

# Intelligent Interpretation of Veterinary Musculoskeletal X-rays Trained with Human Thoracic Limb X-rays

Shijia Zhou (✉ [shijia.zhou@sydney.edu.au](mailto:shijia.zhou@sydney.edu.au))

The University of Sydney

Euijoon Ahn

The University of Sydney

Michael Fulham

The University of Sydney

Jinman Kim

The University of Sydney

---

## Research Article

**Keywords:** Artificial Intelligence, Automatic Medical Image Interpretation, Deep Neural Network

**Posted Date:** June 10th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-598900/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Deep learning (DL) using convolutional neural networks (CNNs) is being widely applied to assist in the interpretation of medical images in modern healthcare but there is a paucity of ‘artificial intelligence’ being currently applied to veterinary medicine. Most veterinary musculoskeletal (MSK) x-ray imaging is done in a community setting and there has not been the development of large image repositories that are available in human healthcare. Domestic animals – cats and dogs – however, have similar skeletal anatomies to humans. Hence, we hypothesized that annotated human MSK x-rays (Xs) could perhaps be used as a surrogate for the lack of adequate veterinary data to develop an automated system to help interpret veterinary MSK Xs. We refer to this as ‘intelligent interpretation of Veterinary Musculoskeletal Xs (*iVetMSK-Xs*). Our *iVetMSK-Xs* has an x-ray classifier built on EfficientNet and a lesion localizer built on gradient-weight class activation mapping (Grad-CAM). We used the human Musculoskeletal Radiograph (MURA) dataset (40,005 thoracic limb Xs) and a small veterinary x-ray dataset (500 thoracic and pelvic limb Xs from 141 cats and dogs) downloaded from online case repositories. Our results show that using the human thoracic limb Xs and the veterinary dataset to train a CNN improved diagnostic accuracy threefold. We suggest that our *iVetMSK-Xs* is an important first step in developing automated image interpretation for veterinary imaging.

## Introduction

Deep learning (DL) has been a major advance in the development of computer-assisted diagnosis (CAD) systems for human diagnostic imaging <sup>1</sup>. These DL algorithms are primarily enabled by convolutional neural networks (CNNs) <sup>2</sup>. CNNs are image feature extractors that can separate different features in images after being trained with annotated images <sup>3</sup>. CNNs have been applied to medical image segmentation, detection, classification, and diagnosis <sup>2,4</sup>. As an example, Lindsey et al. reported that DL improved clinician’s detection of wrist fractures on plain x-rays (Xs), especially where the fractures were subtle and small <sup>4</sup>. In veterinary medicine, limb musculoskeletal (MSK) Xs are frequently carried out for the investigation of lameness. These Xs are usually carried out in community practices where the initial interpretation is done by the community veterinarian (vet). The skillsets of MSK-Xs interpretation in community practices vary. So, with the rapid advances in applying DL in human healthcare, we were motivated to develop a system that could help community vets to rapidly focus on an abnormality in thoracic or pelvic limbs (limbs attached to the thorax or pelvis) of common domestic pets – cats and dogs. There are few published data on the application of artificial intelligence (AI) to veterinary imaging. Recently, Banzato et al. reported on using CNNs to detect brain tumors in dogs, with magnetic resonance imaging (MRI), and degenerative liver disease in dogs using ultrasound <sup>5-7</sup>. They concluded that their approach had the potential to be reliable tools in separating meningiomas and gliomas and diagnosing degenerative liver disease. However, the datasets were small with < 100 cases and there was also an imbalance between the different types of diseases. In addition, there was a lack of transparency in how the algorithms made decisions and were explained. In contrast, regulatory bodies in Europe and the US have been promoting a strong campaign to ensure that any CNN-driven algorithms are transparent and

explainable if they are used in clinical diagnosis<sup>8,9</sup>. This principle should also be applied to veterinary medicine in the authors opinion.

DL algorithms require a large amount of annotated data for training. Lindsey et al. annotated over 135,000 Xs of human wrists to detect fractures with a diagnostic sensitivity of 93.9% and specificity of 94.5%<sup>4</sup>. Such datasets, however, are not available in veterinary medicine. In veterinary healthcare, clients generally take their pets to local community vet practices for investigation of lameness. The Picture Archive and Communication systems (PACs) of community vet practices are generally disjointed and independent. Hence the ability to pool large amounts of vet images from multiple community practices is not possible.

Fortunately, large annotated and public datasets of MSK-Xs exist in human medicine. We believe it is possible to use the large number of human MSK-Xs to train an image interpretation system for animals. The reasons for this possibility are the anatomical similarities between human and animal limbs. In humans, the scapula, humerus, and clavicle form the shoulder joint; the humerus is the single bone that forms the arm and it articulates with the ulna and radius at the elbow. The ulna and radius are the two bones of the forearm and they articulate with the bones at the wrist. In dogs and cats, the scapula is rotated laterally so that the positioning of thoracic limbs is more appropriate for crawling. The shoulder joints are formed by the articulation between the glenoid cavity of the scapula and the head of the humerus which are universal in humans and domestic pets. In dogs and cats, the clavicle is reduced to a fibrous intersection between the two components of *brachiocephalicus*. The ulna and radius are the two bones of the forearm and they articulate with the carpus via the antebrachio-carpal joints. There is no functional difference between the ulna and radius of humans, dogs, and cats. Carpus of humans, dogs, and cats compose two rows of bones and form the base of the palm. The noticeable difference between humans, dogs, and cats' carpus is that in humans, the radial carpal bone and intermediate carpal bone are unfused for the increased flexibility of palm movements. In dogs and cats, the radial and intermediate carpal bones are fused for increased stability under the stress of running. Distal to the carpal bones, both humans and dogs, and cats have 5 sets of proximal, middle, and distal phalanges. The difference is that human has a more robust first metacarpal for the ability to grab with the thumb. In addition, domestic animals suffer from similar conditions that affect humans – trauma, fractures, tumors, osteoarthritis, and treatment can include metallic plates, pins, and screws. Hence, we hypothesized that labeled human MSK Xs could perhaps be used as a surrogate for the lack of adequate vet data to develop an automated system to help interpret veterinary MSK Xs. We refer to this approach as the 'intelligent interpretation of Veterinary Musculoskeletal Xs (*iVetMSK-Xs*) with an x-ray classifier built on EfficientNet and a lesion localizer built on gradient-weight class activation mapping (Grad-CAM). EfficientNet is the state-of-the-art CNN classifier for many computer vision tasks including image classification and object detection<sup>10</sup>. The second component, the Grad-CAM provides rapid visualization of the results.

