International journal of application or innovation in engineering and management

Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/902-909

ISSN: 2319-4847

Evaluating the Effectiveness of Machine Learning Algorithms for Enhancing Kyphosis Disease Prediction

[1] V. Vasudha, [2] P.V. Devika,

[1]Assistant Professor, Megha Institute of Engineering & Technology for Women, Ghatkesar.
[2] MCA Student, Megha Institute of Engineering & Technology for Women, Ghatkesar.

Article Info

Received: 30-04-2025 ***Revised: 06 -06-2025*** ***Accepted: 17-06-2025*** ***Published:28/06/2025***

Abstract:
Common colloquialisms for kyphosis, which is defined as an inward arching of the upper back, include "roundback" and "hunchback" when the curvature is more pronounced. In most cases, compression or fractures in the spine cause this illness to manifest. Spinal anomalies or a gradual twisting of the spinal bones may cause additional types of kyphosis in children or teens. Kyphosis is more common in adolescence; however, it may appear at any age. It may be caused by a variety of things, including bad posture, developmental problems, and spinal anomalies. This study presents a machine learning strategy for kyphosis illness prediction, with the goals of bettering patient outcomes via earlier diagnosis. The goal of this study is to examine and compare various granularity levels of machine learning algorithms applied to biological data, including Decision Trees and Random Forests. The results highlight the importance of ML as a useful tool for dealing with biological issues in a wider sense. Various terms related to spinal curvature and kyphosis include decision trees, random forests, and machine learning.

I. INTRODUCTION

The "machine learning" branch of AI finds hidden patterns in datasets and uses them to build algorithms. Without explicit task programming, these algorithms use learned patterns to predict new data that resembles past data. When paired with statistical approaches, the results predicted by traditional machine learning provide useful insights. Picture and audio identification, recommendation engines, language analysis, automated chores, fraud detection, portfolio optimization, and many more disciplines are just a few of the many different applications of machine learning. Autonomous vehicles, drones, and robots are all examples of machine learning models that can adapt to their surroundings and gain intelligence.

International journal of application or innovation in engineering and management
Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/902-909
ISSN: 2319-4847

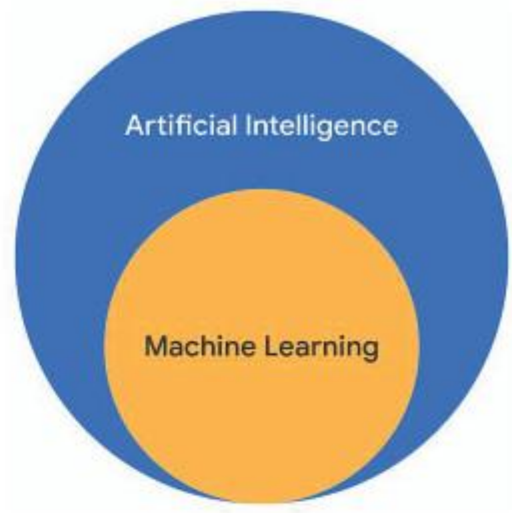

Figure 1: Artificial Intelligence in conjunction with Machine Learning

Reinforcement learning, supervised learning, unsupervised learning, and semi supervised learning are the four main categories of machine learning algorithms. Supervised learning, in which some of the training data acts as a teacher and directs the algorithm to determine the model, is the main emphasis of this study. [4]. a) Supervised learning: In this method, one computer knows the correct inputs and outputs (the ground truth) and learns to predict the outputs using these inputs. When using supervised learning, the best way to find the input-output mapping is to minimize the loss function, which shows how far the machine's predictions deviate from the ground truth. Medical research often makes advantage of this form of learning. b) In unsupervised learning, the system learns from its incoming data without reference to a known ground truth. Using the inputs as a starting point, this learning exercise may extract additional knowledge by identifying patterns and traits. Unsupervised learning has several uses, one of which is clustering. c) Reinforcement Learning: Instead of beginning with ground truth data, this technique involves providing feedback on the task's execution accuracy after it has been completed. This kind of criticism may either encourage or discourage further action. A growing number of clinical decision-makers are turning to reinforcement learning, which has already shown success in dynamic or interactive settings like video games. Research tools for studying how animals and people learn the causal structure of events and tasks include reinforcement learning models. The third.

