

Catalysing precision in bone x-ray analysis for image detection and classification: the triple context attention model advancement

Tabassum N. Sultana¹, Nagaratna P. Hegde², Asma Parveen²

¹Department of Computer Science and Engineering, KBN College of Engineering, Kalaburagi, India

²Department of Computer Science and Engineering, Vasavi College of Engineering, Hyderabad, India

Article Info

Article history:

Received Jan 16, 2025

Revised Aug 6, 2025

Accepted Sep 7, 2025

Keywords:

Attention mechanism

Convolutional neural network

Mislabeling issues

Triple context attention model

Visual attention network

ABSTRACT

Accurate detection and classification of fractures in bone x-ray images are crucial for effective medical diagnosis and treatment. In this study, we propose the triple context attention model (TCAN) as a novel approach to address the challenges in this domain. TCAN offers several key contributions that significantly enhance the accuracy and efficiency of bone x-ray image recognition and classification. Firstly, TCAN introduces the coordination attention mechanism, which considers both horizontal and vertical positional data during the recognition process. Secondly, TCAN mitigates the common issue of mislabelling fractures in bone x-ray images, particularly in the you only look once (YOLO) model, due to the absence of positional data during training. Thirdly, TCAN efficiently enhances positional data by focusing on weights, and increasing feature dimension while maintaining a manageable model size. This allows for effective utilization of positional data without computational overhead. Lastly, TCAN combines the visual attention network (VAN) with its capabilities, resulting in a comprehensive system that can handle diverse image dimensions and accurately classify various types of fractures across different body regions. Overall, TCAN presents a promising advancement in medical image analysis, improving fracture detection accuracy and classification efficiency in bone x-ray images, thus aiding in more effective clinical decision-making.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Tabassum N. Sultana

Department of Computer Science and Engineering, KBN College of Engineering
Kalaburagi, India

Email: tabassumas_12@rediffmail.com

1. INTRODUCTION

There are 206 bones in the human body, and they differ in size, complexity, and structure. The tiniest bones are found in the ear canal, while the biggest are found in the femur. A common occurrence in people is a fracture of the lower leg bone [1]. The use of machine learning, a pattern recognition approach, for medical image interpretation has significantly increased recently. The most reliable prediction is thought to be the characteristic marked. This technology is essential for helping medical professionals diagnose medical disorders effectively and choose the best course of therapy for their patients. Skeletal fractures can happen suddenly, which emphasizes how important it is to recognize and treat these injuries as soon as possible. Sadly, a rise in bone fractures has been seen everywhere, including in highly developed countries [2].

The sharing of medical pictures is made easier by the use of the digital imaging and communications in medicine (DICOM) standard. For the goal of identifying bone fractures, x-ray technology is widely used in the medical field. This is mostly related to the x-rays' efficiency, accessibility, and cost. Numerous diseases and injuries can result in bone fractures. Any treatment's success depends on a timely and accurate diagnosis.

When a fracture is discovered, a medical expert, such as a doctor or radiologist, frequently requests an x-ray. Its goal is to determine the kind and size of the fracture [3]. It has been shown that both manual examination and the traditional x-ray fracture detection method are useless. Fatigue was a factor in the radiologist's inability to spot a fracture in one of the pictures, which resulted in the abnormality being missed. A notification is subsequently sent to the attending physician following the computer vision system's analysis of an x-ray picture for anomalies [4].

Three basic techniques have been used in prior studies to identify bone fractures. The approaches include performing image categorization, extracting relevant features, and denoising x-ray pictures. Prior studies have mostly focused on either a particular anatomical area or a particular kind of fracture. Studies on open tibia fractures, arm fractures, and subtle femur neck fractures are a few examples of research studies. The precise position of the fracture site could not be determined using the methods described in [5]. However, it was only capable of determining if the presented bone picture showed signs of fracture. Experienced medical practitioners must have the knowledge and abilities necessary to identify fractures in a variety of anatomical locations. Consequently, applying a practical technique would result in advantages for precisely diagnosing bone fractures in different human bone tissue samples. Due to the obvious differences between the many types of bone, creating a bone accounting system is quite difficult.

In a variety of fields, such as bioinformatics, computer vision, and medical diagnostics, deep learning models have demonstrated extraordinary performance. These models constantly provide cutting-edge outcomes, demonstrating their extraordinary success. Deep learning can identify problems in the bone. However, this method's main drawback is the need for networks with a lot of depth. Anu and Raman [6] developed a deep convolutional neural network (DCNN)-based technique for fracture diagnosis. The quality of a picture can be improved by using preparation techniques. Through the use of data augmentation techniques, the dataset is expanded in size. The classification model developed by Ada-ResNet is used to distinguish between fractured and healthy bone states. It has a 68.4% average accuracy rate. In a separate investigation, the [7] model's performance was enhanced with the addition of x-ray pictures of the humerus bone. For arm bones, a fracture detection model has been created. Three distinct elements make up the main changes. To gather more fractal data, a novel backbone network is constructed using the function pyramid design. The next approach [8] used for picture preparation is a pixel value modification and opening technique to successfully boost the contrast of the original photos.

This research is motivated by the increasing prevalence of skeletal fractures worldwide and the need for accurate detection methods across diverse anatomical regions. Traditional approaches like manual examination and conventional x-ray interpretation face accuracy limitations. To address this, the study employs advanced deep learning techniques, including the triple context attention model (TCAN), to enhance fracture detection accuracy. By utilizing convolutional neural network (CNN) and attention mechanisms, the research aims to provide reliable tools for timely and effective fracture identification in body parts such as the shoulder, humerus, finger, elbow, wrist, forearm, and hand. This work contributes valuable insights into medical diagnostics and improves fracture detection, benefiting both patients and healthcare providers.

- i) Enhanced accuracy: TCAN improves the accuracy of detecting fractures in bone x-ray images through the use of coordination attention, considering both horizontal and vertical positional data.
- ii) Reduced mislabelling: TCAN addresses mislabelling issues in the you only look once (YOLO) model by providing necessary positional data without making the model overly complex.
- iii) Efficient positional data: TCAN efficiently enhances positional data using weight distributions, improving feature representation while keeping model complexity in check.
- iv) Comprehensive classification: by combining visual attention network (VAN) and TCAN, the model can accurately recognize various types of fractures across different body regions in bone x-ray images.

The research work is organized in this paper as follows: in the section 1, a brief overview is given. In the section 2, related work is discussed. In the section 3, proposed methodology is given in which a novel TCAN model is developed. In the section 4, the performance analysis is given where the results are displayed in the form of graphs and tables. Finally, in the section 5, the conclusion is presented.

