
ISSN: 2321-2152
IJMECE
International Journal of modern electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 1, 2025

Improving Kyphosis Illness Prognosis by Assessing the Performance of Machine Learning Systems

[1] Ankith Patel, [2] K. Rakesh, [3] Ch. Varun, [4] M. Harshit, [5] Y.V.D. Siddhanth, [6] Dr. D. Murali, [7] Mr. Guguloth Raju,

[1,2,3,4,5]UG Scholar, Dept. of CS, Narsimha Reddy Engineering College, Maisammaguda, Kompally, Secunderabad, India.

[6] Professor, Dept. of CSE, Narsimha Reddy Engineering College, Maisammaguda, Kompally, Secunderabad, India.

[7] Assistant Professor, Dept. of EEE, Narsimha Reddy Engineering College, Maisammaguda, Kompally, Secunderabad, India.

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. Although kyphosis may appear at any age, adolescents are the most common age group affected. Its occurrence may be caused by a variety of circumstances, including challenges with growth and posture, as well as structural abnormalities in the spine. In order to improve patient outcomes and early detection rates, this study presents a machine learning strategy for kyphosis disease prediction. 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

One branch of AI called "machine learning" creates algorithms by finding patterns in datasets that were previously unknown. 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.
Machine learning has many different uses; some examples include image and audio recognition, recommendation engines, language analysis, automated chores, optimization of investment portfolios, detection of fraud, and more. Machine learning models also include autonomous vehicles, drones, and robots, which gain intelligence and adaptation to their surroundings.

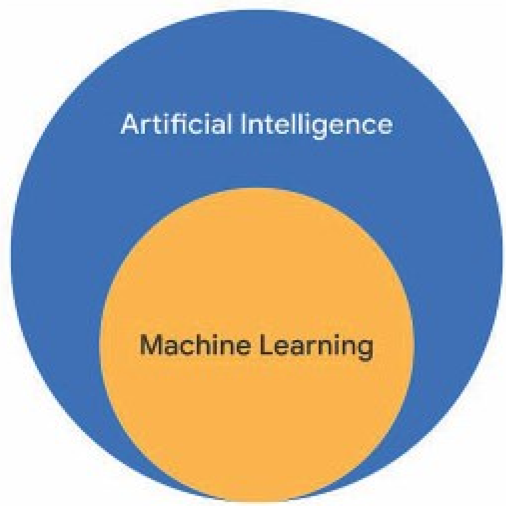

Figure 1: Artificial Intelligence in conjunction with Machine Learning

The four main types of machine learning algorithms are reinforcement learning, supervised learning, unsupervised learning, and semisupervised learning. 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 learns to use a set of inputs for which the correct result (the ground truth) is known to make predictions about the output. When using supervised learning, the best way to find the input-output mapping is to minimize the loss function, which

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 1, 2025

shows how far the machine's predictions deviate from the ground truth. Medical research often makes advantage of this form of learning.
b) Unsupervised Learning: In this kind of learning, the system does not rely on a preexisting ground truth to guide its decisions. 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. More and more, dynamic or interactive environments, such as gaming, are using reinforcement learning for healthcare decision-making. When studying how creatures, both human and others, understand the relationship between events and actions, reinforcement learning models are helpful research tools. [3].

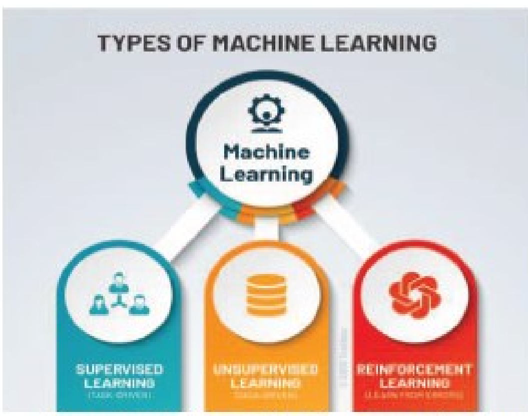

Figure2: Different Machine Learning Types

When it comes to diagnosing kyphosis, a spinal disorder, machine learning (ML) is a lifesaver [8]. To optimize predictive features, ML starts with feature selection and engineering, which involves gathering demographic data, medical records, and spine measurements. Model selection takes dataset complexity into account, as does training, which changes parameters to reduce error. Examples of such models include decision trees and random forests. Legal healthcare system integration follows model deployment, and evaluation metrics assure model performance. The capacity to adapt to changing demographics and medical 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. To lessen the impact of this disorder, it is essential to build prediction models in light of the importance of early identification and care.
As we age, the spinal bones frequently weaken and compress or fracture, causing kyphosis, which is characterized by an excessive inward arching of the upper back[6]. Other types of kyphosis may develop in children and teenagers as a result of spinal deformities or the progressive twisting of the spinal bones [5].