## Materials And Methods

# Image Data

All MSK-Xs were from public datasets where animal and human ethics approvals had already been obtained and additional was not required for this study<sup>11,12</sup>. VetMSK-Xs were downloaded from DVMInsight built by IDEXX and Radiology Cases from the Radiological Society of North America (RSNA)<sup>11,12</sup>. We used the search term “limb” to identify appendicular MSK-Xs. We excluded all Xs of the skull, chest, abdomen, and spine. We used the associated report for each case to identify the lesions that were reported. There were 500 Xs from 141 animals. Each animal on average has 3–4 different views Xs of a lame limb. Furthermore, Xs of the normal limb are sometimes taken to be compared against the lame limb if the lesions are subtle. Among the 141 animals, there were 123 dogs (mean age 4.7 years), and 18 cats (mean age 5.3 years); of the dogs, 68 were male, 50 female and in 5 the gender was not specified. For the cats 13 were male and 4 were female and 1 has unspecified gender. 447 Xs are done in dogs and 53 Xs are done in cats. Out of the 500 Xs, 143 Xs were normal Xs without any lesion reported; 357 Xs were abnormal Xs with reported lesions. The distribution of the Xs was as follows: a) for dogs – 34 Xs of shoulder joint, 17 Xs of the arm, 53 Xs of elbow joint, 52 Xs of the forearm, 19 Xs of both arm and forearm, 10 Xs of carpus joint, 32 Xs of front paw, 3 Xs of phalanges, 43 Xs of the pelvis, 13 Xs of the thigh, 96 Xs of knee joint, 21 Xs of the leg, 25 Xs of both thigh and leg, 9 Xs of tarsus joint, 20 Xs of back paw; b) for cats – 1 x-ray of shoulder joint, 1 x-ray of arm, 2 Xs of elbow joint, 13 Xs of the forearm, 1 x-ray of both arm and forearm, 1 x-ray of phalanges, 1 x-ray of the pelvis, thigh and leg, 13 Xs of the pelvis, 2 Xs of the thigh, 3 Xs of knee joint, 3 Xs of the leg, 6 Xs of both thigh and leg, 3 Xs of tarsus joint, 3 Xs of the back paw. We list detailed descriptions of 500 Xs we used in our study in Table 1. We show examples of Xs without and with lesions are shown in Fig. 1.

We used the Musculoskeletal Radiograph dataset (MURA) for the human data. The MURA has 40,005 Xs from 1,863 patients. The Xs are from the upper (thoracic) limbs – shoulder, arms, elbow, forearm, wrists, and fingers<sup>13</sup>. In the MURA, the authors defined ‘lesions’ as fractures, subluxations, degenerative joint disease, and destructive abnormalities, and examples are shown in Fig. 2.

## iiVetMSK-Xs

The architecture and workflow for our approach are shown in Fig. 3. The first component is the Xs classifier. It’s built on EfficientNet. An EfficientNet in its B5 configuration is trained to classify whether an x-ray has a lesion or not. The proposed *iiVetMSK-Xs* is trained in two different ways: i) Direct Finetuning (DF-*iiVetMSK-Xs*) – finetuning the EfficientNet using the veterinary dataset; ii) MURA-Veterinary finetuning (MVF-*iiVetMSK-Xs*) – finetuning the EfficientNet using the MURA dataset and then the veterinary dataset sequentially. The training was done using PyTorch version 1.1.0 following the guide provided by Pytorch<sup>14</sup>. EfficientNet was loaded from EfficientNet PyTorch developed by Melas-Kyriazi<sup>15</sup>. All input Xs were padded to be square-shaped and resized to 1024 by 1024 pixels. We applied random rotation and contrast adjustment as data augmentation techniques. We then normalized the images using the mean and standard deviation derived from the MURA dataset. The second component was the Lesion Localizer.

We used Grad-CAM<sup>16</sup>. It can visualize areas in an image that contribute the most to the target class' score predicted by the classifier as a heatmap<sup>17</sup>. In our *iVetMSK-Xs*, the target class is the “with lesion” class. Therefore, the “hot” (red) areas in the heatmaps are areas that are the most likely to be associated with lesions identified by the classifier.

## Evaluation

We evaluated the x-ray classifiers trained by the DF and MVF by comparing their diagnostic accuracy, sensitivity, and specificity for correctly classifying the Xs with lesions. For the veterinary dataset, we performed 5-fold cross-validation with the dataset divided into different training and testing set pairs five times. When preparing these, we performed a random sampling of Xs for testing without replacement. Therefore, the 5 testing subsets were mutually exclusive and did not share any common x-ray and also maintained the 7:3 ratio between Xs with and without lesions. We calculated the average diagnostic accuracy, sensitivity, and specificity over the 5 testing subsets and used these metrics to assess which training strategy performed better. The accuracy of lesion localization was determined by the findings on the heatmaps when compared to the imaging report and was carried out by the registered Vet (the first author of this manuscript) in Australia. We display an x-ray reported with osteolysis of the humeral condyle and examples of incorrect and correct lesion localization in Fig. 4.A, B, and C respectively.

