International Journal of application or innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

Kyphosis Disease Prediction: Evaluating Machine Learning Algorithms Effectiveness

Gudla Sai Krishna[1],Peddi Sion Raj[2] , Kadam Sainandana Rao[3],M.Ramesh [4]

[1,2,3] UG Scholar, Department of IT, St. Martin's Engineering College, Secunderabad, Telangana ,India– 500100

[4]Assistant Professor, Department of IT, St. Martin's Engineering College, Secunderabad, Telangana ,India– 500100

saikrishna5780@gmail.com

Article Info

Received: 28-03-2025 ***Revised: 05 -04-2025*** ***Accepted: 16-04-2025*** ***Published:27/04/2025***

Abstract:

Kyphosis, characterized by the inward arching of the upper back, is often colloquially referred to as "roundback" or "hunchback" when a noticeable curvature is present. This condition typically arises due to weakened spinal bones, leading to compression or fractures. In children or adolescents, other forms of kyphosis may be attributed to spinal abnormalities or a progressive twisting of the spinal bones. While kyphosis can manifest at any age, it is most prevalent in teenagers. Multiple factors, ranging from poor posture and developmental issues to structural abnormalities in the spine, can contribute to its occurrence. This research introduces a machine learning approach aimed at predicting kyphosis disease, with a focus on enhancing early detection and improving patient outcomes. This research intends to incorporate different machine- learning approaches to biological data, such as Decision Trees and Random Forests, evaluate and assess the granularity of these algorithms. The findings underscore the significance of machine learning (ML) as a valuable tool for addressing biological problems in a broader context.

Keywords: *Machine Learning, Decision Trees, Random Forest, Kyphosis, Spinal Curvature*

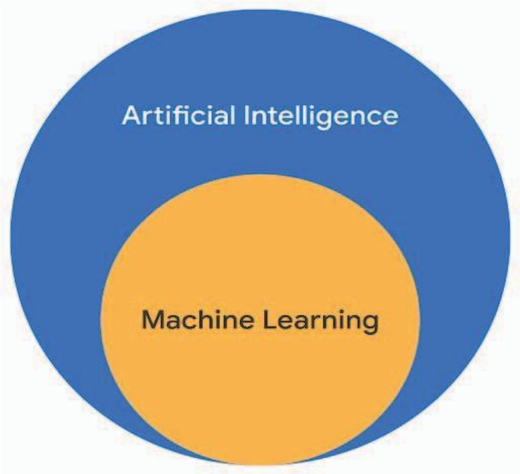
1. INTRODUCTION

In the realm of artificial intelligence, the subfield known as "machine learning" develops algorithms by uncovering hidden patterns within datasets. These algorithms utilize discovered patterns to anticipate new data, resembling old data, without explicit task programming. Conventional machine learning predicts outputs that offer practical insights when combined with statistical methods. Applications of machine learning are diverse, encompassing areas such as picture and audio recognition, recommendation engines, linguistic analysis, fraud detection, portfolio optimization, automated tasks, and more. Robots, drones, and autonomous cars also fall under the category of machine learning models, benefiting from increased intelligence and environmental adaptability. Machine learning algorithms are broadly categorized as supervised, unsupervised, semi- supervised, and reinforcement learning. This work primarily focuses on supervised learning, where a Machine learning algorithms are broadly categorized as supervised, unsupervised, semi- supervised, and reinforcement learning. This work primarily focuses on supervised learning, where aportion of the training data serves as an instructor, guiding the algorithm in determining the model.

a) **Supervised learning:** A single computer learns to predict the output using a set of inputs for which the right outcome (ground truth) is known. In supervised learning, the ideal mapping of inputs to outputs is typically discovered by reducing the value of a loss function that represents the discrepancy between the ground truth and the machine's predictions. This kind of learning is frequently used in studies on medicine.

b) **Unsupervised Learning:** In this scenario, a known ground truth is not used when the system learns from input data. Finding patterns and characteristics in the inputs allows this learning activity to glean new information from the available data. Clustering is one instance of an unsupervised learning application.

c) **Reinforcement Learning:** This method entails giving feedback regarding the accuracy of the task execution once it has been finished, as opposed to starting with ground truth data. This feedback functions as a motivator or deterrent. Reinforcement learning is becoming more popular in clinical decision-making and is used in dynamic or interactive contexts like gaming. Models of reinforcement learning are useful research instruments for examining how humans and nonhuman animals pick up the causal structure of events and tasks.



