# DEVELOPMENT AND VALIDATION OF MACHINE LEARNING ALGORITHMS FOR PREDICTING SPONDYLOLISTHESIS: NOVEL APPROACH.

**Sadhana Mishra[1*], Dr. Kapil Kumar Nagwanshi[2]**

**Abstract**
spondylolisthesis refers to the slippage of one vertebral body over the adjacent one. It is a chronic condition that requires early detection to prevent unpleasant surgery. The paper presents an optimized deep learning model for detecting spondylolisthesis in X-ray radiographs. The present review focuses on key advances in machine and deep learning, allowing for multi-perspective pattern recognition across the entire information set of patients in spine disease problems. The techniques discussed could become important in establishing a new approach to decision-making in spine problems based on three fundamental pillars: (1) patient-specific, (2) artificial intelligence-driven, (3) integrating multimodal data. The findings reveal promising research that already took place to develop multi-input mixed-data hybrid decision-supporting models. Their implementation in spine surgery may hence be only a matter of time.

**Index Terms**—: Keywords: prognosis, diagnosis, classification tree algorithms, machine learning approach, etc.

---

[1*]Department of Computer Science & Engineering, Guru Ghasidas University Bilaspur, Central University, Chhattisgarh, India
[2]Associate Professor, Department of Computer Science & Engineering, Guru Ghasidas University Bilaspur, Central University, Chhattisgarh, India

***Corresponding Author:** Sadhana Mishra
*Department of Computer Science & Engineering, Guru Ghasidas University Bilaspur, Central University, Chhattisgarh, India

**DOI:** 10.53555/ecb/2023.12.Si13.297

## I INTRODUCTION

Artificial intelligence (AI) and machine learning (ML) are rapidly becoming integral components of modern healthcare, offering new avenues for diagnosis, treatment, and outcome prediction. This review explores their current applications and potential future in the field of spinal care. From enhancing imaging techniques to predicting patient outcomes, AI and ML are revolutionizing the way we approach spinal diseases. AI and ML have significantly improved spinal imaging by augmenting detection and classification capabilities, thereby boosting diagnostic accuracy. Predictive models have also been developed to guide treatment plans and foresee patient outcomes, driving a shift towards more personalized care. Looking towards the future, we envision AI and ML further ingraining themselves in spinal care with the development of algorithms capable of deciphering complex spinal pathologies to aid decision making. Despite the promise these technologies hold, their integration into clinical practice is not without challenges. Data quality, integration hurdles, data security, and ethical considerations are some of the key areas that need to be addressed for their successful and responsible implementation. In conclusion, AI and ML represent potent tools for transforming spinal care. Thoughtful and balanced integration of these technologies, guided by ethical considerations, can lead to significant advancements, ushering in an era of more personalized, effective, and efficient healthcare.

Our major goal is to build a trustworthy, efficient, and real-time Computer-Aided Diagnosis (CAD) system using a convolutional Learning model to diagnose spondylolisthesis in X-ray images. The model can assist radiologists and can be used in routine clinical practice to avoid chronic situations.

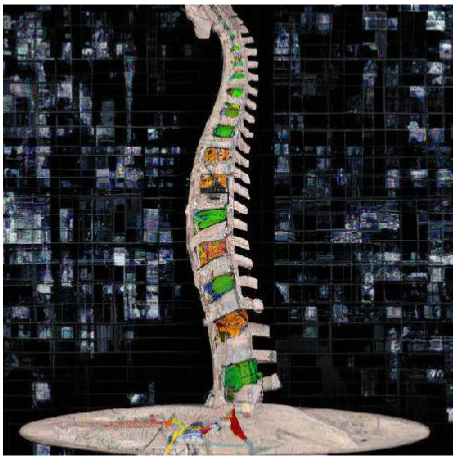


**Fig 1.0** Spinal problems issues

Out of a 7Billion population in the world, only about 10% can access adequate medical services. Even the developed countries suffer from the growing expenses and excessive wait times to avail of medical services. The medical field is confronted with new challenges such as new diseases, costs, novel medicines, and quick decisions. Because medical decision-making necessitates pinpointing accuracy in diagnosis, it is a time-consuming, demanding, complex, and demanding duty for clinicians. An automated system that helps in disease detection and prognosis, benefits the medical field. This has enticed researchers to create highly accurate medical decision support systems.

## II. MEDICAL DATA APPROACH

Human beings are the most composite organisms on this globe. It is hard to envision how billions of microscopic parts, each one with its own identity, work together in a planned manner for the profit of the total being. A system is an union of various organs arranged together so that they can carry out complex functions for the body. Body functions are the physiological or psychological functions of body systems. Survival is the body's most important ambition and it depends on the body's maintaining homeostasis. Homeostasis is a situation of relative constancy of body’s internal environment and it depends on the body's carrying out many actions in a structured manner continuously. Its major actions or functions are responding to changes in the body's environment, exchanging materials between the environment and cells, metabolizing foods, and integrating all of the body's diverse actions. In case there is any transform in the homeostasis diseases set in[1]. A disease is an abnormal situation affecting any part of the body. Diseases are roughly classified into communicable and non-communicable disease.

**Predictive tasks**: The goal of this task is to predict the value of one particular attribute, based on values of other attributes. The attributes used for making the prediction is named the explanatory or independent variable and the other value which is to be predicted is commonly known as the target or dependent variable [7].

**Descriptive task**: The purpose of this task is surmise underlying relations in data. This task of data mining, values are independent in nature and frequently require post-processing to validate results [7].

## MACHINE LEARNING APPROACH

**Association:** Association Rule Learning (Dependency modeling) is a technique that describes related features in data, searching for relationships between variables. As an example, Web pages which are accessed together can be recognized by association analysis[12].

**Clustering:** Separating objects in meaningful groups of objects or classes (cluster) based on common characteristic, play an important role in how people analyze and describe the world. For an example, even children can speedily label the object in a photograph, such as buildings, trees, people and so on.

In the ground of understanding data we can say clusters are potential classes, and cluster analysis is a technique to find classes. Before discussing about clustering technique we need to give a necessary explanation as a background for understanding the topic. First we term cluster analysis and the motive behind its difficulties, and also give details about its relationship to other techniques that group data. Then clarify two subjects, dissimilar ways of grouping a set of objects into a set of clusters and cluster types [12].

**Clustering Algorithm:** Cluster analysis is the general task to be solved which means that, it is not one specific algorithm. It is a result of various algorithms itself, in order to be proficient at clustering. It is notable by various type of clustering: Hierarchical (nested) versus partitioned (unnested), Exclusive versus overlapping versus fuzzy, complete versus partial.

**Hierarchical versus partial:** It will be discussed more among different clusters, whether the set of clusters is nested or unnested. In more conventional terminology, it has known as hierarchical or partitioned. A partitional is a division of one data set closely in one subset.

If the cluster has sub clusters it obtains the hierarchical clustering, which is a set of nested clusters which is organised as a tree. The main node (root) is a cluster and each node is sub cluster, excluding leaves which sometimes are singleton cluster of individual data objects [12].

