



Classification of spinal curvature types using radiography images: deep learning versus classical methods

Parisa Tavana¹ · Mahdi Akraminia² · Abbas Koochari¹ · Abolfazl Bagherifard³

Published online: 10 April 2023

© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract

Scoliosis is a spinal abnormality that has two types of curves (C-shaped or S-shaped). The vertebrae of the spine reach an equilibrium at different times, which makes it challenging to detect the type of curves. In addition, it may be challenging to detect curvatures due to observer bias and image quality. This paper aims to evaluate spinal deformity by automatically classifying the type of spine curvature. Automatic spinal curvature classification is performed using SVM and KNN algorithms, and pre-trained Xception and MobileNetV2 networks with SVM as the final activation function to avoid vanishing gradient. Different feature extraction methods should be used to investigate the SVM and KNN machine learning methods in detecting the curvature type. Features are extracted through the representation of radiographic images. These representations are of two groups: (i) Low-level image representation techniques such as texture features and (ii) local patch-based representations such as Bag of Words (BoW). Such features are utilized by various algorithms for classification by SVM and KNN. The feature extraction process is automated in pre-trained deep networks. In this study, 1000 anterior–posterior (AP) radiographic images of the spine were collected as a private dataset from Shafa Hospital, Tehran, Iran. The transfer learning was used due to the relatively small private dataset of anterior–posterior radiology images of the spine. Based on the results of these experiments, pre-trained deep networks were found to be approximately 10% more accurate than classical methods in classifying whether the spinal curvature is C-shaped or S-shaped. As a result of automatic feature extraction, it has been found that the pre-trained Xception and mobilenetV2 networks with SVM as the final activation function for controlling the vanishing gradient perform better than the classical machine learning methods of classification of spinal curvature types.

Keywords Scoliosis · Machine learning · Transfer learning · Xception · MobileNetV2 · Radiography images

Abbreviations

AEC-Net	Adaptive error correction network
AIS	Adolescent idiopathic scoliosis
AP	Anterior–posterior

✉ Abbas Koochari
koochari@gmail.com; koochari@srbiau.ac.ir

Extended author information available on the last page of the article

API	Application programming interface
BOW	Bag of words
CBIR	Content-based image retrieval
CNN	Convolutional neural network
CT scan	Computed tomography scan
DCNN	Deep convolutional neural network
DNN	Deep neural network
DoG	Difference of Gaussians
DR	Diabetic retinopathy
Faster RCNN	Faster region-based convolutional neural network
FN	False negative
FP	False positive
GLCM	Gray level co-occurrence matrix
Image-CLEF	Image cross-language evaluation forum
KNN	K nearest neighbors
LAT	Lateral
LBP	Local binary pattern
MRI	Magnetic resonance imaging
MVE-Net	Multi-view extrapolation network
PE	Pulmonary embolism
RBF	Radial basic function
SVM	Support vector machine
TN	True negative
TNN	Traditional neural network
TL	Transfer learning
TP	True positive

1 Introduction

The health issue is essential in today's world with approximately 8 billion individuals. A healthy musculoskeletal structure and the absence of any abnormalities in it is a critically important indicator of health (Foley 2013). The effects of musculoskeletal disorders on an individual's appearance and beauty may result in mental and emotional issues. These abnormalities may lead to heart and lung disease and slight imbalances for the patient over time, causing joint erosion and muscle cramps that result in pain, decreased spinal mobility, nervous system disorders, and asymmetry. These disorders often result in a limp on both sides of the body and may even lead to death. Therefore, doctors need to diagnose this disease as early as possible.

Adolescent Idiopathic Scoliosis (AIS) is one of the spinal disorders. Generally, this condition is characterized by a disorder in the spine structure, resulting in an abnormal body weight distribution. It is one of the most common abnormalities and deformities of the spine in the world (Driscoll et al. 2009; Han et al. 2015; Komeili et al. 2014). There are two types of scoliosis: C-shaped and S-shaped. Scoliosis C-shape occurs when the spine bends in one direction with a C-shaped curve, resulting in a C-shaped curve throughout the spine (Fig. 1). Scoliosis of this type can manifest in several ways: 1. the lumbar curve occurs in the lower back. 2. The thoracolumbar curve begins at the

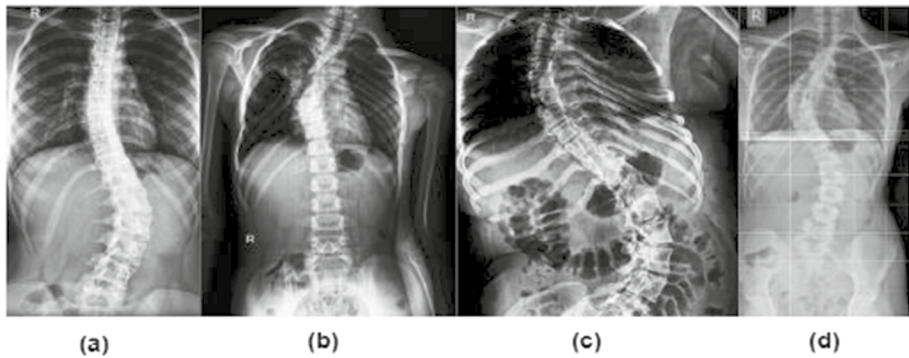


Fig. 1 The anterior–posterior images of the spine with a C-shaped curve (a, b) and the anterior–posterior images of the spine with an S-shaped curve (c, d)

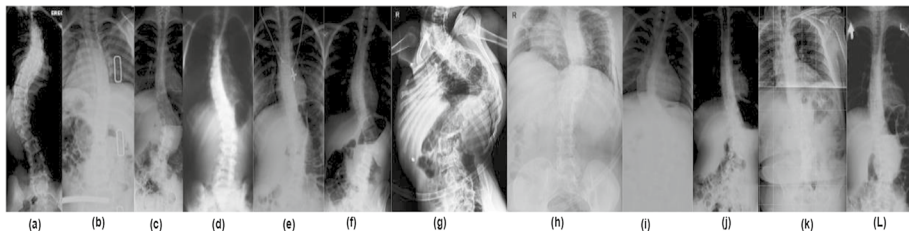


Fig. 2 It is challenging to determine the type and severity of spinal curvature from scoliosis X-rays due to high ambiguity and variability (a–L)

upper area of the waist and ends at the lower area of the waist. 3. The thoracic curve is the type of curvature scoliosis that occurs in the upper part of the spine.

In S-shaped scoliosis, the spine curves in two places—one at the top of the chest and the other at the waist. Interestingly, when these curves are in the opposite direction, they make the spine resemble the letter S (Fig. 1). Since these curves may balance each other at different times, diagnosing types of scoliosis may be more challenging (Fig. 2).

In addition, the diagnosis of the type of spinal curvature is often influenced by many factors such as the selection of vertebrae, observer bias, and image quality. When evaluating the type and progression of scoliosis, such variables have a significant impact on the diagnosis. It can also be difficult to detect the severity and type of curvature due to the high level of X-rays-related ambiguity and variability. As shown in Fig. 2, large anatomical variability and low tissue contrast may make identifying the interesting vertebrae difficult. Accordingly, it may be difficult to detect the type of curvature in the spine, and the patient may eventually go through a more complicated stage of the disease, which will require braces as well as a more complex surgical procedure. As a result, determining the type of abnormality and its severity is very critical both for the doctor and the patient (Chen et al. 2019).

Therefore, depending on the type, shape and severity of the deviation (C-shaped or S-shaped), sports techniques, the type of brace, and, in severe cases, surgery specific to the same kind of deviation are considered. Consequently, diagnosing the type of spinal

curvature plays a significant role in developing spinal evaluation systems that will assist the physician in formulating a treatment plan for the scoliosis patient.

Radiograph images, spiral CTs, and MRIs of the spine may be used to examine the spinal structure. The gold standard for diagnosing this disease is currently a spinal radiograph images (Cherian et al. 2005; Franquet 2001).

The private dataset used in this study consists of two classes. Based on the type of curvature of the spine, this dataset has been classified by two specialists. There is a challenge in classifying X-ray images (radiographs) due to their visual similarity between the different classes. A set of features plays a significant role in image classification. The following section discusses a collection of various features that other researchers have used to classify medical images (A Mueen et al. 2007).

The classification of images is the process of assigning them to predefined classes. Machine learning algorithms can perform image classification in both a classical and an automatic method. There are some basic principles underlying the classical method, such as representation in which the visual feature of the image is extracted and used as input by the classifier model. In contrast, in automatic machine learning algorithms, such as deep neural networks, there is no need to extract features and represent images (Zare et al. 2013).

In recent years, researchers have developed automated algorithms to analyze radiographic images due to their complexity of interpretation. The first articles describing an automatic anomaly detection system on radiographic images were published in the 1960s (Bhattad and Jain 2020).

Deep learning has become the preferred technique for image analysis tasks in recent years, and it has made a tremendous impact in medical imaging. It is well known that deep learning is very data-hungry. Recent years have witnessed the release of numerous labeled radiographic image databases primarily produced by automated radiology report analysis. The radiographic imaging community has greatly benefited from such databases.

Among deep learning techniques, Convolutional Neural Networks (CNN) demonstrate to significantly improve image segmentation and classification and provide a rapid diagnosis for developing assessment systems to assist doctors.

Large amounts of labeled data are usually required to train Convolutional Deep Neural Network (DNN) models. Transfer learning has successfully applied deep neural networks to real-world problems with a small dataset that traditional machine learning cannot solve, including image processing (Niu et al. 2020; Ravishankar et al. 2016).

This paper aims to classify the types of spine curvature (C-shaped and S-shaped) using classical machine learning algorithms, namely SVM and KNN, and automatic methods of deep convolutional neural networks. Classical machine learning methods for image classification require feature extraction and image representation. To improve the performance of deep neural networks, SVM has been used here as the final activation function, which prevents the vanishing gradient. In addition, it aims to help doctors diagnose patients quickly and accurately with the minimum amount of radiation in the classification of spinal curvatures.

This paper is organized as follows: Sect. 2 reviews the related literature. Section 3 presents the proposed approach in more detail. Section 4.2 provides a review of methods and experiments and experimental datasets. Lastly, Sect. 5.2 is the conclusion section.

2 Literature review

The first component in any classical classification system is the representation of images. There are two main approaches to image representation: (i) low-level image representation and (ii) patch-based image representation (Zare et al. 2013).

Several researchers have employed low-level image representations in various applications (Chang-ming et al. 2008; Deselaers and Ney 2008; Dimitrovski et al. 2011; Kim et al. 2010; Ahmed Mueen et al. 2008; Mustaffa et al. 2008; Sohail et al. 2010; Tian et al. 2008; Tommasi et al. 2008). Kim et al. proposed a method for classifying medical X-ray images that use LBP to extract texture information from the x-ray images, finding that the classification accuracy obtained was better than that of others who used edge histogram descriptors (Kim et al. 2010). Combining a block-based local binary pattern with an edge histogram was used as a medical image representation for Image-CLEF 2007's task of automatic medical image annotation (Tian et al. 2008).