Artificial Intelligence in conjunction with Machine Learning

LITERATURE SURVEY

Kyphosis, a disorder characterized by an abnormal forward curve of the spine, poses both physical and psychological challenges. Recognizing the significance of early detection and intervention, the development of predictive models becomes crucial to mitigate the effects of this condition. Kyphosis, marked by excessive inward arching of the upper back, is often caused in older individuals by the weakening of spinal bones, leading to compression or cracking[6]. In children or adolescents, other forms of kyphosis may manifest due to spinal deformity or the gradual twisting of the spinal bones [5].

Congenital Kyphosis:

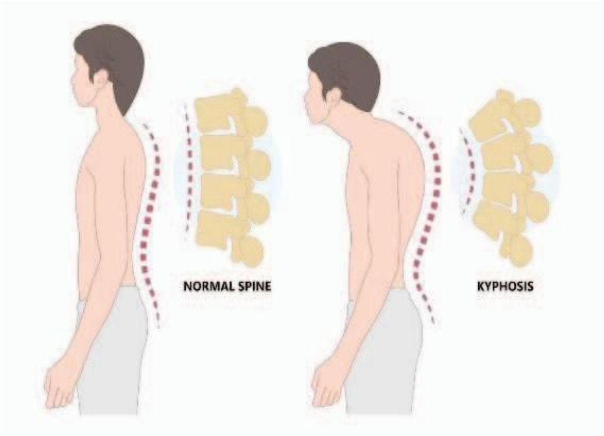
Congenital kyphosis is a spinal condition that arises from abnormal vertebral development during fetal development. It is characterized by an abnormal forward rounding of the thoracic (upper) vertebrae and is present from birth. The condition can result from anomalies in the shape or size of the vertebrae, or segmentation.

Postural Kyphosis:

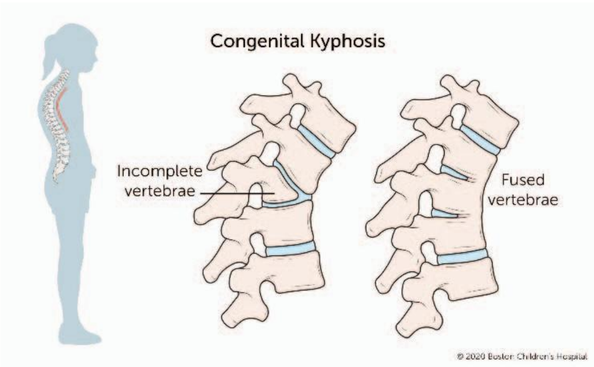
Postural kyphosis is characterized by an exaggerated rounding of the upper back, often referred to as postural roundback or postural hunchback. This type of kyphosis is typically reversible and is associated with poor posture rather than structural changes in the anatomy of the spine, as seen in structural kyphosis. While individuals of all ages can develop postural kyphosis, it is more commonly experienced by teenagers and young adults

The spinal condition kyphosis poses diagnostic challenges, effectively addressed by machine learning (ML)[8]. ML begins with feature selection and engineering, optimizing predictive factors by collecting demographics, medical records, and spine measurements. Model selection, like random forests and decision trees, considers dataset complexity, while training adjusts parameters to minimize error.

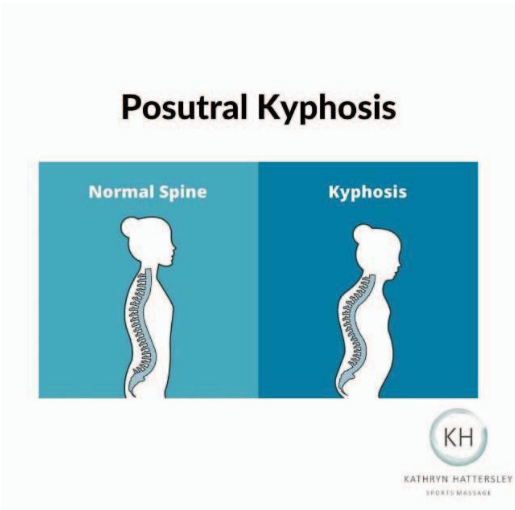
Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847



