed and unobserved nodes in Naïve Bayes [31]. A previous review of the literature revealed that these models were used in a variety of fields, including business, agriculture, health, general (text documents), and education [32]. Regretfully, it is currently unknown which algorithm can be applied to classify respiratory muscle weakness. The best results of the area under the curve (AUC), classification accuracy (CA), weighted average of precision and recall (F1), precision, recall, and the Matthews correlation coefficient (MCC) should be compared. Previous data reported that the AUC ranges from less than 0.6 (very poor model) to more than 0.9 (excellent model) as well as the sensitivity or specificity of the model from recall and F1 results. In addition, the MCC is a special case of the coefficient for binary classification that presents the perfect classification [33]. Therefore, a high percentage in three models; Logistic Regression, Decision Tree, and Random Forest showed an excellent model to use to classify the respiratory muscle as either weak or non-weak.

As a result, the analysis of the classification with various models such as logistic regression, Random Forest, Decision Tree, Naïve Bayes, KNN, and Neural Network must be included. Numerous research has demonstrated the need to highlight each algorithm's accuracy percentage. For instance, the Decision Tree proved to be the most effective classifier for Dengue disease, heart disease, and diabetes [34]. Naïve Bayes was used to propose an approach to chronic disease with 90% accuracy [35]. The three most accurate algorithms, as determined by the study's results, were Decision Tree, Random Forest, and Logistic

Regression. Therefore, these classifiers or algorithms can be used to classify inspiratory muscle weakness.

The criteria of data mining analysis on a supervised approach is the prediction of the target attribute value as the IMW or non-IMW class from the attribute values such as characteristic and lung function. Using a confusion matrix or error matrix between the actual and predicted outcomes, the prediction analysis findings for the IMW and non-IMW classes are determined. According to a prior study using datasets for type-2 diabetes and the general population, the confusion matrix expressing the true positive/negative and false positive/negative yields the greatest results on the supervised machine learning algorithms [36]. The study's findings demonstrated that, when compared to Random Forest, Logistic Regression, KNN, Neural Network, and Naïve Bayes, among other models, the Decision Tree model provided the best prediction at 100% in both IMW and non-IMW between predicted and actual values. The Decision Tree, which is essentially a complex relationship between input factors and target variables, has been shown in a prior study to be a potent statistical tool for categorization, prediction, interpretation, and data manipulation in medical research [37]. Predictive accuracy reacts to the overall ensemble of numerous decision trees, and Random Forest is an ensemble of many decision trees. This study demonstrated a lesser percentage of prediction on Random Forests when compared to Decision Trees because of the low datasets, but the Decision Tree is suitable for small datasets [38]. Therefore, this is consistent with a previous report that documented the advantages of Decision Tree as simple, easy to display, capable of handling both numerical and categorical data, requiring little data preparation, and performing well [39].

The study's final step was to assess the factors' correlation. The findings of the association indicated a substantial ($p<0.05$) relationship between PImax and the FVC, height, weight, waist, and BMI. Furthermore, a moderately positive connection ($r = +0.45$) was obtained from the statistical analysis of the FVC and PImax data. Furthermore, statistical analysis revealed a significant difference ($p<0.01$) in FVC and PImax between the IMW and non-IMW groups. The multiple regression analysis and correlation data from earlier studies were validated by these findings. Previous research revealed that numerous reference PImax equations for healthy people have been published [5, 6]. Numerous factors that are associated with respiratory muscle strength have been studied, including height, weight, BMI, FVC, age and weight [40], BMI [41], and six-minute walking distance [9]. Therefore, the prediction of the weakness for inspiratory muscle should be performed by measuring a PImax value. Unfortunately, an expansive machine to determine PImax is the clinical barrier. Possibly, classification and prediction with accuracy, sensitivity, precision, false-positive rate, and f-measure from characteristics and lung function can performed under the best data-mining algorithm [13]. This is the practicality used in clinical research, particularly in conditions like blood disorders, skin conditions, and breast cancer diagnosis [42].