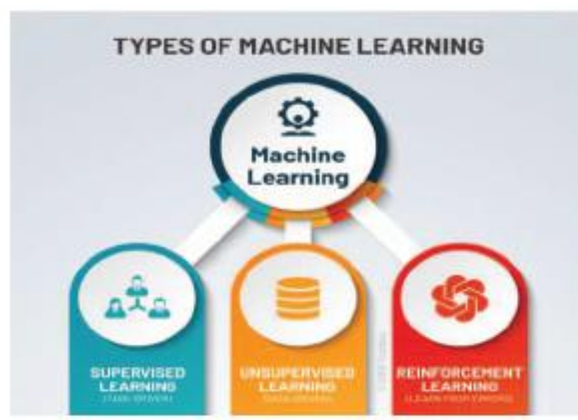

Figure2: Different Machine Learning Types

Machine learning (ML) successfully tackles the diagnostic issues posed by kyphosis, a spinal ailment [8]. Machine learning starts with engineering and feature selection, which optimizes predictive parameters using data collected from patients' demographics, medical records, and spine measures. Training minimizes error by adjusting parameters, but model selection takes dataset complexity into account (e.g., decision trees and random forests). Model performance is ensured by evaluation measures, and deployment follows the integration of healthcare systems in a lawful manner. The capacity to adapt to changing demographics and medical

International journal of application or innovation in engineering and management
Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/902-909
ISSN: 2319-4847

knowledge is crucial for continuous monitoring. It provides accurate insights that may be used to better patient care. The physical and mental tolls of kyphosis, a spinal curvature abnormality, are substantial. The need of developing predictive models to diagnose and intervene early makes them critical for reducing the impact of this illness. Spinal bone deterioration, compression, or cracking is a common cause of kyphosis, an abnormal curvature of the spine that manifests as an abnormally arched upper back, particularly in the elderly [6]. Spinal abnormalities or progressive spinal bending may cause other types of kyphosis to appear in children and teenagers [5].

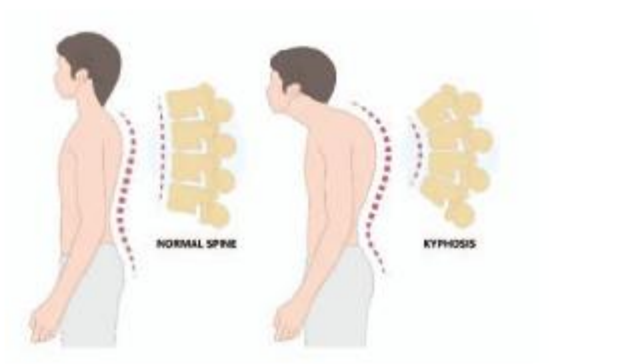

Figure3: Differentiate the spine before kyphosis and

(1) Congenital Kyphosis:
An abnormality in vertebral development that occurs during fetal development is the cause of congenital kyphosis, a spinal disorder. An aberrant forward rounding of the thoracic (upper) vertebrae is a hallmark of this condition, which is present from birth. Segmentation abnormalities, abnormalities in the size or form of the vertebrae, or both may cause the disorder [7].

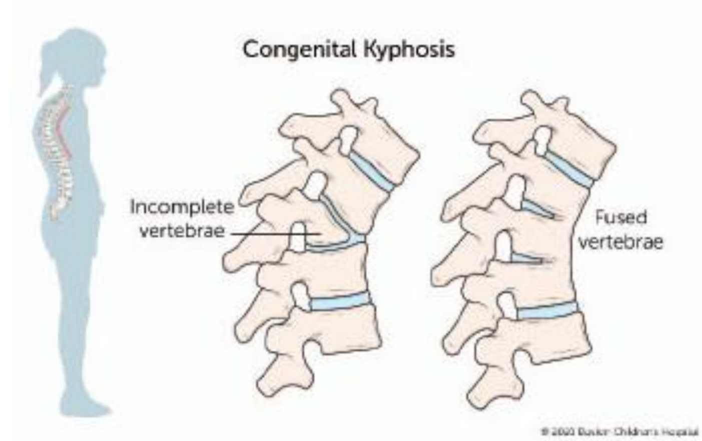

Figure4: Congenital Kyphosis Disease

(2) Postural Kyphosis:
Exaggerated rounding of the upper back, sometimes called postural roundback or postural hunchback, is a hallmark of postural kyphosis. This kind of kyphosis, in contrast to structural kyphosis, is usually reversible and linked to bad posture rather than alterations in the spine's architecture. Although people of any age may acquire postural kyphosis, it mostly affects young adults and teens [7].