## Results

### *Diagnostic Performance:*

DF-*iVetMSK-Xs* correctly classified 383/500 Xs (76.6%). Among the 383 correctly classified Xs, 86 are Xs without lesions and 297 are Xs with lesions. Therefore, the diagnostic specificity of the DF-*iVetMSK-Xs* is 0.60 (86/143) and the diagnostic sensitivity is 0.83 (297/357). MVF-*iVetMSK-Xs* correctly classified 387/500 Xs (77.4%). Among the 387 correctly classified Xs, 82 are Xs without lesions and 305 are Xs with lesions. Therefore, the diagnostic specificity of the MVF-*iVetMSK-Xs* is 0.57 (82/143) and the diagnostic sensitivity is 0.85 (305/357).

### *Lesion Localization:*

After examining all heatmaps of the 357 Xs with reported lesions, we determined that the *iVetMSK-Xs* trained via the DF strategy (DF-*iVetMSK-Xs*) correctly localized lesions in 67 Xs. The *iVetMSK-Xs* trained via the MVF strategy correctly localized lesions in 198 Xs. We also noticed that in some Xs with reported lesions, the heatmaps generated by MVF-*iVetMSK-Xs* are more specific to the lesion areas than those generated by DF-*iVetMSK-Xs*. The heatmaps generated by DF-*iVetMSK-Xs* sometimes highlight image elements that are irrelevant to the lesions such as side or size markers. We show some selected examples in Fig. 5. The distribution of correct heatmaps generated by DF- and MVF-*iVetMSK-Xs* among different types of lesions are listed in Table 2.

We examined heatmaps generated from Xs without reported lesions by MVF-*iVetMSK-Xs* to investigate why its diagnostic specificity was low. We show some sample results in Fig. 6. The remaining heatmaps of Xs with lesions and Xs without lesions that were falsely classified are available on the GitHub page.

## Discussion

The main finding of this study was that using human data when training with MVF- *iVetMSK-Xs* improved its lesion localization accuracy threefold when compared to training with a small veterinary dataset alone.

We noticed that DF- *iVetMSK-Xs* often mistakenly detect a radiopaque marker as a false “lesion”. These phenomena explain why MVF-*iVetMSK-Xs*’s lesion accuracy was three times higher than the accuracy of DF-*iVetMSK-Xs* despite their similar diagnostic sensitivity. An example of this type of error is shown in Fig. 4.B. Our finding indicates that the use of the MURA dataset overcomes this type of error. We suggest that this is because MVF-*iVetMSK-Xs* was able to learn lesion-related features such as sharp lines of fractures, mottled appearance of destructive bone lesions from the large MURA dataset. This learning outcome was not possible with the veterinary dataset alone. The other benefit of using the MURA dataset is that the generated heatmaps are more specific to the lesion areas. In Fig. 5, we show the comparisons of heatmaps of DF- and MVF-*iVetMSK-Xs*. These heatmaps support our hypothesis that the features learned from the MURA dataset can help with the detection of lesions in veterinary Xs.

We also reviewed heatmaps of Xs without lesions that were falsely classified by MVF-*iVetMSK-Xs*. We noticed that our MVF-*iVetMSK-Xs* had a high tendency of detecting normal joints as joints with lesions. We show 2 examples in Fig. 6.B and Fig. 6.D. We suggest the reason for this issue is that small lesions are most frequently present within joints due to limited space and small bone sizes. In total we counted 34 Xs with normal joints has been detected as joints with lesions (8 Xs of elbow, 8 Xs of carpus, 2 Xs of front paw, 1 x-ray of spine, 3 Xs of hip joint, 5 Xs of knee, 5 Xs of tarsus, and 2 Xs of metatarsus). The CNN classifier we chose is not adequately designed for detecting small lesions. Therefore, instead of learning to localize small lesions, our *iVetMSK-Xs* simply overreads all Xs with joints to compensate for this inadequacy. This also explains why MVF-*iVetMSK-Xs* were still poor at detecting osteochondrosis. Since osteochondrosis’s reported prevalence in men aged over 15 in Denmark is 4.1% <sup>18</sup>, it should also be present in the MURA dataset and allowing *iVetMSK-Xs* to learn to recognize osteochondrosis. This is a common shortcoming of CNN based image classifier. Other studies overcome this issue by using specialized CNN architectures that are trained and optimized for small structures. For example, in a study that attempted to detect microaneurysms in color images of human fundus images, the author divided the original images into small patches and oversampled the patches that contained microaneurysms and were able to achieve better accuracy than the standard CNN classifier <sup>19</sup>.

Another cause of false “lesion” detection by MVF- *iVetMSK-Xs* in veterinary Xs is that MURA contains Xs of thoracic body limbs only. However, the 56 Xs of the pelvis also contains the lower abdomen. The texture of certain elements in the abdomen can resemble bone lesions. For example, in Fig. 6.F, the fecal

material with mottled texture was detected as a “lesion” by MVF- *iVetMSK-Xs*. To overcome these inadequacies and further improve our proposed CAD’s clinical usefulness, we envision the following approaches to improve it by (i) adjusting the CNN architecture so that it can better localize small lesions; (ii) identifying Xs datasets of human pelvic limbs such as Xs of the hip, knee, and ankle.

## Conclusion

In this study, we present an intelligent image interpretation system that can first classify between Xs without and with lesions. Then the interpretation system can localize the lesions if the Xs are classified as with lesions. The use of human x-ray data improved the proposed CAD’s ability to localize lesions in Xs of dogs and cats by three folds. This study lays the foundation for developing an intelligent and automatic MSK-Xs interpretation system for veterinary medicine.