Conclusion

In this study, the clinical inspiratory muscle weakness is predicted and analyzed using classification algorithms on a dataset containing 171 subjects. The model was created using a few clinical variables, including FVC and PImax, as well as age, weight, height, waist, and BMI. The greatest algorithms for data mining classifiers are Logistic Regression, Random Forest, and Decision Tree, all of which achieve 100% accuracy, specificity, sensitivity, and excellent models for classifying respiratory muscle weakness or non-weakness. However, regarding clinical concerns, a Decision Tree is the best predictor of whether inspiratory muscle weakening will occur. Additionally, there is a noteworthy variation between the groups and a moderately positive connection between FVC and PImax in the dataset. Therefore, clinical uses of data mining and machine learning technologies for the supervised classification and prediction of inspiratory muscle weakness can be applied in the future.

Limitation of study

The imbalance and low sample sizes between the groups with and without weaknesses could affect the statistical analysis's findings and overfitting in some models. Furthermore, the participants ranged in age from 19 to 50, so the results might not apply to other range years, like being younger than 19 or older than 50. Additionally, the primary goal of this work was to show how to apply particular Orange Data Mining with different machine learning algorithms to the categorization and prediction of two kinds of respiratory weakness; hence, further research is required to determine the therapeutic utility of the formula.

Acknowledgments

We also express our gratitude to the Far Eastern University team in Chiang Mai, Thailand, for their insightful recommendations and excellent guidance.

Contributions from the authors

Leelarungrayub. J. handled the data collecting, analysis, and interpretation as well as drafted the manuscript and submission. Before submitting, Chantaraj, P., and Thipcharoen, S., double-checked the text and gave the important details in the final draft and manuscript.

Completing interest

The authors declare that there is no conflict of interest regarding the publication of this article.

References

1. Laveneziana P, Albuquerque A, Aliverti A, Babb T, Barreiro E, Dres M, Dube BP, Fauroux B, Gea J, Guenette JA, et al. ERS statement on respiratory muscle testing at rest and during exercise. The European Respiratory Journal. 2019; 53(6): 1801214.
2. Laghi F, Tobin MJ. Disorders of the respiratory muscle. American Journal of Respiratory and Critical Care Medicine. 2003; 168(1): 10-48.
3. American Thoracic Society/European Respiratory Society. ATS/ERS. Statement on respiratory muscle testing. American Journal of Respiratory and Critical Care Medicine. 2022; 166(4): 518-624.
4. Lotters F, van Tol B, Kwakkel G, Gosselink R. Effects of controlled inspiratory muscle training in patients with COPD: a meta-analysis. The European Respiratory Journal. 2002; 20 (3): 570-576.
5. Souto-Miranda S, Jácome C, Alves A, Machado A, Paixão C, Oliveira A, et al. Predictive equations of maximum respiratory mouth pressures: a systematic review. Pulmonology. 2021; 27(3): 219-239.