According to Chang-ming et al., the segmentation method was applied to ultrasound medical images based on the texture features obtained from the GLCM (Chang-ming et al. 2008). As a feature extraction method, GLCM was combined with histogram moments to classify ultrasound medical images (Sohail et al. 2010). However, combining different local and global descriptors into one unique feature representation can be seen as another widely used approach. As a global image descriptor, pixel values were combined with other image representation techniques to construct a feature vector to describe the image (Deselaers and Ney 2008; Dimitrovski et al. 2011; Ahmed Mueen et al. 2008; Tommasi et al. 2008; Zare et al. 2013). In several studies, image representations based on local patches have been studied. In the local patch-based image representation method, Bag of Words (BoW) is used to represent images as histograms of quantized appearances of local patches.

Several studies have utilized this feature in classifying medical images (Avni et al. 2010a, b; Deselaers et al. 2006). Because of the increasing size of medical X-ray archives, it is essential to have simple, discrete representations and simple matching methods to ensure computer efficiency. A visual word dictionary was generated to represent the X-ray chest images used in another paper (Avni et al. 2010a). Using a "bag of visual words" approach, (Avni et al. 2010b) proposed classifying and retrieving local patch representations-based X-ray images. The best performance was obtained using dense sampling of features and a kernel-based Support Vector Machine (SVM) based on nonlinear kernels. In Deselaers et al. (2006) extracted features based on local patches of different sizes, which were measured at every position and then scaled down to a common size (Zare et al. 2013).

Researchers and physicians have been considering deep neural networks as an automated machine learning system in recent years since they have the advantage of automatic feature extraction, which helps diagnose and classify diseases in medicine (Szegedy et al. 2015, 2016). Studies have been reported in the articles (Pratt et al. 2016; Shuangling Wang et al. 2015), diagnoses of diabetic retinopathy (DR) or ocular diabetes (Giancardo et al. 2012; Niemeijer et al. 2009; Rahim et al. 2016; Reddy et al. 2018; Vo and Verma 2016), in (Tajbakhsh et al. 2015b; Zhang et al. 2016) diagnoses polyps during colonoscopy, in (Tajbakhsh et al. 2015a) correctly diagnoses Pulmonary embolism in CT scan images, in (Zheng et al. 2015) classify lung diseases, in (Rajinikanth et al. 2022) lung pneumonitis is diagnosed with the InceptionV3 network (Rajinikanth et al. 2022) demonstrated the lung segmentation performance of the U-Net scheme with a one-fold and two-fold training process. Experimental verification of the proposed scheme is conducted by segmenting the lung section from chest X-rays in (Nagi

et al. 2022). This research involves an evaluation of the performance of deep learning models for COVID-19 diagnosis using chest X-ray images from a dataset containing the largest number of COVID-19 images ever used in the literature, according to the best of the author's knowledge. In (Mohammed et al. 2021) an automatic prediction of COVID-19 identification is proposed to automatically discriminate between healthy and COVID-19-infected subjects in X-ray images using two successful modern are traditional machine learning methods (e.g., artificial neural network (ANN), support vector machine (SVM), linear kernel and radial basis function (RBF), k-nearest neighbor (k-NN), Decision Tree (DT), and CN 2 rule inducer techniques) and deep learning models (e.g., MobileNets V2, ResNet50, GoogleNet, DarkNet, and Xception) and in (Dubost et al. 2019; Khanal et al. 2019; Lin et al. 2019; Wang et al. 2019a, b; Zhao et al. 2019) study spinal segmentation and measurement of Cobb angle by a deep neural network.

According to Lin et al., Seg4Reg is a framework that consists of two deep neural networks for segmenting the spine and measuring Cobb angles (Lin et al. 2019). Based on the deep neural network, Zhao et al. used a top-down approach to determine spine curvature automatically (Zhao et al. 2019). According to Khanal et al., an automated method using the Faster RCNN network detects vertebrae as objects, followed by a landmark detector using Dense-Net to estimate the four landmark corners of each vertebra separately. Cobb Angles were calculated using the slope of each vertebra obtained from the predicted landmarks (Khanal et al. 2019). Chen et al. presented a new adaptive error correction network (AEC-Net) to formulate the estimation of Cobb angles from spine X-ray images as a high-accuracy regression task. AEC-Net consists of two networks that calculate landmarks and Cobb angles separately, which robustly solves the disadvantage of ambiguity in X-rays since these networks focus on more features (Chen et al. 2019). Wang et al. proposed a multi-view extrapolation network (MVE-Net) that provides accurate automated scoliosis estimation in multi-view (both AP and LAT) X-rays (Wang et al. 2019a). According to the studies conducted for the automatic spine assessment articles, the only focus is on measuring the Cobb angles of the spine.

Studies indicate that there is no article regarding the classification of spinal curvature type (C-shaped and S-shaped). Due to the biased observer and the poor quality of the radiographic images, determining the type of curvature is a challenging task. Because there are few doctors, spinal deformity assessment takes considerable time, and as this disease progresses, the patient experiences more severe stages of the disease, accordingly. Thus, automated methods of diagnosing and categorizing the curvature type and severity of curvatures are becoming increasingly important to aid patients as quickly as possible to provide them with the best possible care.

Although there is an importance and necessity for the automatic detection of the type of curvature of the spine, no research has been conducted to date in this area. The objective of this study was to automatically detect the type of curvature using machine learning techniques for the first time. In this investigation, the traditional machine learning method, including support vector machines (SVM) and K-nearest neighbor (KNN) in front of the pre-trained deep networks, are used, studied, and compared.

The classical machine learning method first extracts the feature vector from the images, as shown in Fig. 3a. Following the training process, classification is performed. In contrast, in the deep learning method (Fig. 3b), the images are fed directly into the proposed pre-trained deep networks. After fine-tuning the networks, the image classification is carried out automatically.

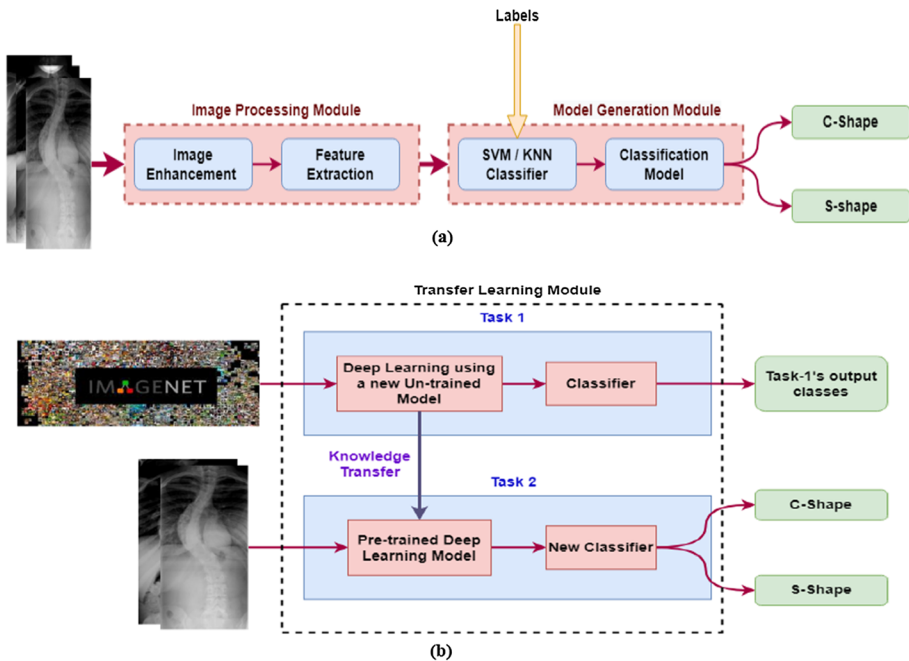


Fig. 3 A classification framework based on SVM and KNN machine learning algorithms and pre-trained networks

3 The proposed general approach

Automated image classification is an active field of research in pattern recognition and machine learning. An image classification method based on supervised learning is used here. There are two steps in the classification process: training and testing. It is well known that in supervised learning, a training set is given with a category label and a set of image features. The learning objective is to compute the classifier from the training set and create the classification model. During the test step, this model is used to classify new images that are not labeled to determine the model's accuracy (Mueen et al. 2007). Figure 3 illustrates the proposed classification training process.

3.1 The image processing module (image enhancement and feature extraction)

The training step in the Support Vector Machine (SVM) and the K-nearest neighbor (K-NN) algorithms includes two modules of image processing and model production. As shown in Fig. 3a, an image processing module consists of several components, including image enhancement and feature extraction. Image enhancement methods are used to improve the quality of the images. In classical machine learning, feature extraction methods are used to determine the appropriate model inputs.

The image classification process begins with extracting appropriate features from the images. To determine image classification accuracy, extracting a set of features from the image is primarily necessary. Adding more distinctive characteristics will improve the

classification result. It is possible to extract features at three levels: global, local, and pixel. An image's most superficial visual properties are directly related to pixel values. As a result of scaling the images to a common size, the Euclidean distance and distortion model is used to compare them (Keysers et al. 2004). In this process, local features are extracted from small sub-images from the original image. The global features can be extracted to describe the whole image in an average way. Color, texture, and shape are the low-level features extracted from images and local patches (Ma and Zhang 1998; Mueen et al. 2007).

3.1.1 Image enhancement

Histogram equalization is one of the image enhancement techniques applied to improve the quality of the image, which in turn increases the contrast of the image. As a result of the contrast adjustment, the gray intensities distribution on the histogram is improved. This method is appropriate for images with bright or dark backgrounds and foregrounds. Notably, the technique can enhance the interpretation of X-ray images of bone structure (Ahmed Mueen et al. 2008; Zare et al. 2013).

3.1.2 Feature extraction

Feature extraction is the second component of the image processing module. It plays a crucial role in the performance of any image classification system since it can impact the classification results. However, since X-ray images are gray-level and do not contain color information, the related Content-Based Image Retrieval (CBIR) systems deal primarily with textures and shapes to extract features. This study uses various methods to extract features from images, including texture, shape, pixel value, and the bag of words, as illustrated in Fig. 4 (Zare et al. 2013).

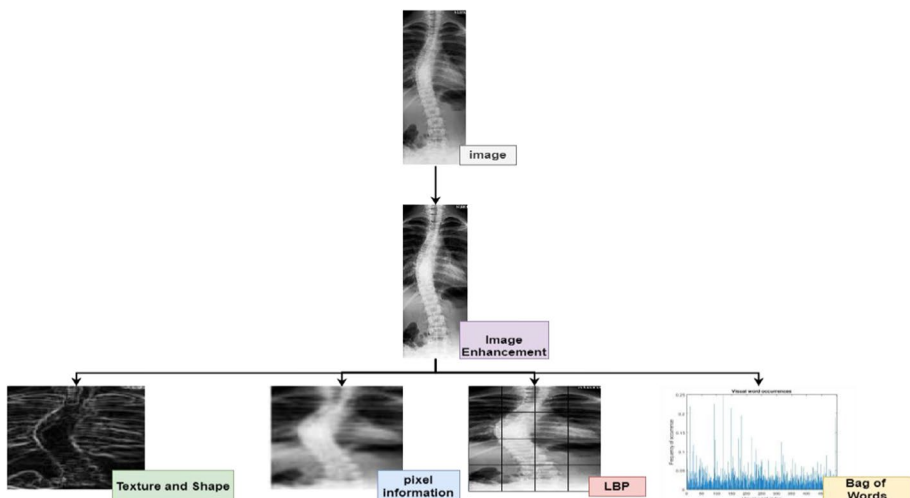


Fig. 4 Feature extraction methods for radiographic images

3.1.3 Texture features

Texture features can be used to distinguish categories of medical images with distinct textures. The texture contains information essential to comprehend the underlying structural arrangement of the surfaces within an image. In this study, the commonly used feature extraction techniques, Gray Level Co-occurrence Matrix (GLCM) (Haralick et al. 1973) and Local Binary Patterns (LBP) proposed by Ojala et al. (1996), were used to extract texture features (Zare et al. 2013).

