Bilge International Journal of Science and Technology Research
ISSN: 2651-401X
e-ISSN: 2651-4028
Research Article
2023, Volume: 7, Issue: 1, 63-69
Received: 09.06.2022; Accepted: 23.01.2023
DOI: 10.30516/bilgesci.1128622

The Effects of the Traditional Data Augmentation Techniques on Long Bone Fracture Detection

Gülnur Begüm Cangöz[*1], Selda Güney[2]

Abstract: Image collection and preparation phases are highly costly for machine learning algorithms. They require the majority of labeled data. Hence, the image pre-processing method, data augmentation, is commonly used. Since there are so many proposed methods for the augmentation task, this comparison study is presented to be a supporting guide for the researchers. In addition, the lack of studies with animal-based data sets makes this study more valuable. The study is investigated on a comprehensive medical image data set consists of X-ray images of many different dogs. The main goal is to determine the fracture of the long bones in dogs. Many traditional augmentation methods are employed on the data set including flipping, rotating, changing brightness and contrast of the images. Transfer learning is applied on both raw and augmented data sets as a feature extractor and Support Vector Machine (SVM) is utilized as a classifier. For the classification task, the experimental study shows that changing the contrast is the outstanding method for accuracy manner, while the rotation method has the best sensitivity value. The classification accuracy of the raw data, which was 0.817, improved to 0.845 with augmented data by changing the contrast values. The findings of the study also demonstrate that the transfer learning method is highly effective on the animal-based data set.

Keywords: biomedical image processing, bone fractures, convolutional neural networks, deep learning, data augmentation.

[1]**Address:** ASELSAN A.Ş. Gölbaşı, Ankara/Turkiye.
[2]**Address:** Başkent University, Electrical -Electronics Engineering Faculty, Ankara/Turkiye.

***Corresponding author**: bcangoz@aselsan.com.tr

Citation: Cangöz, G. B., Güney, S. (2023). The Effects of the Traditional Data Augmentation Techniques on Long Bone Fracture Detection. Bilge International Journal of Science and Technology Research, 7(1): 63-69.

1. INTRODUCTION

The recent development of deep learning studies has allowed the use of deep learning methods in many various fields. One of the most popular field is biomedical (Li, 2022). In the area of orthopedics, these deep architectures are utilized for assisting in fracture detection (Jangid et al., 2021), bone disease detection (Rajesh et al., 2022) and age assessment (Deshmukh et al., 2021).

Despite of all advances, deep architectures require majority of labeled data for learning phase. Accessing a comprehensive data set, especially in the field of medicine, is an important problem for deep learning analyzes. That problem leads us to one of the most preferred pre-processing methods, data augmentation.

Data augmentation is a transition from limited data to more data. On one hand, it helps to reduce overfitting problem, on the other hand, it equalizes unbalanced data sets. In order to prevent overfitting problem, it is possible to modify the network structure. Batch normalization and drop-out can be given as examples of the modifications. Data augmentation techniques are different from them inasmuch as they are basically a pre-processing step (Shorten & Khoshgoftaar, 2019).

In the literature, data augmentation methods are frequently used in deep learning analyzes. In a work, authors explored and compared the data augmentation methods for image classification by applying simple techniques, such as cropping, rotating, and flipping the images. Also, they experimented Generative Adversarial Neural Networks (GANNs), and a proposing method named neural augmentation. Their experiments show that the traditional augmentation methods are more effective than the others (Perez & Wang, 2017).

Shijie et al. used some data augmentation methods in their paper include: GAN/WGAN, flipping, cropping, shifting, PCA jittering, color jittering, noise, rotation, and some combinations. According to the results of the study, the four individual methods (cropping, flipping, WGAN, rotation) perform generally better than the others, and some appropriate combination methods are slightly more effective than the individuals (Shijie et al., 2017).

In the recent years, GANNs have become very popular for synthesizing images (Calimeri et al, 2017, Frid-Adar et al, 2018, Shin et al., 2018). GANNs are one of the machine learning method purposed by Ian Goodfellow (Goodfellow et al., 2014). GAN Network is formed by two neural networks competing with each other. The network has an ability to generate new data from given training set.