**Exclusive versus overlapping versus fuzzy:** The definition of overlapping or non- exclusive is that one or more than one data objects belongs to more than one class, e.g. one person at university that is can be both an enrolled student and an employee of the university. In a fuzzy clustering, each object in each class has membership weight between 0 to 1 which means that it does not fit in and belong respectively. With a diverse definition the cluster is treated as a fuzzy set[12].

**Classification:** Classification is the method of finding a model or function that describes and differentiates data classes or concepts. The model is based on the analysis of a set of training data i.e., data objects for which the class labels are known. The model is used to predict the class label of objects for which the class label is unknown. Classification has number of applications, including fraud detection, target marketing, performance prediction and medical diagnosis.

Data classification is a two step process, consisting of a learning step where a classification model is build and a classification step where the model is used to predict class labels for the given data[13].

**Building the Classifier or Model**

- It is the learning step or the learning phase.
- In this step the classification algorithms constructs the classifier.
- The classifier is built from the training set invented of database tuples and their associated class labels.
- Each tuple that constitutes the training set is referred to as a category or class. These tuples can also be referred to as sample, object or data points[13].

## SPONDYLOLISTHESIS PREDICTION: MACHINE LEARNING APPROACH

Spondylolisthesis prediction from spine x-ray images is more challenging because of the different modalities in the image dataset. The research conducted in this paper is the result of a thorough investigation into the use of DL methods and efforts made in establishing a novel model for Lumbar Spondylolisthesis diagnosis. Efforts made in establishing a novel model for Lumbar

Spondylolisthesis diagnosis are explained in this paper. This paper makes several contributions and the main contributions of this research work are described in this section. The first contribution of this paper is to develop an innovative classification model based on a hybrid clustering algorithm, which can be used to classify Spondylolisthesis. Some prior studies have offered solutions but these models were less accurate. This paper covers a classification model which can be used to diagnose Spondylolisthesis accurately. The model's effectiveness is evaluated in the context of prior studies. Second, this paper contributes to implementing small, low-cost devices readily available off-the-shelf for the diagnosis of Lumbar Spondylolisthesis. This unique strategy is designed for tackling model complexity and parameters to utilize models on both small and low-cost devices. This paper looked at two different approaches to parameter reduction: TFLite quantization and the Pruning method. Prior research based on TFLite-based quantization is not available in Spondylolisthesis diagnosis. Few reviewed researches have provided adequate designs or addressed real-time exposure to identify the disease. On the other hand, this paper focuses on procedures that require no modification or minimal modification because the majority of previous models have a lot of variables.

Therefore, two strategies based on parameter and complexity reduction

i) an optimized pre-trained, and ii) a pruning-based light-weight DL model are developed in this paper, which can be easily deployed to mass consumers using low-cost devices.
Third, this paper contributes to creating an integrated model to overcome the individual model's shortcomings in terms of accurate prediction. Implemented model does the real-time, accurate, and fast diagnosis of Spondylolisthesis using innovative DL techniques. In contrast to the above models, the implemented model is deployed online on the gradio platform. This could help us get results in less than 20 seconds. It can also be used as a real-time diagnosis to assist clinicians in everyday clinical practices due to its high accuracy. Finally, this paper contributes to the evaluation of novel DL techniques and the findings. Each proposed technique is assessed to determine its interaction capabilities and experience, validate the performance, and understand the feasibility and limitations. To demonstrate the techniques for various circumstances, several types of applications, such as low-cost devices and online web application tools, are used. Although these techniques are independent of each other, one can combine these to achieve more expressive capability.

## PROBLEM SPECIFICATION

Spondylolisthesis which is pronounced as “Spohn-di-low-less-THEE-sis” is made up of two Greek words i) spondylos, which means "spine" or "vertebra,", and ii) lispaper, which means "slipping, sliding, or moving". It is a condition in which one of the vertebrae becomes anteriorly misaligned (slips forward) in relation to the vertebrae below [18]-[21], i.e., Spondylolisthesis is a condition in which the vertebrae move more than they should that results in spine instability. A vertebra slips out of place and lands on the one below it. It could exert pressure on a nerve and cause lower back or leg pain. A high percentage of the population is affected by Lumbar Spondylolisthesis. Spondylolisthesis (a major cause of Low Back Pain) is the leading cause of disability, back and neck problems, and paralysis, according to the reports of the following organizations:

- Global Burden of Disease Study (GBDS): According to GBDS, lumbar Spondylolisthesis (a major cause of LBP) is the main cause of disability, neck and back pain, and can cause paralysis (1990 - 2015). It affects all age groups (Starting from 25 years old). According to the recent data, an average of 24.8% (1990 - 2006) and 18% (2006 - 2016) of the population in the developed and developing countries lived with this disability (YLDs).
- World Health Organization (WHO): WHO has declared this disease as the most common ailment in occupational-based disease after getting that around 37% of the patients suffering from this disease].
- European Agency for Safety and Health at Work (EU-OSHA): EU-OSHA has concluded that the existence of the disease varies from 59 - 90 percent, with 15 - 42 percent of point prevalence. The severity of the problem is based on the studied population and the level of vertebra slippage.
- National Health Interview Survey: According to a survey conducted in 2011– 2013 (US), back and neck disorders (paralysis) were also the top and most common reported cause of work disability among working-age (18–64 years) persons of both genders.

Spondylolisthesis identification is commonly done qualitatively in today's clinical practice. Although Meyerding grading provides for a better quantitative assessment of Spondylolisthesis, it still relies on time-consuming and inaccurate physical measurement. As a result, there is a need to involve technological breakthroughs for automated Spondylolisthesis detection.

The emergence of DL has made significant progress in automated medical diagnosis. The development of novel DL models to examine Spondylolisthesis-related issues comes from interdisciplinary collaboration, with promising results and significant potential [30] In several applications, the reported findings are promising and outperformed the prior studies; for example, DL approaches currently allow for an exact and perfectly reproducible grading of Spondylolisthesis in X-ray images. As a result, the CAD system becomes desirable for boosting measuring efficiency substantially.

**RESEARCH OBJECTIVES**

The main objective of this paper is to develop precise, near perfect, simple to use method, cost effective procedures and methods for supporting medical practitioners. The advancement in computer technology has encouraged the researchers to build up a predictive method for assisting medical experts, psychologists, special educators and occupational therapists in improved assessment of spine disease. In this paper various techniques like K-means, M-tree techniques is proposed to predict spine abnormalities. A new predictive method using data clustering techniques has been proposed in this paper to identify the spine abnormal disease and its types from the clinical spine abnormal database. This technique uses a collection of clustered data set that is more accurate than the normal method. The proposed method can serve as a helpful method to aid medical experts and to train medical students and nurses to diagnose spine abnormal disease.

The major objectives of directing the research are as follows:

- To assess the significance of CNN techniques in the medical domain.
- To investigate several approaches for reliable prediction of lumbar Spondylolisthesis using M-tree algorithm
- To determine an appropriate CNN Learning model for identifying similar patterns in test images by utilizing the extracted features from the spine x-ray dataset.
- To improve the accuracy of lumbar Spondylolisthesis diagnosis by implementing novel CNN Learning algorithms.
- To evaluate the performance of the suggested Deep Learning Model in terms of evaluation metrics such as accuracy, precision, and so on.
- To assess the model's suitability for genuine or remote applications, and incorporate it into routine clinical practices.