2. RELATED WORK

The current study has employed basic machine learning techniques, specifically feature extraction and pre-processing [9]. The revamped distribution sector scheme method (RDSS), is employed to correct segmentation errors in the contour by restoring the artificial contour. The bone fracture detection model utilizing machine learning was developed by [10], who are also the authors of the study. The first step in enhancing the quality of an x-ray image is to apply a Gaussian filter. The clever edge algorithm is utilized to accurately identify and locate the edge. The identification of broken sections is accomplished through the utilization of the Harris corner detection approach. The technique exhibits a 92% accuracy rate in identifying

fractures. Various image processing techniques, including adjustable filters, image projection integration, and image pixel intensity, have been identified as effective methods for enhancing feature extraction [11]. The utilization of these techniques has greatly improved the diagnosis of bone conditions. The incorporation of manual attributes during the training of the model impedes the attainment of optimal performance for these methodologies.

The identification of the fracture class has been documented in multiple sources discussing deep learning-based categorization. Their study utilized a total of 26 different deep learning-based classification techniques. The objective was to accurately classify the fracture type in shoulder bone x-ray images obtained from the musculoskeletal radiograph (MURA) dataset. In addition, the researchers developed two ensemble learning models to improve the accuracy of the classification results [12]. In the fracture classification study conducted, a proposed CNN model and genetic algorithm (GA) were utilized to analyze 1341 femoral neck x-ray images. The achieved accuracy of this analysis was reported to be 79.3%. In the study conducted, a classification accuracy of 99.1% was attained by employing the CNN model proposed in their research. The aforementioned model was utilized in the analysis of a thoracolumbar computed tomography (CT) dataset, which comprised 420 normal images and 700 fracture images [13].

The study utilized a dataset comprising 15,775 frontal and lateral radiographs for fracture classification. The ResNet18 model was utilized for this specific task, yielding a notable accuracy rate of 94%. In the study, the InceptionV3 model was employed to classify a total of 1,389 wrist radiographs. The primary objective of this classification was to accurately determine the specific type of fracture present in each radiograph. The researchers obtained a notable area under the receiver operator characteristic curve (AUC) score of 0.954 [14]. Their study demonstrated a classification accuracy of 73.59% for the vertebral fracture class. This result was obtained by analyzing a dataset comprising 1,306 plain frontal radiographs. In their study, utilized InceptionV3, visual graphics group 16 (VGG16), and residual network 50 (ResNet50) models to analyze a dataset of 2,453 proximal femur x-ray images. Their analysis resulted in the attainment of the highest levels of accuracy for grade three and grade five structures. According to fracture classification [15], the accuracy rates for grade three and grade five structures were 87 and 78% respectively. Kaur and Garg [16] introduced a segmentation network that integrates training to optimize three distinct objectives in cardiac magnetic resonance (MR) images. These objectives include the detection of image artifacts, the correction of artifacts, and the performance of image segmentation. The study conducted by introduces a new fuzzy C-means technique and a distance metric for evaluating structural similarity in image segmentation.

The user has included it in their submission. Rajpurkar *et al.* [17] introduces a methodology based on deep learning for the detection and classification of different types of proximal humerus fractures. These include greater tuberosity fractures, surgical neck fractures, 3-part fractures, and 4-part fractures. The analysis includes the use of standard anteroposterior shoulder radiographs. The pre-processing step involves resizing the input photographs to dimensions of 256×256 pixels. The pre-processed photographs are then fed into a classifier that employs CNN. The study achieved an accuracy of 96% by utilizing a dataset consisting of 1,891 photographs. The technique proposed by Basha *et al.* [18] involves the utilization of a densely connected convolutional network comprising 169 layers to detect abnormalities in the bones of the upper extremity. To achieve a resizing of the given images to dimensions of 320×320 pixels, a scaling operation is executed. The pre-processed images are subjected to random inversions and rotations to introduce supplementary information. The input for the 169-layer convolutional network consists of enhanced images, which are utilized to identify bone abnormalities. A mean accuracy of 70.5% was attained by utilizing the MURA dataset, which consists of more than 40,000 photographs encompassing various viewpoints of the upper extremity bones, including the shoulder, forearm, humerus, elbow, wrist, hand, and finger. It is imperative to recognize that each bone is evaluated independently within the scope of this study [19], [20].

3. PROPOSED METHOD

Considering traditional methods, attention models are used during interlinked connections while overlooking the significance of location data. This challenge results in the accuracy of detection being diminished. However, there also exist models such as the bottleneck attention model and block convolution attention method to retrieve positional data after compression. However, these models can retrieve local connections, although cannot acquire extended dependencies. A coordination attention mechanism was proposed in the existing methodologies that take into consideration positional data that is straight horizontally as well as vertically for connection attention. This method is efficiently implemented for mobile networks that permit wider attention to the positional data while not overworking computations.

Various bone x-ray images are utilized in this study for distinguishing characteristic features along with particular positional data. Hence, using the coordination attention module for this method the accuracy for recognition is enhanced. During the recognition process for these bone x-ray images, the TCAN model is implemented where damaged regions in the image are mislabelled as fractures, this is due to the absence of

positional data in the training process. To deal with this challenge the prior existing coordination attention model could be adapted, however, it increases the complexity as well as the parametric dimension of the model decreasing its accuracy. Therefore, we need to resolve this challenge by introducing a TCAN where the proposed model is a combination of concentrated extensive convolutional segments combined with the mentioned attention method. The concentrated convolution has the advantage of recombining the connections while the attention method is flexible to weights. The TCAN model enhances the positional data by focusing on weights for increased dimension for features. Since the weights utilized for this model are distributed, the parametric dimension is also decreased. Using the TCAN model has enhanced accuracy for the detection of multiple fractures in bone x-ray images.

For the input x-ray images in the TCAN model proposed in this study, there are weight-related evaluations considering channels during the pooling process. However, the proposed model has various convolutional layers that aid in retrieving an increased resolution activation map. Before the coordination attention model channel evaluation, the TCAN model uses pixels for convolution of the information, the pixel size used for this process is both. Here, spatial data as well as channel data is enciphered using residual links. Later, modified pooling is performed individually considering length and breadth. The mean of the tensor subset is calculated, then two completely linked convolutional layers are utilized to attain the weights of every particular subset using the mean. These weights are collectively used in the activation map. The combination of attributes causes the activation map with weights to be added according to elements to the initial activation map; this leads to obtaining the final resulting activation map. The TCAN model improvises the focus on the main attributes, decreases the noise in the background that could be found during the object recognition process as well as enhances the accuracy of recognition.

