Research Inventy: International Journal of Engineering And Science
Issn (e): 2278-4721, Issn (p):2319-6483, PP: 29-32
www.researchinventy.com

Orthopedic Patient Classification withMachine Learning Algorithms Based On Biochemical Features

[1]Kabita Patel

[1]*Department of Computer Science, Research Scholar, G.M. University, Sambalpur, Odisha, India*

[2]Sudipta Priyadarshinee

[2]*Department of Computer Science, Research Scholar, G.M. University, Sambalpur, Odisha, India*

Abstract
Due to the increased frequency of traffic accidents, cycling, exercises, and other similar behaviours, the number of orthopaedic patients is on the rise. Data from lab tests and medical equipment is utilised to make diagnoses and respond to clinically demanding issues. Machine learning classifiers were used to classify orthopaedic patients based on their biomechanical characteristics in this study. Where to forecast categorical outcome, the classification algorithm uses one or more feature values. A target variable's categorization outcome can be binary or multinomial in nature. Medical data, like any other dataset, can be utilised to train machine learning algorithms for issue dataset classification. Seven classification algorithms applied in this paper were Random Forest, logistic model tree (LMT), Logistic regression, Linear discriminant analysis (LDA), Stochastic Gradient Descent (SGD), alternating decision tree (AD Tree), Quadratic Discriminant Analysis (QDA). The suggested work outperforms previous strategies for Biochemical characteristics are used to classify orthopaedic patients, according to the results.

Keywords – Orthopaedic; Machine Learning; Prediction

I. Introduction

Machine learning algorithms are used to solve two different types of problems: prediction and categorization. A continuous result is predicted by a prediction algorithm using more than one special values. To forecast category result, the system of categorization algorithm uses one or more feature values. A target variable's categorization outcome might be multinomial or binary in nature. The Logistic Regression approach may be Modelling a generalised result for classification [1], and many of its applications can be establish in the medical area. For binary classification, logistic regression is commonly utilised, but it may also be used for multinomial classification [2]. Decision trees [3] [4] are another frequently used machine learning technique for classification. They are excellent in order to classify binary data but may also be applied to multiclass problems. The supervised machine learning method KNN – K-Nearest Neighbour [5] may be used to address both regression and classification problems. By calculating distance between two points, KNN performs categorization based on similarity found in close proximity.

The healthcare sector has requested improvements [6] computer diagnosis is used in electronic record management, and data transformation as a result of the problems it experiences in embracing technology. The healthcare industry has embraced machine learning algorithms for categorization. This study uses a machine learning system to classify orthopaedic patients as normal or abnormal. On the same dataset, different machine learning algorithms can be used; however, which ML method would perform better in comparison relies on the data's nature [7] and the pre-processing given to the data. Here Seven classification techniques Random Forest, logistic model tree (LMT), Logistic regression, Stochastic Gradient Descent (SGD), Linear discriminant analysis (LDA), alternating decision tree (AD Tree), Quadratic Discriminant Analysis (QDA) are applied on the dataset and compared.

II. Literature Review

Machine learning has made a huge impact on the medical area, according to many research. Many complexities may be avoided by using machine learning to predict illnesses early on. Machine learning algorithms are the most complex method for intelligently addressing various stages of illnesses and making the best conclusion.

One study took data from magnetic resonance imaging (MRI) and categorised it into three categories: disc hernia, spondylolisthesis, and normal. They produced better results by using SVM and other classifiers, The best results were obtained using a feed forward back propagation neural network [8]. The authors relied on decision trees, random forests, naive Bayes, and logistic regression, classifiers as a foundation. The suggested categorization method identified severe instances within each patient. In [9], a technique for early diagnosing chronic kidney disease was proposed. The authors employed K-means clustering to prepare the data. The classification algorithms K-Nearest Neighbour, Support vector machine, and Naive Bayes were employed on progressed features. The highest accuracy of 97.8% was achieved using classification algorithms. Almasoud and Ward [10] published CKD research that included random forest, logistic regression, SVM, and gradient boostingapproaches.

Machine learning methods including logistic regression, decision trees, and K-NN are utilised to categorise biochemical characteristics of orthopaedic patients into three categories: hernia, normal, and spondylolisthesis. Data is scaled through standardisation before being exposed to machine learning algorithms. The data is divided 70:30 into train and test data. The confusion matrix and measures such as recall, precision, F1-score, and accuracy are used to assess the performance of each of the three approaches. The parametric technique is used in logistic regression, but the non-parametric approach is used in decision trees and KNN. The accuracy score of logistic regression is found to be higher than that of the decision tree and the K-NN technique for this dataset [11].

