

Classifying Humerus Fracture Using X-Ray Images



Tahmina Akter Sumi , Nanziba Basnin , Md. Shahadat Hossain , Karl Andersson , and Md. Sazzad Hoassain

Abstract Bone is the most important part of our body which holds the whole structure of human body. The long bone situated in the upper arm of human body between the shoulder and elbow junction is known as “Humerus”. Humerus works as a structural support of the muscles and arms in the upper body which helps in the movement of the hand and elbow. Therefore, any fracture in humerus disrupts our daily lives. The manual fracture detection process where the doctors detect the fracture by analyzing X-ray images is quite time consuming and also error prone. Therefore, we have introduced an automated system to diagnose humerus fracture in an efficient way. In this study, we have focused on deep learning algorithm for fracture detection. In this purpose at first, 1266 X-ray images of humerus bone including fractured and non-fractured have been collected from a publicly available dataset called “MURA”. As a deep learning model has been used here, data augmentation has been applied to increase the dataset for reducing over-fitting problem. Finally, all the images are passed through CNN model to train the images and classify the fractured and non-fractured bone. Moreover, different pretrained model has also been applied in our dataset to find out the best accuracy. After implementation, it is observed that our model shows the best accuracy which is 80% training accuracy and 78% testing accuracy comparing with other models.

Keywords Humerus fracture · Deep learning algorithm · Data augmentation · Convolution neural network

T. A. Sumi · Md. S. Hossain (✉)
University of Chittagong, Chittagong, Bangladesh
e-mail: hossain_ms@cu.ac.bd

N. Basnin
International Islamic University Chittagong, Sonaichhari, Bangladesh

K. Andersson
Lulea University of Technology, Skelleftea, Sweden

Md. S. Hoassain
University of Liberal Arts Bangladesh, Dhaka 1209, Bangladesh

1 Introduction

A humerus fracture is a break in the upper arm's humerus bone. Plain radiography is used to identify humerus fractures [22]. To determine the type of fracture, its anatomical location followed by fragmentation and displacement levels is taken into account. Albeit experienced surgeons may misdiagnose because humerus fracture has various representations [5]. As a result, an effective, as well as accurate method, is required to classify the fracture. Deep learning is a subset of artificial intelligence that extracts features and creates transformations using a cascade of many layers of nonlinear processing units. It is based on the learning of several levels of features or representations of the data. With the success of employing a deep learning model to identify and categorize images, there has been interest in using deep learning in medical image analysis in a variety of domains, including skin cancer diagnosis, diabetic retinopathy, mammographic lesions, and lung nodules identification. Trials in orthopedic surgery and traumatology, on the other hand, are extremely rare, despite their relevance to public health. Therefore, for this research, a CNN model is employed to classify humerus fractures as it proved to perform better in case of image classification [4, 9, 17, 23]. A dataset is collected and preprocessed. Data augmentation is applied to increase the robustness of the dataset. A CNN model is applied to train and test the dataset.

The following sections of this paper are organized as follows: Sect. 2 describes the previous works based on ALL detection. Section 3 depicts the entire methodology where whole CNN model used in this paper has been illustrated broadly. Section 4 is about the the result of our research, and finally, Sect. 5 is about conclusion and future work.

2 Literature Review

This section basically covers the previous research works related to the detection of bone fracture. Sasidhar et al. [20] in 2021 applied 3 pretrained model (VGG16, Densenet121 and Densenet169) on humerus bone images collected from the MURA dataset for detecting the fractured bones . They showed their best accuracy 80% using DenseNet161. Demir et al. [7] introduced an exemplar pyramid feature extraction method for classification of humerus fracture. They employed HOG and LBP for feature generation. They achieved a outstanding accuracy which is 99.12%. However, in this research, only 115 images images have been used which is very poor. Chung et al. [5] in 2018 developed a deep convolution neural network for classifying the fracture types where they gained a promising result with 65–86% top-1 accuracy. Negrillo-Cárdenas et al. [14] in 2020 proposed a geometrically-based algorithm for detecting landmarks in the humerus for the purpose of reducing supracondylar fractures of the humerus. In this research, they calculated the distance between two corresponding landmarks which is 1.45 mm ($p < 0.01$). Sezer and Sezer [21] shoulder images had been evaluated by CNN for the feature extraction and the classification of the head of

humerus into three categories, for example, normal, edematous and Hill-Sachs lesion with 98.43% accuracy. Negrillo-Cárdenas et al. [15] in 2019 considered a geometrical approach and spatial for detecting the landmarks on distal humerus. In this paper, they calculated 6 points for each bone. All these research took humerus as parameter.