6. Bairapareddy KC, Augustine A, Alaparthi GK, Hegazy F, Shousha TM, Ali SA. Nagaraja R, Chandrasekaran B. Maximal respiratory pressures and maximum voluntary ventilation in young Arabs: Association with anthropometrics and physical activity. Journal of Multidisciplinary Healthcare. 2021; 14: 2923-2930.
7. Hautmann H, Hefele S, Schotten K, Huber RM. Maximal inspiratory mouth pressure (PIMAX) in health subjects-what is the lower limit of normal?. Respiratory Medicine. 2000; 94(7): 689-693.
8. Sriboonreung T, Leelarungrayub J, Yanakai A, Puntumetakul R. Correlation and Predicted Equations of MIP/MEP from the Pulmonary Function, Demographics and Anthropometrics in Healthy Thai Participants aged 19-50 Years. Clinical Medicine Insights: Circulatory, Respiratory and Pulmonary Medicine. 2021; 15: 11795484211004494.
9. de Souza Y, Suzana ME, Medeiros S, Macedo J, da Costa CH. Respiratory muscle weakness and its association with exercise capacity in patients with chronic obstructive pulmonary disease. The Clinical Respiratory Journal. 2022; 16(2): 162-166.
10. Dissanayake K, Johar MGM, Ubeysekara NH. Data Mining Techniques in Disease Classification: Descriptive Bibliometric Analysis and Visualization of Global Publications. International Journal of Computing and Digital Systems. 2023; 13(1): 289-301.
11. Tan H. Machine Learning Algorithm for Classification. Journal of Physics: Conference Series. 2021; 1994: 012016.
12. Al-Sammarraie MAJ, Gierz L, Przybyl K, Koszela J, Szychta M, Brzykcy J, Baranowska HM. Predicting Fruit's Sweetness Using Artificial Intelligence-Case Study. Orange. Applied Sciences. 2022; 12(16): 8233.
13. Kaur P, Singh M, Josan GS. Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector. Procedia Computer Science. 2015; 57: 500-508.
14. Enright PL, Kronmal RA, Manolio TA, Schenker MB, Hyatt RE. Respiratory muscle strength in the elderly: correlates and reference values. American Journal of Respiratory Critical Care Medicine. 1994; 149(2 Pt 1): 430-438.
15. Lista-Paz A, Langer D, Barral-Fernandez M, Quintela-del-Rio A, Gimeno-Santos E, Arbillaga-Etxarri A, Torres-Castro R, Casamitjana JV, de la Fuente AB, et al. Maximal Respiratory Pressure Reference Equations in Healthy Adults and Cut-off Points for Defining Respiratory Muscle Weakness. Archives de Bronconeumologia. 2023; 59(12): 813-820.
16. Camhi SM, Bray GA, Bouchard C, et al. The relationship between of waist circumference and BMI to visceral, subcutaneous, and total body fat: sex and race differences. Obesity (Silver Spring). 2011; 19(2): 402-408.
17. Culver BH, Graham BL, Coates AL, Wanger J, Berry CE, Clarke PK, Hallstrand TS, Hankinson JL, Kaminsky DA, Maclntyre NR, et al. Recommendations for a standardized pulmonary function report. An Official American Thoracic Society Technical Statement. American Journal of Respiratory and Critical Care Medicine. 2017; 196(11): 1463-1472.
18. ATS/ERS statement on respiratory muscle testing. ATS/ERS statement on respiratory muscle testing. American Journal of Respiratory and Critical Care Medicine. 2002; 166 (4): 518-624.
19. Caro LD, Frias-Martinez V, Frias-Martinez E. Analyzing the Role of Dimension Arrangements for Data Visualization in Radviz. In: Zaki, M.J., et al (Eds). PAKDD. Part II, LNAI .2010; 6119: 125-132.
20. Marchette DJ, Solka, J.L. Using data images for outlier detection. Computational Statistics & Data Analysis. 2003; 43(4): 541-552.
21. Park CH. A comparative study for outlier detection methods in high dimensional test data. JAISCR. 2023; 13(1): 5-17.
22. Padhiyar H, Rekh P. An Improved expectation maximization based semi-supervised email classification using Naïve Bayes and K-Nearest Neighbor. International Journal of Computer Applications. 2014; 101(6): 7-11.
23. Kuang I, Zhao L. A practical GPU based kNN algorithm. In Proceedings of the International Symposium on Computer Science and Computational Technology (ISCSCI 2009), Hyderabad, India, 15-17 January 2009; Academy Publisher: Orlando, FL, USA. P.151.
24. Ahmad HM, Sohail M, Ahmad MM, Iqbal S, Sarfaraz A, Noor K. Predictions of Pneumonia Disease using Image Analytics in Orange Tool, In Proceedings of the GS International Conference on Computer Science and Engineering 2020 (GSICCSE 20), Beijing China, 22-24 May 2020.
25. Che D, Liu Q, Rasheed K, Tao X. Decision tree and ensemble learning algorithms with their applications in bioinformatics. Advances in Experimental Medicine and Biology. 2011; 696: 191-199.
26. Kesavaraj G, Sukumaran S. A Study on Classification Techniques in Data Mining. 4th ICCCNT. IEEE-31661.2013.
27. Padmavaty V, Geetha C, Priya N. Analysis of data mining tool orange. International Journal of Modern Agriculture. 2020; 9(4): 1146-1150.
28. Tiwari R, Kumar G, Gunjan VK. Effect of Environment on Students Performance Through Orange Tool of Data Mining. Proceedings of the 4th International Conference on Data Science, Machine Learning and Applications. ICDSMLA 2022. 2023; 1038: pp.283-292.
29. Ventocilla E, Riveiro M. A comparative user study of visualization techniques for cluster analysis of multidimensional data sets. Information Visualization. 2020; 19(4): 318-338.
30. Jeatrakul P, Wong KW, Fung CC. Data Cleansing for Classification Using Misclassification Analysis. Journal of Advanced Computational Intelligence and Intelligent Informatics. 2010; 14(3): 297-302.
31. Good I J. Probability and the Weighing of Evidence[R]. London: C. Griffin, 1950.
32. Wati M, Haeruddin, Indrawan W. Predicting Degree-Completion Time with Data Mining. 3rd International Conference on Science in Information Technology (ICSITech), Bandung, 25-26 October 2017.
33. Chicco D, Jurman G. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. BioData Mining. 2023; 16: 4.
34. Pandey, Anand Kishor, And Dharmveer Singh Rajpoot. "A Comparative Study of Classification Techniques By Utilizing Weka." Signal Processing and Communication (Icsc), 2016 International Conference On. IEEE. 2016.