As one of the well-known texture extraction techniques, GLCM is a technique that assesses second-order texture characteristics. In other words, the gray level coordinate matrix, or GLCM, of an $N \times N$ image, which consists of pixels with gray levels 0, 1, 2, ... $G - 1$, is a matrix $C(i, j)$ which depicts the probability of co-occurrence of intensity levels i and j at a given distance and angle θ . The four occurrence matrixes of GLCM are presented from four different directions ($\theta \in \{0^\circ, 90^\circ, 45^\circ, 135^\circ\}$) at a global level (Zare et al. 2013).

LBP is another important texture feature used in this study. The original LBP describes only global texture information. On the other hand, local features are more robust to occlusion and clutter. The image was divided into 4×4 or 2×2 non-overlapping sub-images to produce this result. Then, each sub-image was further divided into a non-overlapping square image block with the same size as the 'neighbor set' itself. The LBP distribution can be represented in each sub-image as a histogram by combining the local histograms for each square block. The final histogram of each sub-image is then generated by combining local histograms (Zare et al. 2013).

3.1.4 Shape features

The shape of an object in an image provides geometrical information which remains the same despite the location, scale, and orientation changes. Accordingly, the shape information of an image is described by the edges of the image. For each image and image patch, a histogram of the edge directions is used to represent the shape attribute. The Canny edge operator (Ojala et al. 1996) is used to generate edge histograms. The algorithm of the Canny edge detector is divided into five stages:

1. *Smoothing*: The process of removing noise using a Gaussian filter.
2. *Finding gradients*: Finding the gradient of the intensity of an image using finite-difference approximations for partial derivatives to calculate gradient magnitude and orientation.
3. *Suppression of local maxima*: Only local maxima should be marked as edges. The non-maxima suppression will be applied to the gradient magnitude to obtain these results.
4. *Double thresholding*: Potential edges are determined by the thresholding process.
5. *Edge tracking by hysteresis*: Final edges are determined by suppressing all edges that are not connected to a very specific (strong) edge (Zare et al. 2013).

3.1.5 Pixel value

The simplest method for representing images is the pixel value method. A feature vector contains the values of image pixels in an algorithm that scales down the image to a common size and represents it as a feature vector containing the image pixels. It is argued that

image pixel values do not give satisfactory classification or retrieval results. When comparing two images, it is not always easy to determine which two pixels should be evaluated. Nevertheless, this method is a reasonable basis for classifying and retrieving simple images with few objects, such as medical X-rays and character recognition (Keyzers et al. 2007; Zare et al. 2013).

3.1.6 Combining visual features

One thousand images (800 training images and 200 test images) exist in the dataset. The images have all been scaled down to 100×100 pixels. A global feature vector combines texture and shape features into a 53-dimensional feature vector (16 textures + 37 edges). In experiments two and three, local features are obtained by dividing the images into sixteen and four pieces, respectively, without overlapping. The original image has been resized to 15×15 for pixel-level information, and a feature set has been created with 225 features. The outcome of experiments 1, 2, and 3 have been 278, 944, and 750 dimension features, respectively (Fig. 5).

3.1.7 Bag of words

The process of BoW is initiated with the detection of local interest points. These point detectors should find specific points and regions in an image invariant to certain geometric and photometric transformations. One of the popular approaches for detecting local interest points, which was applied in this experiment, is the Difference of Gaussians (DoG). The DoG detector proposed by Lowe (1999) is designed to be invariant to translation, scale, rotation, and illumination changes and to be able to sample images from different locations and scales (Zare et al. 2013).

Secondly, some distinctive features are extracted that characterize an image's set of key points. For a bag of visual words to function properly, the next step is constructing a codebook, which requires that 128-dimensional local image features be quantized into

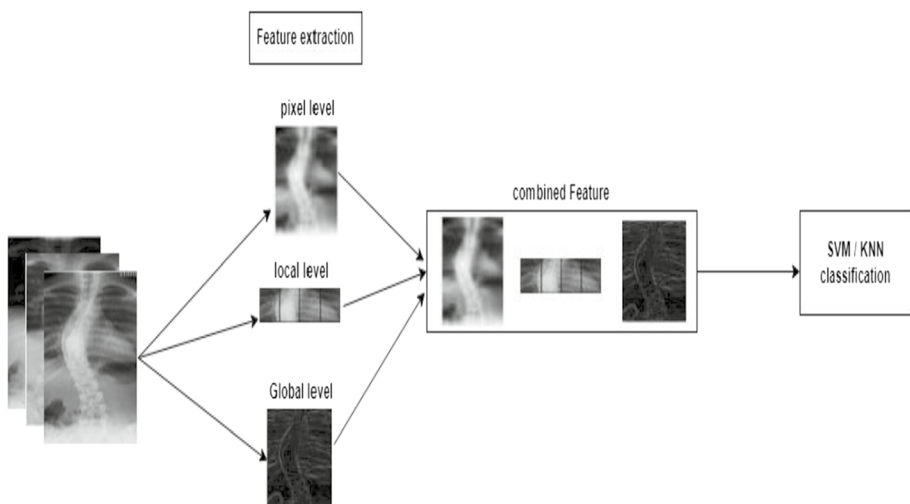


Fig. 5 The three-level features-based classification (Mueen et al. 2007)

discrete visible words. A clustering algorithm or a vector quantization algorithm is used for this task. Usually, the k-means clustering method is used in this step, along with the cluster center as a visual vocabulary term. Following the identification of cluster centers, each image is presented as a histogram of those cluster centers by simply counting the frequency with which the words appear in the image. Every feature vector within an image is assigned to a cluster center to accomplish this task by giving it a Euclidean distance to its nearest neighbor (Zare et al. 2013).

3.2 Model generation module

After the image processing module, the next module in the process is the model generation module to detect the type of curvature of the spine (C-shaped or S-shaped). The training set, the test set, and the label of each image are specified. After extracting the features from the training set, the extracted features are entered into the classifier along with the label of each image in the dataset to construct a classification model. The Support Vector Machine (SVM) and K-nearest neighbor have been classified better than other classification techniques based on experimental results and several applications in this field (Eljamassi and Maghari 2020; Hasoon et al. 2021; Zare et al. 2013). As a result, these two methods are used in this study to investigate the problem.

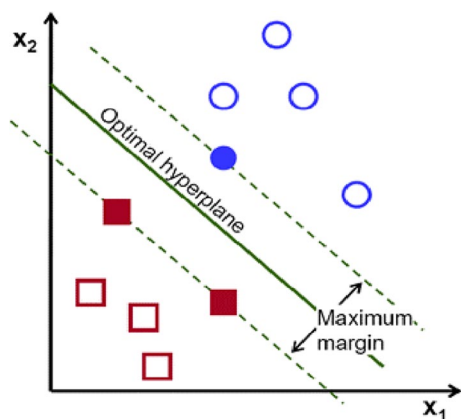
3.2.1 Support vector machine (SVM)

The SVM approach is exceptionally appealing for image classification because it produces a hyperplane when separating relevant and irrelevant vectors and thus maximizes the margin between them. Optimal hyperplanes (Fig. 6) have the most significant margin toward sample objects; the greater the margin, the less likely any feature vector will be misclassified (Wasule and Sonar 2017; Zare et al. 2013).

3.2.2 K-nearest neighbors

The KNN classifier is the simplest method to classify data in machine learning. A KNN classification is based on the majority vote among the neighboring classes expressed as k-nearest neighbors. Classification can be achieved by identifying the neighbors closest to

Fig. 6 Optimal hyperplane margin (Wasule and Sonar 2017)



the query example and using those neighbors to determine the class to which the query belongs (Govil and Chugh 2015).

3.3 Transfer learning module

The pre-trained mobileNetV2 and Pre-trained Xception modules are used in Fig. 3b. The MobileNetV2 and Xception are networks built using convolutional neural networks (CNNs). In recent years, CNN has achieved excellent results in object recognition and image classification, which is one of the most established algorithms among the various models of deep learning networks (Krizhevsky et al. 2012; Russakovsky et al. 2015).

3.3.1 Convolutional neural network (CNN)

Convolutional neural networks (CNNs) are a subset of artificial neural networks. Like traditional neural networks (TNNs), CNNs consist of neurons, layers, and weights. The most significant difference between CNNs and TNNs is that CNNs have unique ability to detect and classify images. That is, the input to deep neural networks can include information matrices with higher dimensions (Krizhevsky et al. 2012).

The CNN architecture has several building blocks, such as convolution, pooling, and fully connected layers. An architectural structure typical of the process uses a stack of several convolution layers and a pooling layer, followed by one or more fully connected layers. As a result of these layers, input data are transformed into output data, known as forward propagation (Fig. 7) (Yamashita et al. 2018).

3.3.2 Training-based transfer learning

In medical imaging, it is desirable to have an abundance of well-labeled data, but it is not readily available because of the cost and the load of radiologists. Two techniques can be used to train a model on a smaller dataset: data augmentation and transfer learning. A common and effective strategy for training a network on a small dataset is transfer learning, which involves pre-training a network on a vast dataset such as the ImageNet

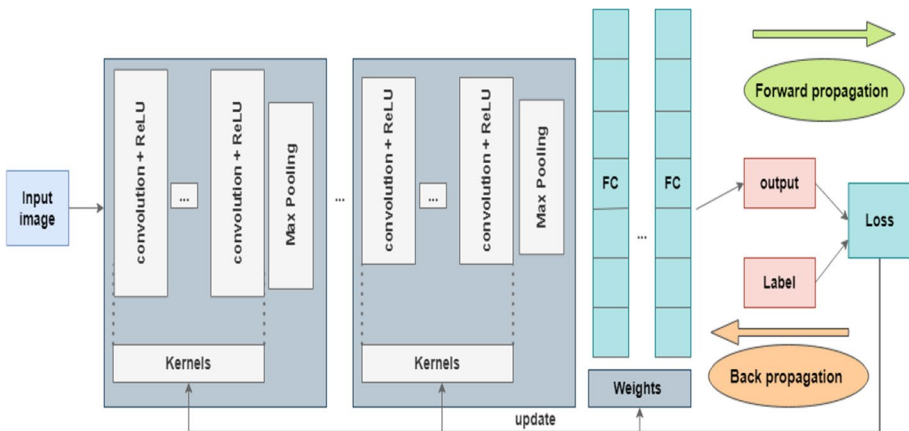


Fig. 7 A representation of the architecture and training process of a convolutional neural network (CNN) (Yamashita et al. 2018)