**IV. RELATED WORK:**

Ansari et al. have used the UCI ML dataset to assess the effectiveness of Artificial Neural Network (ANN) and Support Vector Machine (SVM) for the diagnosis and categorization of vertebral column disorders. Three types of MRI image datasets Normal, Spondylolisthesis, and disc hernia were gathered. Half of the training data with ten-fold cross-validation procedures are used to train classifiers (different models, activation, and kernel functions were tested). Experimental findings reveal that ANN outperforms SVM (accuracy of 93.87%)

Al-Shayea et al. used ANN to classify and grade Spondylolisthesis based on 310 datasets. The results reveal an 82% in Spondylolisthesis staging . Single Decision Tree (SDT), Binary Decision Tree (BDT), and Decision Tree Forest (DTF) classifiers were used by Azar et al. to develop a decision support tool for detecting vertebral column disease. For simulation, the UCI ML Repository database is utilized, as well as the DTREG software package for decision trees and regression. During performance evaluation, BDT (84.84%) outperformed over SDT (81.94%) and DTF (84.19%). The results reveal that the BDT classifier is the most effective To grade Spondylolisthesis, Karabulut and Ibrikci used the Synthetic Minority Over Sampling TEchnique (SMOTE) and the Logistic Model Tree (LMT). Throughout 310 datasets, the outcome was 89% accurate . The classification model used by Indriana et al. is an ensembled decision tree (J48) and bagging. The model is important in the categorization of spinal diseases. The ensemble model attained a maximum accuracy of 85% during the experiments [.

Cai et al. have proposed a new Spondylolisthesis detection technique that involves immediately detecting the abnormal spine section and generating the relevant grade. Random sub-images (50 CT + 50 MR) from the training set are used to train detectors. Positive (150 CT + 150 MR) synthesized and negative (200 CT + 200 MR) images from the training data were used to train the SVM. In genuine instances, the predicted Spondylolisthesis grading (grades 0, 1, 2) was 85.3% accurate on both MR and CT scans .

Oyedotun et al. suggested methods that accurately diagnose distinct classes. BackPropagation Neural Networks (BPNN) and Radial Basis Function Networks (RBFNs) are trained using a public

database. Two diagnostics techniques were used with similar biomechanical characteristics: Three-class diagnostic system, and a Twoclass diagnostic system. The BPNNs and RBFNs have the best overall classification accuracies, with 99.05% and 96.67%, respectively. The findings indicate that neural networks may be utilized in ES to detect Spondylolisthesis.
Jamaludin et al. have developed a method that automatically grades the lumbar Spondylolisthesis using MRIs and locates the diseases. This is demonstrated using a Convolutional Neural Network (CNN) framework which accepts intervertebral disc volumes as inputs and is trained exclusively on disc-specific class labels. The proposed method gives accurate results near-human for each of the gradings, as well as visualizes the gradings on the original scans.
Jamaludin et al. created a machine learning-based grading software to classify intervertebral discs, Spondylolisthesis, and other spinal disorders. On MRI data the method has achieved 79% accuracy.

Ghogawala et al. presented an SVM-trained automatic classification system that was applied to 268 MRI images from a Nationwide Inpatient Population. Standardized methods for extracting data from electronic medical records, as well as the ability to capture radiographic imaging and incorporate Patient-Reported Outcomes (PROs), will eventually lead to the creation of modern, structured, data-filled registries that will serve as the foundation for ML .

## IV. RESEARCH METHODOLOGY:

### k-means Algorithm

The simple definition of k-means clustering, as discussed previously, is to classify data to groups of objects based on attributes/features into K number of groups. K is positive integer number. K-means is Prototype-based (center-based) clustering technique which is one of the algorithms that explain the well-known clustering problem. It creates a one-level partitioning of the data objects [40].
k-means (KM) defines a model in terms of a centroid, i.e. the mean of a group of points and is applied to dimensional continuous space. Another technique as prominent as k-means is k-medoid, which defines a model i.e. the most representative point for a group and can be applied to a ample range of data since it needs a proximity measure for a pair of objects. The difference with centroid is the medoid corresponds to an actual data point. **k-means Algorithm**
The simple definition of k-means clustering, as discussed previously, is to classify data to groups of objects based on attributes/features into K number of groups. K is positive integer number. K-means is Prototype-based (center-based) clustering technique which is one of the algorithms that explain the well-known clustering problem. It creates a one-level partitioning of the data objects [40].
k-means (KM) defines a model in terms of a centroid, i.e. the mean of a group of points and is applied to dimensional continuous space. Another technique as prominent as k-means is k-medoid, which defines a model i.e. the most representative point for a group and can be applied to a ample range of data since it needs a proximity measure for a pair of objects. The difference with centroid is the medoid corresponds to an actual data point [40].
As reward using k-means, there is some idea which find in one paper that referenced to J. MacQueen:
"The process, which is called "k-means", appears to give partitions which are reasonably efficient in the sense of within-class variance, corroborated to some extend by mathematical analysis and practical experience. Also, the k-means procedure is easily programmed and is computationally economical, so that it is feasible to process very large samples on a digital computer."
And the another one is Likewise idea which summarized in introduction part of his work benefits of using k-means:
"k-means algorithm is one of first which a data analyst will use to investigate a new data set because it is algorithmically simple, relatively robust and gives "good enough" answers over a wide variety of data sets [41]."

**Completely the analysis in aspect of optimality of k-means describes into two different components:**

- Optimal Content for given (cluster membership optimal): Each point will be a member of the cluster to whose representative point, it is closest.
- Optimal Intent for given Content: Each cluster's representative point will be the centroid of its member points, for more, the similarity define according to point that select.

In the cluster memberships optimal, the concept of optimization is straightforward where any feature still out of original dataset is measured as a member of the cluster.
The k-means clustering technique is one of the easiest algorithms; we begin with a description of the basic algorithm:
Let suppose we have some data point, D=(X1… Xn), first we choose from this data points, K initial centroid, where k is user - parameter, the number of clusters desired. Each point is then assigned to

nearest centroid. For many, we want to identify group of data points and we assign each point to one group.
The scheme is to choose random cluster centres, one for each cluster. The centroid of each cluster is then rationalized based on means of each group which assign as a new centroid.