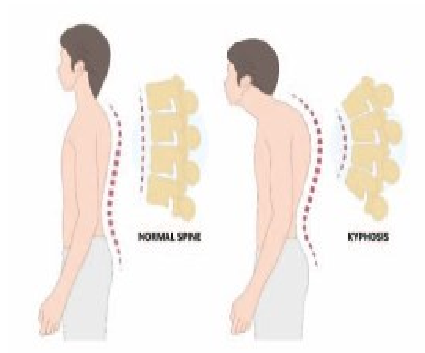

Figure3: Differentiate the spine before kyphosis and after

(1) Congenital Kyphosis:
A spinal disorder known as congenital kyphosis results from aberrant vertebral development that occurs during fetal development. The condition, which manifests at birth, is defined by an irregular curvature of the thoracic (upper) spine. Segmentation abnormalities, abnormalities in the size or form of the vertebrae, or both may cause the disorder [7].

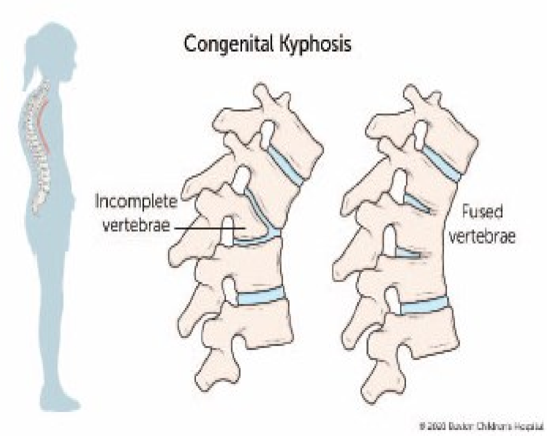

Figure4: Congenital Kyphosis Disease

(2) Postural Kyphosis:
Exaggerated rounding of the upper back, sometimes called postural roundback or postural hunchback, is a hallmark of postural kyphosis. As opposed to structural kyphosis, which causes permanent alterations to the spine's architecture, this form of the

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 1, 2025

condition is usually curable and is linked to bad posture. Postural kyphosis may affect anybody at any age, but it mostly affects young people and teens [7].

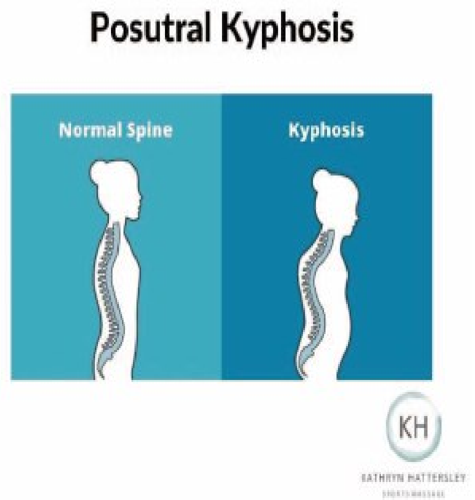

Figure 5: Normal spine (without Kyphosis) vs. Postural Kyphosis spine.

(3) Scheuermann's Kyphosis:

An abnormal curvature of the vertebrae in the upper (thoracic) spine is a hallmark of Scheuermann's kyphosis, which goes by a few other names: juvenile kyphosis and Scheuermann's illness. This issue may worsen with time and usually manifests between the ages of twelve and sixteen, throughout puberty. In Scheuermann's kyphosis, the afflicted vertebrae, which typically appear as stacked rectangles when seen from the side, transform into a triangle or wedge. It might be difficult for people with this kind of kyphosis to stand up straight and straighten their curvature since the spine hunches forward as a result.