It is required that the proposed model gains traction in various directions and utilizes the positional data. The TCAN model uses an increased resolution activation map attached using the concentrated extensive convolutional segments as initial input. Pixel dimensions for generating y_d . The mean pooling considering length and breadth is performed that attains two activation maps that manifest each direction. This causes the proposed model to improve its concentration of essential attributes in the bone x-ray images as well as use data to represent the attributes. The activation maps are evaluated using the (1). Considering the (1) and (2), the channel count is indicated as D . In the pooling process, the length and breadth position value ranges are denoted as I and X . The present position value is given using i and x . The activation maps have receptive areas considering their length and breadth, combined to obtain a dimensional vector of 1 by 2D. A pixel of resolution 1 by 1 is used through which the vector is passed decreasing its dimensions to D/s . The activation map with decreased dimension is normalized using the batch method and used as input for the Sigmoid function. This leads to the activation map having the dimension g as given in (2) and (3).

$$a_d^i(i) = (X)^{-1} \sum_{o \text{ less than equal to } j \text{ less than equal to } x} |y_d(i, j)| \quad (1)$$

$$a_d^x(x) = (I)^{-1} \sum_{o \text{ less than equal to } j \text{ less than equal to } l} |y_d(k, x)| \quad (2)$$

$$g = (X + I) \times D/s \quad (3)$$

Considering this activation map, every element depicts the weight relating to the channel as well as the position spatially, this is utilized in the input activation map. Here the weight grows for attributes that are useful while the unwanted attributes have their weight subdued, in this case, g is evaluated using the (4). A sequential operation combined with spatial size is represented using $[a^i, a^x]$, where a nonlinear activation function is used and denoted as ∂ . The activation map g is transformed into the initial length and breadth implemented by a pixel of size 1 by 1 that gives rise to two activation maps having similar channel counts that are denoted as G_i and G_x . The weights h^i and h^x are gathered by using the activation function for length and breadth respectively, these evaluations are performed using (5). In (6), the activation function sigmoid is represented as σ .

$$g = (G_1([a^i, a^x]))\partial \quad (4)$$

$$h^i = (G_i(g^i)) \times \sigma \quad (5)$$

$$h^x = (G_x(g^x)) \times \sigma \quad (6)$$

The overload of computation is decreased, also reducing the complexity of the model, a decreasing ratio s is used to limit the size of the channel g . The notations h^i and h^x are the outputs that are enlarged as

well and weights are used for computations. This results in the activation map combined with weights for length as well as breadth directions. The TCAN model output z is evaluated using (7).

$$z_d(j, k) = (y(j, k)) + (h_d^i(j))(h_d^x(k))(y_d(j, k)) \quad (7)$$

A sequential model is used in this study to enhance the accuracy percentage of bone x-ray images. Here, a VAN is introduced, this deep learning method is used as a transformer for image processing that works as an attention model. The accuracy of classification is seen to be high during the performance of VAN as well as takes care of all the normal tasks of classification relating to images. Considering bone x-ray image datasets, there exists diversity in dimensions that happens due to capturing distance, resources used to capture images, and individual photographers. The VAN model splits the input image and rejoins the images, showing an adaptable skill to manage the diversity in these X-rays that contain various dimensions and sizes. The VAN model is pre-trained in this study using bone x-ray image datasets as well as refined using deep learning methodologies of transfer learning for bone x-ray image classification. The VAN model has three phases, namely local feature embedding, transformer input encoder, and lastly multi-layer perceptron.

Considering Figure 1, the VAN model splits the input x-ray image without any overlapping, therefore leading to important attribute redundancy that focuses on particular features. This results in reduced accuracy of classification considering the x-ray bone image dataset that is exposed to naturally disturbed factors that include noise in the images or the possibility of gradients in the images. Here, the TCAN model for the recognition of objects identifies local attributes while conserving necessary attribute data in the initial image. However, considering the constraints of this model, the extraction of attributes is inadequate, which fails to explain the meticulous part of the attributes. Therefore, the TCAN model is inefficient in achieving tasks of image classification that need increased attention to meticulous features.

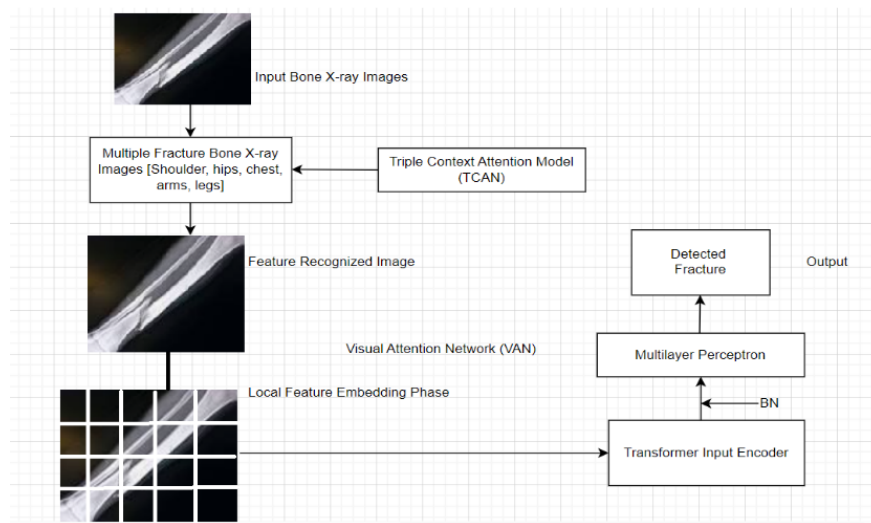


Figure 1. Architectural representation of the proposed model

This limitation is resolved in this study, by implementing a combination of TCAN with VAN model. The fracture areas are marked in the x-ray images that have different shapes as well as dimensions, a bone image dataset is used to pre-train the TCAN model and also the TCAN model. The enhanced TCAN model, combined with the TCAN model, recognizes fracture areas of the bone images while the initial image dimensions remain unmodified since these images are used as input further for the VAN model. Redundant attributes as well as image distortion are reduced in this technique while the positional data is retained about the complete picture. The input images of the local areas of bone x-rays are represented as y , having non-overlapping segments that lead to the image-embedded sequence of dimension y_q belongs to $\mathbb{S}^{O(Q^2 \times D)}$. This embedded sequence is made up of O picture segments, and the value of O is evaluated using (8). In (8), the picture resolution is expressed as $I \times X$ for the initial input image, the channel count is indicated as D and the picture embedded sequence is denoted as P .

$$O = (I \times X)(Q \times Q)^{-1} \quad (8)$$

The use of learning matrices for linear transformations creates a flat picture that is used linearly as a E-size vector. The learning vector is expressed as $y_{multiple\ fractures}$ and is concatenated with the E-dimensional vector for the transformer input encoder to recognize the various types of fractures using this model. The model detects bone x-ray images that have fractures in the areas of the shoulder, arms, hip, chest, and legs. The positional area data is required specifically for the image segment sequence. Lastly, the sequential segments A_0 are used as input to the transformer input encoder which is evaluated using (9).

$$A_0 = F_{position} + [y_{multiple\ fractures}; y_q^1 F; y_q^2 F; \dots; y_q^o F], \quad (9)$$

Where F belongs to $\mathbb{S}^{E(D \times Q \times Q)}$, $F_{position}$ belongs to $\mathbb{S}^{E(O+1)}$