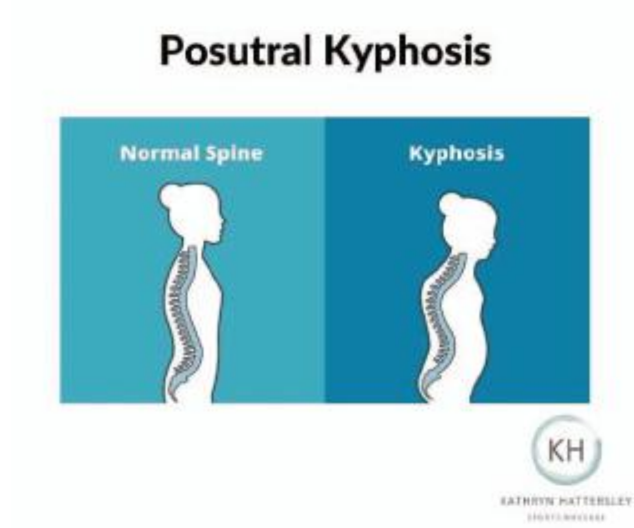

Figure 5: Normal spine (without Kyphosis) vs. Postural Kyphosis spine.

(3) Scheuermann's Kyphosis:
A spinal malformation that manifests as an uneven curvature of the upper (thoracic) spine is known as Scheuermann's kyphosis, juvenile kyphosis, or

International journal of application or innovation in engineering and management

Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/902-909

ISSN: 2319-4847

Scheuermann's disease. This syndrome normally manifests throughout adolescence, which usually occurs between the ages of 12 and 16, and it might worsen with time. When looking at the spine from the side, the vertebrae typically form a stack, but with Scheuermann's kyphosis, they form a triangle or wedge. As a result, the spine sags forward, and those who suffer from this kind of kyphosis could have trouble standing up straight and adjusting their posture.

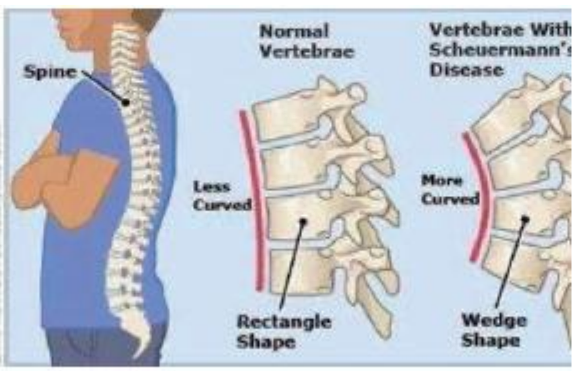

Figure6: Normal spine (without Kyphos Scheuermann's Kyphosis spine.

This work introduces a machine learning method for kyphosis illness prediction, with the aim of improved early detection and better patient outcomes. The main goal of this study is to apply several machine learning techniques, such Decision Tree and Random Forest, to biological data and evaluate the algorithms' accuracy.

II. MATERIALS AND METHODS

This section provides an overview of the dataset, data pretreatment procedures, and algorithmic software. The model implementations and data preparation were carried out in the Python environment, more precisely using Python 3.8. The use of the Google Collab notebook while developing the code led to this version being selected as the development platform. Part A: Dataset Kaggle (https://www.kaggle.com/code/data855/kyphosis-disease classification) is where we got the Kyphosis dataset. The records of patients who had spinal correction surgery are documented in four columns over eighty-one rows. In Table 1 you can see all the details about the dataset's properties.