Differentiate the spine before kyphosis and after



Congenital Kyphosis Disease

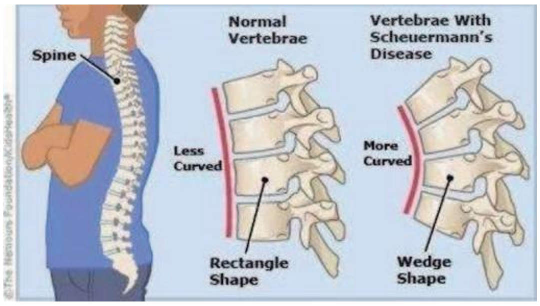


Normal spine (without Kyphosis) vs. Postural Kyphosis spine.

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

Scheuermann's Kyphosis:

Scheuermann's kyphosis, also known as juvenile kyphosis or Scheuermann's disease is a structural deformity of the spine characterized by an irregular curvature of the upper (thoracic) spine's vertebrae. Typically emerging between the ages of 12 and 16 during adolescence, this condition may progress with age. The affected vertebrae, usually arranged like stacked rectangles when viewed from the side, take on a triangle or wedge shape in Scheuermann's kyphosis. Consequently, the spine hunches forward, and individuals with this type of kyphosis may find it challenging to stand up straight and straighten their curves.



**Normal spine (without Kyphosis) vs.
Scheuermann's Kyphosis spine.**

With the goal of better early identification and improving patient outcomes, this study presents a machine learning approach to predict kyphosis disease. Applying different machine learning approaches, such as Random Forest and Decision Tree, to biological data and assessing the algorithms' correctness is the primary objective of this work.

MATERIALS AND METHODS

In this segment, we introduce the dataset, outline the data preprocessing steps, and delve into the algorithmic program. The data preprocessing and model implementations were conducted within the Python environment, specifically utilizing Python 3.8. This version was chosen as the coding platform due to the utilization of the Google Collab notebook during the code-writing process.

Dataset

The Kyphosis dataset was sourced from Kaggle (https://www.kaggle.com/code/data855/kyphosis- disease- classification). It comprises eighty-one rows, with four columns documenting the records of patients who underwent corrective spinal surgery. The attributes of the dataset are detailed in Table 1.

Table1: Dataset Description

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

International Journal of application or innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

Numerous individuals undergo surgery to correct spine curvature, yet the challenge lies in the persistence of the condition post-procedure. The primary goal is to predict, based on diverse patient characteristics, whether individuals will continue to experience spinal curvature issues after surgery. Framed as a classification problem, an effective approach to address this challenge involves implementing the Random Forest Algorithm.

Data Preprocessing

The data has been processed, adopting a format suitable for frequent model training. The preprocessing of the data was carried out using the Scikit-Learn library. As illustrated in Figure 8, the kyphosis column underwent transformation into binary values (0s and 1s) using the Label Encoder imported from the sklearn package. In this representation, (1) signifies the presence of the disorder, while (0) indicates its absence[1].

	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

Screenshot of the Kyphosis data (First 10 records)

A. The Unpredictable Forest along with Decision Tree architectural designs

Decision Tree:

In machine learning, decision trees are a prevalent supervised learning technique utilized for modeling and predicting outcomes based on input. The structure of a decision tree resembles a tree, where each internal node tests an attribute, each branch denotes an attribute value, and each leaf node represents the conclusion or forecast. This method falls within the realm of supervised learning and is applicable to addressing both classification and regression issues. Machine learning's fundamental components include decision trees, which form the basis for techniques like Random Forests[11]. Decision trees analyse patient data, such as age and spine curvature, to predict diseases like kyphosis. This structured approach helps classify patients based on their risk of developing kyphosis. Decision trees are valued for their versatility and simplicity in handling various data types. They offer transparency, allowing healthcare providers to understand prediction rationale and potentially discover new disease insights. While decision trees may not always outperform more complex methods, they can enhance accuracy when used in ensemble techniques like Random Forests.

Random Forest:

Indeed, Random Forest employs an ensemble learning technique to improve overall predicted accuracy and robustness. This is achieved by amalgamating the predictions of multiple independent models or decision trees. The fundamental concept involves training a group of diverse and uncorrelated decision trees and consolidating their outcomes. This versatile algorithm applications in both classification and regression scenarios, showcasing its adaptability across various predictive modeling tasks[13].