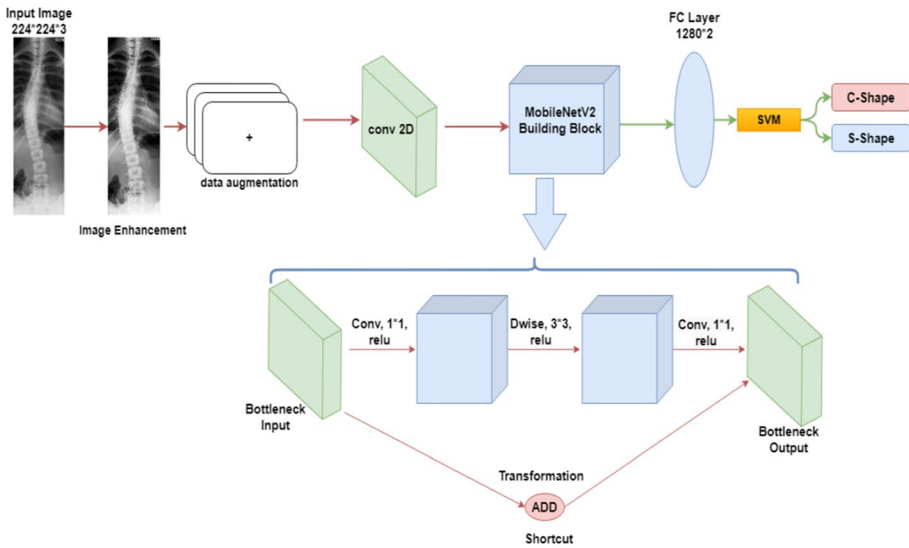


Fig. 8 The MobileNetV2 architecture consists of bottlenecks and depth-separable convolutions (Hashmi et al. 2020)

dataset, which contains 1.4 million images with 1000 classes, and then reused to apply to a specific problem (Yamashita et al. 2018). While this method reduces the cost of training models, it also enhances the effectiveness of machine learning in a significant way. As a result, transfer learning can enable people to handle some new or different application scenarios when there is not enough labeled data available to allow machine learning to be performed (Xiang et al. 2019).

3.3.3 Pre-trained MobileNetV2

A group of Google researchers developed Mobile-Nets as a small and low-power model (Sandler et al. 2018). The MobileNetV2 is an advanced version of the MobileNetV1. It uses depth-wise separable convolutions as efficient building blocks. The architecture has two types of blocks: residual blocks with a stride of one, and downsizing blocks with a stride of two. In addition, it has linear bottlenecks between the layers, which should be avoided by using linear layers to prevent nonlinearities from damaging a large volume of data. These bottlenecks in the model can help by encoding intermediate inputs and outputs. As a result of the inner layer, lower-level concepts such as pixels are transformed into higher-level descriptors such as image categories. Additionally, bottlenecks are connected by shortcuts. As shown in Fig. 8, a pre-trained MobileNetV2 was used (Hashmi et al. 2020). In this network, SVM is used instead of the sigmoid activation function to improve the accuracy of the classification as well as prevent the vanishing gradient in the last layer of the network.

3.3.4 Pre-trained Xception

Francois Chollet proposed the Xception model (Chollet 2017) which is an extension of the Inception architecture and uses modified depth-wise separable convolutions instead of the standard Inception modules, which makes it better for both the ImageNet ILSVRC and JFT datasets. This modified depth-wise separable convolution initiates with a pointwise convolution and ends with a depth-wise convolution. Res-Net architecture also proposes residual/skip connections (He et al. 2016). There are three blocks in the Xception architecture: entry flow, middle flow, and exit flow. In addition, eight repetitions of the middle flow block are available. A modified depth-wise separable convolution in the architecture has no intermediate non-linearity. As shown in Fig. 9, the pre-trained model of Xception has been used (Hashmi et al. 2020). To improve the performance and increase the accuracy of the network in the classification of spinal curvature types, SVM is used as the activation function of the last layer.

4 Materials and methods

In this section, several experiments have been presented that analyzed the performance of pre-trained mobileNetV2 and pre-trained Xception networks classification against SVM and KNN methods using feature extraction methods originating from related radiographic image classification articles (Mueen et al. 2007; Zare et al. 2013).

In this article, 1000 anterior–posterior radiographs of the spine with C-shaped and S-shaped scoliosis are used. This private dataset was collected from the Shafa Hospital in Tehran, Iran, and labeled by two medical specialists. In this dataset, the anterior–posterior images of the spine relate to different genders, age groups, and geographical locations who have visited the hospital with scoliosis in the last 20 years and represent the population of patients with scoliosis.

The radiographic images included in this dataset are anterior–posterior images of the spine, which include 17 vertebrae. Such images should be standardized and normalized before usage. For this purpose, the images are improved, which equalized the histogram and adjusted its resolution and size considering to the algorithm and network.

The dataset includes 500 anterior–posterior C-shaped radiographic images and 500 anterior–posterior S-shaped radiographic images (Table 1). Eighty percent of the dataset

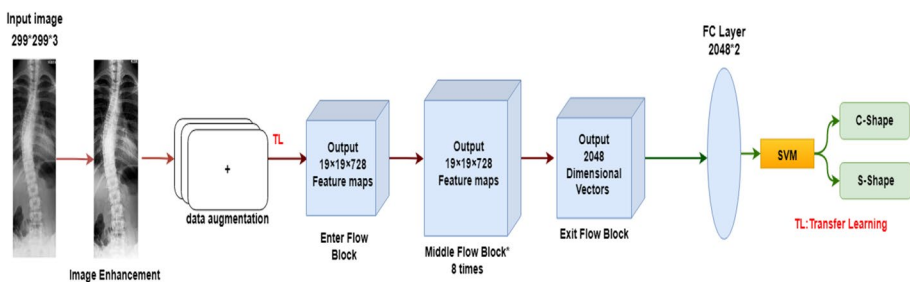
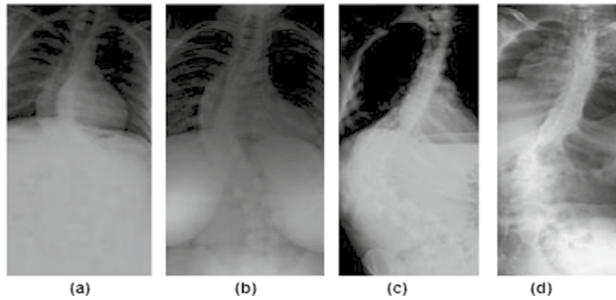


Fig. 9 An Xception architecture comprises three blocks: the entry flow block, the middle flow block, and the exit flow block (Hashmi et al. 2020)

Table 1 Describing the experimental private dataset

Class	Private dataset-based training dataset	Private dataset-based testing dataset
C-shape	400 anterior–posterior images of the spine	100 anterior–posterior images of the spine
S-shape	400 anterior–posterior images of the spine	100 anterior–posterior images of the spine

**Fig. 10** The specimens of the private dataset used to classify C-shaped and S-shaped curvatures (a–d)

is the training set, and twenty percent is the test set. Figure 10 illustrates specimens of the private dataset, and Table 1 describes its characteristics.

The purpose of this article is first to examine the accuracy of classifying spinal anomalies into C-shaped and S-shaped classes by SVM and KNN algorithms using different methods of image representation. Then, the performance of these classical algorithms is analyzed using the two pre-trained MobileNetV2 and pre-trained Xception networks. A general outline of the procedure is shown in Fig. 11.

4.1 The first experiment

A diagram of the first experiment is shown in Fig. 12. This experiment involves the following steps.

1. Equalization of histograms.
2. The texture (co-occurrence matrix) and shape features are extracted as global features from each image (16 textures + 37 edges).
3. The original image is resized to have a 15×15 pixel resolution to get the pixel vector.
4. All the above-extracted features are combined into one vector. The total dimensionality of the feature vector for this experiment reached 278 ($225 + 53 = 278$).
5. A SVM was tested using the Radial basic function (RBF) kernel and the KNN with K values of 3, 5, 7, 9, 11, 13, and 15, and the best results were obtained without the under-fitting or over-fitting model.

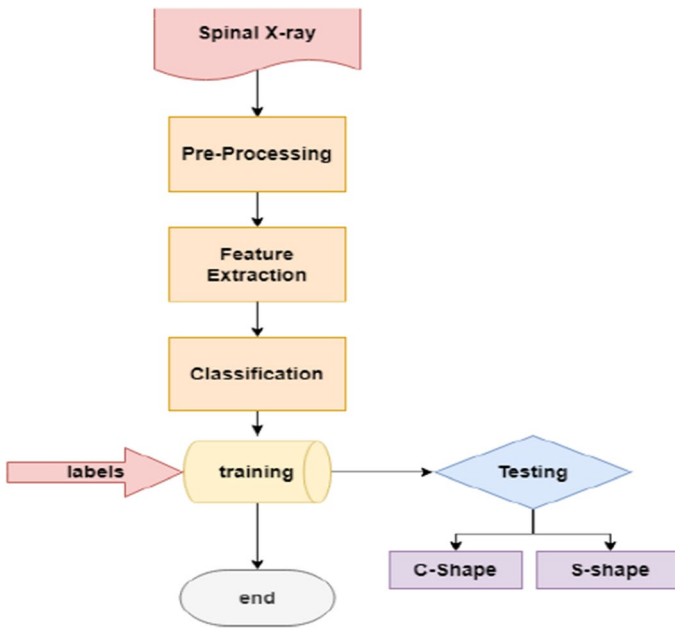


Fig. 11 A general description of the experiments in Experiments 1 to 4

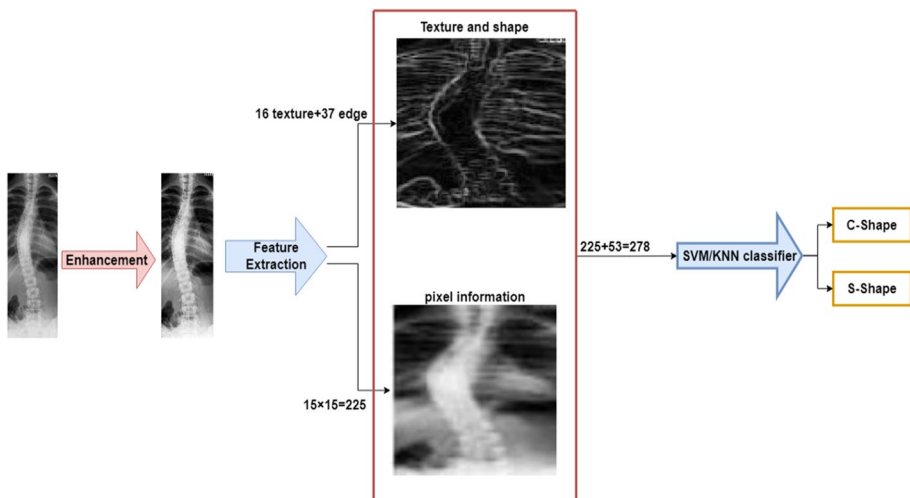


Fig. 12 A schematic representation of the first experiment (Zare et al. 2013)

4.2 The second experiment

A diagram of the second experiment is shown in Fig 13. This experiment involves the following steps.