S. No.	Attribute	Description
1	Kyphosis	Whether the Kyphosis condition was present or absent after the operation
2	Age	Age of the Patient
3	Number	Number of vertebrae involved in the operation
4	Start	Number of the first or topmost vertebrae that was operated on

Table1: Dataset Description

A lot of people have their spines corrected by surgery, but the problem is that the ailment often stays with them even after the treatment. Predicting whether patients will continue to have spine curvature problems following surgery is the main objective, depending on varied patient variables. When approached as a classification issue, the Random Forest Algorithm proves to be a useful tool for tackling this task. B. Preprocessing the Data The data has been transformed into a format that is ideal for regularly training models. The Scikit-Learn software was used to do the data preparation. Figure 8 shows the result of converting the kyphosis column to binary values (0s and 1s) using the Label Encoder that was imported from the sklearn package. The existence of the disease is shown by (1) in this form, whereas its absence is denoted by (0)[1].

International journal of application or innovation in engineering and management

Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/902-909

ISSN: 2319-4847

	Kyphosis	Age	Number	Start
0	0	71	3	5
1	0	158	3	14
2	1	128	4	5
3	0	2	5	1
4	0	1	4	15
5	0	1	2	16
6	0	61	2	17
7	0	37	3	16
8	0	113	2	16
9	1	59	6	12

Figure 7 Screenshot of the Kyphosis data (Fir records)

Part C: The Unpredictable Forest and Decision Tree Plans Hierarchy of requirements: When it comes to machine learning, decision trees are among the most common supervised learning techniques used for input-based modeling and prediction. In a decision tree, each node within the tree stands for an attribute being tested, each branch for an attribute's value, and each leaf node for the conclusion or prediction. Classification and regression problems may be handled using this approach, which is part of the supervised learning domain. Decision trees are essential to machine learning and provide the groundwork for methods such as Random Forests [11]. In order to diagnose conditions like kyphosis, decision trees use patient data including age and spinal curvature. This methodical strategy aids in patient categorization according to kyphosis risk. Because of their adaptability and ease of use with different kinds of data, decision trees are highly regarded. Thanks to their openness, medical professionals can see the reasoning behind predictions and maybe learn something new about diseases. Using decision trees alone won't guarantee better results than more sophisticated algorithms, but they may greatly improve accuracy when combined with ensemble techniques such as Random Forests. Random Forest: To boost the general accuracy and resilience of its predictions, Random Forest does use an ensemble learning approach. To do this, we combine the results from many separate models or decision trees. Training and combining the results of many independent decision trees is the main idea. This flexible technique is adaptable across many predictive modeling tasks, since it finds applicability in both classification and regression situations [13]. Random Forest, a well-known machine learning technique, may be used to accurately predict the occurrence of kyphosis. Random Forest considers a number of patient characteristics, including age, spinal level, and angle of curvature, to forecast the probability of kyphosis onset. Its ability to process high-dimensional information and identify complex relationships between input and output variables has earned it praise and has made it a viable option for use in medical diagnostics. In addition to helping us understand the underlying causes of the illness, Random Forest can provide light on which traits are most significant for predicting Kyphosis.

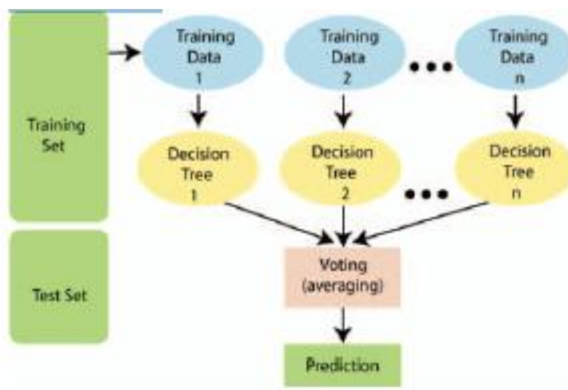

Figure 8: The Random Forest Algorithm's Architectural Framework

D. MODEL EVALUATION

When working with a small sample size, K-Fold cross-validation is typically recommended in the literature on K-Fold cross-validation for Decision Trees and Random Forests. As a result, stratified K-Fold cross-validation was used to assess the planned Random Forest and Decision Tree models. Consistent with empirical data supporting the use of 5-fold or 10-fold cross-validation, the models in this research were evaluated using both approaches. E. The Method of Prediction Get the patient's age, spinal measurements, and medical history straight before you try to predict kyphosis. After that, encode the variables and fill in any blanks to prepare the data for analysis. After that, choose a suitable machine learning approach, such as decision trees or random forests, and choose relevant characteristics for your task. After splitting the data into a training set and a testing set, train the model using the training data. Use metrics for accuracy and precision to evaluate the model's output. Make any required adjustments to

Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/902-909

ISSN: 2319-4847

the model and then put it into action. Maintaining reliable predictions over time requires monitoring the model's output and making adjustments as necessary.