## Abbreviations

AI, Artificial Intelligence; CNN, Convolutional Neural Network; CAD, Computer-assisted Diagnosis; DF, Direct Finetuning; DL, Deep Learning; MURA, Musculoskeletal Radiograph Dataset; MVF, MURA-Veterinary Finetuning; X-rays, Xs;

## Declarations

### Conflict of interest

None.

### Imaging Techniques:

- System brand: Unknown
- Post-processing filter or algorithm: A deep learning-based computer-assisted diagnosis system
- Technique settings: unknown
- The musculoskeletal Xs used in this project were downloaded from publicly available websites: Veterinary Radiology Cases (link: <http://mirc.veterinaryradiology.net/query>) and DVMInsight Image Library BETA (Link: <https://www.dvminsight.com/ImageLibrary/SearchPage.aspx>).

The code used in this study are available at <https://github.com/DasuberVetLeonidas/Veterinary-Radiograph-Abnormality-Detection>

## References

## Tables

Table 1

Caption: Demographic characteristics of the animals in the veterinary dataset.

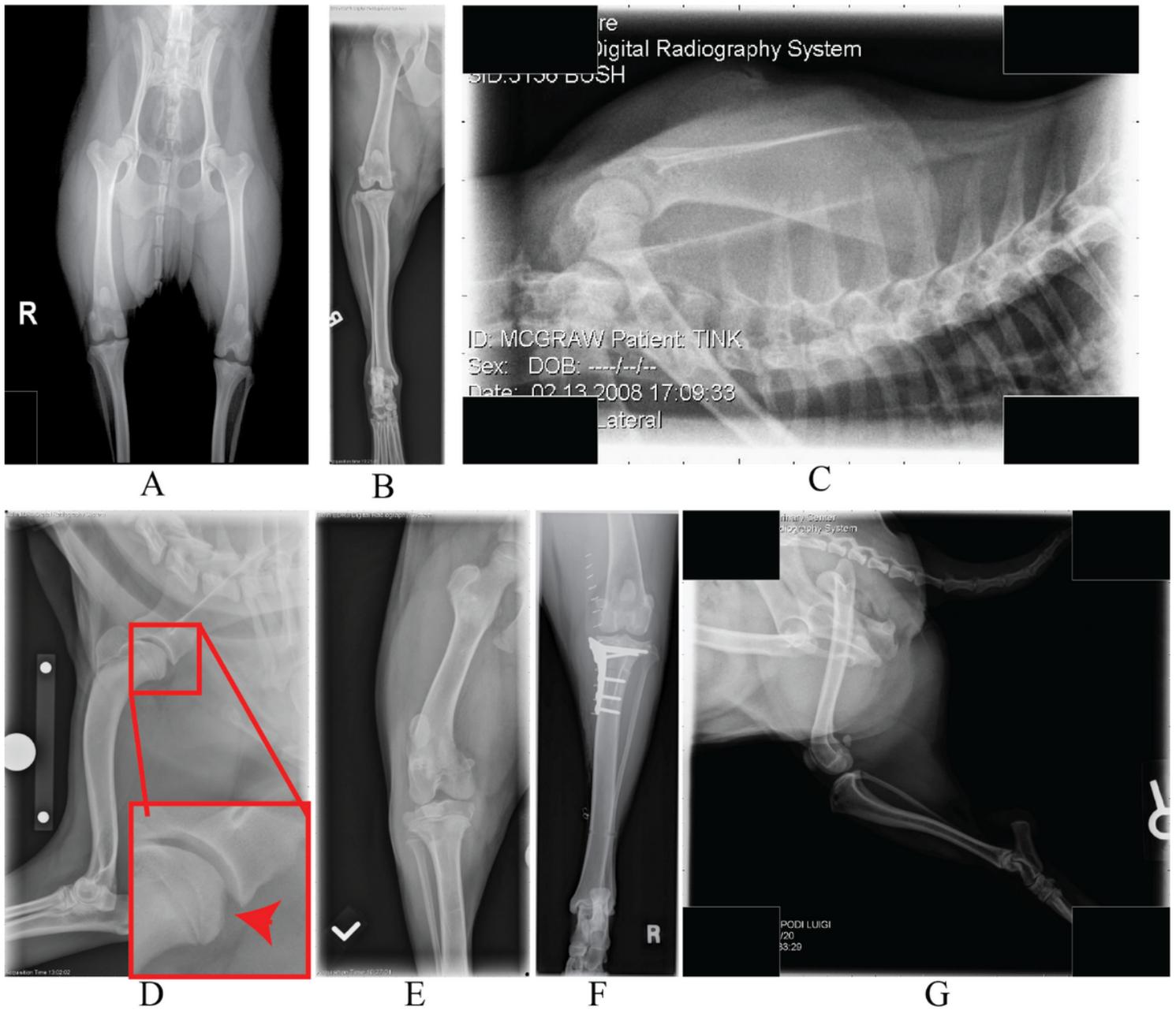
<b>Species</b>	<b>Dog</b>	<b>Cat</b>
No. of Animals	123	18
Mean age (y)	4.74 ± 4.22	5.28 ± 5.44
Gender		
No. of Male	68	13
No. of Female	50	4
No. of Unknown Gender	5	1
Note Mean data with ± standard deviation.		

Table 2

Caption: Number of Xs that are associated with different lesions, and number of Xs correctly interpreted by DF- and MVF-*i*/VetMSK-Xs.