We go over assignment and updated centroid until no point changes, it means no points navigate from each cluster to another or equivalently, each centroid remains the same[42].
The figure 1.1, shows the operation of k-means, which exemplify how starting with 3 centroid the final clusters are originate in four iteration

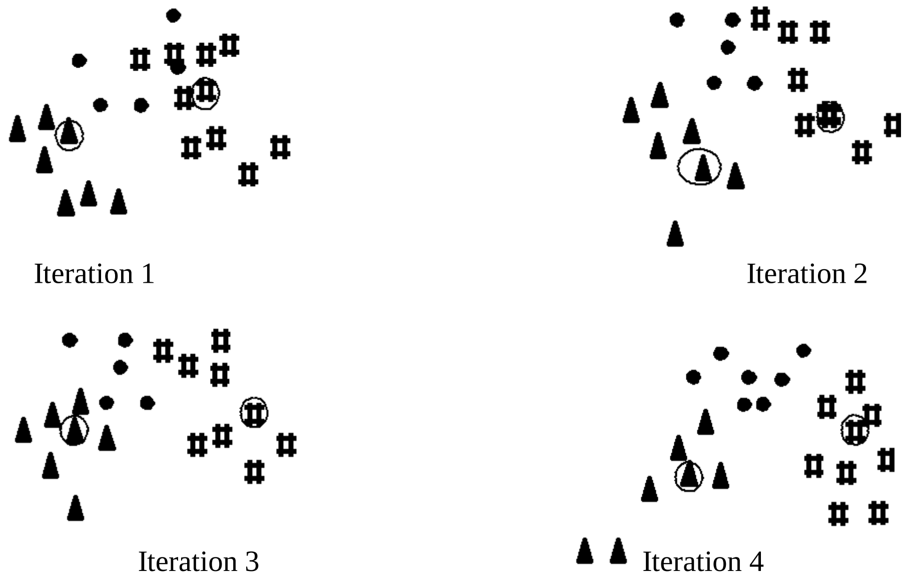



**Figure 1.1:** Using the k-means algorithm to find three clusters in sample data

The figure for each step represents the centroid at the commencement of the step and assignment of point to those centroids and in second step point assigns to updated centroids. In step 2, 3, 4, it is shown in figure 3.1, the centroids shift to the small groups of point at the bottom of the figure. In part 4, the k-means algorithm terminates, because no more change take place.
We think each step of basic k-means algorithm in more detail and then present an analysis of the algorithm's space and time complexity.

**Assigning points to closest centroid:** To assign a point to closest centroid we require a proximity measure that quantifies the closest for the exact data under concern. Euclidean (L2) distance is frequently used for data points. Though, there can be several types of proximity measures that are suitable for a given data. As example Manhattan **(**L1) distance can be used for Euclidean data while **Jaccard** measure is frequently used for documents.

**Centroid and objective function:** Step 4 in algorithm is "Recompute the centroid of each cluster", as the centroid can be variable depending on the goal of clustering. For example, to measuring distance, minimized the squared distance of each point to closest centroid, is the objective of clustering that depends on the proximity of the point to another, which can be expressed by an objective function.

**Data in Euclidean space:** Consider that the proximity measure is Euclidean distance. For our purpose function we use the **Sum of the Squared Error** (SSE) which is known as scatter. For more, we compute the error of each data point. SSE formula is generally defined as follows:

$$\sum_{i=1}^{K}\sum_{x\in c_i} \mathrm{dist}(C_i, x)^2$$

**SSE= dist**. is the standard Euclidean (L2) distance between two objects in Euclidean space.

**M-tree Algorithm (proposed algorithms)**
M-tree is a supervised approach to classification. A M-tree is a simple tree structure where all non-terminal nodes denotes tests on one or more attributes and terminal nodes reflect M outcomes. The basic M-tree induction algorithm has been enhanced. The WEKA classifier package has its own version of known as J4.8. Information gain and gain ratio measures are used by as splitting criterion respectively.

M-tree is a dynamic access method appropriate to index generic "metric spaces", where the function is used to calculate the distance between any two objects satisfies the positivity, symmetry, and triangle inequality postulates. The M-tree design fulfills distinctive necessities of multimedia applications, where objects are indexed using difficult features, and similarity queries can require application of time-consuming distance functions.

**Steps of M-tree Algorithm:**

1. Choose an attribute that best differentiates the output attribute values.
2. Create a separate tree branch for each value of the chosen attribute.
3. Divide the instances into subgroups so as to reflect the attribute values of the chosen node.
4. For each subgroup, terminate the attribute selection process if:

(a) The members of a subgroup have the same value for the output attribute, terminate the attribute selection process for the current path and label the branch on the current path with the specified value.
(b) The subgroup contains a single node or no further distinguishing attributes can be determined. As in (a), label the branch with the output value seen by the majority of remaining instances.

5. For each subgroup created in (3) that has not been labeled as terminal, repeat the above process.

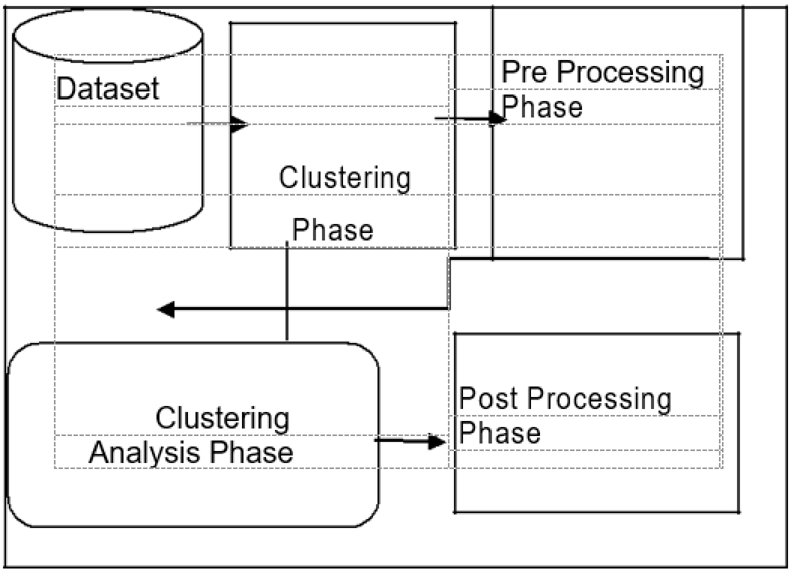



**Fig 1.0** proposed methodology.

**function y = simple_ fitness(x) y = 100 * (x (1) ^2 - x(2)) ^2 + (1 - x(1))^2;**

The formula for calculating accuracy, based on the chart above, is (TP+TN)/(TP+FP+FN+TN) or all true positive and true negative cases divided by the number of all cases.

$$DC = \frac{Total\ Detected\ Attacks}{Total\ Attacks} \times 100$$

$$FP = \frac{Total\ misclassified\ process}{Total\ Normal\ Process} \times 100$$

$$f_i = normalized_{f_i} = I(X, Y_{f_i}) \frac{f_i}{\max(F_i)}$$

$$p_i = \frac{freq(C_j, S)}{|S|}$$

Computer-aided automatic Spondylolisthesis diagnosis and grading receive high demand. Radiologists generally agree that DL is a truly disruptive technology that can deeply transform imaging data and also interpret and exploit it for treatment planning and follow-up. An interdisciplinary collaboration of DL in the medical domain resulted in the development of innovative DL models to investigate Spondylolisthesis-related issues, with promising outcomes and substantial potential. This paper covers a wide range of issues within the Spondylolisthesis domain. This is accomplished by first exploring problems and opportunities through the design space described in "Literature Review", and then solving this problem using different algorithms. In beginning, the survey was started with a broader vision by including the research papers for diagnosis, identification, labelling, etc. Later, the survey was narrowed to diagnose the disease and its classification. By

employing this methodology, DL algorithms have been used to diagnose Spondylolisthesis using Computer Tomography (CT) scans, Magnetic Resonance Imaging (MRI), and X-rays by several researchers and practitioners. CT scans and MRIs are exceedingly expensive and not available in remote areas, and also require a specialist [32]. X-ray is the first line of investigation, but there is very little research on it, and also the DL models proposed in previous studies do not have novelty in diagnosing Spondylolisthesis in the early stages through X-ray images. Section 1.4
"Rational" presents the methodology adopted in this paper.