Deviating from humerus fracture, some researchers also used other bones as parameter, for example, shoulder, femur, calcaneus, etc. De Vries et al. [6] in 2021 worked on predicting the chance of osteoporotic fracture (MOF) where they developed three machine learning models (Cox regression, RSF, and ANN) to show the comparison. Here, they showed that Cox regression outperformed the other models with 0.697 concordance-index. Pranata et al. [16] applied ResNet and VGG for the detection of calcaneus fractures using CT images where they gained a good accuracy 98%. Adams et al. [1] worked on the detection of neck of femur fractures where they implemented deep convolutional neural networks. In this work, the researchers showed an accuracy of 90.5% and AUC of 98%. Devana et al. [8] utilized machine learning model for the prediction of complications following reverse total shoulder arthroplasty (rTSA). Here, they applied five classifiers where XGBoost gained the top AUROC and AUPRC of 0.681 and 0.129.

After analyzing all the previous research work, it can be concluded that, in [7, 20], research on humerus fracture detection has been done, but further improvement is needed here. In [5, 21], humerus fracture classification has been done where they classified the fracture types rather detection. Negrillo-Cárdenas et al. [14, 15] studies were about the detection of landmark on humerus. So, it is clear that there has been very little research on humerus fracture detection. So in this study, we have tried to make a contribution on the detection of humerus fracture using deep learning model (Table 1).

3 Methodology

This section is about the whole network architecture that has been utilized in this study to detect humerus fracture. Figure 1 illustrates the system flowchart of this study.

As we can see in Fig. 1, this study starts with collecting dataset. After the data collection, all the images are sent to the data augmentation section. After that, pre-processing step is applied to the augmented data where the images are resized and then converted to an array. After the completion of pre-processing step, the processed images are sent to the CNN model where the features are extracted, and classification step finally predicts the images as “positive” or “negative”.

3.1 Dataset

The dataset we have used in this research for humerus fracture detection is collected from a public source which is named “musculoskeletal radiographs (MURA)”

Table 1 Related works

Reference	Year	Parameters	Methodology	Performance
[20]	2021	Humerus bone fracture	VGG16, Densenet121, Densenet169	Accuracy: 0.80
[7]	2020	Humerus fractures	HOG, LBP	Accuracy: 99.12
[5]	2018	Proximal humerus fractures	DCNN	Accuracy: 0.65–0.86
[14]	2020	Landmarks in the humerus	Geometrically-based algorithm	Distance between two corresponding landmarks 1.45 mm ($p > 0.01$)
[21]	2020	Diagnosis of proximal humerus	CNN	98.43
[15]	2019	Distal humerus	Geometrical approach	Calculated 6 points for each bone
[6]	2021	Hip, wrist, spine, and humerus fractures	Cox regression, RSF and ANN	Concordance-index 0.697 for Cox regression
[16]	2019	Calcaneus	ResNet and VGG	Accuracy: 0.98
[1]	2019	Neck of femur	Deep convolutional neural networks	0.91 (accuracy) 0.98 (AUC)
[8]	2021	Shoulder arthroplasty	LR, XGBoost, RF	AUROC:681, AUPRC: 0.129 (for XGBoost)

dataset. The MURA dataset is a large dataset containing 40,561 images collected from 12,173 patients which is labeled by board-certified radiologists of the Stanford Hospital whether it is positive or negative. To inquire the different types of abnormalities present in the dataset, the report has been reviewed manually hundred abnormalities where fifty three studies have been marked as fractures, 48 has been marked with hardware, 35 has been marked with degenerative joint diseases, and 29 has been marked with other mixed abnormalities.

Anyway, this dataset contains 7 types of bone fracture images, namely elbow, finger, forearm, hand, humerus, shoulder, and wrist. As we are working with the detection of humerus fracture, so only the X-ray images of humerus fracture have been considered in this study. Total of 1266 (669 negatives and 597 positives) X-ray images of humerus have been used in MURA dataset which are collected from 727 patients.