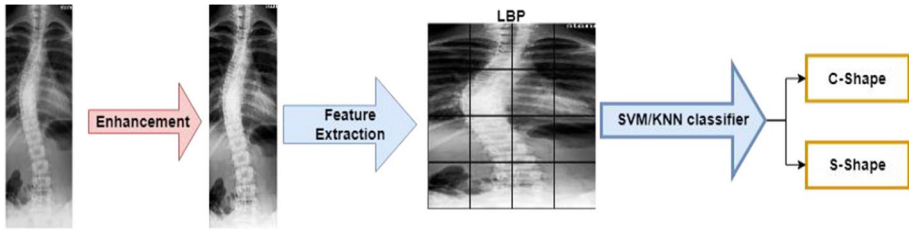


Fig. 13 A schematic representation of the second experiment (Zare et al. 2013)

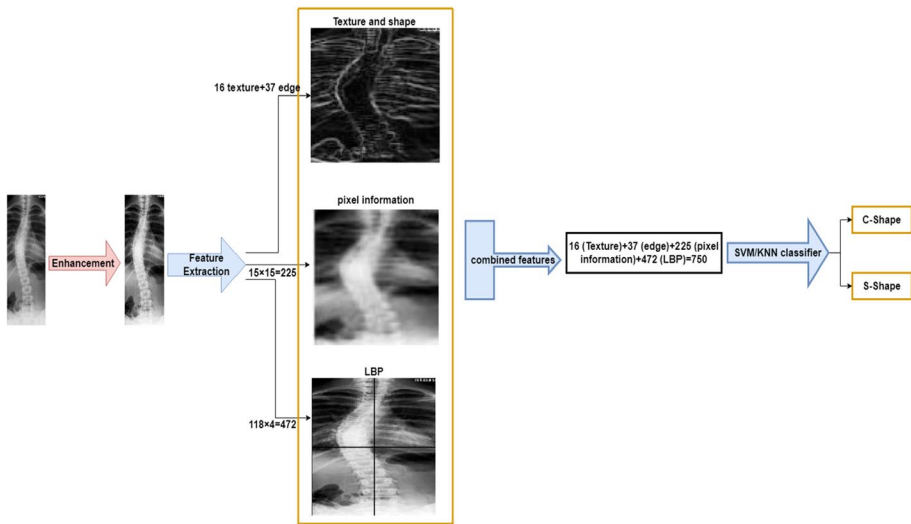


Fig. 14 A schematic representation of the third experiment (Zare et al. 2013)

1. Equalization of histograms.
2. A local binary pattern (LBP) is extracted from each sub-region and then combined at the end. Since the image consists of 16 sub-images, $59 \times 16 = 944$ histogram bins are generated for the image containing the feature vector.
3. A SVM was tested using the RBF kernel and the KNN with K values of 3, 5, 7, 9, 11, 13, and 15, and the best results were obtained without the under-fitting or over-fitting model.

4.3 The third experiment

A diagram of the third experiment is shown in Fig 14. This experiment involves the following steps.

1. Equalization of histograms.
2. The texture (co-occurrence matrix) and shape features are extracted as global features from each image (16 textures + 37 edges).
3. The original image is resized to have a 15×15 pixel resolution to get the pixel vector.

4. All of the above-extracted features are combined in one vector. As a result, the overall dimensions of the feature vector for this experiment were 278 ($278 = 53 + 225$).
5. Each image is divided into four non-overlapping sub-regions. The LBP is extracted from each sub-region.
6. Finally, adding the vectors to a single vector, the feature vector reaches 750.
7. A SVM was tested using the RBF kernel and the KNN with K values of 3, 5, 7, 9, 11, 13, and 15, and the best results were obtained without the under-fitting or over-fitting model.

4.4 The fourth experiment

The BoW is extracted from training images and fed into SVM and KNN classifiers to construct a classification model. Finally, A SVM was tested using the RBF kernel and the KNN with K values of 3, 5, 7, 9, 11, 13, and 15, and the best results were obtained without the under-fitting or over-fitting model.

4.5 The fifth experiment

The two pre-trained MobileNetV2 and pre-trained Xception networks are used in the fifth experiment to compare the accuracy and power of deep neural networks against classical machine learning methods. Figures 8 and 9 show a diagram of the procedure and details for the fifth experiment, and the steps for the experiment are as follows:

4.5.1 Pre-processing and data augmentation

For a pre-trained network to work, radiographic images are used as input. It is essential to pre-process each dataset according to the used deep neural network. Resizing and normalizing are two crucial steps in this field. These images must be resized to make them appropriate for the pre-trained MobileNetV2 and Xception networks. To enter the pre-trained MobileNetV2 network, images are first resized to 224×224 , and to enter the pre-trained Xception network, images are resized to 299×299 . The pixel intensity is transferred to the $[0, 1]$ range and then normalized.

The training of a neural network requires a large dataset to achieve the desired results. With a small dataset, the parameters of the networks weaken, and the trained networks have a feeble generalization ability. Furthermore, noise may affect the accuracy of classification. Generally, data augmentation is the solution usually used to resolve this problem. For this purpose, data augmentation is used to increase the amount of training data and ensure

Table 2 The data augmentation technique used in the proposed method

Technique	Setting
Width_shift_range	0.2
Height_shift_range	0.2
Zoom_range	0.2
Horizontal_flip	True

robustness against noise attacks (Kim et al. 2021). In this dataset, five hundred images have been labeled as C-shaped, and five hundred images have been labeled as S-shaped. The data augmentation process produces 24,000 images. Table 2 below shows the settings used in data augmentation.

Subsequently, the data is divided into training and test sets. About 80% of the data is contained in the training set, and the rest is included in the test set. The training is done on the training set, and the built model is evaluated on the test set (Table 1).

4.5.2 Fine-tuning of the networks

In Sects. 3 and 4.2, all details of the pre-trained MobileNetV2 and pre-trained Xception networks architecture are discussed in detail. This private anterior–posterior spine X-ray images dataset has been used to fine-tune the network after preprocessing and normalization. Then, the data augmentation method is used to improve the efficiency of the processing of the dataset. All the layers of the networks could be trained, and each layer extracted different visual features from the images for classification purposes. However, due to the small size of the dataset, the pre-trained mobileNetV2 and pre-trained Xception networks were used through transfer learning. Indeed, the pre-trained mobileNetV2 and pre-trained Xception networks were used on the image-Net dataset for classification purposes (Sandler et al. 2018).

Consequently, after preparing the data and transferring them to the networks at this point, only the last layer of the network is trained to classify the types of curvatures. Before network training begins, several hyper-parameters must be set. Kingma published an interesting paper (Kingma and Ba 2014), which indicates that the adaptive moment estimation method (Adam) performs better than other optimizers in generalization. Adam has been used as the optimizer, and this model has been trained for 50 epochs. Another hyper-parameter, 0.001, is required to control the learning rate (Table 3). For the last layer, SVM is used instead of the sigmoid activation function to avoid the vanishing gradient problem. To achieve the best accuracy, the networks must be fine-tuned to detect whether the spine has a C or S shape.

4.6 Performance criteria for classification

After completing the training, the developed models should be evaluated using test datasets. The models' performance is assessed based on accuracy, recall, precision, and the F1 score. In this section, all the performance metrics that have been used in this paper are discussed. In the following definitions and equations, True-Positive (TP) represents the number of C-shaped images as C-shaped in classifying patients with C-shaped and S-shaped curvature. True-Negative (TN) indicates the number of S-shaped images specified as S-shaped. False-Positive (FP) represents the number of S images incorrectly known as C images, and False-Negative (FN) indicates the number of C-curved X-ray images that are incorrectly identified as S:

Table 3 Hyper-parameters used to fine-tune the model

Architecture	Image Size	Epochs	Optimizer	Learning rate
Pre-trained MobileNetV2	224×224	50	Adam	0.001
Pre-trained Xception	299×299			

- *Accuracy*: Accuracy is the ratio of the total number of correct predictions and the total number of predictions.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

- *Precision*: Precision is the ratio between the True Positives and all the Positives.

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

- *Recall*: The recall is the measure of the model correctly identifying True Positives.

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

- *F1-score*: F1-score is the harmonic mean of precision and recall

$$F1 = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{4}$$

5 Experimental results

To evaluate the performance and implementation of the machine learning algorithms, a standard computer with 8 GB of RAM, and a seventh-generation NVIDIA GeForce 940 MX processor was used. As the backend of the pre-trained MobileNetV2 and pre-trained Xception networks, the Keras Open Source Deep Learning Framework with Tensorflow is used.

The results of the analysis of C-shaped or S-shaped spinal deformities with classical machine learning methods, i.e. SVM and KNN and pre-trained MobileNetV2 and pre-trained Xception networks, are described in this section, and the results are compared.

5.1 The results of performance criteria

Parameters and hyper-parameters were set during the experimentation process. The test dataset images were pre-processed in the same manner as the training images.

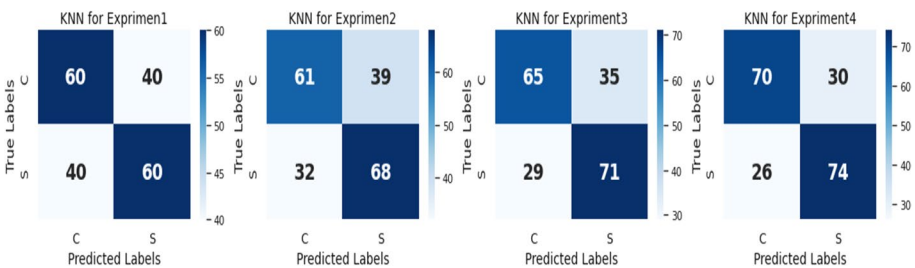


Fig. 15 C-shaped and S-shaped spinal deformity classification confusion matrix results for the first, second, third, and fourth experiments using KNN

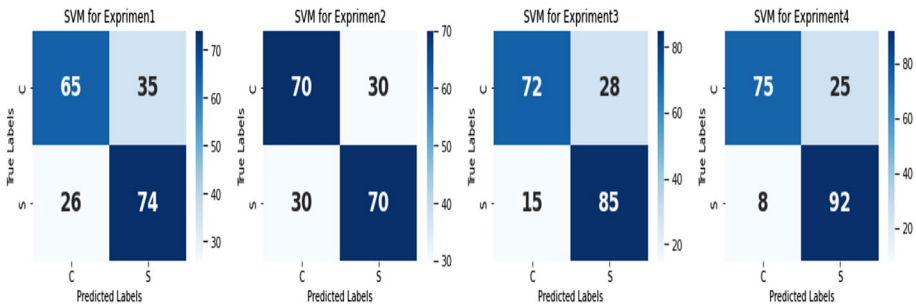


Fig. 16 C-shaped and S-shaped spinal deformity classification confusion matrix results for the first, second, third, and fourth experiments using SVM

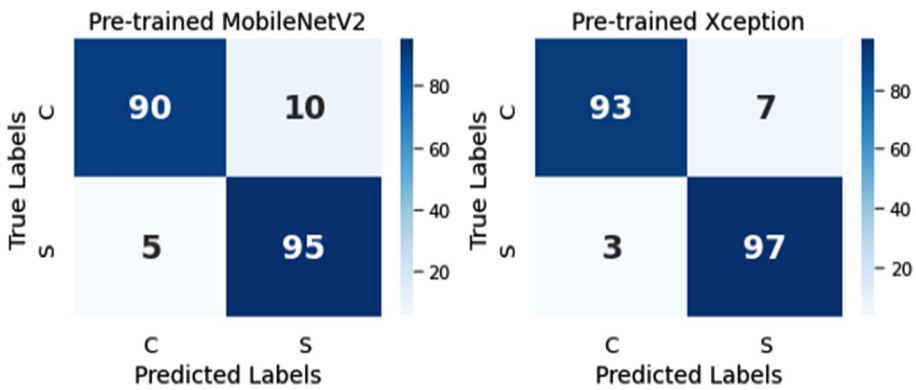


Fig. 17 C-shaped and S-shaped spinal deformity classification confusion matrix results for pre-trained MobileNetV2 and pre-trained Xception

The confusion matrix is also used to obtain the classification models of the experiments (Figs. 15, 16, and 17). The confusion matrix allows the number of correct and incorrect diagnoses for the C-shaped and the S-shaped curvature for Scoliosis to be analyzed for the desired models.