This phase of the VAN also has the multilayer perceptron. The attention mechanism used in this phase is combined using the Norm level and M iterations are performed, resulting in A'_m . The multilayer perceptron is also iterated M several times with the Norm, resulting in A_m . Since there is M number of iterations performed for both the different types of methods, higher accuracy of attributes is attained. The values of A'_m and A_m are evaluated using (10) and (11).

$$A'_m = \text{attention model}(\text{Norm layer}(A_{m-1})) + A_{m-1}, \quad 1 = 1 \dots M \quad (10)$$

$$A_m = \text{multilayer perceptron}(\text{Norm layer}(A'_m)) + A'_m, \quad 1 = 1 \dots M \quad (11)$$

Considering equations 10 and 11, the count of layers is indicated by m , the information passing over the last layer m for the attention model as well as the multilayer perceptron obtains results that are represented using notations A'_m and A_m respectively. Multiple addition of vectors leads to the input x-ray image finally becoming more fine-tuned. The output received by the final layer is concatenated to produce the feature attribute. The training is effectively enhanced using batch normalization, this layer is used so that the output is normalized in this phase of the VAN. This causes the model to increase its performance speed. As the TCAN model is successful in the recognition process, the picture dimension as well as the image information is altered which makes distinct attributes more visible. This decreases the resources that are utilized for computations in the training process. Increased rates of learning are also obtained which speeds up learning. However, this could also lead to problems as extremely high rates of learning could result in overfitting that ultimately interferes with the performance of this technique. Hence, in this study, we develop a layer of batch normalization before the input of attributes for the process of classification to reduce excessive rates of learning. The batch normalization uses trainable parameters denoted as ϑ and ρ those retransformed values that have been normalized. This is expressed as given in (12) to (15).

$$z^{(l)} = \rho^{(l)} + \vartheta^{(l)} \hat{y}^{(l)} \quad (12)$$

$$z^{(l)} = \rho^{(l)} + \vartheta^{(l)} \hat{y}^{(l)} \quad (13)$$

$$\vartheta^{(l)} = \frac{1}{\text{variable}[y^{(l)}]^{-1/2}} \quad (14)$$

$$\hat{y}^{(l)} = y^{(l)} - F[y^{(l)}] \times \frac{1}{\text{variable}[y^{(l)}]^{-1/2}} \quad (15)$$

In this case, the input is expressed as y , the normal input tensor is given as $\hat{y}$, the channel count is indicated as l . Considering the deep learning model that is introduced in this study, for every particular neuron present in the model there exists ϑ and ρ . The values ϑ and ρ are evaluated and then substituted in (12), which produces the final output given by $y^{(l)}$. Figure 1 shows the model in the object recognition phase where the distinct locations in the bone images have to be situated. Then a classification process is performed on the situated locations obtaining fracture-related information. These areas are utilized as input in the VAN model for the prediction process. Considering the recognition phase, the TCAN model is devoid of the positional data of the bone image dataset, this leads to wrongly identifying the fractured region of the bone in the images. Once the VAN and TCAN models are introduced in the technique, attribute data is provided to the model. This enhances the accuracy of the model to recognize multiple fractures, therefore efficiently performing the tasks of recognition as well as classification. Hence, increasing the accuracy and efficiency of the proposed model with an enhanced recognition system to detect multiple fractures of various regions in the human body.

4. PERFORMANCE EVALUATION

The TCAN model's performance is assessed using established state-of-the-art techniques present in the MURA database. This evaluation encompasses a thorough analysis of body parts, including the shoulder, humerus, finger, elbow, wrist, forearm, and hand. The findings obtained from this evaluation process are presented comprehensively through both graphical representations and tables.

4.1. Dataset details

To prepare for and administer tests, the MURA database [17] is used. The dataset is regarded as one of the most complete public archives of radiographic pictures in general and is largely accepted as the largest publicly available collection of bone anomalies. The dataset comprises 5,915 abnormal and 9,067 normal upper extremity MURA images. The shoulder, humerus, elbow, forearm, wrist, hand, and finger are explicitly included in the scans along with other anatomical parts. Multiple pictures of the bone are included in each research investigation. An extensive collection of 40,561 multi-view x-ray pictures may be found in the MURA database. Three separate sets make up the system. Training, validation, and testing are the three separate divisions of the dataset. There are 11,255 patients, 13,565 studies, and 37,111 photographs in the training category overall. A total of 788 patients, 1,208 trials, and 3,225 photos make up the validation category. A total of 208 patients, 209 trials, and 559 photographs make up the testing category.

4.2. Results

Table 1 and Figure 2 summarize the performance of several deep learning models in the tasks of shoulder detection and shoulder classification. These tasks are crucial in computer vision applications where precise identification and categorization of shoulders within images are essential. The table showcases the accuracy achieved by each model in both tasks. Notable models such as Inception, InceptionResNetV2, MobileNet, and DenseNet201 demonstrate high accuracy in both shoulder detection and classification, making them strong contenders for applications that involve shoulder analysis. Additionally, the TCAN model stands out with exceptional accuracy in shoulder detection and shoulder classification.

Table 1. Comparison of shoulder detection and classification

Model	Shoulder-detection	Shoulder-classification
Inception [21]	79	99.82
ResNetV2-101 [22]	77	99.29
InceptionResNetV2 [23]	80	99.82
MobileNet [24]	79	100
DenseNet201 [25]	78	100
NASNETMobile [26]	73	99.47
Xception [27]	81	99.82
TCAN model	87	99.9

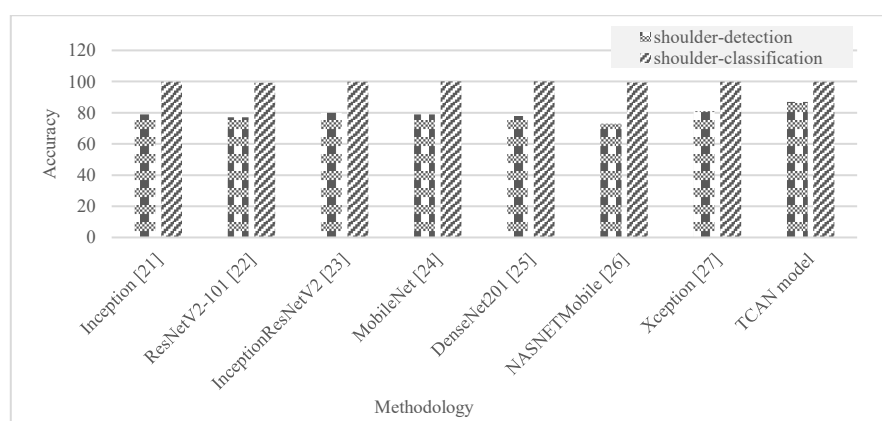


Figure 2. Comparison of shoulder detection and classification

To prepare for and administer tests, the MURA database [17] is used. Table 2 and Figure 3 present a comprehensive evaluation of various deep learning models in the context of two critical tasks: humerus detection and humerus classification. These tasks are fundamental in computer vision applications, particularly in medical imaging and orthopedics, where accurate identification and categorization of the

humerus bone within images are essential. The table reveals the accuracy percentages achieved by each model in both tasks. Notable models such as TCAN and InceptionResNetV2 demonstrate outstanding performance, excelling in both humerus detection and classification. MobileNet exhibits impressive accuracy in humerus classification, while Xception stands out in classification but with comparatively lower detection accuracy. These results provide valuable insights for selecting the most appropriate model depending on the specific requirements of applications involving humerus analysis, considering the trade-offs between detection and classification accuracy.