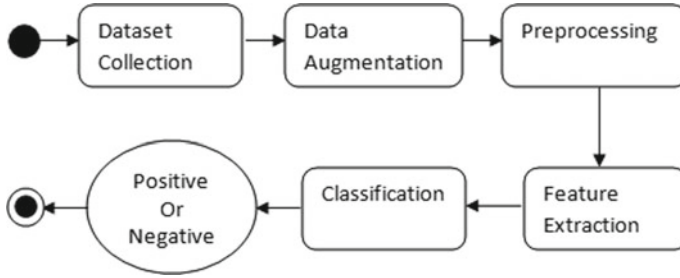


Fig. 1 System flowchart

3.2 Data Augmentation

Enhancing the performance of a model is an all time challenging task. Data augmentation helps to a great extent in this case. Generally by the data augmentation, the dataset is increased by creating altered copies of data which are already exist. It is about fabricating more data from the existing dataset without losing the data information. Data augmentation lessens the risk of over-fitting problem by increasing dataset. However, data augmentation can boost the model accuracy when the dataset is insufficient.

As stated earlier, in this study, we have used total of 1266 X-ray images of humerus including positive case and negative case which is relatively smaller for deep learning method. That is why we have implemented the data augmentation process on our dataset before training the model. For this augmentation, ImageDataGenerator has been utilized here. The settings for image augmentation that we utilized have been illustrated in Table 2.

Table 2 Data augmentation settings

Augmentation techniques	Range
Rotation	25
Width shift	0.10
Height shift	0.10
Zoom	0.40
Horizontal flip	True
Fill mode	Nearest

3.3 Pre-processing

Image pre-processing creates a positive effects on image by increasing the quality of feature extraction. Image pre-processing is needed to clean the image dataset prior to train the model and inference. It generally works by removing background noise, reflection of light, normalizing the images, and preparing the images for better feature extraction.

In this study, the pre-processing step starts by resizing the images into 96×96 dimension. After that, the pixel values of the images are converted to an array form where each of the element of array indicates the value of pixel. The pixel value range is in between 0 and 255 which creates a great variation among the array elements. To reduce this difference of this array elements, each of the pixel values are divided by 255 so that all the array elements come to a range between 0 and 1.

3.4 Convolutional Neural Network Model

Convolutional neural network model (CNN) and a deep learning neural network model have been utilized in this study for training the image dataset as CNNs are effective in learning the raw data automatically [2, 3, 19, 26]. The architecture of this model has been illustrated in Fig. 2.

According to Fig. 2, the model consists of three convolution layers where there are 32 filters in first convolutional layer, 64 filters in second convolutional layer, and 128 filters filters in third convolutional layer with 3×3 kernel size. In Eq. (1), the mathematical expression of convolution layer over an image $s(x, y)$ is defined using filter $t(x, y)$.

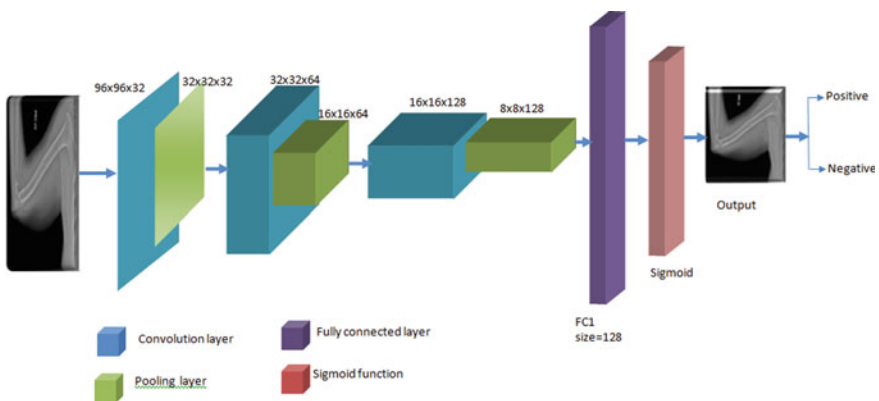


Fig. 2 Model architecture

$$t(x, y)s(x, y) = \sum_{y=-j}^j \sum_{y=-k}^k w(m, n) f(x - m, y - n) \tag{1}$$

Each of the convolution layer is followed by batch normalization and ReLU activation function. The ReLU activation function is expressed mathematically which is shown in Eq. (2)

$$\text{Relu}(z) = \max(0, z) \tag{2}$$

A max pooling layer with 2×2 pool size has been included in each convolution layer. The max pooling layer efficiently downsizes the tight scale images and mitigates total numbers of parameters. Thus, it reduces the size of the dataset.