III. ANALYSIS

Figure 9 shows that 21% of patients reported having the kyphosis problem, whereas 79% said they did not. This information was derived from an exploratory investigation. The correlation between the number (the range of damaged vertebrae), which is 0.36 in Figure 10, and kyphosis illness is worth considering. Figure 11 demonstrates that the presence or absence of the kyphosis sickness is often indicated by the patient's characteristics. Based on the patterns that have been identified, it seems that splitting the two groups would be a straightforward operation. Figures 12, 13, and 14 show the results of the outlier identification process using a box plot. Figure 14 shows that several data points were considered outliers. The dataset is normalized to fix this problem.

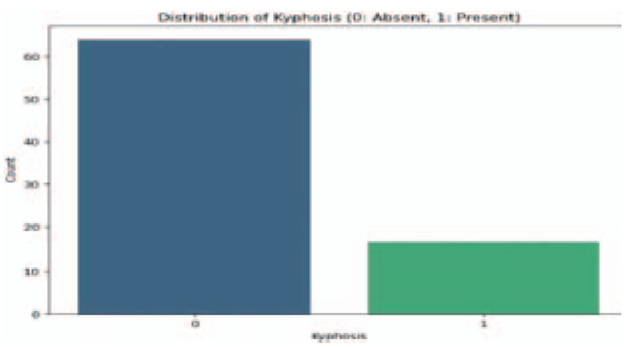

Figure 9: Present or Absent: The distribution Kyphosis disease

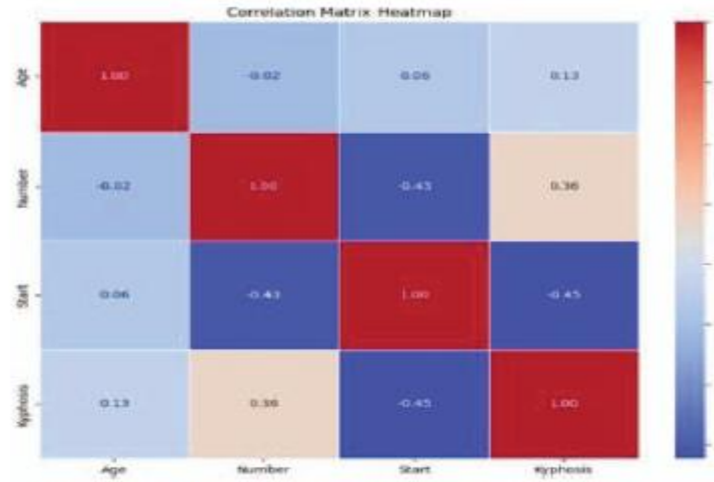

Figure 10: Features in the data and their correlations

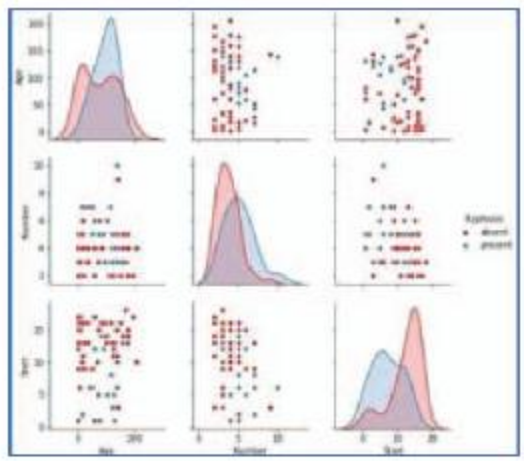

Figure 11: The three kyphosis patterns were recognized input features(Age, Start, and Number)

International journal of application or innovation in engineering and management
Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/902-909
ISSN: 2319-4847