In another research paper, a variety of augmentation strategies, horizontal flips, random crops, and principal component analysis (PCA) are investigated. Their work shows that augmentation strategy greatly affects classification performance (Hussain et al., 2018).

The works related traditional methods can be exemplified further (Jia et al., 2017; Hernández-García & König, 2018; Sajjad et al., 2019; Shunjiro et al., 2020). Nevertheless, these methods can be highly impacted by data sets, and it is difficult to find studies in the literature using comprehensive data sets of animal-based X-Ray images.

In the literature there are few studies based on X-Ray images of pets. In a research, 143 X-ray images of dogs are used due to detect the tibia fractures. They used SSD MobileNet-v2 and obtained the F-score value as 0.68 (Baydan & Ünver, 2020). After that publication, the researchers applied different deep learning architectures for the same task by increasing the number of X-Ray images. They emphasized that the results of the study are promising for fracture detection of the tibia bone (Baydan et al., 2021).

Another paper based on animal medical image analysis is a lesion identification problem. Arsomngern et al. examined 2862 thoracic X-ray images obtained from both dogs and cats to classify lung lesion. They achieved 79.6% success with the CNN model they used (Arsomngern et al., 2019).

McEvoy et al used partial least square discriminant analysis and artificial neural networks as machine learning methods for the classification problem of canine pelvic radiographs. Their dataset consists of 256 images of dogs. Due to the results, their study can be useful in the veterinary field. (McEvoy et al., 2013).

The literature review showed that augmentation methods can be remarkably beneficial for the classification and detection tasks. In addition, it is noticed that related works are inadequate in the veterinary field. Therefore, this comparison study can be useful for the researchers.

In this study, a comprehensive data set created from dogs in Ankara Metropolitan Municipality Stray Animals Temporary Care Home is employed. For more detail about the data set, readers can be referred in our previous studies (Ergün et al., 2021; Ergün & Güney, 2021). The aim is to investigate the effect of the traditional augmentation techniques on detecting long bone fractures of the dogs. Deep architectures are utilizied for realization the task. Since the task needs a lot of labeled data, data augmentation techniques are investigated including flipping, rotating, changing brightness and contrast of the images. These methods are traditional data augmentation techniques and simple to apply. After performing the pre-processing step, deep neural models are carried out for the feature extraction, then SVM is used for the classification.

2. MATERIAL AND METHOD

2.1. Long Bones of Canine (Dog)

The canine's skeleton is made up from an average of 319 bones. These bones are divided into five types depending on their function: short, long, flat, irregular and sesamoid (The Skeleton, Ch. 4, 2016). In this work, the data set is created by X-Ray images of long bones. Long bones of dogs are also divided into six classes: femur, humerus, radius, ulna, tibia and fibula. The types of the long bones are given in Fig.1.

Although long bones of the dogs are investigating in six types, four classes are employed in this paper. According to the orthopedist veterinarian, any bone fracture occurred on radius or ulna bones, occur on both bones. Thus, radius-ulna is considered as a single bone class named 'Radius-Ulna'. Besides, the fibula bone fractures are not considered in the work. Because fibula bone is an accessory bone and carrying body weight is not a duty for it.

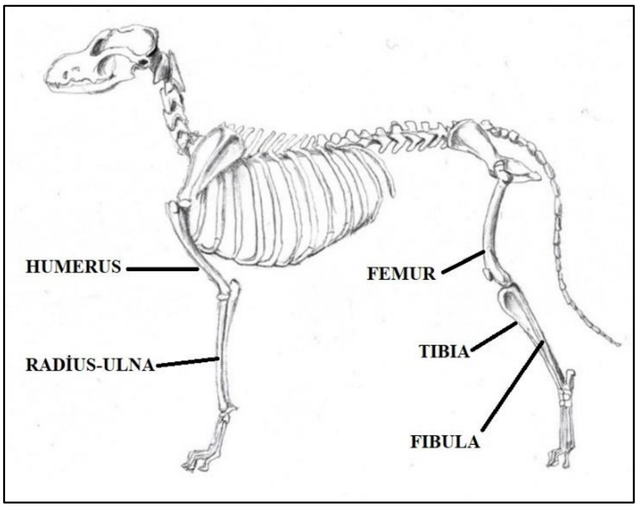

Figure 1. Labelled Long Bones of Canine (Lewis, 2019).