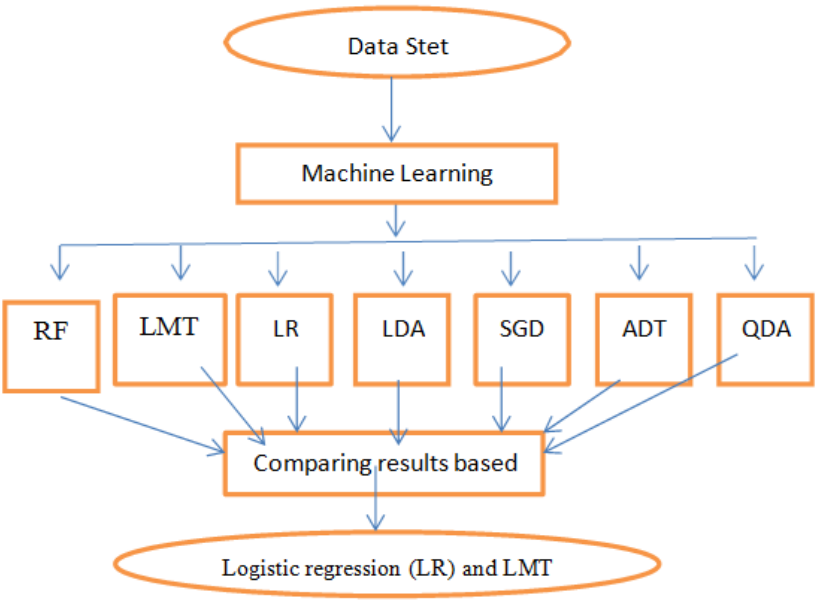
III. Data and Processing of Data

The biological properties of an orthopaedic patient dataset are used to categorise patients as normal or abnormal, based on the form and alignment of the pelvis and lumbar spine. There are three hundred ten rows and seven columns in the dataset: pelvic incidence, pelvic tilt numeric, lumbar lordosis angle, sacral slope, pelvic radius, degree spondylolisthesis, and class. Except for the column-class, which is a string, all columns are numeric. There are two distinct attributes in the column class- 'Abnormal', and 'Normal'. The dataset contains 200 patients from the Abnormal class and 100 individuals from the Normal class. The target variable is represented by column class, while the feature vectors are represented by the remaining columns. There are no duplicate values or null values in this dataset.

IV. Proposed Framework

In this research we apply seven classification models: Random Forest, logistic model tree (LMT), Stochastic Gradient Descent (SGD), Logistic regression, Linear discriminant analysis (LDA), alternating decision tree (AD Tree), Quadratic Discriminant Analysis (QDA) to predict the orthopaedic disease. Then there are the algorithms for machine learning employed on the dataset and estimated rate of accuracy. Each experiment is subjected to 10-fold cross validation to ensure that the results are free of bias. The primary goal was to find the method that could be the best classifier.

Figure 1. Methodology workflow



V. Machine Learning AlgorithmsRandom Forest –

Random Forest is a well-known supervised learning model approach used in machine learning. A classifier is Random Forest that improves the dataset's prediction precision by averaging the outcomes of several on different subsets of the dataset, decision trees. Rather than having a single decision tree, the random forest takes into account the predictions of each tree and forecasts the final result focused on the amount of votes.

logistic model tree (**LMT**) –

A logistic model tree (LMT) is a classification model that incorporates both logistic regression and decision tree learning, as well as supervised training. In the logistic variant, the Logit Boost technique is accustomed to create an LR dummy at each branch in the tree, again bust using the C4.5 criterion. The results of the parent node are used to warm up each Logit Boost execution. Finally, the tree is pruned.

Logistic regression –

Logistic regression explores all datasets for the hyper plane that best matches the data for class identification. The "logistic function" lies at the heart of logistic regression. The sigmoid function is another name for the logistic function. This function was designed to describe the feature of population expansion in bioecology, like speedy increase and reaching the environment's bring capacity. It's a 'S'-shaped slope capable of converting a genuine integer to a range between 0 and 1. The function given below:

1 / (1 + e^ − value)

The base of natural logarithms is 'e,' and value is the numerical value we would like to convert.