The architecture is the followed by fully connected (FC) layer where all the inputs of 1 layer are linked to each of the activation unit of the subsequent layer. A dropout layer is attached to the FC layer, and here, the size of the FC layer is 128. The dropout layer helps to reduce the over-fitting problem by randomly disconnecting nodes. Here, we have selected 0.2 as dropout range for this layer.

Finally, a sigmoid function has been applied after the FC layer which acts as classifier to classify images as “positive” or “negative”. The mathematical expression of sigmoid function has been shown in the following Eq. (3).

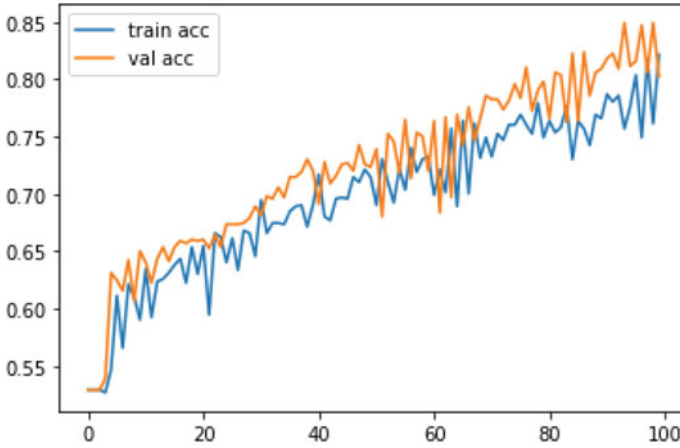
$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-\theta^T x}} \tag{3}$$

4 Experimental Results and Evaluation

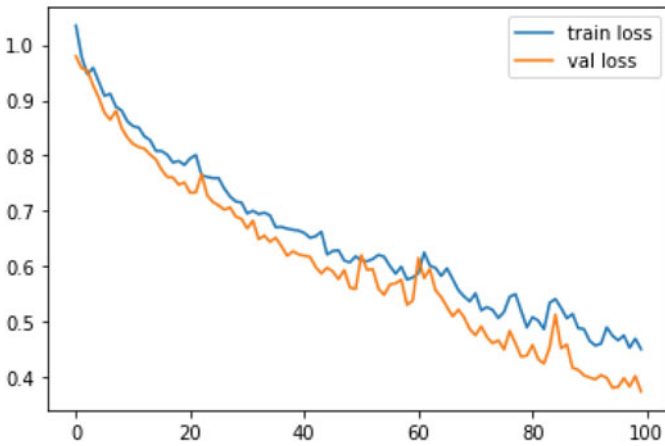
The outcomes after implementing our model on the humerus bone images collected from MURA dataset have been explained in detail in this section. The result after applying five fold cross validation has been illustrated. In addition, we have also shown the comparison of our proposed model with other traditional models.

4.1 Results and Discussion

Figure 3 graphically represents the learning curve of our model. According to Fig. 3, we can see that the X axis denotes the total epoch used to train dataset, whereas the Y axis denotes the accuracy and loss, respectively. From Fig. 3a, it can be observed that the training accuracy increases from 52 to 71%, and testing accuracy increases from 53 to 72%, respectively, after the completion of 50 epochs. Finally, the training accuracy reaches to 82%, and testing accuracy reaches to 80% after the completion of final epoch.



a



b

Fig. 3 Accuracy and loss curve: **a** accuracy versus epoch **b** loss versus epoch

Accordingly, from Fig. 3b, it can be observed that the training loss decreases from 1.03 to 0.62, and testing loss decreases from 0.97 to 0.56, respectively, after the completion of 50 epochs. Finally, the training loss reaches to 45%, and testing loss reaches to 37% after the completion of final epoch.

In this research, we have attempted five fold cross validation for assessing our model. In Table 3, the training and testing accuracies after applying five fold cross validation have been demonstrated where individual accuracy for each fold is added. The average (77.20% training and 73.60% testing) and best (80% training and 78% testing) accuracies are also added in the table which are calculated from the five fold cross validation.