Reported Lesion	No. of Xs	No. of Xs correctly interpreted by DF- iiVetMSK-Xs	No. of Xs correctly interpreted by MVF- iiVetMSK-Xs
Fracture	53	11	28
Destructive Lesion	50	8	34
DJD	43	8	27
Osteochondrosis	33	2	8
Hardware	30	11	25
Mixed Lesion	23	8	14
Elbow Dysplasia	20	5	13
Hypertrophic Osteodystrophy	20	1	12
Soft Tissue Swelling	19	4	9
Luxation	18	3	10
Deformity	15	1	8
Avulsion	9	0	3
Productive Lesion	6	0	1
Retained Cartilage Core	6	1	1
Soft Tissue Mass	6	2	0
Radial Hypoplasia	4	1	3
Bone Cyst	2	1	2
Note: DJD = Degenerative Joint Disease			

## Figures



**Figure 1**

A) ventrodorsal view of a dog showing pelvis, hip joints, humerus, knee joints, and proximal halves of tibia and fibula, there is no reported lesion; B) craniocaudal view of a dog showing right humerus, knee joint, tibia and fibula, tarsus and proximal metatarsals, there is no reported lesion; C) lateral view of a cat showing C6-7 and T1-12 vertebrae, ribs, scapula, shoulder joint, and proximal humerus; D) lateral view of a dog showing the shoulder joint with a defect in the humeral head (red arrowhead), it is most likely due to osteochondrosis; E) craniocaudal view of a dog showing the left knee joint, the patella is subluxated; F) craniocaudal view of a dog showing the right knee joint, the presence of hardware is due to tibial plateau leveling osteotomy; and G) lateral view of a cat pelvis showing the pubis, humerus, patella, tibia and fibula and tarsus, the hip joint is luxated.

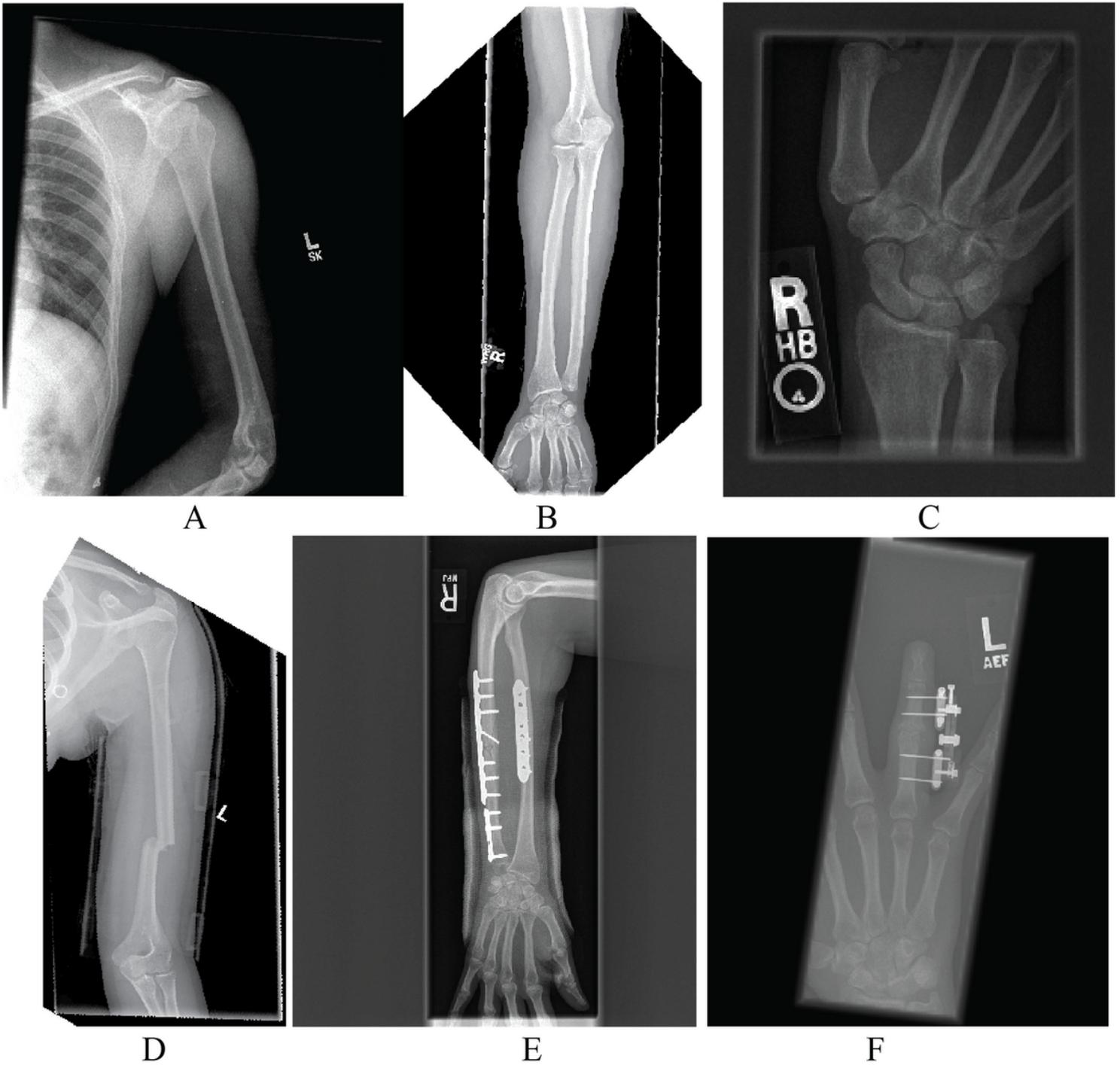
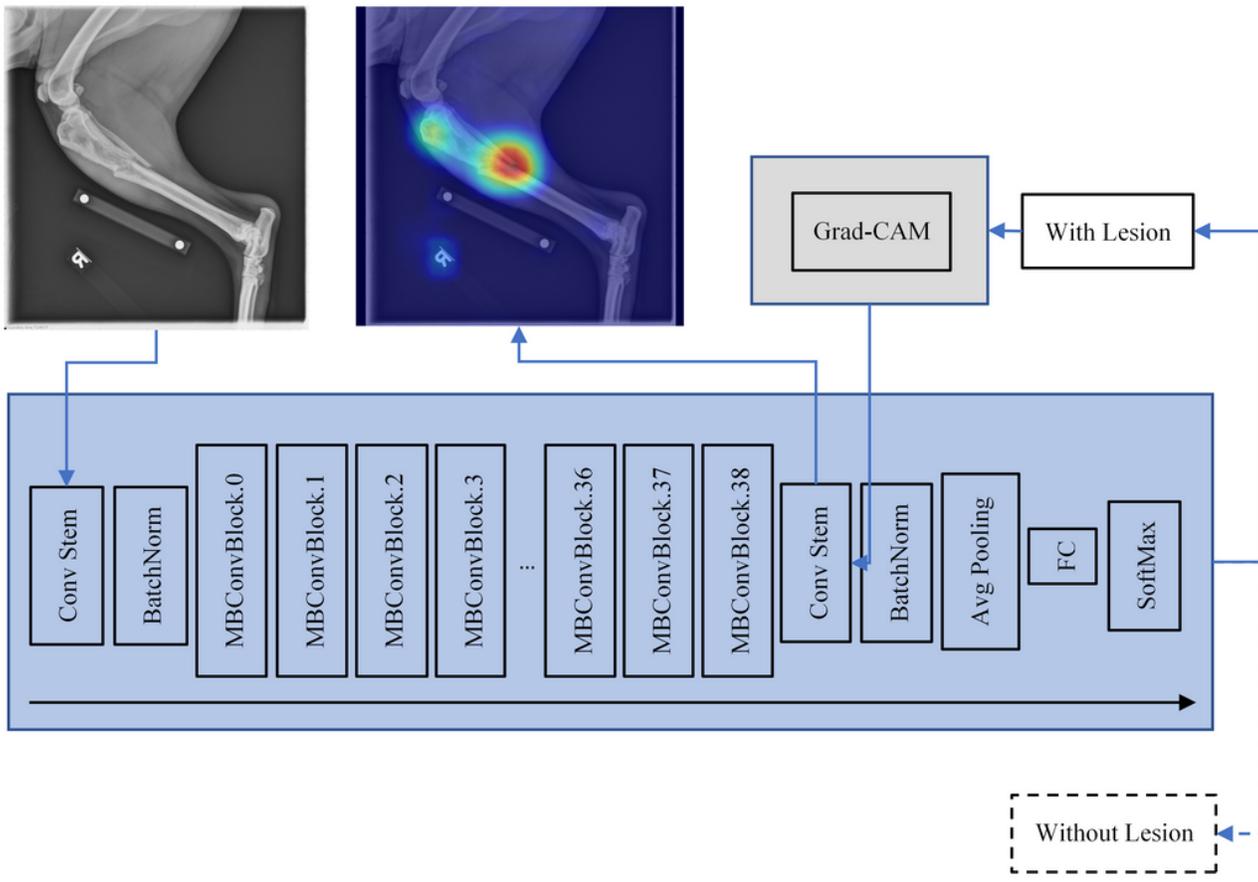