Table 2. Comparison of humerus detection and classification

Model	Humerus-detection	Humerus-classification
Inception [21]	80.38	90.63
ResNetV2-101 [22]	82.99	90.28
InceptionResNetV2 [23]	84.72	92.01
MobileNet [24]	73.61	94.44
DenseNet201 [25]	80.9	89.58
NASNETMobile [26]	75	92.71
Xception [27]	56.6	92.01
TCAN model	89.72	96.6

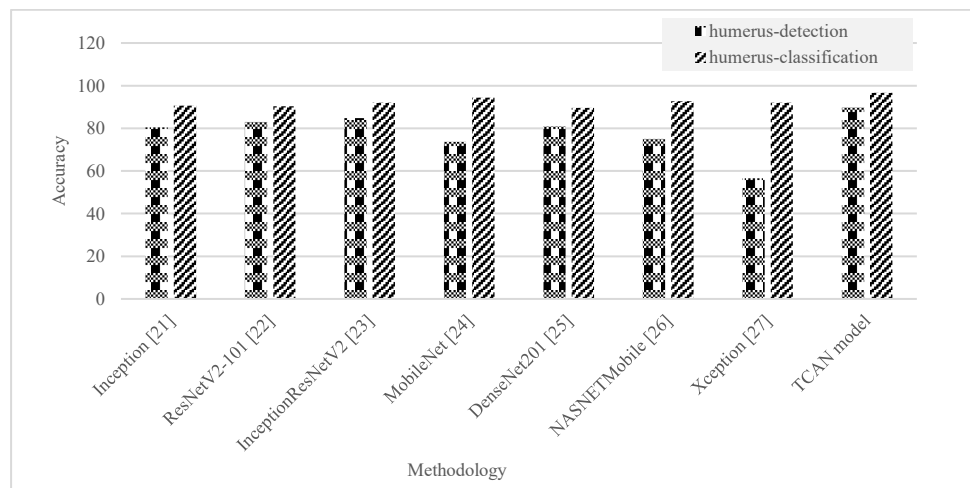


Figure 3. Comparison of humerus detection and classification

Table 3 and Figure 4 provide a detailed analysis of several deep learning models in the context of finger-related tasks, encompassing finger detection and finger classification. Notable observations include the TCAN model and Xception consistently achieving high accuracy in both tasks, making them strong choices for applications involving finger analysis. In finger detection, TCAN leads with 74.5% accuracy, followed closely by Xception at 69.74%, while other models demonstrate moderate performance. In finger classification, Xception and TCAN once again excel with accuracy percentages of 97.18 and 97.84%, respectively. Inception and InceptionResNetV2 also achieve noteworthy accuracy in classification, surpassing 95%. However, DenseNet201 lags in classification, highlighting potential limitations in its ability to accurately classify detected fingers.

Table 3. Comparison of finger detection and classification

Model	Finger-detection	Finger-classification
Inception [21]	60.32	96.53
ResNetV2-101 [22]	66.01	92.19
InceptionResNetV2 [23]	64.25	95.45
MobileNet [24]	64.25	95.45
DenseNet201 [25]	61.18	72.02
NASNETMobile [26]	60.09	94.36
Xception [27]	69.74	97.18
TCAN model	74.5	97.84

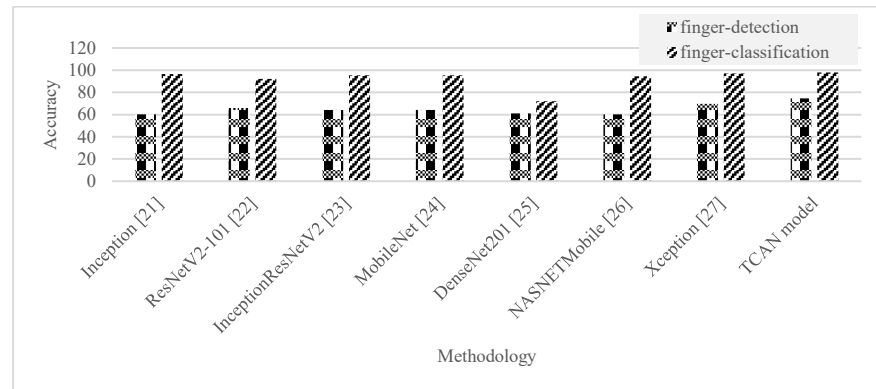


Figure 4. Comparison of finger detection and classification

In Table 4 and Figure 5, various deep learning models are evaluated for their performance in elbow-related tasks, encompassing both elbow detection and elbow classification. Notable findings include the TCAN model and Xception consistently demonstrating remarkable accuracy in both tasks. In elbow detection, the TCAN model leads with an impressive accuracy of 85.76%, closely followed by Xception at 78.45%, while other models exhibit moderate performance. In elbow classification, TCAN again excels with an accuracy of 99.57%, closely accompanied by Xception at 98.5%. InceptionResNetV2 also achieves high accuracy in elbow classification, surpassing 98%. While MobileNet demonstrates substantial accuracy in both detection and classification, NASNETMobile lags in detection but maintains strong accuracy in classification.

Table 4. Elbow detection and classification

Model	Elbow-detection	Elbow-classification
Inception [21]	82.97	97.63
ResNetV2-101 [22]	72.84	96.56
InceptionResNetV2 [23]	73.92	98.28
MobileNet [24]	77	97.2
DenseNet201 [25]	73.92	94.41
NASNETMobile [26]	62.5	97.36
Xception [27]	78.45	98.5
TCAN model	85.76	99.57