2.2. Data Set

The data set consists of 2028 X-Ray images of long bones of many different dogs taken from Ankara Metropolitan Municipality Stray Animals Temporary Nursing Home. 479 images of the data set were labeled as fractured, and the remaining 1549 images were labeled as no fracture. For better understanding of the data set, an example is given in Fig.2. Both images in the figure belong to radius-ulna.

2.3. Methods for Data Augmentation

Some of classical image data augmentation techniques can be divided into two categories. These are:

a. Position augmentation: cropping, flipping, padding, rotation.
b. Color augmentation: brightness, contrast, saturation, hue.

(a) (b)

Figure 2. Sample images from the data set.
a) fractured, b) no fracture

Although, there are various methods for augmentation, four common data augmentation methods are investigated in this study because of the simplicity.

2.3.1. Flipping: Images can flip horizontally and vertically. In this study, three flipping options are applied: horizontally, vertically and both.

2.3.2. Rotation: It is rotating the images at specific angles. In this study, three different angles are applied: +15, -15 and -30 degrees. Positive angles rotate images to the left and vice versa.

2.3.3. Brightness: Another way for augmentation is changing the brightness of the image. The resultant images become lighter or darker depends on the application.

In a gray-level image matrix, pixel values can be between 0-255. Values closer to 0 mean darker, while values closer to 255 mean lighter.

Changing the brightness is done by increasing or decreasing the value of each pixel of the image. Therefore, adding a coefficient to each pixel value of the image increases the brightness of the image. Likewise, the process of subtracting the coefficient has a reducing effect on the brightness.

In this study, three brightness settings are applied:

a. Adding 50 of each pixel value of the images.
b. Adding 75 of each pixel value of the images.
c. Subtracting 50 of each pixel value of the images.

2.3.4. Contrast: Changing the contrast of the image is another method for color augmentation. The resultant images become more or less distinguishable.

In gray-level input images, the desired contrast to match the values in the output image must be between (0-1). These limits were processed in three different ways by trial and error method in this study:

a. Contrast Setting 1: taking the contrast limits of the images as (0.1 - 0.9).
b. Contrast Setting 2: taking the contrast limits of the images as (0.15 - 0.85).
c. Contrast Setting 3: taking the contrast limits of the images as (0.2 - 0.8).

After the process, size of the data set increased from 2028 (479 broken, 1549 non-broken bones) to 8108 (1916 broken, 6192 non-broken bones) for each technique. After augmentation process, the outputs from all techniques are given in Fig.3- Fig.6.

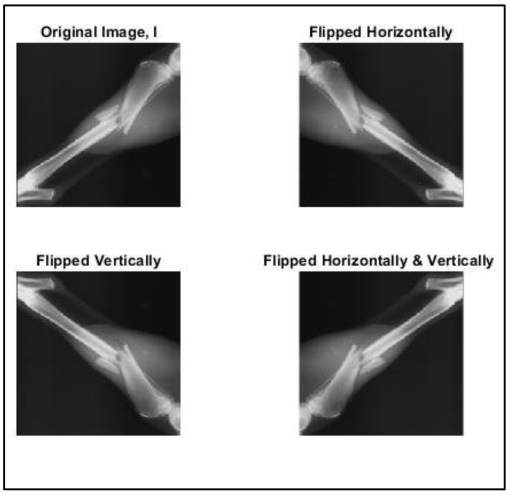

Figure 3. Output images after augmentation process for flipping.