Linear discriminant analysis (LDA) –

For supervised classification techniques, Linear Discriminant Analysis (LDA) is a dimensionality reduction technique. It's used to show team differences like splitting two or more classes. It's a technique for projecting higher-dimensional properties onto a lower-dimensional area.

Stochastic Gradient Descent (SGD) -

Stochastic gradient descent is a machine learning optimization technique for finding pattern specifications that better match predicted and proper results. It's a clumsy but productive technique. Stochastic gradient descent is apopular technique in machine learning.

Alternating decision tree (AD Tree) –

An alternate decision tree (AD Tree) is a classification machine learning approach. It is related to boosting and generalises decision trees. An AD Tree is made up of a series of decision nodes that indicate a condition and prediction nodes that carry a single integer.

Quadratic Discriminant Analysis (QDA) -

The only difference between quadratic and linear discriminant analysis is that we relaxed the implication that the mean and covariance of all classes were same. As a result, we had to compute it separately.

VI. Results and Discussion

The experiment is run on the given dataset using a variety of machine learning methods. In this section, we evaluate the efficacy of all algorithms in terms of occurrences that have been properly categorised, incorrectly classified instances, and accuracy. Table 1 displays the outcomes.

Table 1 Classification Accuracy of Classifiers

Algorithm	Correctly classified instances	Incorrectly classified instances	Accuracy (%)
Random Forest	84.83	263	47
LDA	84.19	261	49
Logistic regression	85.48	265	45
SGD	84.19	261	49
LMT	85.48	265	45

AD Tree	82.58	256	44
QDA	82.9	257	53

From Table 1, it is found that all classifier performed well and logistic regression and LMT achieved the highest accuracy of 85.48%.

..

VII. Conclusion

The number of orthopaedic patients is on the rise as a result of an increase in traffic accidents, cycling, exercises, and other similar behaviours. This study employed machine learning classifiers to categorise orthopaedic patients based on their biomechanical properties. This study looked at the seven most important machine learning arrangement processes for forecasting the orthopaedic patients. From the experiment it is found that logistic regression and LMT achieved the highest accuracy of 85.48%.

References

[1]. Audrone Jakaitiene, Nonlinear Regression Models , in Encyclopedia of Bioinformatics and Computational Biology, 2019
[2]. Joseph T. Hefner, Kandus C. Linde, Analytical Methods, in Atlas of Human Cranial Macromorphoscopic Traits, 2018
[3]. T. Mitchell, "Decision Tree Learning", in T. Mitchell, Machine Learning, The McGraw-Hill Companies, Inc., 1997, pp. 52-78.
[4]. P. Winston, "Learning by Building Identification Trees", in P. Winston, Artificial Intelligence, Addison-Wesley Publishing Company, 1992, pp. 423-442.
[5]. M. Goldstein, Kn -nearest Neighbor Classification, Mathematics, Computer Science, IEEE Trans. Inf. Theory (1972), DOI:10.1109/TIT.1972.1054888, Corpus ID: 30697197
[6]. Sumeet Dua, U. Rajendra Acharya, Prerna Dua, Machine Learning in Healthcare Informatics, DOI https://doi.org/10.1007/978-3-642-40017-9, Springer-Verlag Berlin Heidelberg 2014, Print ISBN978-3-642-40016-2
[7]. M. Bichler, C. Kiss, A Comparison of Logistic Regression, k-Nearest Neighbor, and Decision Tree Induction for Campaign Management, Published in AMCIS 2004
[8]. P. Srimani and M. Koti, "Medical Diagnosis Using Ensemble Classifiers - A Novel Machine-Learning Approach", Journal of Advanced Computing, 2013.
[9]. Akben, S. Early stage chronic kidney disease diagnosis by applying data mining methods to urinalysis, blood analysis and disease history. IRBM 2018, 39, 353–358.
[10]. Almasoud, M.; Ward, T.E. Detection of chronic kidney disease using machine learning algorithms with least number of predictors. Int. J. Soft Comput. Appl. 2019, 10, 89–96.
[11]. Desai, Chitra. "Classification of Orthopaedic Patients Based On Biochemical Features Using Machine Learning Algorithms." Journal of Critical Reviews 7 (2020): 3823-3828.
[12]. Dhamodharan, S. (2014, May). Liver disease prediction using bayesian classification. In 4th National Conference on Advance Computing, Application Technologies.
[13]. Joshi, J., Doshi, R., & Patel, J. (2014). Diagnosis and prognosis breast cancer using classification rules. International Journal of Engineering Research and General Science, 2(6), 315-323.