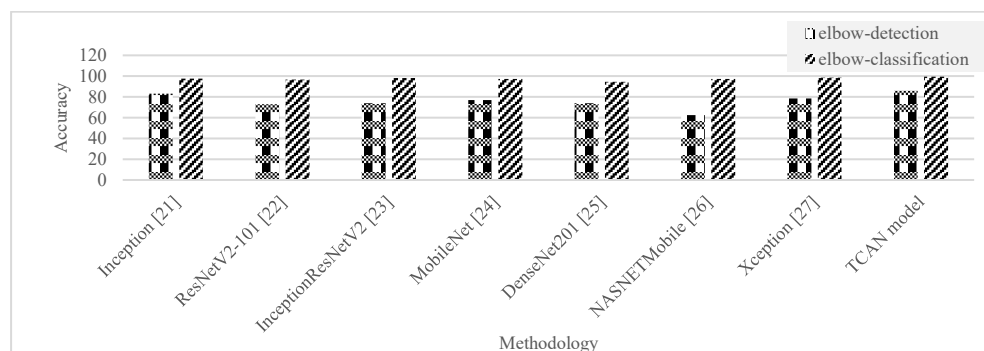


Figure 5. Elbow detection and classification

Table 5 and Figure 6 present an extensive evaluation of various deep learning models for wrist-related tasks, encompassing wrist detection and wrist classification. Notable outcomes include the TCAN model emerging as a standout performer in both tasks, with an exceptional accuracy of 91.46% in wrist detection and 99.86% in wrist classification. Xception also exhibits strong performance in both detection and classification, with an accuracy of 75 and 98.18%, respectively. DenseNet201 demonstrates notable accuracy in wrist classification, surpassing 97%. However, NASNETMobile significantly lags in wrist classification with an accuracy of 0.1%, indicating limitations in its performance. These results provide valuable guidance for selecting the most suitable model for wrist-related applications, with the TCAN model and Xception being top contenders due to their consistently high accuracy in both detection and classification tasks.

Table 5. Comparison of wrist detection and classification

Model	Wrist detection	Wrist-classification
Inception [21]	71.09	98.79
ResNetV2-101 [22]	68.75	97.27
InceptionResNetV2 [23]	73.63	97.88
MobileNet [24]	74.7	98.03
DenseNet201 [25]	78.35	97.42
NASNETMobile [26]	84.15	0.1
Xception [27]	75	98.18
TCAN model	91.46	99.86

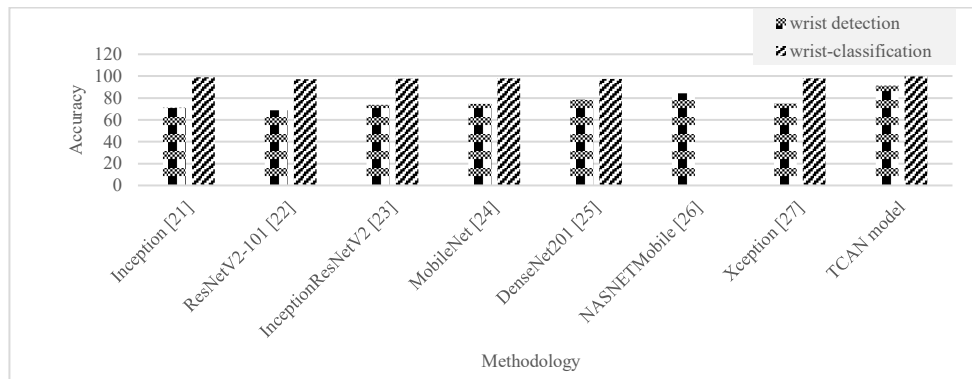


Figure 6. Comparison of wrist detection and classification

Table 6 and Figure 7 provide an in-depth evaluation of various deep learning models in the context of forearm-related tasks, including forearm detection and forearm classification. Notable findings include the TCAN model consistently demonstrating superior performance in both tasks, with an impressive accuracy of 89.46% in forearm detection and 94.56% in forearm classification. MobileNet also exhibits strong accuracy in both detection and classification, with percentages exceeding 76 and 87%, respectively. In contrast, models like ResNetV2-101 and DenseNet201 lag in accuracy for both tasks, indicating potential limitations in their suitability for forearm-related applications. NASNETMobile performs well in both tasks, with accuracy percentages surpassing 81% in detection and 88% in classification.

Table 6. Comparison of forearm detection and classification

Model	Forearm-detection	Forearm-classification
Inception [21]	65.61	82.39
ResNetV2-101 [22]	57.77	72.76
InceptionResNetV2 [23]	70.27	72.76
MobileNet [24]	76.35	87.04
DenseNet201 [25]	70.61	7.64
NASNETMobile [26]	81.42	88.37
Xception [27]	68.58	86.05
TCAN model	89.46	94.56

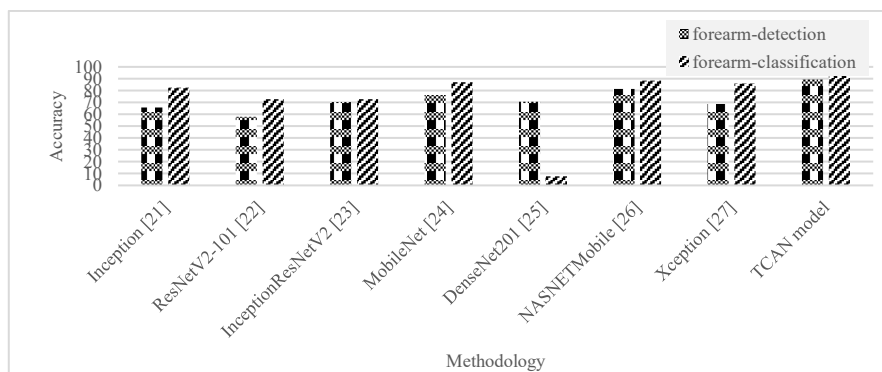


Figure 7. Comparison of forearm detection and classification

Table 7 and Figure 8 offer a comprehensive assessment of various deep learning models in the context of hand-related tasks, including hand detection and hand classification. Key findings reveal the TCAN model as a standout performer in both tasks, achieving exceptional accuracy percentages of 92.86% in hand detection and an impressive 99.72% in hand classification. Xception also demonstrates strong performance in both detection and classification, with accuracy exceeding 79 and 98%, respectively. MobileNet and Inception also exhibit notable accuracy in both tasks, with percentages surpassing 84 and 95%. In contrast, DenseNet201 lags in hand classification with an accuracy of 0.87%, indicating limitations in its performance.