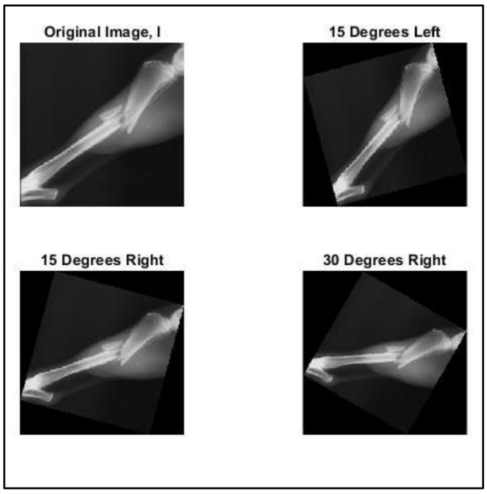

Figure 4. Output images after augmentation process for rationing.

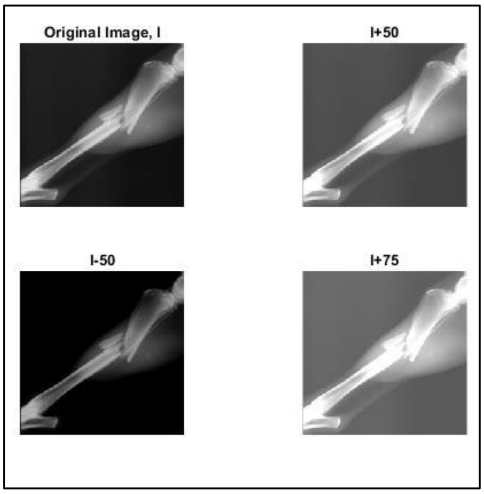

Figure 5. Output images after augmentation process for brightness.

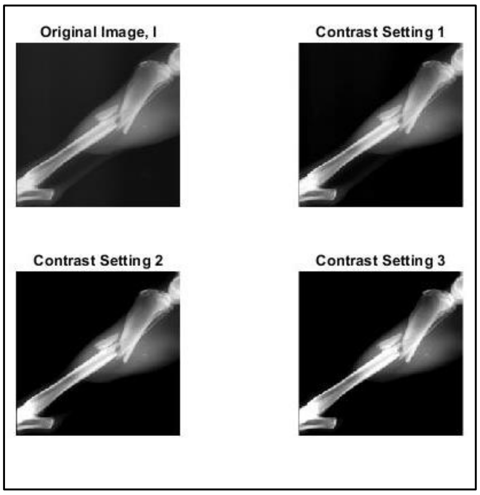

Figure 6. Output images after augmentation process for contrast.

In this study, because of the high performances on image processing problems, convolutional neural network(CNN) is used for both training and validation. (Guo et al., 2016). A convolutional neural network consists of 5 primary layers: An input layer, convolution layers, pooling layers, fully connected layers and an output layer (LeCun et al., 1989). The purpose of the convolution layer is to extract features from the input image by performing a dot product between images and filters. Pooling layer reduces dimension of image obtained from the previous layer. The matrix obtained by passing through all the determined layers is turned into a flat vector in the fully connected layer.

For the classification process, transfer learning is applied. One of the most popular deep architecture, Inception-v3 is employed for the feature extraction. Afterwards, Support Vector Machines (SVMs) is used for the classification task.

Inception-v3 is a deep convolutional neural network for assisting in detection and classification tasks. The architecture is the third edition of Google's Inception Convolutional Neural Network (Szegedy et al., 2015.). Inception-v3 has 316 deep layers with 350 connections. A faster training process is achieved by choosing number of smaller convolutional filters. The network is presented in Fig.7.

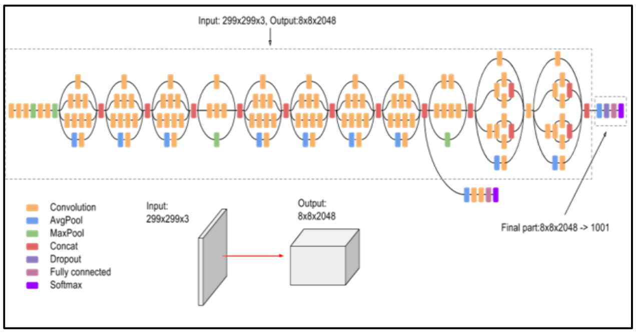

Figure 7. Inception-v3 Architecture (Szegedy et al., 2015).