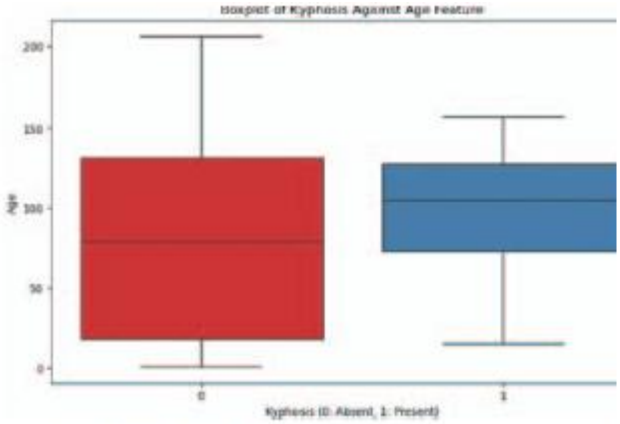

Figure 12: To identify anomalies, use the kyphosis against the age feature.

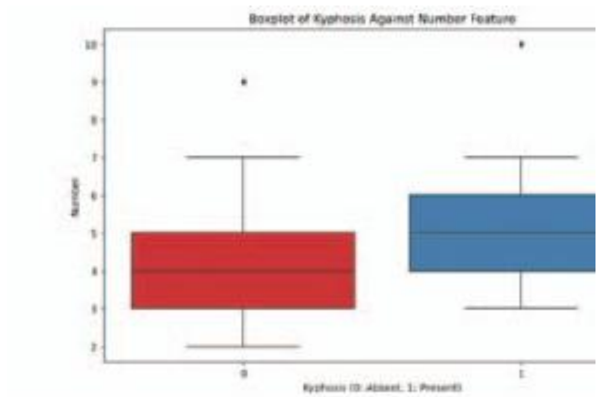

Figure 13: The method of boxplot kyphosis against n feature for outlier detection

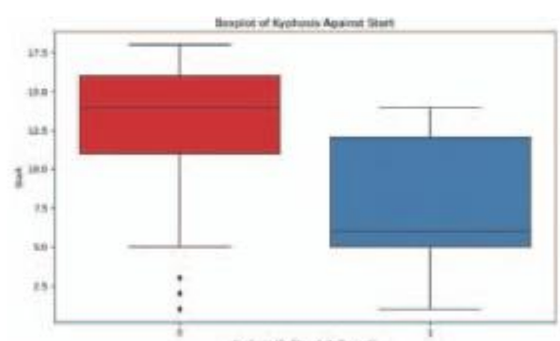

Figure 14: The kyphosis boxplot versus the Start feat

RESULT AND DISCUSSION

To keep training the model, we use machine learning techniques like Random Forest, Support Vector Machines, Logistic Regression, and Decision Trees. Determining the best possible accuracy and creating decision trees. To get the best results, you should compare all of these strategies. Accuracy rates for Logistic Regression are 75%, Support Vector Machines are 79%, Decision Trees are 80%, and Random Forests are 85.79%.

Algorithms	Accuracy
Logistic Regression	75%
Support Vector Machine(SVM)	79%
Decision Trees	80%
Random Forest	85.79%