Figure 4 illustrates accuracy versus epoch curve for each fold.

Table 3 Data augmentation settings

Fold	Training accuracy (%)	Testing accuracy (%)
1	80	78
2	77	75
3	78	71
4	73	70
5	78	74
Average accuracy	77.20	73.60
Best accuracy	80	78

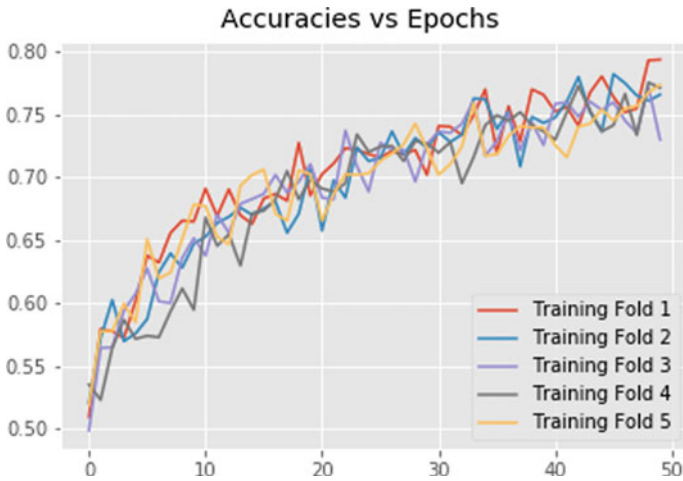


Fig. 4 Accuracy versus epoch for five fold cross validation

Table 4 Comparison of results

Models	Train acc	Train loss	Test acc	Test loss
CNN	0.80	0.4505	0.78	0.3739
VGG19	0.5830	0.6700	0.5825	0.6790
VGG16	0.6290	0.6510	0.6310	0.6500
InceptionV3	0.7100	0.5500	0.6530	0.6460
DenseNet121	0.7600	1.100	0.5490	0.5400
MobileNetV2	0.8400	0.3800	0.5490	1.10
ResNet50	0.7879	0.4493	0.4724	0.7093

4.2 Comparison of Different Models

We have trained our dataset using some pretrained model (VGG16, VGG19, ResNet50, MobileNetV2, InceptionV3, DenseNet121) in the purpose of showing