In the study, the training and test sets are randomly selected 0.8 and 0.2 from the data set, respectively. Dimensions of each image in the data set are set to 200x200.

2.4. Performance Metrics

Many performance metrics are utilized due to measure the success of the models. In this study, accuracy and sensitivity are preferred. The accuracies of all models are calculated using (1).

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

where the true negative (TN) parameter shows the number of negative examples classified accurately. Similarly, true positive (TP) indicates the number of positive examples classified accurately. False positive (FP) means the number of actual negative examples classified as positive; false negative (FN) is the number of actual positive examples classified as negative.

Sensitivity parameter is calculated using (2). This metric is often used when false negatives are more attention than false positives.

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \tag{2}$$

3. RESULTS

The objective of the study is to classify the images as fractured or no fracture. In order to achieve this goal, Inception-v3 network and SVMs are utilized. While realizing the objective, some image augmentation methods are applied and compared their effect of the classification success. The results of the classification are given in Table 1 for each augmentation technique.

Despite the diversity of the augmentation techniques, in this study four of them are studied. It is because, these methods are easy to implant and they have very low cost.

From Table 1, it can be seen that the classification accuracy of the raw data is 0.8177. After the augmentation methods, the accuracy is increased except for the rotation process. The most effective method is changing the contrast of the images in data set, as it has 0.8450 classification accuracy. Although the rotation process seems to decrease the classification success, it is observed that it increases the sensitivity.

In medical cases, false negatives are more substantial than false positives. Consequences can be more severe if actual positive cases are missed. Therefore, sensitivity comes to the front. Although the accuracy of rotation method is low, the sensitivity rates are more promising.

Table 1. Classification accuracy for raw and augmented data sets

Data Set	Accuracy	Sensitivity
Raw data	0.8177	0.8210
Augmented data with flipping	0.8425	0.8374
Augmented data with rotating	0.7700	**0.9029**
Augmented data with brightness	0. 8225	0.8524
Augmented data with contrast	**0.8450**	0.8136

4. DISCUSSION AND CONCLUSIONS

After literature review, it is easily seen that the veterinary medicine problems are not adequately studied using deep neural networks. Although deep neural architectures need majority of data sets, it is hard to find comprehensive animal-based medical image data set. In order to make the deep models well trained and more robust, data augmentation is used.

Recently, data augmentation has been made using deep models. Using deep models can achieve great success, however they have complex structures as well as being highly dependent on hardware systems. For these reasons, traditional methods are applied in this study. For the future studies it would be intriguing to implement the modern approaches based on deep learning.

Despite recent advances in deep models, classical augmentation methods still remain popular. In a study which exemplifies this popularity, the authors aimed to classify different types of fractures in the proximal humerus bone of humans. 1891 plain shoulder radiographs with five labels were investigated in the study. The size of the training data set was increased by using traditional augmentation methods such as shifting, scaling and rotations (90°, 180°, 270°). After the pre-processing step, the authors used ResNet-152 network and obtained promising performance. It is claimed that using 90°, 180° and 270° rotations for augmentation process might lead to overfitting and they suggested changing the degree slightly. Furthermore, using JPEG images may affect the image quality because of the lossy compression process. For this reason, they also suggested to use lossless compression images, such as PNG and TIFF (Chung et al., 2018.).

In this study, using PNG images and rotating the images with small angles seems to be assisting to the study of Chung et. al. Nevertheless, it was observed that the rotation method reduced the classification accuracy. It appears that some image information may have lost somehow during the rotation process.