**V. simulation details**:
WEKA is an open hotspot programming issued under general masses permit those. Information record consistently utilized Toward Weka may be On ARFF record for-tangle, which incorporates from attesting remarkable marks, will exhibit particular things in the information record first: quality names, quality sorts, Also quality qualities and the information. To working from ensuring WEKA we not persuading reason. UI of the customer and gives various workplaces. The GUI Chooser contains four gets one for each of the four critical Weka applications.

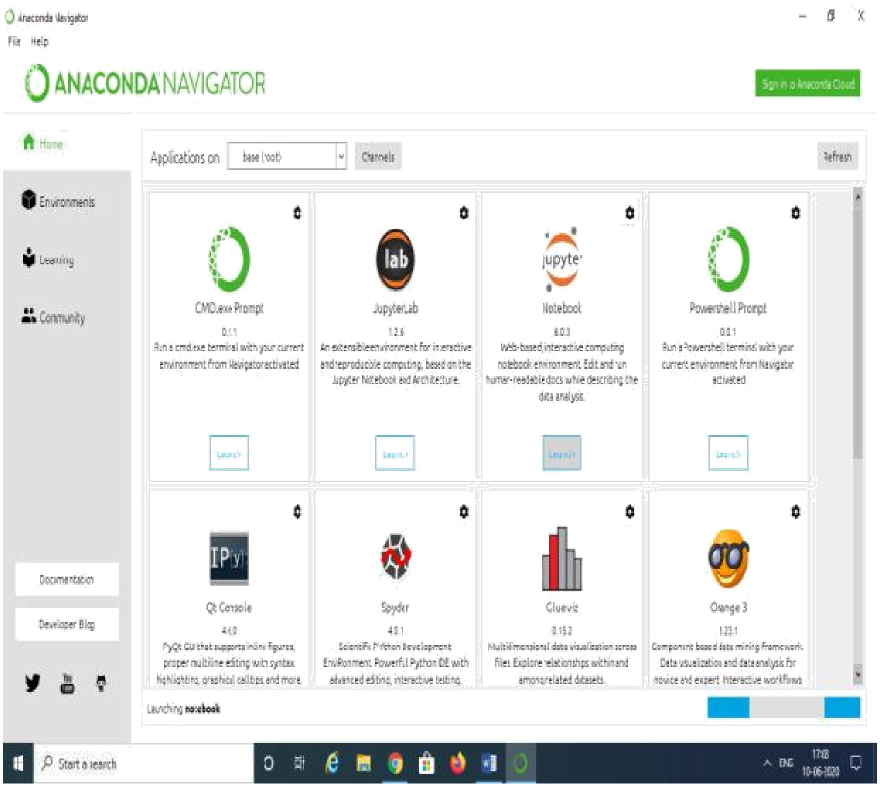



**Fig 2.0** jupyter simulation tool

**4.1.2 Features of Weka**

- Weka is completely actualized in java programming dialects, it is stage autonomous and compact.
- It is openly accessible under GNU General Public License.
- Weka s/w contains exceptionally graphical UI, so the framework is anything but difficult to get to.
- There is extensive gathering of various information mining calculations [47].

In order to assess the performance of prediction of Spondylolisthesis the outcome of M-tree has been compared with the outcomes of clustering algorithms like k-Means. The result has been compared in terms of Sum of Squared Errors, Number of Iterations, Execution Time, Goodness of Fit(accuracy).

The training dataset used for our research work is Spondylolisthesis identification is commonly done qualitatively in today's clinical practice. Although Meyer ding grading provides for a better quantitative assessment of Spondylolisthesis, it still relies on time-consuming and inaccurate physical measurement [29]. As a result, there is a need to involve technological breakthroughs for automated Spondylolisthesis detection. This dataset contains 768 record samples; each contains 8 attributes. The attributes descriptions are shown in below Table 2:

**Table 1.2:** Dataset Attribute Description

| S. No. | Attribute Name | Relabeled Values |
|---|---|---|
| 1 | X ray scan snap shot | x-rays |
| 2 | Plasma glucose concentration | Plas |
| 3 | Diastolic blood pressure (mm Hg) | Pres |
| 4 | Triceps skin fold thickness (mm) | Skin |
| 5 | Lumbar Spondylolisthesis | Lumb |
| 6 | Body mass index (kg/m$^2$) | Mass |
| 7 | DBSCAN /, Magnetic Resonance Imaging | MRI |
| 8 | Age (years) | Age |
| 9 | Class Variable (0 or 1) | Class |

**Proposed Algorithm**

Procedure M-tree (x1, x2, …, xN; K; F)
// K – the number of clusters
// F– filling factor
for (i=1, i<N, i++) {

Ci = Find Centroid (centroids, xi);
if (#Leaf [Ci] has *F* instances)
if (#we have k clusters)
#put xi in Leaf [Ci]
else
#split Leaf [Ci]
Else
#put xi in Leaf[Ci]
Recompute Centroids(Leaf[Ci])
}//end for

The following table shows the assignment of cluster using k-means and M-tree. The dataset contains 768 attributes which divides into two cluster which has been shown in the table. The experimental outcome of proposed M-tree based Spondylolisthesis prediction model are as follows:

**Table 1.1** Cluster Assignment using M-tree

| Division of cluster | Cluster 1 | Cluster 2 |
|---|---|---|
| Objects | 477 | 291 |
| Percentage | 62% | 38% |

.

**Table 1.2 Cluster Assignment using k-means**

| Division of cluster | Cluster 1 | Cluster 2 |
|---|---|---|
| Objects | 515 | 253 |
| Percentage | 67% | 33% |

**1.3 Comparative Analysis**

In order to assess the performance of prediction of diabetes the outcome of M-tree has been compared with the outcomes of clustering algorithms like k-Means. The result has been compared in terms of Sum of Squared Errors, Number of Iterations, Execution Time, Goodness of Fit(accuracy).

**Table 1.3** Comparison of Sum of Squared Errors, Number of Iterations, Execution Time, Goodness of Fit (accuracy).

| **Performance Evaluation Parameters** | **k-means** | **M-tree** |
|---|---|---|
| Sum of Squared Errors | 255.000 | 69.00 |
| Number of Iterations | 9 | 5 |
| Execution Time (in Sec) | 0.05 | 1.19 |
| Goodness of fit (in %) | 67.318 | 92.125 |

To assess the performance, the outcome of proposed model has been compared with the outcome existing clustering algorithms like k-means.