Kyphosis disease prediction can be successfully achieved with the help of a popular machine learning algorithm, Random Forest. Patient attributes such as age, vertebral level, and angle of curvature are among the factors evaluated by Random Forest to predict the likelihood of kyphosis development. It is noted for its capacity to handle large datasets with high dimensionality and to capture intricate correlations between input variables and the target variable, making it suitable for medical diagnosis tasks. Random Forest can also provide insights into which features are most important for predicting Kyphosis, aiding in understanding the disease's underlying mechanisms.

PROPOSED METHODOLOGY

MODEL EVALUATION

According to the literature on K-Fold Cross-Validation for Decision Trees and Random Forests, it is often beneficial to utilize K-Fold cross-validation, especially when dealing with a limited sample size. Consequently, the envisioned Decision Tree and Random Forest models were evaluated using stratified K-Fold cross-validation. In the current study, both 5- fold and 10-fold cross- qvalidation procedures were conducted to assess the models, aligning with empirical evidence favoring the use of 5-fold or 10-fold cross- validation.

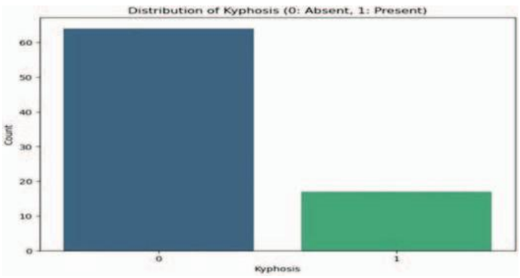
Prediction Process

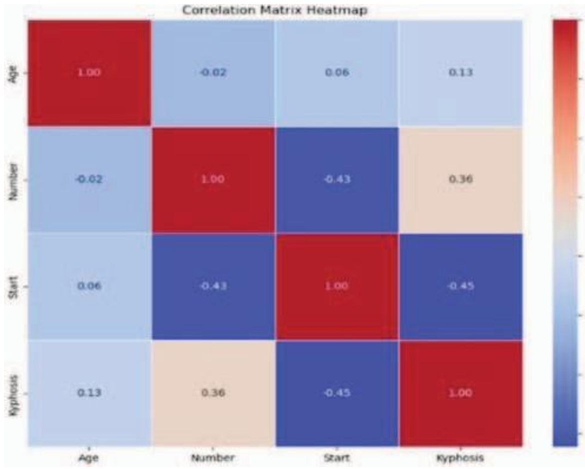
First, collect patient information such as age, spine measurements, and medical history to anticipate kyphosis. Next, take care of any missing values and encode the variables to get the data ready for analysis. Next, pick features that are pertinent to your problem and select an appropriate machine-learning technique, like random forests or decision trees. Train the model using the training data after dividing the data into training and testing sets. Analyze the model's performance with precision and accuracy measures. Adjust the model as necessary, then put it to practical use. To guarantee accurate forecasts over time, keep an eye on the model's performance and change it as needed.

ANALYSIS

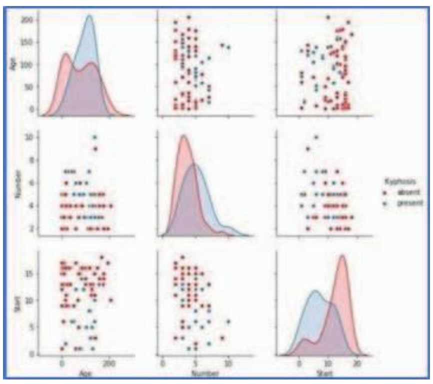
An exploratory analysis revealed that 21% of patients reported having the kyphosis disorder, as depicted in Figure 9, while 79% of patients reported the disorder to be absent. Figure 10 suggests a potential Vassociation between the number (the range of affected vertebrae), signifying 0.36, and kyphosis disease. Based on the patient's features, Figure 11 frequently shows patterns of the kyphosis illness being present or absent. It is observed that dividing the two categories would be an easily discernible process backed by the recognized patterns. Additionally, outlier detection was carried out using a box plot, as illustrated in Figures 12, 13, and 14. Certain outliers were identified in the data, as seen in Figure 14. To address this issue, the dataset is normalized.