For purposes of assessing the efficacy of the methods, the accuracy, precision, recall, and F1 score for the SVM, KNN methods and pre-trained mobileNetV2 and pre-trained Xception networks for classifications of Scoliosis have been calculated (see Table 4 and Figs. 18, 19, 20, 21).

A comparison of accuracy, precision, recall, and F1 score for C-shaped and S-shaped spine curvature classification based on classical machine learning methods, i.e. SVM and KNN, against the two pre-trained networks proposed by MobileNetV2 and Xception paid is presented in Figs. 18, 19, 20 and 21. As shown, the pre-trained networks significantly outperform classical machine learning methods due to automatic extraction.

In Table 4, the performance of the proposed methods is compared with those of radiographic image classification articles published in recent years.

Considering Table 4, pre-trained deep networks outperform classical machine learning methods due to their automatic feature extraction attribute. Furthermore, pre-trained Xception and pre-trained MobileNetV2 networks have better performance than the

Table 4 An examination of accuracy, precision, recall, and F1 score using KNN, SVM, pre-trained MobileNetV2, and pre-trained Xception to classify spinal deformities based on recent research methods

Experiment	Classifier	Type of curve	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
First experiment in (texture and shape) + pixel	KNN	C	60	60	60	60
		S	60	60	60	60
	SVM	C	70	71	65	68
Second experiment in LBP		S	70	68	74	71
	KNN	C	65	66	61	63
		S	65	64	68	66
	SVM	C	70	70	70	70
Third experiment in (texture and shape) + pixel information + LBP		S	70	70	70	70
	KNN	C	68	69	65	67
		S	68	67	71	69
Fourth experiment in BoW	SVM	C	79	83	72	77
		S	79	75	85	80
	KNN	C	72	73	70	71
		S	72	71	74	72
Pre-trained AlexNet (Rahman et al. 2020)	SVM	C	84	90	75	82
	S	84	79	92	85	
	Sigmoid	C	90	91	88	89
	S	90	88	91	89	
Pre-trained ResNet18 (Minaee et al. 2020; Rahman et al. 2020)	Sigmoid	C	90	90	90	90
	S	90	90	90	90	
Pre-trained ResNet50 (Minaee et al. 2020)	Sigmoid	C	90	91	89	90
	S	90	89	91	90	
Pre-trained VGG-16 (Dansana et al. 2020)	Sigmoid	C	90	93	86	89
	S	90	87	94	90	
Pre-trained Squeeze-Net (Minaee et al. 2020; Rahman et al. 2020)	Sigmoid	C	91	91	91	91
	S	91	91	91	91	

Table 4 (continued)

Experiment	Classifier	Type of curve	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Pre-trained DenseNet201 (Rahman et al. 2020)	Sigmoid	C	92	96	87	91
Fifth experiment (pre-trained MobileNetV2)	SVM	S	92	88	96	92
		C	93	95	90	92
Fifth experiment (pre-trained Xception)	SVM	S	93	90	95	92
		C	95	97	93	95
		S	95	93	97	95

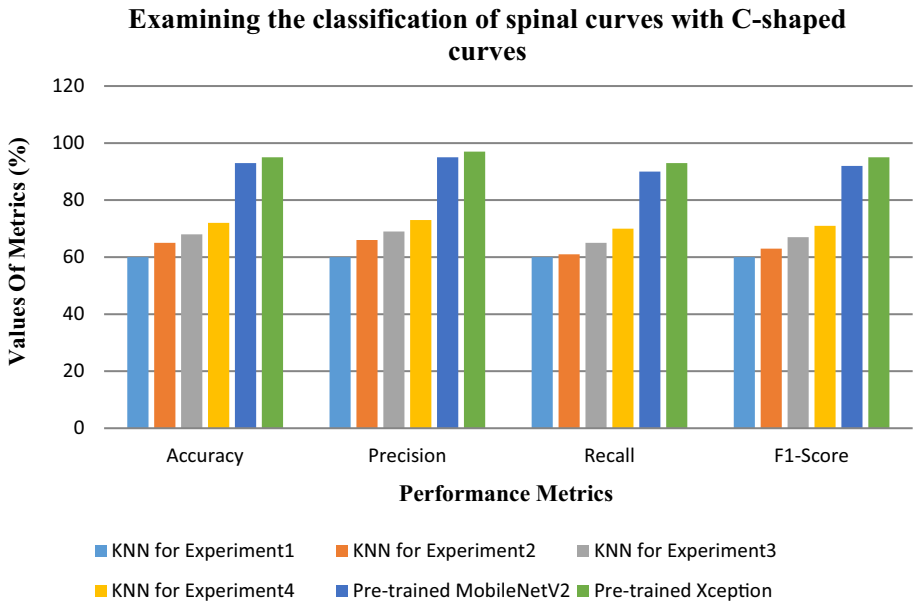


Fig. 18 Examining accuracy, precision, recall, and F1 score using the KNN method and pre-trained networks to detect the C-shaped curve of the spine

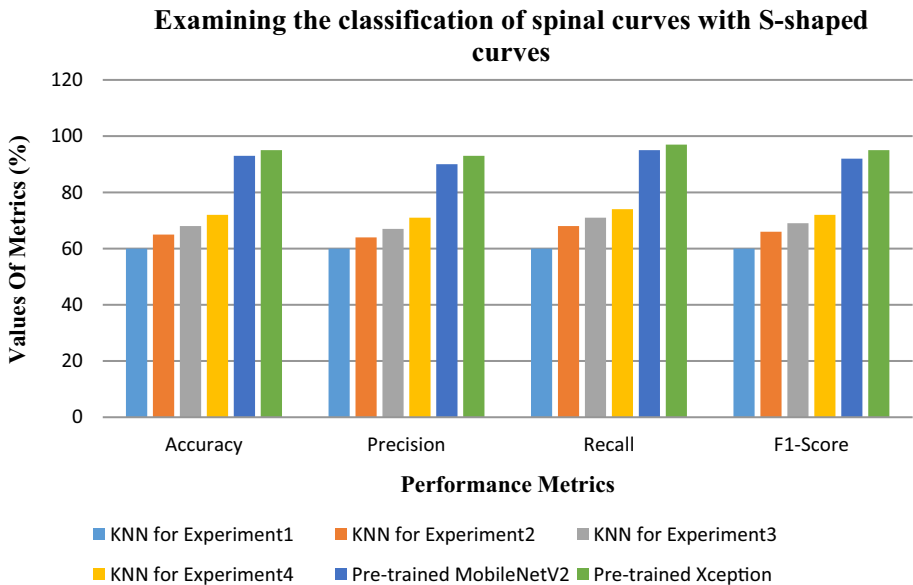


Fig. 19 Examining accuracy, precision, recall, and F1 score using the KNN method and pre-trained networks to detect the S-shaped curve of the spine

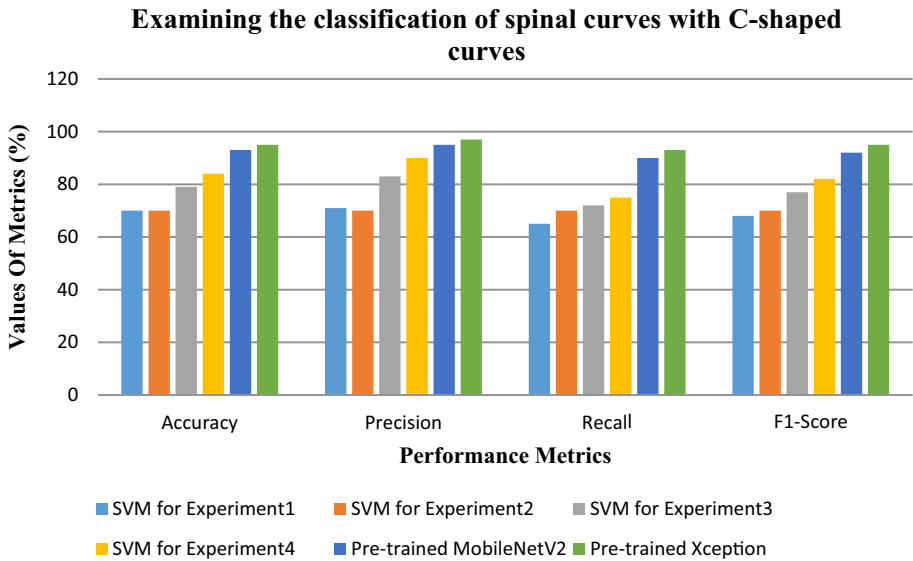


Fig. 20 Examining accuracy, precision, recall, and F1 score using the SVM method and pre-trained networks to detect the C-shaped curve of the spine

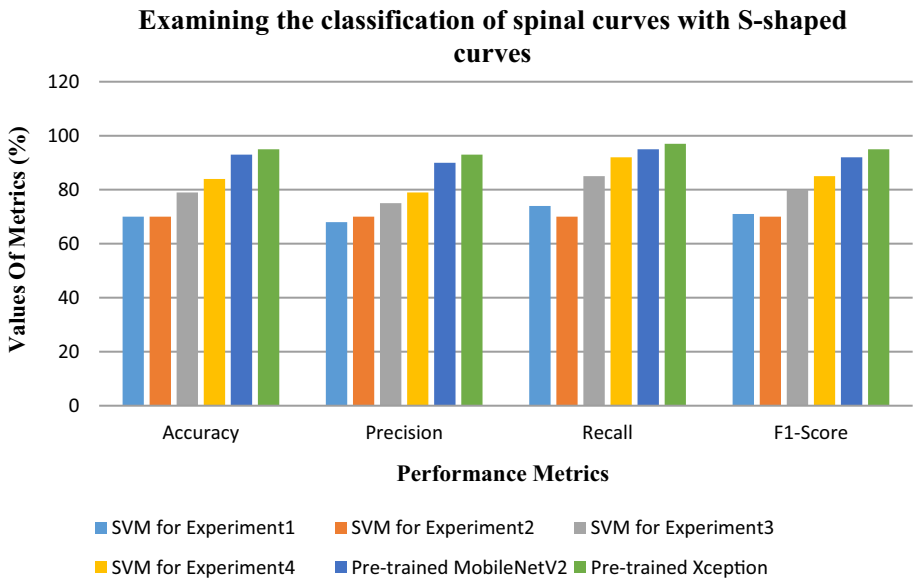


Fig. 21 Examining accuracy, precision, recall, and F1 score using the SVM method and pre-trained networks to detect the S-shaped curve of the spine

aforementioned methods of articles (Dansana et al. 2020; Minaee et al. 2020; Rahman et al. 2020) due to the use of depth-wise separable convolution and pointwise convolution.