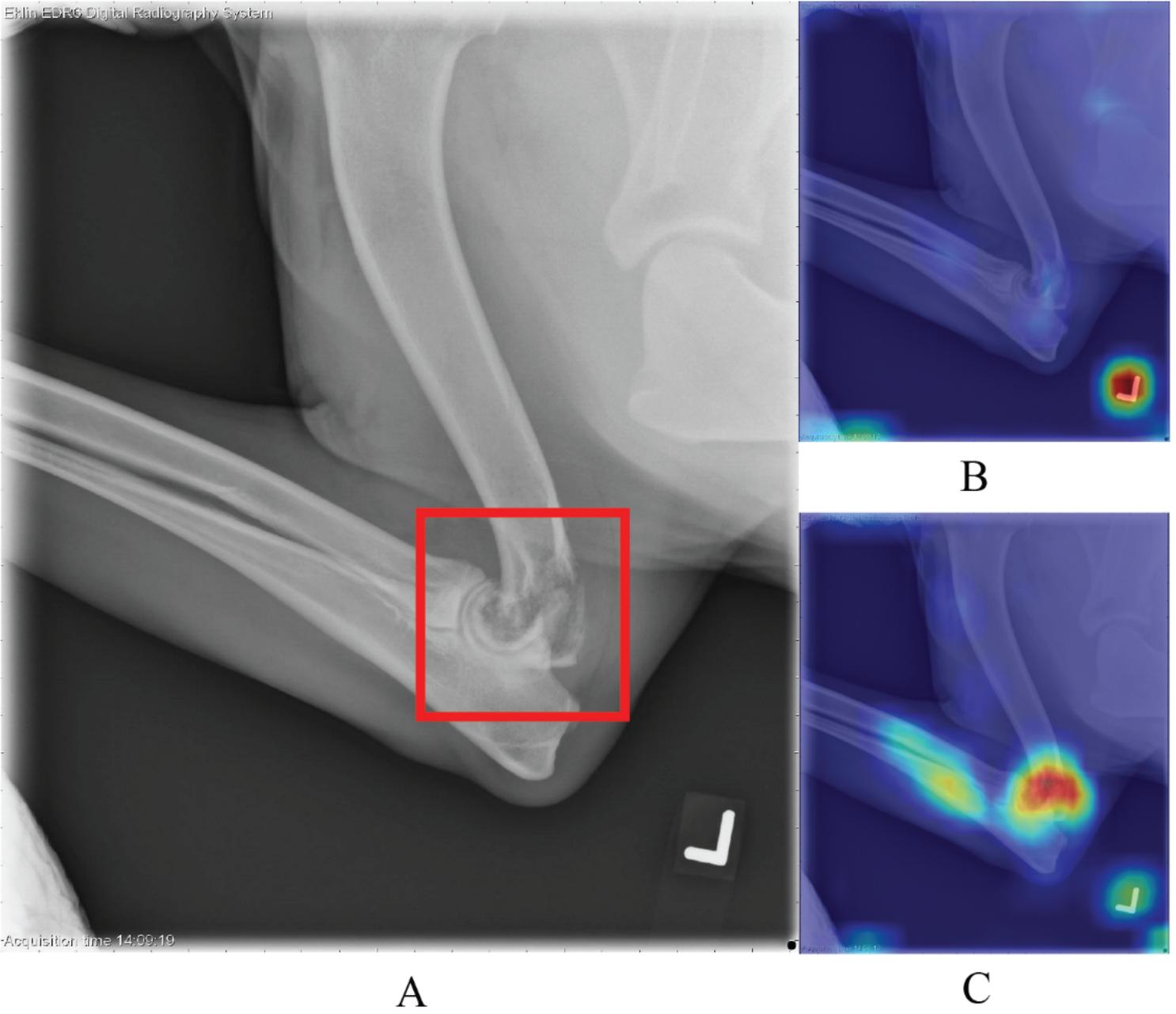
Figure 2

A-C) MSK Xs without lesion from MURA dataset; D) fractured left humerus; E) fractured right radius and ulna fixed with plates and pins; F) fractured finger fixed by an external device



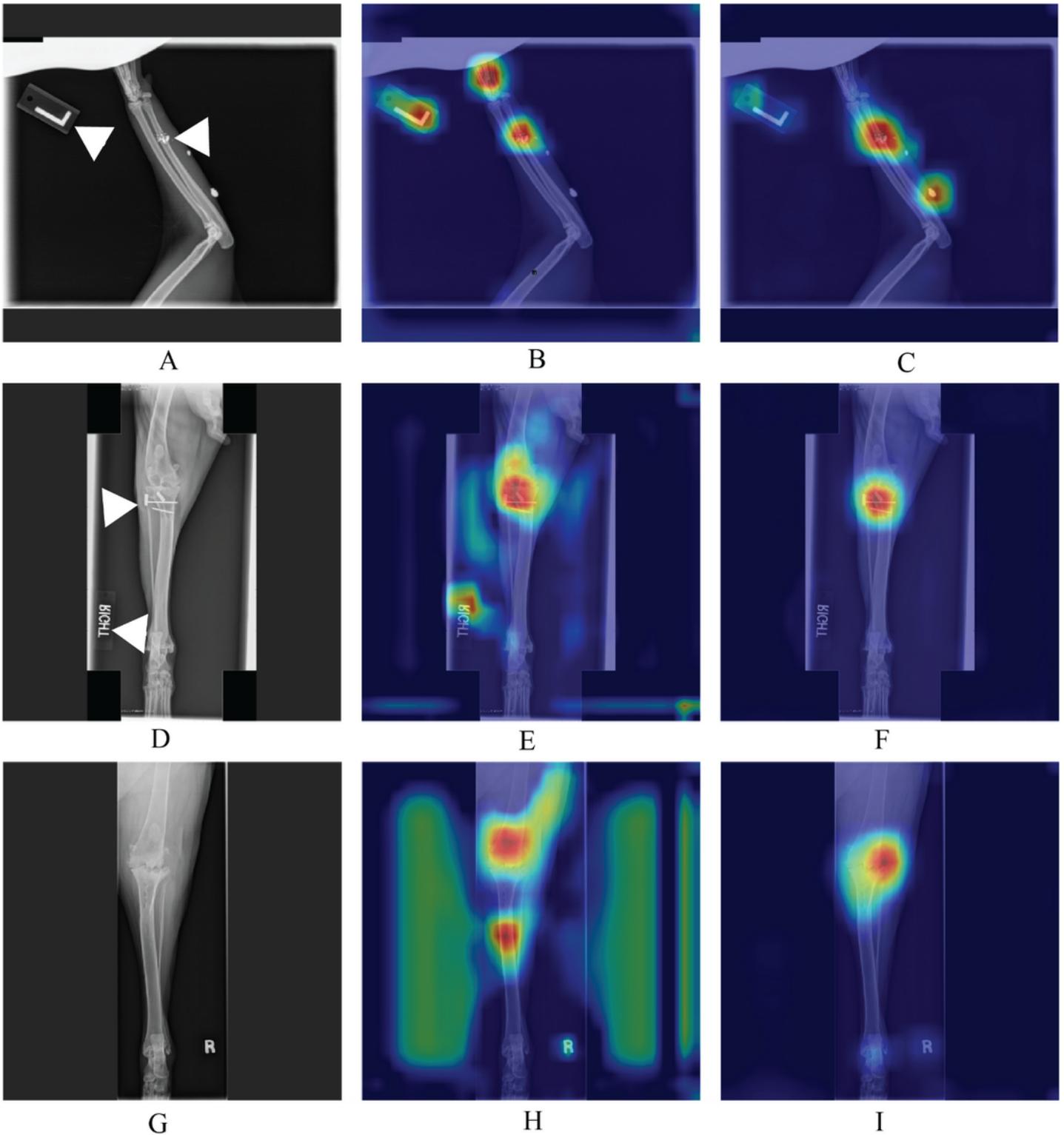
**Figure 3**

A schematic of the proposed iiVetMSX-Xs composed of the x-ray classification component based on EfficientNet and lesion localizer based on Grad-CAM. An input x-ray is first classified by the x-ray classifier. If the score of “with lesion” is higher than the score of “without lesion”, the lesion localizer generates a heatmap indicating where the lesion was detected.



**Figure 4**

Lateral view of the left thoracic limb of a dog, showing distal humerus, elbow joint, and proximal radius and ulna; there is osteolysis of the distal humerus with a fracture of the caudal cortex. The reverse 'L' indicates the markers to identify the limb. A) osteolysis (red square) of the humeral condyle; B) incorrect localization of lesion; the hot area is over the reverse "L" marker; and C) correct localization of lesion; the hot area is over the humeral condyle.