One of the keys to the success of augmented data sets is the selection of the appropriate augmentation method according to the data set. Because, application of some augmentation techniques may add misleading information to the data. In the study, the authors emphasized that using shear, strain or spot noise augmentation can result in a misclassification situation. Thus, they prefer to perform mirroring, sharpness, brightness and contrast augmentation (Yahalomi et.al., 2019.). It can be thought that a similar situation may occur for the rotation method in this study.

What's more, the researchers of the regarding study aim to classify wrist fractures as broken or unbroken. Their data set was composed of 695 fractured and 694 unbroken labeled images. They augmented images by applying flipping, rotation, shifting, shearing and zooming. They used InceptionV3 as a transfer learning method and demonstrated that transfer learning from deep CNN pre-trained on non-medical images are effective just as this study. They computed offline data augmentation in their study, and the same augmented images were used to train deep model. The authors implied that it could be better to use different augmented images for each epoch (Kim and MacKinnon, 2018.). Hence, it may be considered to conduct a study for online data augmentation in future studies.

In conclusion, data augmentation is a very powerful technique for creating more comprehensive data sets. In the literature, numerous related works have been done, however all of these methods depend on the data sets and to the best of authors' knowledge, no similar research exists on the data set consisting canine's X-Ray images. For this reason, authors wanted to contribute to the literature with the original data set. The results of this study are promising for future works.

Ethics Committee Approval
N/A

Peer-review
Externally peer-reviewed.

Conflict of Interest
The authors have no conflicts of interest to declare.

Funding
The authors declared that this study has received no financial support.

REFERENCES

Arsomngern, P., Numcharoenpinij, N., Piriyataravet, J. Teerapan, W., Hinthong W. and Phunchongharn, P. "Computer-Aided Diagnosis for Lung Lesion in Companion Animals from X-ray Images Using Deep Learning Techniques," 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST), Morioka, Japan, pp. 1-6, 2019.

Baydan, B., Ünver, H. M.: "Dataset creation and SSD mobilenet V2 performance evaluation for dog tibia fracture detection", II. International Ankara Congress of Scientific Research, Ankara, Turkey, 6-8 March, 2020.

Baydan, B., Ünver, H. M., Barışçı, N.,"Determining the Location of Tibial Fracture of Dog and Cat Using Hybridized Mask R-CNN Architecture", Kafkas Üniversitesi Veteriner Fakültesi Dergisi, vol. 27, 3, pp. 347 – 353, 2021.

Calimeri, F., Marzullo, A., Stamile, C., Terracina, G. (2017). Biomedical Data Augmentation Using Generative Adversarial Neural Networks. In: Lintas, A., Rovetta, S., Verschure, P., Villa, A. (eds) Artificial Neural Networks and Machine Learning – ICANN 2017. ICANN 2017. Lecture Notes in Computer Science, vol 10614. Springer, Cham. doi:10.1007/978-3-319-68612-7_71.

Christian S., Vincent V., Sergey I., Jonathon S., Zbigniew W., "Rethinking the Inception Architecture for Computer Vision", Dec. 2015.

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. 2018, 89, 468–473.

Deshmukh S. and Khaparde A., "Segmentation Using CNN for BAA using Modified TW3 method," 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET), 2021, pp. 1-8, doi: 10.1109/ICECET52533.2021.9698478.

Ergün G. B., Güney S., Ergün T.G., Köpeklerdeki Uzun Kemiklerin Evrişimsel Sinir Ağları Kullanılarak Sınıflandırılması, Fırat Üniversitesi Fen Bilimleri Dergisi, vol. 33, pp. 125-132, 2021.

Ergün G. B. and Güney, S. "Classification of Canine Maturity and Bone Fracture Time Based on X-Ray Images of Long Bones," in IEEE Access, vol. 9, pp. 109004-109011, 2021, doi: 10.1109/ACCESS.2021.3101040.

Frid-Adar, M., Klang, E., Amitai, M., Goldberger J., and Greenspan, H. "Synthetic data augmentation using GAN for improved liver lesion classification," 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 289-293, doi: 10.1109/ISBI.2018.8363576

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., (2014). Generative Adversarial Nets (PDF). Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014). pp. 2672–2680.