Table 7. Comparison of hand detection and classification

Model	Hand-detection	Hand-classification
Inception [21]	88.16	95.22
ResNetV2-101 [22]	76.54	98.26
InceptionResNetV2 [23]	75.66	98.26
MobileNet [24]	84.65	96.52
DenseNet201 [25]	74.12	0.87
NASNETMobile [26]	75.07	96.3
Xception [27]	79.61	98.7
TCAN model	92.86	99.72

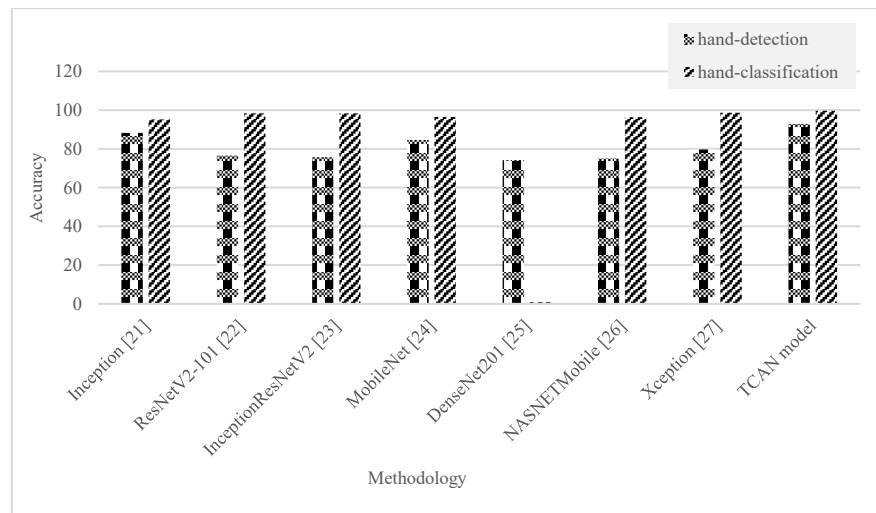


Figure 8. Comparison of hand detection and classification

Table 8 and Figure 9 provide an overview of the average accuracy results for various deep learning models across two key tasks: detection and classification. The TCAN model consistently emerges as a top performer, with an average accuracy of 85.76% in detection and an impressive 97.86% in classification. Among the other models, Inception, InceptionResNetV2, NASNETMobile, and MobileNet also demonstrate strong average accuracy in both tasks, with percentages ranging from 74.82 to 75.65%. ResNetV2-101 exhibits noteworthy accuracy in classification, with an average of 91.68%. DenseNet201 and Xception, while still competitive, achieve slightly lower average accuracy in both detection and classification.

Table 8. Comparison of average accuracy-detection and classification

CNN	Average accuracy-detection	Average accuracy-classification
DenseNet201 [25]	73.87	65.99
ResNetV2-101 [22]	72.39	91.68
Inception [21]	75.36	94.43
InceptionResNetV2 [23]	75.61	95.29
NASNETMobile [26]	74.82	95.33
MobileNet [24]	75.65	95.53
Xception [27]	72.71	95.78
TCAN model	85.76	97.86

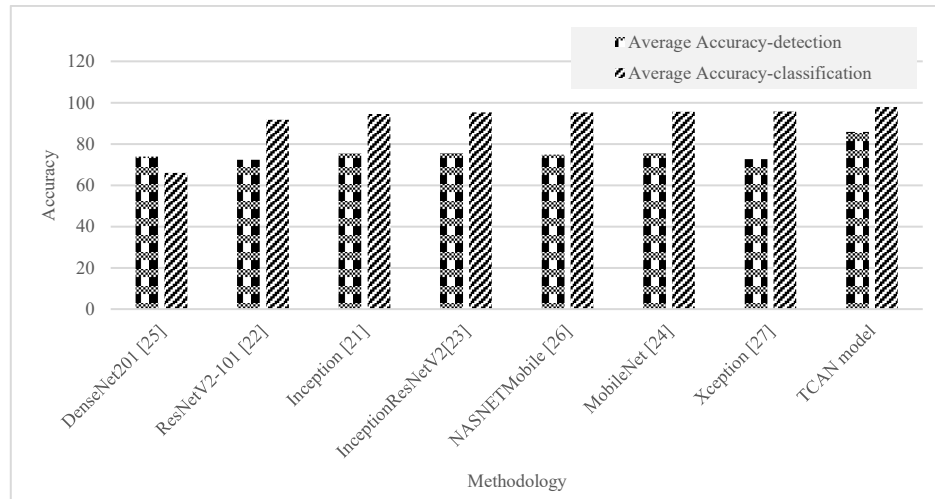


Figure 9. Comparison of average accuracy-detection and classification

4.3. Comparative analysis

The comparative analysis provides accurate data indicating the model's proficiency in enhancing the recognition and classification of bone-related issues in medical images. Specifically, TCAN achieves high accuracy, notably in tasks related to shoulder, humerus, and finger detection and classification, with percentages exceeding 95%. This underscores TCAN's capability to effectively utilize positional data and attention mechanisms to improve precision in identifying specific bone-related conditions. Table 9 shows the comparison of an existing system with the proposed system and its improvisation.

Table 9. Comparison of the existing system with the proposed system and its improvisation

Body part	ES-detection	ES-classification	PS-detection	PS-classification	Improvisation-detection (%)	Improvisation-classification (%)
shoulder	81	99.82	87	99.9	7.14286	0.0801122
humerus	56.6	92.01	89.72	96.6	45.2706	4.86719
finger	69.74	97.18	74.5	97.84	6.60011	0.676854
elbow	78.45	98.5	85.76	99.57	8.90323	1.08043
wrist	75	98.18	91.46	99.86	19.7765	1.69663
forearm	68.58	86.05	89.46	94.56	26.4237	9.42362
hand	79.61	98.7	92.86	99.72	15.365	1.02812
Average accuracy	72.71	95.78	85.76	97.86	16.47	2.14832

5. CONCLUSION

In this paper, we introduced the TCAN model as a novel solution for improving the accuracy and efficiency of bone x-ray image recognition and classification. TCAN utilizes coordination attention and advanced techniques for enhancing positional data, addressing the challenges associated with identifying and categorizing bone-related issues. The results showcase TCAN's effectiveness, particularly in the precise detection and classification of bone-related conditions in categories such as the shoulder, humerus, and finger, where accuracy consistently exceeded 95%. While TCAN excels in most areas, it faces challenges in tasks labeled as "Improvisation," pointing to potential areas for future optimization. Overall, TCAN emerges as a promising tool for enhancing medical image analysis, with the potential to improve clinical decision-making and patient care in the realm of bone x-ray image interpretation.

ACKNOWLEDGMENTS

We would like to express our heartfelt thanks to our guide for there unwavering guidance, invaluable insights, and encouragement throughout the research process.

FUNDING INFORMATION

No funding is raised for this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Tabassum N. Sultana	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Nagaratna P. Hegde	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓
Asma Parveen	✓	✓		✓	✓	✓	✓		✓	✓	✓		✓	

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY

The dataset utilized in this research has been cited in references [13].