**Figure 5**

MSK Xs and lesion localization heatmaps generated by DF- and MVF- iiVetMSK-Xs. A) lateral view of a cat's fractured left thoracic limb due to gunshot and embedded bullet fragments (arrowheads). There is an "L" marker (arrow) to identify the limb.; B) The "L" marker, the normal metacarpus, and the fracture are detected as lesions by DF- iiVetMSK-Xs; C) The fracture and the radiopaque fragment are detected as "lesions" by MVF- iiVetMSK-Xs; D) caudocranial view of a dog's right pelvic limb with the presence of

hardware (arrowhead) and a marker spelling "RIGHT" (arrow); E) The implant and the marker are detected as "lesions" by DF- iiVetMSK-Xs; F) The implant is detected as "lesion" by MVF- iiVetMSK-Xs; G) craniocaudal view of a dog's right knee with evidence of DJD; H) The area surrounding the patella bone, part of the proximal tibia and the image pads on both sides are detected as "lesions" by DF- iiVetMSK-Xs; I) The stifle joint is detected as "lesions" by MVF- iiVetMSK-Xs.

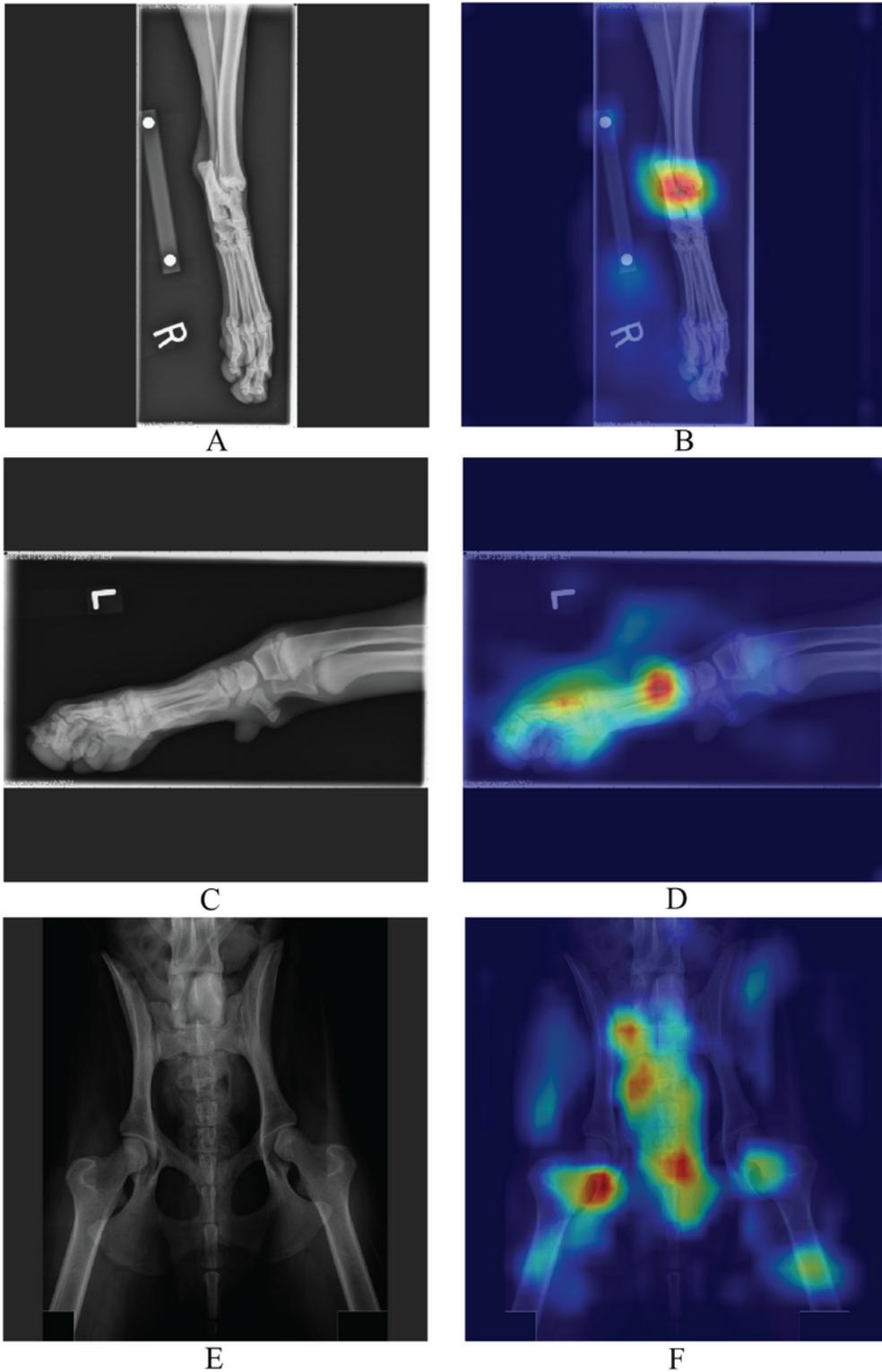


Figure 6

Xs without lesions that were falsely classified as with lesion by MVF- iiVetMSK-Xs, and lesion localization heatmap generated by MVF- iiVetMSK-Xs. A) dorsocranial view of a dog's right pes, there is no reported lesion; B) the normal tibiotarsal joint and the calcaneus are detected as a "lesion"; C) lateral view of a cat's left tarsus; D) the normal metatarsal bones and phalanges are detected as "lesions"; E) ventrodorsal view of a dog's pelvis, there is no reported lesion; F) the fecal materials along the spine, one of the femoroacetabular joints, and part of the other femur are detected as "lesions".