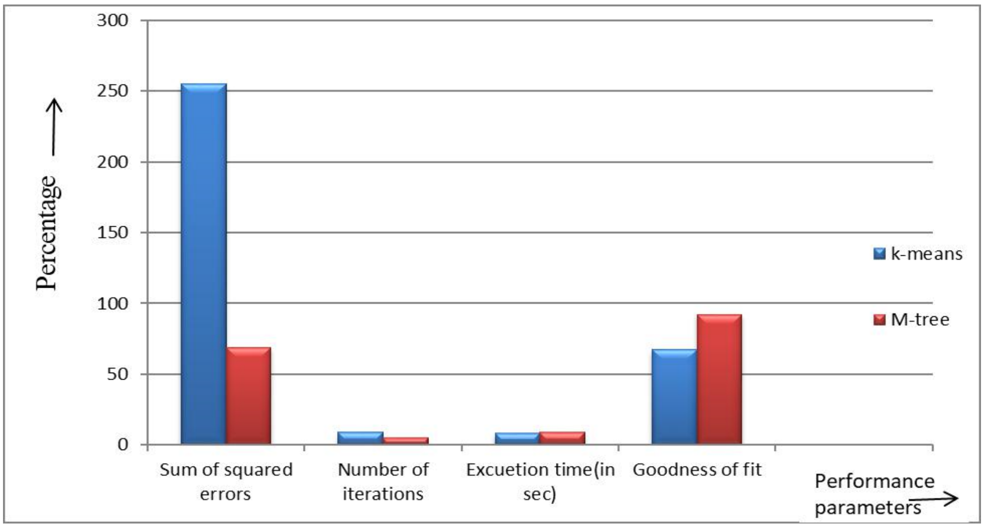



**Figure 4.1** Comparison of Time Execution, sum of errors, accuracy, iteration in k-means and M-tree through graph representation.

Now, We will show the result of M-tree algorithm of predictive value of spondylolisthesis data set the details off result shown below.

**Table 1.4** Analysis report of M-tree model for spondylolisthesis data set

| **0test negative** | **1 test positive** | **Test done by model** |
|---|---|---|
| 454 | 46 | Negative |
| 23 | 245 | Positive test |

**Table 1.5 Analysis report of k-means model for spondylolisthesis data set**

| **0test negative** | **1 test positive** | **Test done by model** |
|---|---|---|
| 380 | 120 | Negative |
| 135 | 133 | Positive test |

The following tables show the analysis of the negative and positive test of the patients, unpredicted value of the instances and accuracy of the M-tree algorithm on spondylolisthesis

In these tables,
0 represents negative test.
1 represents positive test.

**Table 4.6** M-tree and k-means model detect for unpredicted value of instances and accuracy

| **Factors** | **Sum of errors/ Incorrected instances** | **Percentage of errors** | **%(accuracy)** |
|---|---|---|---|
| M-tree | 69 | 8.999 | 92.125 |
| k-means | 255 | 33.023 | 66.977 |

This table shows the final result of our analysis. It shows sum of errors/Incorrected instances, Percentage of errors and accuracy.

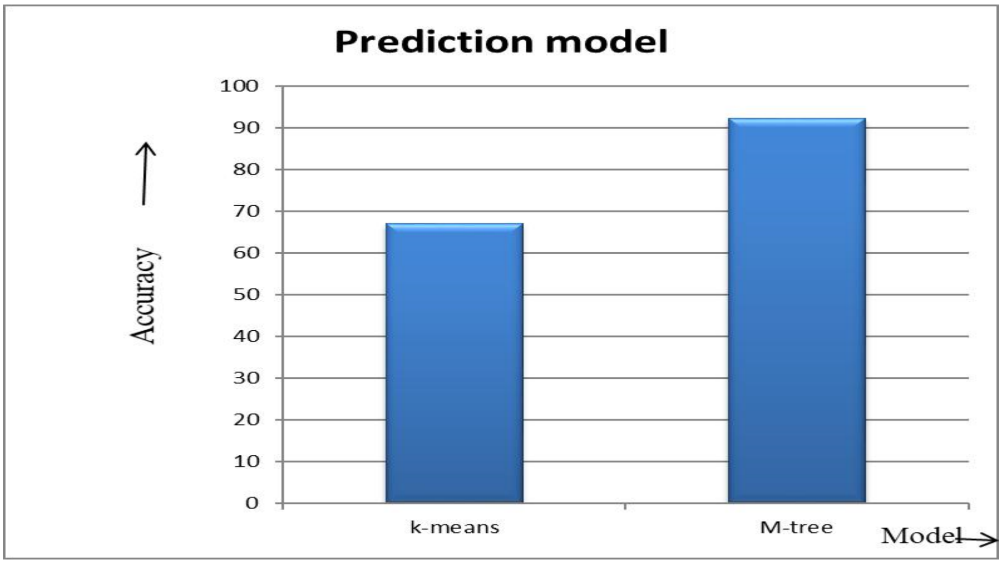
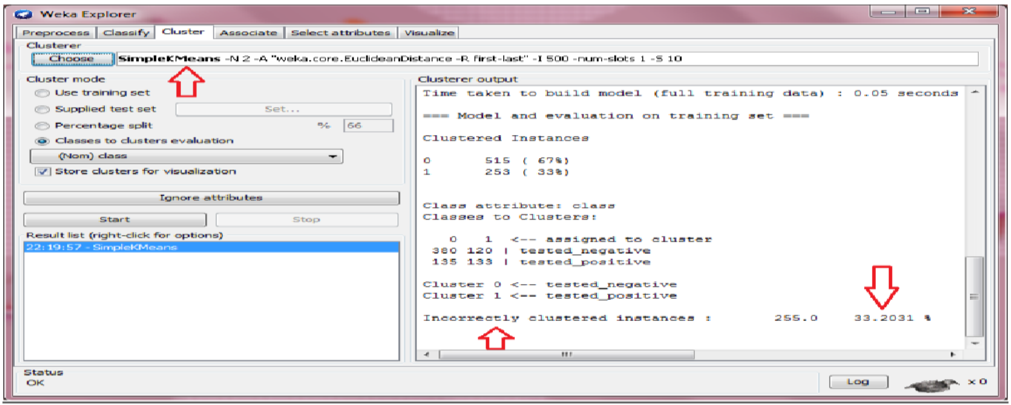
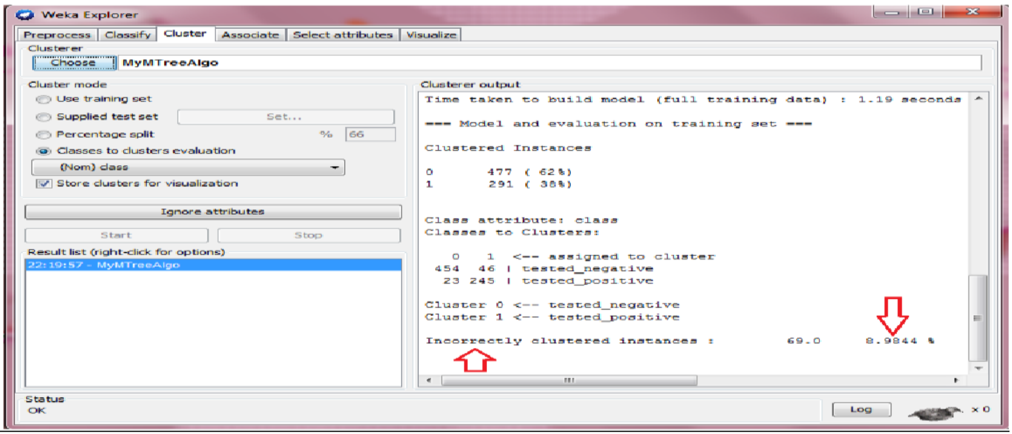