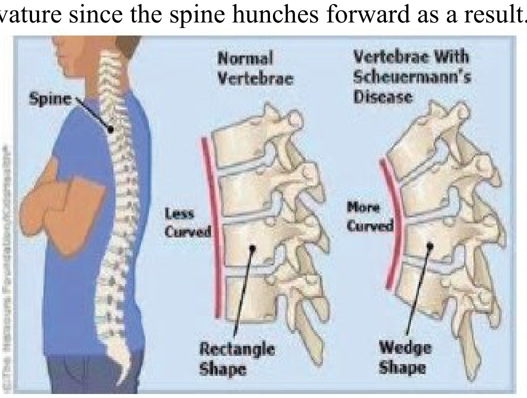

Figure6: Normal spine (without Kyphosis) vs. Scheuermann's Kyphosis spine.

Using a machine learning method, this research aims to predict kyphosis illness for improved early detection and better patient outcomes. The main aim of this study is to evaluate the efficacy of various machine learning algorithms by applying them to biological data. These algorithms may include Random Forest and Decision Tree.

II. MATERIALS AND METHODS

Here we provide the dataset, describe the procedures for prepping the data, and dive into the algorithmic code. Python, and more especially Python 3.8, was used for the data preparation and model implementations. The use of the Google Collab notebook while developing the code led to this version being selected as the development platform.
A. Matrix Kaggle (https://www.kaggle.com/code/data855/kyphosis-diseaseclassification) is where we got the Kyphosis dataset. It has 81 rows and 4 columns that provide information on people who had spinal correction surgery. 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

The continuation of the issue post-procedure is a difficulty, and many persons have surgery to address spine curvature. The main objective is to determine, using a variety of patient attributes, if patients will still have problems with spine curvature after surgery. When approached as a classification issue, the Random Forest Algorithm proves to be a useful tool for tackling this task.
B. Preprocessing the Data We have transformed the data into a format that allows us to train our models often. The Scikit-Learn software was used to do the data preparation. The kyphosis column was transformed into binary values (0s and 1s) using the Label Encoder that was imported from the sklearn package, as shown in Figure 8. Whereas (0) denotes

IJMECE International Journal of modern electronics and communication engineering

ISSN 2321-2152

www.ijmece.com

Vol 13, Issue 1, 2025

the absence of the disorder, (1) denotes its existence in this depiction [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

Figure 7 Screenshot of the Kyphosis data (First 10 records)

C. The Unpredictable Forest along with Decision Tree architectural designs Decision Tree:

One common supervised learning approach in machine learning is decision trees, which are used for input-based modeling and result prediction. Like a tree, a decision tree has nodes that test attributes, branches that indicate attribute values, and nodes that offer conclusions or forecasts. 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]. Diseases like kyphosis may be predicted using decision trees by analyzing patient data like age and spine curvature. Patients may be categorized according to their risk of developing kyphosis using this organized method. Because of their adaptability and ease of use with different kinds of data, decision trees are highly regarded. They are open and honest, which helps doctors to comprehend the reasoning behind the predictions and may even lead to the discovery of new disease insights. Ensemble approaches, such as Random Forests, may improve accuracy, even while decision trees don't always beat more complicated methods on their own. Fuzzy Logic:

Actually, Random Forest uses an ensemble learning method to strengthen and enhance the overall accuracy of predictions. 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 adaptable technique is useful for a variety of predictive modeling tasks, since it may be used in both classification and regression situations [13]. Using the well-known machine learning method Random Forest, it is possible to accurately forecast the occurrence of kyphosis. Patient variables such as age, vertebral level, and angle of curvature are among the criteria analyzed by Random Forest to predict the chance of kyphosis development. It is noted for its ability to handle huge datasets with high dimensionality and to capture subtle correlations between input factors and the target variable, making it suited for medical diagnostic jobs. Additionally, Random Forest may help us understand the mechanics of kyphosis by revealing which traits are most significant for making predictions about the condition.