35. Salama Gouda I, MB. Abdelhalim, And Magdy Abd-Elghany Zeid. "Experimental Comparison of Classifiers for Breast Cancer Diagnosis." In Computer Engineering & Systems (ICCES), 2012 Seventh International Conference On, pp.180-185. IEEE. 2012.
36. Ebrahim OA, Derbew G. Application of supervised machine learning algorithms for classification and prediction of type-2 diabetes disease status in Afar regional state, Northeastern Ethiopia 2021. Scientific Reports. 2023; 13(1): 779.
37. Song YY, Lu Y. Decision tree methods: applications for classification and prediction. Shanghai Archives of Psychiatry. 2015; 27(2): 130-135.
38. Touw WG, Bayjanov JR, Overmars L. Data mining in the Life Sciences with Random Forest: a walk in the part or lost in the jungle?. Briefings in Bioinformatics. 2012; 14(3): 315-326.
39. Matzavela V, Apepis E. Decision tree learning through a Predictive Model for Student Academic Performance in Intelligent M-learning environments. Computers and Education: Artificial Intelligence. 2021; 2(6): 100035.
40. Wilson SH, Cooke NT, Edwards RH, Spiro SG. Predicted normal values for maximal respiratory pressures in Caucasian adults and children. Thorax. 1984; 39(7): 535-538.
41. Sanchez FF, da Silva CDA, de Souza Pereira Gama Maciel MC, Marques JRD, de Leon EB, Goncalves RL. Predictive equations for respiratory muscle strength by anthropometric variables. The Clinical Respiratory Journal. 2018; 12(7): 2292-2299.
42. Mia MR, Hossain SA, Chhotan AC, Chakraborty NR. A Comprehensive Study of Data Mining Techniques in Healthcare, Medical, and Bioinformatics. International Conference on Computer, Communication, Chemical Material and Electronic Engineering (IC4ME2), Rajshahi, 8-9 February 2018.