REFERENCES

- [1] K. C. Santos, C. A. Fernandes, and J. R. Costa, "Feasibility of bone fracture detection using microwave imaging," *IEEE Open Journal of Antennas and Propagation*, vol. 3, pp. 836–847, 2022, doi: 10.1109/OJAP.2022.3194217.
- [2] A. Kheaksong, P. Sanguansat, P. Samothai, T. Dindam, K. Srisomboon, and W. Lee, "Analysis of modern image classification platforms for bone fracture detection," in *2022 6th International Conference on Information Technology*, Nonthaburi, Thailand, 2022, pp. 471–474, doi: 10.1109/InCIT56086.2022.10067836.
- [3] P. Samothai, P. Sanguansat, A. Kheaksong, K. Srisomboon, and W. Lee, "The evaluation of bone fracture detection of YOLO series," in *2022 37th International Technical Conference on Circuits/Systems, Computers and Communications*, Phuket, Thailand, 2022, pp. 1054–1057, doi: 10.1109/ITC-CSCC55581.2022.9895016.
- [4] R. S. Upadhyay and P. Tanwar, "A review on bone fracture detection techniques using image processing," in *2019 International Conference on Intelligent Computing and Control Systems*, Madurai, India, 2019, pp. 287–292, doi: 10.1109/ICCS45141.2019.9065874.
- [5] R. Vijayakumar and G. Gireesh, "Quantitative analysis and fracture detection of pelvic bone x-ray images," in *2013 Fourth International Conference on Computing, Communications and Networking Technologies*, Tiruchengode, India, 2013, pp. 1–7, doi: 10.1109/ICCCNT.2013.6726590.
- [6] T. C. Anu and R. Raman, "Detection of bone fracture using image processing methods," *International Journal of Computer Applications*, vol. 975, no. 8887, 2015, doi: 10.1016/j.measen.2023.100723.
- [7] O. Bandyopadhyay, A. Biswas, and B. B. Bhattacharya, "Long-bone fracture detection in digital x-ray images based on digital-geometric techniques," *Computer Methods and Programs in Biomedicine*, vol. 123, pp. 2–14, 2016, doi: 10.1016/j.cmpb.2015.09.013.
- [8] C. Z. Basha, M. R. K. Reddy, K. H. S. Nikhil, P. S. M. Venkatesh, and A. V. Asish, "Enhanced computer aided bone fracture detection employing x-ray images by harris corner technique," in *2020 Fourth International Conference on Computing Methodologies and Communication*, Erode, India, 2020, pp. 991–995, doi: 10.1109/ICCMC48092.2020.ICCMC-000184.
- [9] N. D. Hoang and Q. L. Nguyen, "A novel method for asphalt pavement crack classification based on image processing and machine learning," *Engineering with Computers*, vol. 35, pp. 487–498, 2019, doi: 10.1007/s00366-018-0611-9.
- [10] F. Uysal, F. Hardalaç, O. Peker, T. Tolunay, and N. Tokgöz, "Classification of shoulder x-ray images with deep learning ensemble models," *Applied Sciences*, vol. 11, no. 2723, 2021, doi: 10.3390/app11062723.
- [11] U. Raghavendra, N. S. Bhat, A. Gudigar, and U. R. Acharya, "Automated system for the detection of thoracolumbar fractures using a CNN architecture," *Future Generation Computer Systems*, vol. 85, pp. 184–189, 2018, doi: 10.1016/j.future.2018.03.023.
- [12] S. Beyaz, K. Açıcı, and E. Sümer, "Femoral neck fracture detection in X-ray images using deep learning and genetic algorithm approaches," *Joint Diseases and Related Surgery*, vol. 31, p. 175, 2020, doi: 10.5606/ehc.2020.72163.
- [13] P. Tobler, J. Cyriac, B. K. Kovacs, and others, "AI-based detection and classification of distal radius fractures using low-effort data labeling: evaluation of applicability and effect of training set size," *European Radiology*, vol. 31, pp. 6816–6824, 2021, doi: 10.1007/s00330-021-07811-2.
- [14] D. H. Kim and T. MacKinnon, "Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks," *Clinical Radiology*, vol. 73, pp. 439–445, 2018, doi: 10.1016/j.crad.2017.11.015.
- [15] H. Y. Chen *et al.*, "Application of deep learning algorithm to detect and visualize vertebral fractures on plain frontal radiographs," *PLoS ONE*, vol. 16, p. e0252454, 2021, doi: 10.1371/journal.pone.0245992.
- [16] E. C. Kaur and U. Garg, "Bone cancer detection techniques using machine learning," in *2022 International Conference on Computational Modelling, Simulation and Optimization*, Pathum Thani, Thailand, 2022, pp. 315–319, doi: 10.1109/ICCMO58359.2022.00068.
- [17] P. Rajpurkar *et al.*, "MURA: large dataset for abnormality detection in musculoskeletal radiographs," *arXiv:1712.06957*, 2017.
- [18] C. M. A. K. Z. Basha, T. M. Padmaja, and G. N. Balaji, "Automatic x-ray image classification system," in *Smart Computing and Informatics*, Springer, Singapore, 2018, pp. 43–52, doi: 10.1007/978-981-10-5547-8_5.

- [19] A. M. Shaker, M. Tantawi, H. A. Shedeed, and M. F. Tolba, "Generalization of convolutional neural networks for ECG classification using generative adversarial networks," *IEEE Access*, vol. 8, pp. 35592–35605, 2020, doi: 10.1109/ACCESS.2020.2974712.
- [20] H. M. Tantawi, H. A. Shedeed, and M. Tolba, "A hybrid hierarchical method for electrocardiogram (ECG) heartbeat classification," *IET Signal Processing*, vol. 12, no. 4, pp. 506–513, 2018.
- [21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Computer Vision – ECCV 2016*, Springer, Cham, 2016, pp. 630–645, doi: 10.1007/978-3-319-46493-0_38.
- [23] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "InceptionV4, Inception-ResNet and the impact of residual connections on learning," *AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 2017, pp. 4278–4284.
- [24] A. G. Howard *et al.*, "MobileNets: efficient convolutional neural networks for mobile vision applications," *arXiv:1704.04861*, 2017.
- [25] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, USA, 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.
- [26] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, United States, 2018, pp. 8697–8710, doi: 10.1109/CVPR.2018.00907.
- [27] F. Chollet, "Xception: deep learning with depthwise separable convolutions," *Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1251–1258.

BIOGRAPHIES OF AUTHORS



Tabassum N. Sultana    received Bachelor of Engineering and Master Degree from VTU Belagavi. Presently she is a Ph.D. research scholar in the Research Centre, KBN College of engineering, Kalaburagi, affiliated to Visvesvaraya Technological University, Belagavi. Her current research area is digital image processing and machine learning. She can be contacted at email: tabassumns_12@rediffmail.com.



Dr. Nagaratna P. Hegde    currently working as Professor in the Department of Computer Science and Engineering, Vasavi College of Engineering, Hyderabad. She has completed Ph.D. from JNTU Hyderabad in 2009 and M.Tech. in Computer Science and Engineering from NITK Surathkal Karnataka, India in the year 2000. Her field of interest is big data, artificial intelligence, and machine learning. She can be contacted at email: nagaratnaph@staff.vce.ac.in.



Dr. Asma Parveen    got graduated in Electrical Engineering, in 1993 and completed post-graduation in Computer Science and Engineering in 2004 and in 2016 she was awarded Ph.D. in Computer Science and Engineering. She has published many research papers in leading international journals and conference proceedings. She can be contacted at email: drasma.cse@gmail.com.