Table2: Accuracy of the algorithms used in machine learning

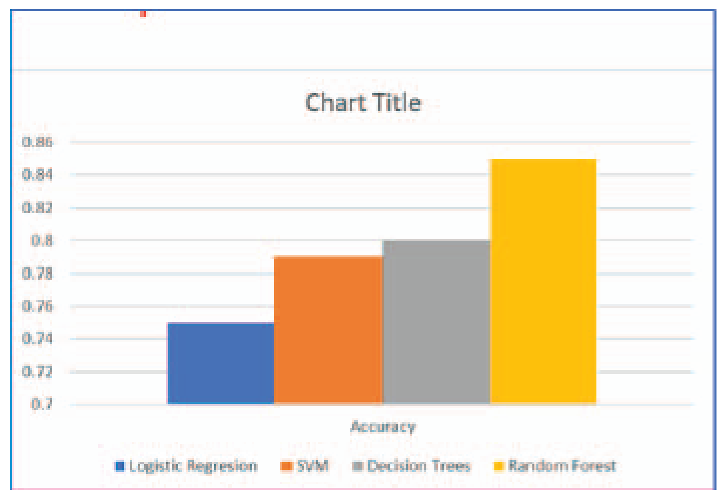

Figure 15 : Accuracy levels attained by the algorithms

In order to forecast the occurrence of kyphosis illness, the dataset made use of the proposed methods, which included Support Vector Machine (SVM), Logistic Regression, Random Forest, and Decision Trees. Following the implementation of 5-fold and 10-fold cross-validation, the following accuracy levels were attained: 79% for SVM, 75% for Logistic Regression, 85.79% for Random Forest, and 80% for Decision Trees.

V. CONCLUSION

In order to predict the onset of kyphosis illness, this research used the Random Forest (RF) model. With

Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/902-909

ISSN: 2319-4847

87.79% cross-validation accuracy, the method proved its worth. In terms of predicting the results of kyphosis disease after surgery, the model outperforms previous research. Consequently, the Random Forest method is suggested for use in identifying and predicting kyphosis in patients who have had surgery or an operation. Additional study may be conducted by other researchers to enhance the accuracy of this work and explore the prediction capabilities of other machine-learning algorithms. Additional machine-learning approaches should be explored in future studies to improve the accuracy of kyphosis illness forecasts. While admitting the possibility for modification to provide alternative data-driven clinical predictions, this research limits the efficacy of future machine learning observation and comparison models.

VI. REFERENCES

[1] A. S. Chauhan, "Kyphosis Disease Prediction Using Deep Neural Networks," in Deep Learning, 2021.

[2] H. R. Takkala et al., "Kyphosis Disease Prediction with help of RandomizedSearchCV and AdaBoosting," in Proc. ICCCNT, 26-12-2022, DOI: 10.1109/ICCCNT54827.2022.9984343.

[3] F. Galbusera, G. Casaroli, and T. Bassani, "Artificial Intelligence and Machine Learning in Spine Research," in Proc. of Spine Research, March 5th, 2019, doi: 10.1002/jsp2.1044.

[4] S. Dankwa and W. Zheng, "Special Issue On using Machine Learning Algorithms In the Prediction of Kyphosis Disease," 13-08-2019, doi: 10.3390/app9163322.

[5] M. R. Al-Masud and M. R. H. Mondal, "Data-driven diagnosis of spinal abnormalities using feature selection and Machine learning," 6 February 2020, DOI: https://doi.org/10.1371/journal.pone.0228422.

[6] Y. Y. Hussein and M. M. Khan, "Using Artificial Intelligence to Predict the Development of Kyphosis Disease: A Systematic Review," 2023 Nov 6, doi: [Insert DOI]

[7] W. S. Bartynski et al., "Severe Thoracic Kyphosis in the Older Patient in the Absence of Vertebral Fracture: Association of Extreme Curve With Age," 2005 September, DOI: http://www.ajnr.org/content/26/8/2077.

[8] A. A. Reshi et al., "Diagnosis of vertebral column pathologies using concatenated resampling with machine," [Insert Additional Information].

[9] R. S. Joshi et al., "Artificial Intelligence for Adult Spinal Deformity," [Insert Additional Information].

[10] K. N. Nithya and P. Suresh, "Effective Prediction of Vertebral Column Pathologies Using RandomForest," 6 Nov- Dec 2020, DOI: 10.21817/indjces2020/v11i6/201106084.

[11] A. A. Abdullah et al., "Prediction of Spinal Abnormalities using Machine Learning Techniques," August 2018, DOI: 10.1109/ICASSDA.2018.8477622.

[12] P. Ernst, "CNN-Based Spinal Curvature Reconstruction/Extraction in MRI for Statistical Analysis in Population Studies," October 2018.

[13] M. Shahhosseini and G. Hu, "Improved Weighted RandomForest for Classification Problems," June 2021.

[14] D. Kapoor et al., "Spinal Cord Injury AIS Predictions Using Machine Learning," January 2023, doi: https://doi.org/10.1523/ENEURO.0149-22.2022.

[15] G. T. Fon et al., "Thoracic Kyphosis: Range in Subjects," December 2017.