Most commonly, the sigmoid function is used as the activation function of the last layer in binary classification; however, this function may suffer from vanishing gradient. The purpose of this article is to prevent this problem and improve the performance of networks by using SVM as the activation function of the last layer.

Based on the results of five experiments (Table 4 and Figs. 18, 19, 20, 21), pre-trained Xception performed better than the pre-trained MobileNetV2 network and classical SVM and KNN methods for the classification of spinal curvature types. After pre-trained Xception, pre-trained MobileNetV2, experiment 4, experiment 3, and finally experiments 2 and 1 are listed in order of accuracy. The results indicate that the SVM algorithm is more powerful than the KNN algorithm since the SVM algorithm is more accurate at detecting the type of spinal curvature.

According to Table 4, the proposed pre-trained networks with the SVM activation function have significant performance. Table 5 compares the classic machine learning algorithms (experiments 1 to 4) and the proposed pre-trained networks (experiment 5), highlighting the accuracy, computational complexity, the number of features extracted, and how the feature extraction was investigated in each experiment.

Table 5 shows that pre-trained networks perform significantly better than classical methods in terms of accuracy and speed. Similarly, due to the use of automatic feature extraction and data augmentation, they are more noise-resistant and robust. In addition, they have data augmentation for the problem of lack of data.

5.2 Analyzing the generalizability of the results

Having reviewed recent articles (Mohammed et al. 2021; Nagi et al. 2022; Rajinikanth et al. 2022), the lung pneumonia dataset in (Mohammed et al. 2021) was used to check the generalizability of the proposed deep pre-trained networks to other datasets and populations. There are 800 lung radiographs in this dataset (400 healthy cases and 400 COVID cases). In this dataset, 80% is considered for the train set and 20% for the test set, the results of which are presented in Table 6.

On the lung pneumonia dataset, the proposed pre-trained networks with SVM activation function perform well (Table 6).

5.3 Discussion

Due to its automatic feature extraction, deep neural networks can be more powerful than classical machine learning algorithms. Generally, deep convolutional neural networks (DCNNs) require large amounts of datasets to reach significant accuracy, and this training typically takes hours or even days. It is evident, however, that this paper's objective is the classification and analysis of the pre-trained mobileNetV2 and pre-trained Xception models and classic methods. Therefore, the parameters of the pre-trained Xception and pre-trained MobileNetV2 models can be shared via transfer learning with the new models and dataset to accelerate and optimize the training of the new model. Furthermore, Depth-wise convolution, pointwise convolution, and Inverted Residual Blocks reduce the number of network parameters, thereby pre-trained MobileNetV2 and pre-trained Xception with SVM are much easier to use, show substantial accuracy even with small datasets, and can even be used on mobile devices. For example, the TensorFlow lite open-source API can be used to

Table 5 The accuracy, time complexity, and feature extraction methods of classical machine learning algorithms and pre-trained neural networks

No.	ML techniques	Accuracy (%)	Number of features	Time complexity	features extracted
1	First experiment + KNN	60	278	7 ms	Texture and shape + pixel information
2	First experiment + SVM	70	278	29 ms	Texture and shape + pixel information
3	Second experiment + KNN	65	944	29 ms	LBP
4	Second experiment + SVM	70	944	105 ms	LBP
5	Third experiment + KNN	68	750	23 ms	(Texture and shape) + pixel information + LBP
6	Third experiment + SVM	79	750	86 ms	(Texture and shape) + pixel information + LBP
7	Fourth experiment + KNN	72	7,540,256	12700 s	BoW
8	Fourth experiment + SVM	84	7,540,256	38114 s	BoW
9	Fifth experiment (pre-trained MobileNetV2 + SVM)	93	1280	53 s	Automatic
10	Fifth experiment (pre-trained Xception + SVM)	95	2048	153 s	Automatic

Table 6 Analyzing the generalizability of the proposed networks using lung pneumonia data

No.	ML techniques	Evaluation metrics			
		Accuracy	Precision	Recall	F1-score
1	ANN (Mohammed et al. 2021)	0.92	0.90	0.91	0.92
2	SVM (Mohammed et al. 2021)	0.95	0.93	0.92	0.93
3	KNN (Mohammed et al. 2021)	0.92	0.89	0.90	0.90
4	The proposed method (pre-trained MobileNetV2 + SVM)	0.96	0.96	0.95	0.95
5	The proposed method (pre-trained Xception + SVM)	0.96	0.96	0.96	0.96

deploy this article model on an Android smartphone for deep learning, making this method readily applicable to physicians and spine specialists.

5.4 Conclusion

One of the most complex spinal deformities is scoliosis. There are two types of abnormality: C-shaped and S-shaped. Each of these abnormalities requires a different treatment procedure. The vertebrae can sometimes balance themselves and it can be difficult to determine the type of curvature. Often, the bias of the observer and the low quality of the image may lead to an incorrect diagnosis of the type of abnormality. Secondly, there are not enough doctors and radiologists specializing in this field to diagnose and guide patients. The evaluation of abnormalities, however, takes time. Furthermore, this disease is progressive, and the patient may experience more severe stages of the disease over time, requiring braces and complicated surgery. The importance of automated assessment systems can be seen here. In order to evaluate spine abnormalities promptly, it is necessary to use intelligent methods and machine learning algorithms. Since deep networks automatically extract features from medical images, they are useful for the analysis of medical images and the development of spinal deformity evaluation systems.

This study aims to examine the type of spinal deformity through the integration of technology and health care. The automatic system acts as an expert radiologist in diagnosing the type of abnormality and assessing many patients in a short period of time. However, the automatic evaluation system does not display any bias in its evaluation.

This study compares the classical SVM and KNN machine learning algorithms and the pre-trained MobileNetV2 and pre-trained Xception networks to classify anterior–posterior radiology images of C-shaped or S-shaped scoliosis. Sometimes, the type of deformation cannot be initially detected. It is essential to identify the type of curvature to develop systems for assessing spinal deformities. This article aims to provide an overview of the classification of the types of spinal curvature due to the importance of the topic and the lack of research in this field.

Various image representations are used in this article to extract features from images. Generally, image representations can be categorized into two groups: low-level image representations and local patch-based image representations. This study uses texture features as a low-level representation of the image. A BoW represents an image using the local patch. In various experiments, these features have been employed to classify X-ray images of spinal deformities.

Due to automatic feature extraction, the pre-trained Xception and pre-trained Mobile-NetV2 network with SVM as the final activation function perform significantly better than the classical methods. Furthermore, the results indicate that [(texture and shape)+pixel information+LBP] and BoW-based classification methods of feature extraction perform better than other models produced with other representation techniques. Finally, the SVM algorithm is more powerful and has greater accuracy than KNN.

A limitation of this article is the small dataset of anterior–posterior spine images and the lack of data. Deep networks require large datasets with correct labeling to be effective. As a solution to this problem, data augmentation and transfer learning are employed in this article. In future research, a larger dataset could be utilized, or transfer learning and data augmentation could be used.

Additionally, since X-rays negatively affect a patient's quality of life, the evaluation of the patient should be conducted with the minimum amount of radiation to diagnose the type and severity of the abnormality.

One of the potential ethical principles when using radiographic images is to avoid using personal information about patients. During the collection of the dataset, only radiographic images without any personal information were collected, and the machine learning algorithms and deep networks of the images, along with their curvature type labels, are used as the ground truth.

As a practical tool, this evaluation system can be used to help the doctor make a clinical diagnosis and to assist him in making a final decision. To fully evaluate the abnormalities in the spine structure automatically in future research, it is suggested that segmenting the vertebrae, measuring the Cobb angle, and detecting the type of curvature should be incorporated into the spinal automatic evaluation system with significant accuracy. In the end, the doctor makes the final decision for the patient based on his diagnosis and the information provided by the automated evaluation system.

It is noteworthy that since radiation can affect the quality of life of patients, and patients have to be exposed to radiation regularly during frequent visits, this evaluation system is highly valuable since it can determine the type of curvature of the spine, the segmentation of the vertebrae, the degree of curvature of the spine and calculate Cobb's angle based on the least radiation required.

As a last point, to advance the research and development of the proposed method, a larger dataset can be used to detect C-shaped curvature of the spine and S-shaped curvature of the spine to improve the classification accuracy of the spinal curvature type. In addition to these neural networks, other networks can be used to classify spine curvatures. It is also possible to improve the performance of networks with activation functions, which are crucial to the prevention of vanishing gradient.

Acknowledgements The authors are sincerely grateful to Dr. M. Chehrassan, Dr. A. Mahnama, H. Mohseni, A. Mokhtari, S. Mir Azimi, P. Shirmia, R. Kordi, and M. Taghavi for collecting data and answering numerous questions related to scoliosis and spinal disorders with great patience.

Author contributions PT: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing—original draft. MA: Conceptualization, Methodology, Validation, Formal analysis, Resources, Data curation, Supervision, Project administration, Visualization, Writing—original draft, Writing—review & editing. AK: Conceptualization, Methodology, Validation, Formal analysis, Supervision, Writing—original draft, Writing—review & editing. AB: Validation, Resources, Supervision.

Data availability The private dataset used in this study includes 1000 anterior–posterior radiographic images obtained from Shafa Hospital in Tehran, Iran. The private dataset used during this study is not available due to the possibility of compromising individual privacy.