**Figure 4.2** Graph of Prediction models



## VI. CONCLUSION WORK:

The execution of grouping calculation relies upon the nature of information. The k-means grouping is utilized to recognize and take out inaccurately ordered occasions. Assist the nonstop information is changed over to straight out information by counseling therapeutic master's recommendation. The effectively arranged example by k-means is utilized as contribution to M- tree after transformation of persistent information to straight out information. work will be founded on different classifiers that can be connected on the informational index and furthermore to apply other information mining apparatuses on the informational collection with the end goal that as well as can be expected be recognized. Above calculations can be connected to different datasets keeping in mind the end goal to watch whether a similar calculation gives the most noteworthy exactness. This work can be additionally improved and extended for the computerization of spondylolisthesis disease prediction.

## VII. REFERENCES


1. Yakdan, S., Botterbush, K., Xu, Z., Lu, C., Ray, W. Z., & Greenberg, J. K. (2023). Machine learning and lumbar spondylolisthesis. Seminars in Spine Surgery, 35(3), Article 101048. https://doi.org/10.1016/j.semss.2023.101048.
2. Bydon M, Alvi MA, Goyal A. Degenerative lumbar spondylolisthesis: definition, natural history, conservative management, and surgical treatment. Neurosurg Clin N Am. (2023) 30(3):299–304. doi: 10.1016/j.nec.2019.02.003.
3. S. Tenny and C. C. Gillis, “Spondylolisthesis,” in The Growing Spine: Management of Spinal Disorders in Young Children, Second ed., pp. 415–448, 2017. Accessed: Oct. 11, 2023, http://europepmc.org/books/ NBK430767.
4. N. Ebraheim, H. Elgafy, P. Gagnet, K. Andrews and K. Kern, “Spondylolysis and Spondylolisthesis: A review of the literature,” Journal of Orthopaedics, vol. 15, no. 2, pp. 404–407, 2023. https://doi.org/10.1016/ j.jor.2018.03.008.
5. F. Varcin, H. Erbay, E. Cetin, I. Cetin and T. Kultur, “Diagnosis of lumbar Spondylolisthesis via convolutional neural networks,”in Proc. 2019 Int. Conf. on Artificial Intelligence and Data Processing Symp. (IDAP), Malatya, Turkey, pp. 1 –4, 2019. https://doi.org/10.1109/IDAP.2019.8875988.
6. V. M. Ravindra, S. S. Senglaub, A. Rattani, M. C. Dewan, R. Härtl et al., Degenerative lumbar spine disease: Estimating global incidence and worldwide volume, Los Angeles, CA: SAGE Publications, 2023.
7. Jamaludin, Automated analysis of spinal MRI using deep learning, England: University of Oxford, 2022.
8. S. Liao, Y. Zhan, Z. Dong, R. Yan, L. Gong et al., “Automatic lumbar Spondylolisthesis measurement in CT images,” IEEE Transactions on Medical Imaging, vol. 35, no. 7, pp. 1658–1669, 2022. https://doi. org/10.1109/TMI.2016.2523452.
9. L. Saba, M. Biswas, V. Kuppili, E. C. Godia, H. S. Suri et al., “The present and future of deep learning in radiology,” European Journal of Radiology, vol. 114, no. February, pp. 14–24, 2019. https://doi.org/10.1016/ j.ejrad.2019.02.038.
10. K. Singh, B. Sharma, J. Singh, G. Srivastava, S. Sharma et al., “Local statistics-based speckle reducing bilateral filter for medical ultrasound images,” Mobile Networks and Applications, vol. 25, pp. 2367–2389, 2020.
11. D. Koundal, B. Sharma and Y. Guo, “Intuitionistic based segmentation of thyroid nodules in ultrasound images,” Computers in Biology and Medicine, vol. 121, pp. 103776, 2020.
12. S. Jalal, W. Parker, D. Ferguson and S. Nicolaou, “Exploring the role of artificial intelligence in an emergency and trauma radiology department,” Canadian Association of Radiologists Journal, vol. 72, no. 1, pp. 167–174, 2021. https://doi.org/10.1177/0846537120918338
13. D. Saravagi, S. Agrawal and M. Saravagi, “Opportunities and challenges of machine learning models for prediction and diagnosis of Spondylolisthesis: A systematic review,” International Journal of Engineering Systems Modelling and Simulation, vol. 12, no. 2–3, pp. 127–138, 2021. https://doi.org/10.1504/ IJESMS.2021.115534.
14. Z. Ahmed, K. Mohamed, S. Zeeshan and X. Dong, “Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine,” Database, vol. 2020, no. 104, pp. 23, 2020. https://doi.org/10.1093/database/baaa010.
15. M. Sethi, S. Ahuja and V. Kukreja, “An empirical study for the deep learning models,” Journal of Physics: Conference Series, vol. 1950, no. 1, pp. 012071, 2021. https://doi.org/10.1088/1742-6596/1950/1/012071.