Guo K. et al., "Angel-Eye: A Complete Design Flow for Mapping CNN onto Customized Hardware," 2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), Pittsburgh, PA, 2016, pp. 24-29.

Hernández-García A., König P. "Further Advantages of Data Augmentation on Convolutional Neural Networks," Artificial Neural Networks and Machine Learning ICANN 2018. Lecture Notes in Computer Science, vol 11139. Springer, Cham. https://doi.org/10.1007/978-3-030-01418-6_10.

Hussain Z., Gimenez F., Yi D., Rubin D., "Differential Data Augmentation Techniques for Medical Imaging Classification Tasks," AMIA Annu Symp Proc. pp. 979-984, 2018.

Jangid R. G, Naik K., N. and Francis R. M., "Detection of Abnormality in Human Hard Tissue using Edge Detection Operators," 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), 2022, pp. 1-5, doi: 10.1109/ICACTA54488.2022.9753141.

Jia S, Wang P, Jia P, Hu S. "Research on data augmentation for image classification based on convolutional neural networks" Chinese automation congress, 2017. p. 4165–70.

Kim, D.H.; MacKinnon, T. Artificial intelligence in fracture detection: Transfer learning from deep convolutional neural networks. Clin. Radiol. 2018, 73, 439–445.

LeCun, Y.; Boser, B.; Denker, J. S.; Henderson, D.; Howard, R. E.; Hubbard, W.; Jackel, L. D. (December 1989). "Backpropagation Applied to Handwritten Zip Code Recognition". Neural Computation. 1 (4): 541–551.

Lewis, G., (2019). Musculoskeletal Development of the Puppy: Birth to Twelve Months. Winter. 41-44.

Li Y., "Research and Application of Deep Learning in Image Recognition," 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA), 2022, pp. 994-999, doi: 10.1109/ICPECA53709.2022.9718847.

McEvoy, F. J., Amigo, J. M. "Using machine learning to classify image features from canine pelvic radiographs: evaluation of partial least squares discriminant analysis and artificial neural network models," Veterinary Radiology & Ultrasound, vol. 54 (2), pp. 122-126, 2013.

Pharmacology, Toxicology & Therapeutics, "The Skeleton, Cheapter 4", https://veteriankey.com/the-skeleton/ access: 21.04.2022, 15.42.

Perez, L., and Wang. J., "The Effectiveness of Data Augmentation in Image Classification using Deep Learning." ArXiv abs/1712.04621 (2017).

Sajjad M. et al. "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," Journal OF Computational Science, vol. 30, pp.174-182, 2019.

Shijie, J., Ping W., Peiyi J. and Siping, H., "Research on data augmentation for image classification based on convolution neural networks," Chinese Automation Congress (CAC), pp. 4165-4170, 2017.

Shin HC. et al. "Medical Image Synthesis for Data Augmentation and Anonymization Using Generative Adversarial Networks," Simulation and Synthesis in Medical Imaging. SASHIMI 2018. Lecture Notes in Computer Science, vol 11037. Springer, Cham. https://doi.org/10.1007/978-3-030-00536-8_1

Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0

Shunjiro N., Mizuho N., Masahiro Y., Keita N., Kaori T., "Bone segmentation on whole-body CT using convolutional neural network with novel data augmentation techniques,” Computers in Biology and Medicine, vol. 121, 2020.

Tabarestani S. S., Aghagolzadeh A. and Ezoji M., "Bone Fracture Detection and Localization on MURA Database Using Faster-RCNN," 2021 7th International Conference on Signal Processing and Intelligent Systems (ICSPIS), 2021, pp. 1-6, doi: 10.1109/ICSPIS54653.2021.9729393.

Yahalomi, E.; Chernofsky, M.; Werman, M. Detection of Distal Radius Fractures Trained by a Small Set of X-Ray Images and Faster R-CNN. In Intelligent Computing; Arai, K., Bhatia, R., Kapoor, S., Eds.; Springer International Publishing: Cham, Switzerland, 2019; Volume 997, pp. 971–981.