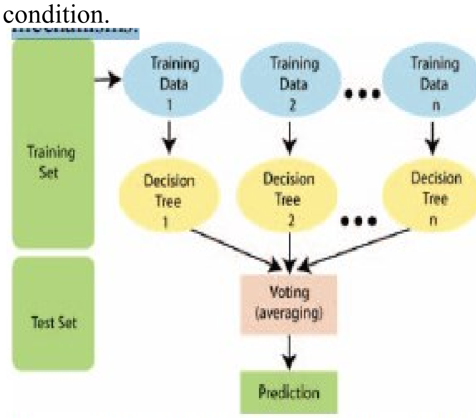

Figure 8: The Random Forest Algorithm's Architectural Framework

D. MODEL EVALUATION

A small sample size is one scenario in which K-Fold cross-validation proves to be very useful, as stated in the literature on K-Fold cross-validation for Decision Trees and Random Forests. This led to the use of stratified K-Fold cross-validation to assess the planned Random Forest and Decision Tree models. To evaluate the models, the present work used both 5-fold and 10-fold cross-validation techniques, which is in line with the empirical data that supports using both methods. The Method of Prediction (E) In order to predict kyphosis, it is necessary to first gather patient data, including age, spine dimensions, and medical history.After that, prepare the data for analysis by encoding the variables and filling in any missing values. The next step is to choose relevant characteristics and a suitable machine learning approach, such as decision trees or random forests. Once you've separated your data into training and testing sets, you may use the former to train your model. Use metrics for accuracy and precision to evaluate the model's output. After making any required adjustments, put the model into practice. Constantly monitoring the model's performance and making adjustments as necessary will ensure accurate projections in the long run.

III. ANALYSIS

Figure 9 shows the results of an exploratory study showing that 21% of patients reported having the kyphosis problem, whereas 79% of patients stated that the disease was not present. Number (the range of afflicted vertebrae), representing 0.36, and kyphosis illness may be associated, according to Figure 10. Depending on the patient's characteristics, Figure 11 often displays patterns that indicate the presence or absence of kyphosis sickness. Based on the patterns that have been identified, it seems that splitting the two groups would be a straightforward operation. Furthermore, as seen in Figures 12, 13, and 14, a box plot was used for outlier identification. Figure 14 shows that several data points were considered outliers. This problem is remedied by normalizing the dataset.

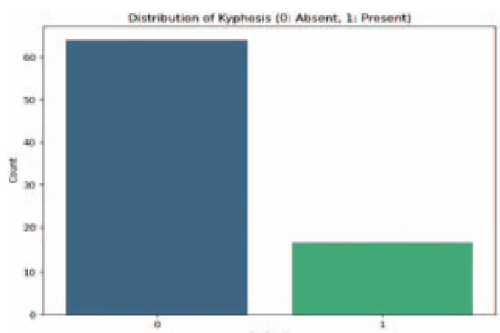

Figure 9: Present or Absent: The distribution of the Kyphosis disease

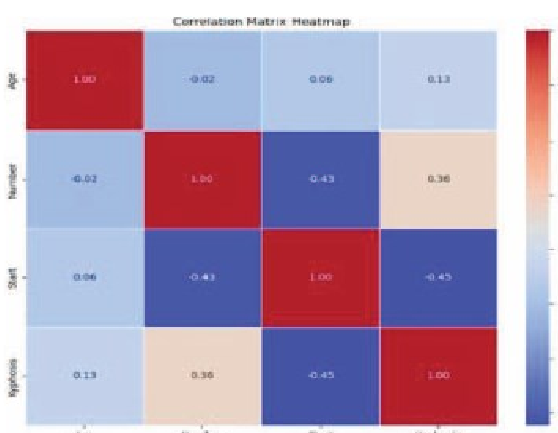

Figure 10: Features in the data and their correlations

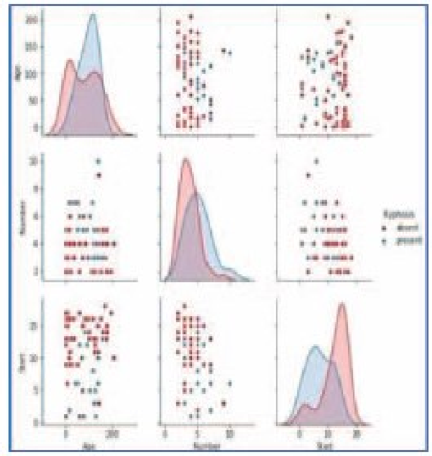

Figure 11: The three kyphosis patterns were recognized input features(Age, Start, and Number)