Declarations

Conflict of interest The authors (Parisa Tavana, Mahdi Akraminia, Abbas Koochari, Abolfazl Bagherifard) declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Avni U, Goldberger J, Sharon M, Konen E, Greenspan H (2010a) Chest X-ray characterization: from organ identification to pathology categorization. In: Proceedings of the international conference on multimedia information retrieval, pp 155–164
- Avni U, Greenspan H, Konen E, Sharon M, Goldberger J (2010b) X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words. *IEEE Trans Med Imaging* 30(3):733–746
- Bhattach PB, Jain V (2020) Artificial intelligence in modern medicine—the evolving necessity of the present and role in transforming the future of medical care. *Cureus* 12(5):841
- Chang-ming Z, Guo-Chang GU, Hai-bo L, Jing S, Hualong Y (2008) Segmentation of ultrasound image based on texture feature and graph cut. In: 2008 International Conference on Computer Science and Software Engineering, vol 1, pp 795–798
- Chen B, Xu Q, Wang L, Leung S, Chung J, Li S (2019) An automated and accurate spine curve analysis system. *IEEE Access* 7:124596–124605
- Cherian T, Mulholland EK, Carlin JB, Ostensen H, Amin R, de Campo M, Greenberg D, Lagos R, Lucero M, Madhi SA (2005) Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies. *Bull World Health Organ* 83:353–359
- Chollet F (2017) Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1251–1258
- Dansana D, Kumar R, Bhattacharjee A, Hemanth DJ, Gupta D, Khanna A, Castillo O (2020) Early diagnosis of COVID-19-affected patients based on X-ray and computed tomography images using deep learning algorithm. *Soft Comput* 1:1–9
- Deselaers T, Ney H (2008) Deformations, patches, and discriminative models for automatic annotation of medical radiographs. *Pattern Recogn Lett* 29(15):2003–2010
- Deselaers T, Hegerath A, Keysers D, Ney H (2006) Sparse patch-histograms for object classification in cluttered images. *Jt Pattern Recogn Symp* 1:202–211
- Dimitrovski I, Kocev D, Loskovska S, Džeroski S (2011) Hierarchical annotation of medical images. *Pattern Recogn* 44(10–11):2436–2449
- Driscoll M, Aubin C-E, Moreau A, Villemure I, Parent S (2009) The role of spinal concave–convex biases in the progression of idiopathic scoliosis. *Eur Spine J* 18(2):180–187
- Dubost F, Collery B, Renaudier A, Roc A, Posocco N, Niessen W, de Bruijne M (2019) Automated estimation of the spinal curvature via spine centerline extraction with ensembles of cascaded neural networks. In: International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging, pp 88–94
- Eljamassi DF, Maghari AY (2020) COVID-19 detection from chest X-ray scans using machine learning. *Int Conf Promis Electron Technol.* 2020:1–4
- Foley JA (2013) UN world population report predicts 8 billion people by 2025, 11 billion by 2100. *Nature World News RSS*
- Franquet T (2001) Imaging of pneumonia: trends and algorithms. *Eur Respir J* 18(1):196–208
- Giancardo L, Meriaudeau F, Karnowski TP, Li Y, Garg S, Tobin KW Jr, Chaum E (2012) Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. *Med Image Anal* 16(1):216–226
- Govil S, Chugh N (2015) Prediction of defects in software using K-nearest neighbour algorithm for cost reduction. *Int J Adv Res Comput Sci Technol* 3(2):1–10
- Han J, Xu Q, Yang Y, Yao Z, Zhang C (2015) Evaluation of quality of life and risk factors affecting quality of life in adolescent idiopathic scoliosis. *Intract Rare Dis Res* 4(1):12–16
- Haralick RM, Shanmugam K, Dinstein IH (1973) Textural features for image classification. *IEEE Trans Syst Man Cybern* 6:610–621
- Hashmi MF, Katiyar S, Keskar AG, Bokde ND, Geem ZW (2020) Efficient pneumonia detection in chest X-ray images using deep transfer learning. *Diagnostics* 10(6):417

- Hasoon JN, Fadel AH, Hameed RS, Mostafa SA, Khalaf BA, Mohammed MA, Nedoma J (2021) COVID-19 anomaly detection and classification method based on supervised machine learning of chest X-ray images. *Results Phys* 31:105045
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 770–778
- Keyers D, Gollan C, Ney H (2004) Classification of medical images using non-linear distortion models. In: *Bildverarbeitung für die medizin 2004*. Springer, New York, pp 366–370
- Keyers D, Deselaers T, Gollan C, Ney H (2007) Deformation models for image recognition. *IEEE Trans Pattern Anal Mach Intell* 29(8):1422–1435
- Khanal B, Dahal L, Adhikari P, Khanal B (2019) Automatic Cobb angle detection using vertebra detector and vertebra corners regression. In: *International workshop and challenge on computational methods and clinical applications for spine imaging*, pp 81–87.
- Kim S-H, Lee J-H, Ko B, Nam J-Y (2010) X-ray image classification using random forests with local binary patterns. *Int Conf Mach Learn Cybern* 6:3190–3194
- Kim E, Kim J, Lee H, Kim S (2021) Adaptive data augmentation to achieve noise robustness and overcome data deficiency for deep learning. *Appl Sci* 11(12):5586
- Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. <https://arxiv.org/abs/1412.6980>.
- Komeili A, Westover LM, Parent EC, Moreau M, El-Rich M, Adee S (2014) Surface topography asymmetry maps categorizing external deformity in scoliosis. *Spine J* 14(6):973–983
- Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 25:1–10
- Lin Y, Zhou H-Y, Ma K, Yang X, Zheng Y (2019) Seg4Reg networks for automated spinal curvature estimation. In: *International workshop and challenge on computational methods and clinical applications for spine imaging*, pp 69–74
- Lowe DG (1999) Object recognition from local scale-invariant features. In: *Proceedings of the seventh IEEE international conference on computer vision*, vol 2, pp 1150–1157
- Ma W-Y, Zhang HJ (1998) Content-based image indexing. *Handb Multim Comput* 5:227
- Minace S, Kafieh R, Sonka M, Yazdani S, Soufi GJ (2020) Deep-COVID: predicting COVID-19 from chest X-ray images using deep transfer learning. *Med Image Anal* 65:101794
- Mohammed MA, Abdulkareem KH, Garcia-Zapirain B, Mostafa SA, Maashi MS, Al-Waisy AS, Subhi MA, Mutlag AA, Le D-N (2021) A comprehensive investigation of machine learning feature extraction and classification methods for automated diagnosis of COVID-19 based on X-ray images. *Comput Mater Contin* 66:3289–3310
- Mueen A, Baba S, Zainuddin R (2007) Multilevel feature extraction and X-ray image classification. *J Appl Sci* 7(8):1224–1229
- Mueen A, Zainuddin R, Baba MS (2008) Automatic multilevel medical image annotation and retrieval. *J Digit Imaging* 21(3):290–295
- Mustaffa MR, Ahmad F, Rahmat RWOK, Mahmud R (2008) Content-based image retrieval based on color-spatial features. *Malays J Comput Sci* 21(1):1–12
- Nagi AT, Awan MJ, Mohammed MA, Mahmud A, Majumdar A, Thinnukool O (2022) Performance analysis for COVID-19 diagnosis using custom and state-of-the-art deep learning models. *Appl Sci* 12(13):6364
- Niemeijer M, Van Ginneken B, Cree MJ, Mizutani A, Quellec G, Sánchez CI, Zhang B, Hornero R, Lamard M, Muramatsu C (2009) Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. *IEEE Trans Med Imaging* 29(1):185–195
- Niu S, Liu Y, Wang J, Song H (2020) A decade survey of transfer learning (2010–2020). *IEEE Trans Artif Intell* 1(2):151–166
- Ojala T, Pietikäinen M, Harwood D (1996) A comparative study of texture measures with classification based on featured distributions. *Pattern Recogn* 29(1):51–59
- Pratt H, Coenen F, Broadbent DM, Harding SP, Zheng Y (2016) Convolutional neural networks for diabetic retinopathy. *Procedia Comput Sci* 90:200–205
- Rahim SS, Jayne C, Palade V, Shuttleworth J (2016) Automatic detection of microaneurysms in colour fundus images for diabetic retinopathy screening. *Neural Comput Appl* 27(5):1149–1164
- Rahman T, Chowdhury MEH, Khandakar A, Islam KR, Islam KF, Mahub ZB, Kadir MA, Kashem S (2020) Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. *Appl Sci* 10(9):3233
- Rajinikanth V, Kadry S, Damaševičius R, Gnanasoundharam J, Mohammed MA, Devadhas GG (2022) UNet with two-fold training for effective segmentation of lung section in chest X-ray. In: *2022 third international conference on intelligent computing instrumentation and control technologies (ICICICT)*, pp 977–981

- Ravishankar H, Sudhakar P, Venkataramani R, Thiruvankadam S, Annangi P, Babu N, Vaidya V (2016) Understanding the mechanisms of deep transfer learning for medical images. In: Deep learning and data labeling for medical applications. Springer, New York, pp 188–196
- Reddy YM, Ravindran RE, Kishore KH (2018) Diabetic retinopathy through retinal image analysis: a review. *Int J Eng Technol* 7(15):19–25
- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M (2015) Imagenet large scale visual recognition challenge. *Int J Comput Vision* 115(3):211–252
- Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C (2018) Mobilenetv2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4510–4520. <https://doi.org/10.1109/CVPR.2018.00474>
- Sohaail ASM, Rahman MM, Bhattacharya P, Krishnamurthy S, Mudur SP (2010) Retrieval and classification of ultrasound images of ovarian cysts combining texture features and histogram moments. *IEEE Int Symp Biomed Imaging Nano Macro 2010*:288–291
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1–9
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826
- Tajbakhsh N, Gotway MB, Liang J (2015a) Computer-aided pulmonary embolism detection using a novel vessel-aligned multi-planar image representation and convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp 62–69.
- Tajbakhsh N, Gurudu SR, Liang J (2015b) A comprehensive computer-aided polyp detection system for colonoscopy videos. In: International conference on information processing in medical imaging, pp 327–338
- Tian G, Fu H, Feng DD (2008) Automatic medical image categorization and annotation using LBP and MPEG-7 edge histograms. In: 2008 international conference on information technology and applications in biomedicine, pp 51–53
- Tommasi T, Orabona F, Caputo B (2008) Discriminative cue integration for medical image annotation. *Pattern Recogn Lett* 29(15):1996–2002
- Vo HH, Verma A (2016) New deep neural nets for fine-grained diabetic retinopathy recognition on hybrid color space. *IEEE Int Symp Multim 2016*:209–215
- Wang S, Yin Y, Cao G, Wei B, Zheng Y, Yang G (2015) Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing* 149:708–717
- Wang L, Xu Q, Leung S, Chung J, Chen B, Li S (2019a) Accurate automated Cobb angles estimation using multi-view extrapolation net. *Med Image Anal* 58:10154
- Wang S, Huang S, Wang L (2019b) Spinal curve guide network (SCG-Net) for accurate automated spinal curvature estimation. In: International workshop and challenge on computational methods and clinical applications for spine imaging, pp 107–112
- Wasule V, Sonar P (2017) Classification of brain MRI using SVM and KNN classifier. In: 2017 third international conference on sensing, signal processing and security (ICSSS), pp 218–223
- Xiang Q, Wang X, Li R, Zhang G, Lai J, Hu Q (2019) Fruit image classification based on Mobilenetv2 with transfer learning technique. In: Proceedings of the 3rd international conference on computer science and application engineering, pp 1–7
- Yamashita R, Nishio M, Do RKG, Togashi K (2018) Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 9(4):611–629
- Zare MR, Seng WC, Mueen A (2013) Automatic classification of medical X-ray images. *Malays J Comput Sci* 26(1):9–22
- Zhang R, Zheng Y, Mak TWC, Yu R, Wong SH, Lau JYW, Poon CCY (2016) Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain. *IEEE J Biomed Health Inform* 21(1):41–47
- Zhao S, Wang B, Yang K, Li Y (2019) Automatic spine curvature estimation by a top-down approach. In: International workshop and challenge on computational methods and clinical applications for spine imaging, pp 75–80
- Zheng Y, Liu D, Georgescu B, Nguyen H, Comaniciu D (2015) 3D deep learning for efficient and robust landmark detection in volumetric data. In: International conference on medical image computing and computer-assisted intervention, pp 565–572

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Parisa Tavana¹  · Mahdi Akraminia²  · Abbas Koochari¹  · Abolfazl Bagherifard³ 

Parisa Tavana
pa.tavana@gmail.com; parisa.tavana@srbiau.ac.ir

Mahdi Akraminia
MAkraminia@nri.ac.ir; Mahdi.akraminia@gmail.com

Abolfazl Bagherifard
a.bagherifard45@gmail.com; bagherifd@gmail.com

¹ Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

² Mechanical Rotary Equipment Research Department, Niroo Research Institute, Tehran, Iran

³ Bone and Joint Reconstruction Research Center, Shafa Orthopedic Hospital, Iran University of Medical Sciences, Tehran, Iran