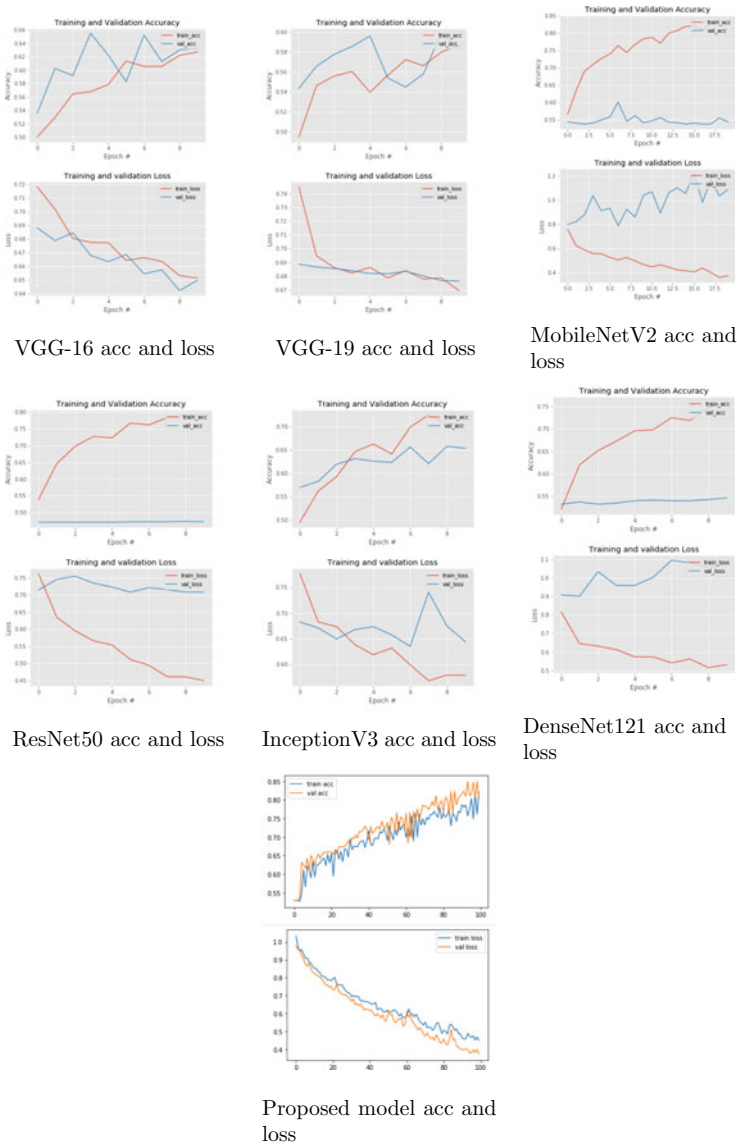


Fig. 5 Accuracy and loss curve of proposed model and other traditional models

the comparison and also for achieving the best accuracy. After implementing all this models, we have observed that our proposed CNN model outperforms the other models. Table 4 illustrates the training and testing accuracy as well as the training and testing loss of the other models.

In Fig. 5, the curve of accuracy and loss function of our models as well as other models are illustrated.

5 Conclusion

The main objective of this research is to detect fractured humerus from X-ray image using deep learning algorithm. This study exhibits the probability of future use of deep learning algorithm in the field of orthopedic surgery. This automated system of detecting humerus fractures has some advantages, for example, quick detection, efficiency, and time saving process. However, we have utilized convolutional neural network (CNN) model as a deep learning approach for training our dataset. The X-ray image of fractured and non-fractured humerus bone was used to perform the experiment where total of 1266 X-ray images have been used. Moreover, data augmentation has been applied to increase the size of the dataset which helps to overcome the over-fitting problem. Finally after classification, the accuracy that we gained from this research is 78%.

In future, this research aims at collecting more images to train our model. Moreover, we intent to improve our system by applying the BRBES-based adaptive differential evolution (BRBaDE). This can ameliorate the accuracy by conducting parameter and structure optimization [10–13, 18, 24, 25].

References

1. Adams, M., Chen, W., Holcdorf, D., McCusker, M.W., Howe, P.D., Gaillard, F.: Computer vs human: deep learning versus perceptual training for the detection of neck of femur fractures. *J. Med. Imaging Radiat. Oncol.* **63**(1), 27–32 (2019)
2. Ahmed, T.U., Hossain, S., Hossain, M.S., ul Islam, R., Andersson, K.: Facial expression recognition using convolutional neural network with data augmentation. In: 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), pp. 336–341. IEEE (2019)
3. Basnin, N., Nahar, L., Hossain, M.S.: An integrated CNN-LSTM model for micro hand gesture recognition. In: International Conference on Intelligent Computing & Optimization, pp. 379–392. Springer, Berlin (2020)
4. Basnin, N., Sumi, T.A., Hossain, M.S., Andersson, K.: Early detection of Parkinson's disease from micrographic static hand drawings. In: International Conference on Brain Informatics, pp. 433–447. Springer, Berlin (2021)
5. Chung, S.W., Han, S.S., Lee, J.W., Oh, K.S., Kim, N.R., Yoon, J.P., Kim, J.Y., Moon, S.H., Kwon, J., Lee, H.J., et al.: Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop.* **89**(4), 468–473 (2018)
6. De Vries, B., Hegeman, J., Nijmeijer, W., Geerdink, J., Seifert, C., Groothuis-Oudshoorn, C.: Comparing three machine learning approaches to design a risk assessment tool for future fractures: predicting a subsequent major osteoporotic fracture in fracture patients with osteopenia and osteoporosis. *Osteoporos. Int.* **32**(3), 437–449 (2021)
7. Demir, S., Key, S., Tuncer, T., Dogan, S.: An exemplar pyramid feature extraction based humerus fracture classification method. *Med. Hypotheses* **140**, 109663 (2020)
8. Devana, S.K., Shah, A.A., Lee, C., Gudapati, V., Jensen, A.R., Cheung, E., Solorzano, C., van der Schaar, M., SooHoo, N.F.: Development of a machine learning algorithm for prediction of complications and unplanned readmission following reverse total shoulder arthroplasty. *J. Shoulder Elbow Arthroplasty* **5**, 24715492211038172 (2021)