Present or Absent: The distribution of the Kyphosis disease

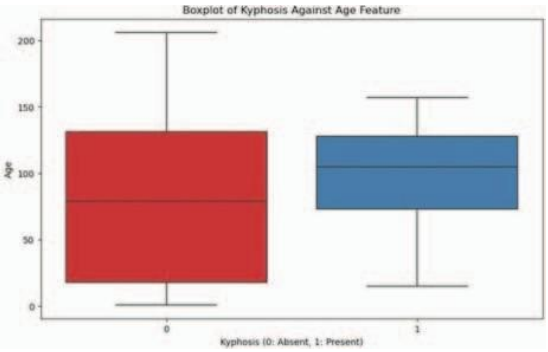




Features in the data and their correlations



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



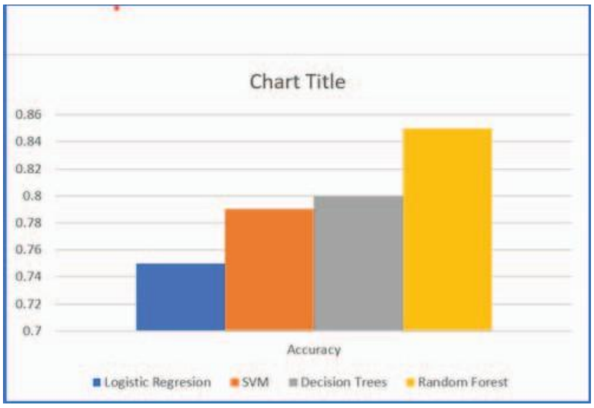
To identify anomalies, use the kyphosis boxplot against the age feature.

RESULT AND DISCUSSION

Utilizing machine learning methods including Random Forest, SVM, Logistic Regression, and Decision Trees to continue the model's training. Reaching optimal precision, and Decision Trees. Achieving the highest accuracy requires comparing each of these methods. The obtained accuracies are 75% for Logistic Regression, 79% for SVM, 80% for Decision Trees, and 85.79%forRandomForests.

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

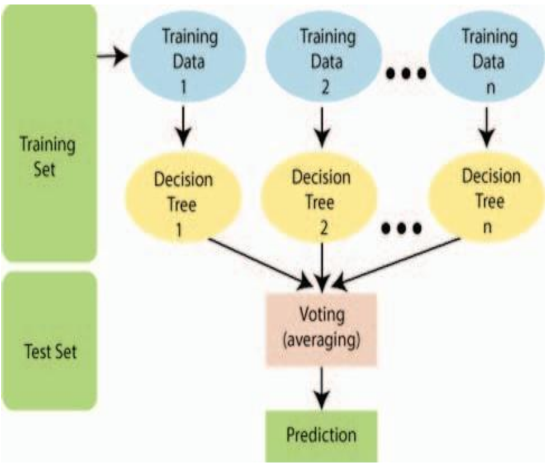
Table2: Accuracy of the algorithms used in machine learning



Accuracy levels attained by the algorithms

The dataset employed the suggested algorithms, including Support Vector Machine (SVM), Logistic Regression, Random Forest, and Decision Trees, to predict kyphosis disease. Utilizing 5-fold and 10-fold cross-validation, SVM achieved 79% accuracy, Logistic Regression yielded 75% accuracy, Random Forest demonstrated 85.79% accuracy, and Decision Trees resulted in 80% accuracy, respectively

The Random Forest Algorithm's Architectural Framework



CONCLUSION

In this paper This study has utilized Random Forest (RF) model to forecast when kyphosis disease may occur. The algorithm demonstrated a cross-validation accuracy of 87.79%. The model forecasts the outcomes of kyphosis disease post- surgery, surpassing existing analyses in scholarly publications. Therefore, it is recommended to incorporate to identify and predict kyphosis in individuals who have had surgery or an operation, the Random Forest machine learning algorithm. Other researchers can build on this work to improve accuracy and investigate the prediction capabilities of various machine-learning techniques. Future research may delve into additional machine-learning techniques to further refine the precision of forecasting the disease kyphosis. Nevertheless, this study restricts future machine learning observation and comparison models' performances, acknowledging the potential for modification to generate alternative clinical predictions with data.

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

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

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