16. M. Agrawal and S. Agrawal, “A systematic review on artificial intelligence/deep learning applications and challenges to battle against covid-19 pandemic,” Disaster Advances, vol. 14, no. 8, pp. 90–99, 2021. https:// doi.org/10.25303/148da9021.
17. R. Tandon, S. Agrawal and P. Goyal, “Sequential convolutional neural network for automatic breast cancer image classification using histopathological images,” Journal of Critical Reviews, vol. 7, no. 15, pp. 6104– 6117, 2020.
18. D. Saravagi, S. Agrawal, M. Saravagi and M. H. Rahman, “Diagnosis of lumbar pondylolispaper using a pruned CNN model,” Computational and Mathematical Methods in Medicine, vol. 2022, no. 6, pp. 1–10, 2022. https://doi.org/10.1155/2022/2722315.
19. S. Sara, F. Nicholas and G. E. Hinton, “Dynamic routing between capsules Advances in Neural Information Processing Systems 30,” in Proc. of the Annual Conf. on Neural Information Processing Systems, pp. 3856– 3866, 2017.
20. G. Zhao, Y. Wu, M. He, J. Bai, J. Cheng et al., “Preprocessing and grading of glioma data acquired from multicenter,” Journal of Medical Imaging and Health Informatics, vol. 9, no. 6, pp. 1236–1245, 2019. https:// doi.org/10.1166/JMIHI.2019.2724.
21. M. Raghu, C. Zhang, J. Kleinberg and S. Bengio, “Transfusion: Understanding transfer learning for medical imaging,” Advances in Neural Information Processing Systems, vol. 32, no. NeurIPS, 2019.
22. V. K. Sakshi and S. Ahuja, “Recognition and classification of mathematical expressions using machine learning and deep learning methods,” in 2021 9th Int. Conf. on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2021, Noida, pp. 1–5, 2021. https://doi.org/10.1109/ ICRITO51393.2021.959616.
23. D. Saravagi, S. Agrawal, M. Saravagi, J. M. Chatterjee and M. Agarwal, “Diagnosis of lumbar Spondylolisthesis using optimized pretrained CNN models,” Computational Intelligence and Neuroscience, vol. 2022, pp. 1–12, 2022. https://doi.org/10.1155/2022/7459260.
24. S. Tiwari and A. Jain, “Convolutional capsule network for COVID-19 detection using radiography images,” International Journal of Imaging Systems and Technology, vol. 31, no. 2, pp. 525–539, 2021. https://doi. org/10.1002/IMA.22566.
25. Mittal, D. Kumar, M. Mittal, T. Saba, I. Abunadi et al., “Detecting pneumonia using convolutions and dynamic capsule routing for chest X-ray images,”Sensors, vol. 20, no. 4, pp. 1068, 2020.
26. D. Yousra, A. B. Abdelhakim and B. A. Mohamed, “A novel model for detection and classification coronavirus (COVID-19) based on chest x-ray images using CNN- CapsNet,” Sustainable Smart Cities and Territories, vol. 253, no. 8, pp. 187–199, 2021. 10.1007/978-3-030-78901-5_17.
27. K. Liakos, P. Busato, D. Moshou, S. Pearson and D. Bochtis, "Machine Learning in Agriculture: A Review", Sensors, vol. 18, no. 8, pp. 1-29, 2018. Available: 10.3390/s18082674
28. J. He, S. Baxter, J. Xu, J. Xu, X. Zhou and K. Zhang, "The practical implementation of artificial intelligence technologies in medicine", Nature Medicine, vol. 25, no. 1, pp. 30-36, 2019. Available: 10.1038/s41591-018-0307-0.
29. Hosny, C. Parmar, J. Quackenbush, L. Schwartz and H. Aerts, "Artificial intelligence in radiology", Nature Reviews Cancer, vol. 18, no. 8, pp. 500-510, 2018. Available: 10.1038/s41568-018-0016-5.
30. T. Davenport and K. Dreyer, "AI Will Change Radiology, but It Won’t Replace Radiologists", Harvard Business Review,2018.
31. M. Korada, "Fruit Classification| Madhu Korada", Madhu Korada, 2022. [Online]. Available: https://madhu.work/project/fruit_classification/. [Accessed: 11- Apr- 2022].
32. T. Balodi, "Introduction to Machine Learning: Supervised, Unsupervised and Reinforcement Learning | Analytics Steps", Analyticssteps.com, 2022. [Online]. Available: https://www.analyticssteps.com/blogs/introduction-machine-learning-supervised-andunsupervised-learning. [Accessed: 11- Apr-2022].
33. V. Kodžoman, "Pseudo-labeling a simple semi-supervised learning method - Data, what now?", Data, what now?, 2017. [Online]. Available: https://datawhatnow.com/pseudo-labelingsemi-supervised-learning/. [Accessed: 11- Apr-2022].
34. Esteva et al., "A guide to deep learning in healthcare", Nature Medicine, vol. 25, no. 1, pp. 24-29, 2019. Available: 10.1038/s41591-018-0316-z.
35. S. Lhessani, "What is the difference between training and test dataset?", Medium, 2021. [Online]. Available: https://medium.com/analytics-vidhya/what-is-the-difference-between-training-andtest-

dataset-d20820e5f632. [Accessed: 11- Apr-2022].

36. M. Lee, M. Grabowski, G. Habboub and T. Mroz, "The Impact of Artificial Intelligence on Quality and Safety", Global Spine Journal, vol. 10, no. 1, pp. 99S-103S, 2020. Available: 10.1177/2192568219878133.
37. F. Galbusera et al., "Fully automated radiological analysis of spinal disorders and deformities: a deep learning approach", European Spine Journal, vol. 28, no. 5, pp. 951-960, 2019. Available: 10.1007/s00586-019-05944-z.
38. Z. Ahmed, K. Mohamed, S. Zeeshan and X. Dong, "Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine", Database, vol. 2020, 2020. Available: 10.1093/database/baaa010
39. F. Galbusera, G. Casaroli and T. Bassani, "Artificial intelligence and machine learning in spine research", JOR SPINE, vol. 2, no. 1, p. e1044, 2019. Available: 10.1002/jsp2.1044.
40. S. Agrawal and S. Jain, "Medical Text and Image Processing: Applications, Issues and Challenges", Learning and Analytics in Intelligent Systems, pp. 237-262, 2020. Available: 10.1007/978-3-030-40850-3_11.
41. S. Jalal, W. Parker, D. Ferguson and S. Nicolaou, "Exploring the Role of Artificial Intelligence in an Emergency and Trauma Radiology Department", Canadian Association of Radiologists Journal, vol. 72, no. 1, pp. 167-174, 2020. Available: 10.1177/0846537120918338.
42. Tang et al., "Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology", Canadian Association of Radiologists Journal, vol. 69, no. 2, pp. 120-135, 2018. Available: 10.1016/j.carj.2018.02.002.
43. Syed and A. Zoga, "Artificial Intelligence in Radiology: Current Technology and Future Directions", Seminars in Musculoskeletal Radiology, vol. 22, no. 05, pp. 540-545, 2018. Available: 10.1055/s-0038-1673383.
44. N. Ebraheim, K. Kern, K. Andrews, H. Elgafy and P. Gagnet, "Spondylolysis and Spondylolisthesis: A review of the literature", Journal of Orthopaedics, vol. 15, no. 2, pp. 404-407, 2018. Available: 10.1016/j.jor.2018.03.008.
45. Merbs, "Spondylolysis and Spondylolisthesis: A cost of being an erect biped or a clever adaptation?", American Journal of Physical Anthropology, vol. 101, no. 23, pp. 201-228, 1996. Available: 10.1002/(sici)1096-8644(1996)23+3.0.co;2-7.
46. S. Butt and A. Saifuddin, "The imaging of lumbar Spondylolisthesis", Clinical Radiology, vol. 60, no. 5, pp. 533-546, 2005. Available: 10.1016/j.crad.2004.07.013
47. Z. Ghogawala, M. Dunbar and I. Essa, "Lumbar Spondylolisthesis: modern registries and the development of artificial intelligence", Journal of Neurosurgery: Spine, vol. 30, no. 6, pp. 729-735, 2019. Available: 10.3171/2019.2.spine18751.
48. Hoy et al., "The global burden of low back pain: estimates from the Global Burden of Disease 2010 study", Annals of the Rheumatic Diseases, vol. 73, no. 6, pp. 968-974, 2014. Available: 10.1136/annrheumdis-2013-204428