9. Islam, M.Z., Hossain, M.S., ul Islam, R., Andersson, K.: Static hand gesture recognition using convolutional neural network with data augmentation. In: 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), pp. 324–329. IEEE (2019)
10. Islam, R.U., Hossain, M.S., Andersson, K.: A deep learning inspired belief rule-based expert system. *IEEE Access* **8**, 190637–190651 (2020)
11. Jamil, M.N., Hossain, M.S., ul Islam, R., Andersson, K.: A belief rule based expert system for evaluating technological innovation capability of high-tech firms under uncertainty. In: 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), pp. 330–335. IEEE (2019)
12. Kabir, S., Islam, R.U., Hossain, M.S., Andersson, K.: An integrated approach of belief rule base and deep learning to predict air pollution. *Sensors* **20**(7), 1956 (2020)
13. Karim, R., Andersson, K., Hossain, M.S., Uddin, M.J., Meah, M.P.: A belief rule based expert system to assess clinical bronchopneumonia suspicion. In: 2016 Future Technologies Conference (FTC), pp. 655–660. IEEE (2016)
14. Negrillo-Cárdenas, J., Jiménez-Pérez, J.R., Cañada-Oya, H., Feito, F.R., Delgado-Martínez, A.D.: Automatic detection of landmarks for the analysis of a reduction of supracondylar fractures of the humerus. *Med. Image Anal.* **64**, 101729 (2020)
15. Negrillo-Cárdenas, J., Jiménez-Pérez, J.R., Feito, F.R.: Automatic detection of distal humerus features: first steps. In: VISIGRAPP (1: GRAPP), pp. 354–359 (2019)
16. Pranata, Y.D., Wang, K.C., Wang, J.C., Idrum, I., Lai, J.Y., Liu, J.W., Hsieh, I.H.: Deep learning and surf for automated classification and detection of calcaneus fractures in CT images. *Comput. Methods Programs Biomed.* **171**, 27–37 (2019)
17. Progga, N.I., Hossain, M.S., Andersson, K.: A deep transfer learning approach to diagnose covid-19 using X-ray images. In: 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), pp. 177–182. IEEE (2020)
18. Rahaman, S., Hossain, M.S.: A belief rule based clinical decision support system to assess suspicion of heart failure from signs, symptoms and risk factors. In: 2013 International Conference on Informatics, Electronics and Vision (ICIEV), pp. 1–6. IEEE (2013)
19. Rezaoana, N., Hossain, M.S., Andersson, K.: Detection and classification of skin cancer by using a parallel CNN model. In: 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), pp. 380–386. IEEE (2020)
20. Sasidhar, A., Thanabal, M., Ramya, P.: Efficient transfer learning model for humerus bone fracture detection. *Ann. Rom. Soc. Cell Biol.* 3932–3942 (2021)
21. Sezer, A., Sezer, H.B.: Convolutional neural network based diagnosis of bone pathologies of proximal humerus. *Neurocomputing* **392**, 124–131 (2020)
22. Shrader, M.W., Sanchez-Sotelo, J., Sperling, J.W., Rowland, C.M., Cofield, R.H.: Understanding proximal humerus fractures: image analysis, classification, and treatment. *J. Shoulder Elbow Surgery* **14**(5), 497–505 (2005)
23. Sumi, T.A., Hossain, M.S., Islam, R.U., Andersson, K.: Human gender detection from facial images using convolution neural network. In: International Conference on Applied Intelligence and Informatics, pp. 188–203. Springer, Berlin (2021)
24. Uddin Ahmed, T., Jamil, M.N., Hossain, M.S., Andersson, K., Hossain, M.S.: An integrated real-time deep learning and belief rule base intelligent system to assess facial expression under uncertainty. In: 9th International Conference on Informatics, Electronics & Vision (ICIEV). IEEE Computer Society (2020)
25. Zisad, S.N., Chowdhury, E., Hossain, M.S., Islam, R.U., Andersson, K.: An integrated deep learning and belief rule-based expert system for visual sentiment analysis under uncertainty. *Algorithms* **14**(7), 213 (2021)
26. Zisad, S.N., Hossain, M.S., Andersson, K.: Speech emotion recognition in neurological disorders using convolutional neural network. In: International Conference on Brain Informatics, pp. 287–296. Springer, Berlin (2020)