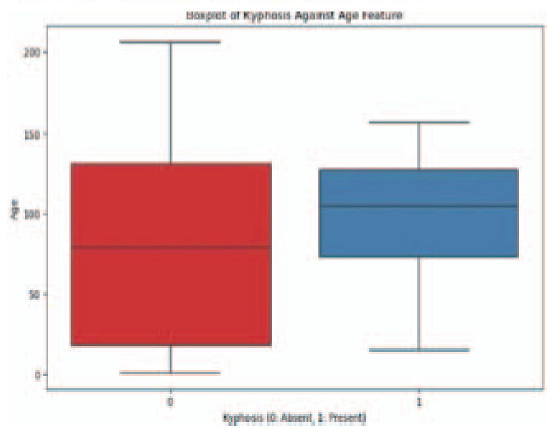

Figure 12: To identify anomalies, use the kyphosis boxplot against the age feature.

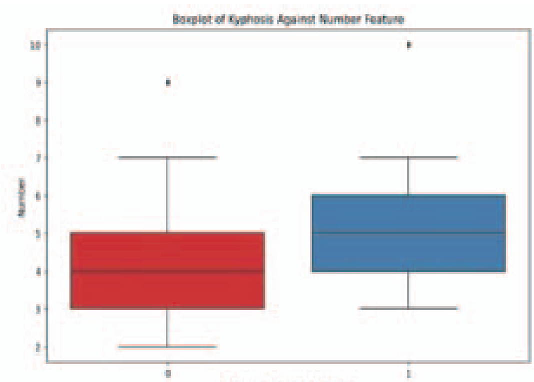

Figure 13: The method of boxplot kyphosis against number feature for outlier detection

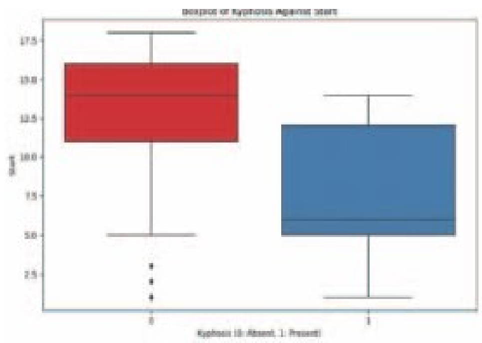

Figure 14: The kyphosis boxplot versus the Start feature.

IV. RESULT AND DISCUSSION

To keep training the model, we use machine learning techniques like Random Forest, Support Vector Machines, Logistic Regression, and Decision Trees. Reaching optimum accuracy, and Decision Trees. You need to compare each of these approaches to get the best accuracy. When it comes to decision trees, 75% is the attained accuracy, SVM is at 79%, Logistic Regression is at 75%, and Random Forests is at 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

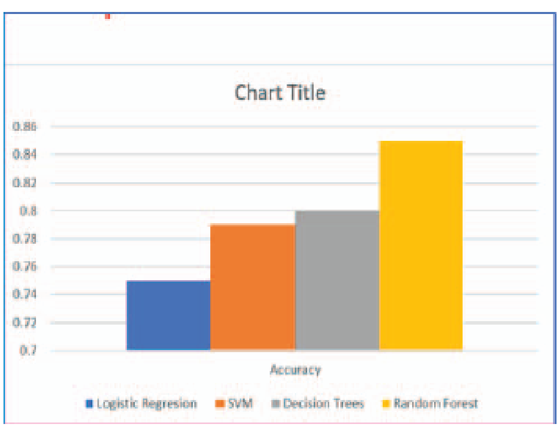

Figure 15 : Accuracy levels attained by the algorithms

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

The researchers in this study used a Random Forest (RF) model to try to predict when kyphosis illness could strike. With 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. This work may be expanded upon by other academics who want to study the predictive power of different machine-learning methods and increase accuracy. In order to improve the accuracy of illness kyphosis forecasts, future studies may investigate other machine-learning methods. However, this research limits the capabilities of future machine learning observation and comparison models, while recognizing that they may be improved to make different clinical predictions using the results.

VI. REFERENCES

Sure, here is the provided bibliography in IEEE format:

[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.

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 1, 2025

[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.