SEVC SOUTHERN EUROPEAN VETERINARY CONFERENCE CONGRESO NACIONAL AVEPA
BARCELONA 9-11 November 2023

ID: 03146 - Comunicación Científicas - Diagnostico por imagen

DEEP LEARNING PREDICTION FOR RADIOGRAPHIC SIGNS OF CANINE APPENDICULAR SKELETON: A COMPARISON STUDY WITH VETERINARIANS

Nam-Gyu Cho[1] , Sooyoung Choi[2] , Hee Chun Lee[3] , Tae Sung Hwang[3] , Jiyoung An[4] , Soeun Chang[4] , Hyounjung Kang[1] , Dongil Chang[4] , Hansu Cho[1] , Jongmin Lee[5] , Youngwon Lee[6]

[1] BioMedical AI Team, SK Telecom, Korea [2] College of Veterinary Medicine, Kangwon National University, Korea [3] Department of Veterinary Medical Imaging, College of Veterinary Medicine, Gyeongsang National University, Korea [4] DX Business Development Team, SK Telecom, Korea [5] Future R&D, SK Telecom, Korea [6] College of Veterinary Medicine, Chungnam National University, Korea

RESUMEN CORTO/RESUME

Objectives of the study
Diagnostic radiology in veterinary medicine using deep learning (DL) demonstrated effectiveness recently. Even in canine musculoskeletal diseases, radiographic detection using DL will be useful but has not been investigated yet. This study aims to demonstrate the efficacy of DL for radiographic detection of canine appendicular skeletal diseases by comparing the agreement between the DL model and veterinarians.

Material and Methods
A convolutional neural network (CNN) was trained with 7,473 canine appendicular skeletal radiographs. The radiographs were localized as a bound box of seven signs: infra-patella fat pad loss, fascial plane deviation, patella luxation, osteophyte and enthesophyte, popliteal lymph node enlargement, subluxation, and fracture. Additional 356 radiographs were annotated twice by groups A and B, consisting of each of four veterinarians, and used as the testing dataset of CNN.

Results
Group B had the average recall (AR) of 0.54 for the detected radiographic signs by group A, and vice versa of 0.56. CNN predicted with ARs of 0.79 and 0.78 for each detection of groups A and B and with ARs of 0.76 and 0.85 for all detection (A?B) and the consensus detections (A?B). Note that CNN prediction for the detections by veterinarians had higher ARs than groups A and B.

Conclusions
The DL showed an efficient prediction for veterinarians' detection on canine appendicular radiographs and will be helpful as an assistive tool in radiographic interpretation.

OBJETIVOS DEL TRABAJO / OBJECTIVES OF THE STUDY

The potential of deep learning (DL), a branch of artificial intelligence, has changed the research landscape in many fields, ranging from computer vision[1,2] and natural language processing[3,4] to healthcare and medicine[5,6], genomics[7], and even transportation and logistics[8]. The field of veterinary medicine has begun to experience the potential impacts of this technology[9,10,11,12,13,14]. Possible benefits of adopting DL include improved diagnostic

accuracy, enhanced efficiency, and the ability to handle a vast amount of data effortlessly. Diagnostic radiology, a crucial component of veterinary medicine, is one such area in which using DL demonstrated effectiveness recently[11,12,13,14].
However, its application in the radiographic detection of canine musculoskeletal diseases still needs to be explored. We focus on investigating the efficacy of DL in the radiographic detection of canine appendicular skeletal diseases. By comparing the agreement between a DL model and detections provided by experienced veterinarians, we aspire to highlight the potential of DL as an efficient and reliable assistant tool in detecting such diseases.

MATERIAL Y MÉTODO / MATERIAL AND METHODS

Ventrodorsal, dorsoventral, and lateral radiographic images of canine appendicular skeletons were retrospectively acquired from five national veterinary medicine universities in Korea. Twelve veterinary radiologists used the VIA annotator tool[15] to annotate these images. Within each radiograph, bounding boxes were drawn to localize and identify distinct eighteen radiographic signs: infra-patella fat pad loss, fascial plane deviation, patella luxation, osteophyte and enthesophyte, popliteal lymph node enlargement, subluxation, fracture, moth-eaten lysis, osteophyte, avascular femoral head necrosis, permeative lysis, spiculated periosteal reaction, amorphous periosteal reaction, cortical destruction, disk space narrowing, facet joint narrowing, intervertebral foramen opacity, and muscle atrophy. The annotation process took around eight months, including cross-review among veterinarians.
After analyzing the distribution of the collected data, it was decided to discard the last 11 signs (from moth-eaten lysis to muscle atrophy), whose sample size is less than a few hundred. Given the small sample size, it is likely that the DL model is overfitted to the samples, memorizing all the samples during training, thus generating inaccurate predictions when applied to test data.
Faster R-CNN[16], a Convolutional Neural Network (CNN) designed for object detection, is employed for training and analysis. This CNN operates by first detecting proposals and then classifying each proposal as one of the remaining seven signs (from infra-patella fat pad loss to fracture). The detection example is depicted in Figure 1 (see the attached file Detection_example.png). This CNN was trained with 7,473 radiographs, of which 5,229 were used for training and the remaining 2,244 for validation. Additional 356 radiographs were obtained from seven SKY Animal Medical Center branches in Korea, then annotated twice by groups A and B, in which each group consisted of four veterinarians, and used as the testing dataset of the CNN.

RESULTADOS / RESULTS

The average recall (AR)[17] was employed as an evaluation metric. The score implies the degree of agreement between a prediction (*x*) and the ground truth (*y*), and it ranges from zero, complete disagreement, to one, absolute agreement. Note that this metric is asymmetric. Six agreements were measured and summarized in Table 1: veterinarian-to-veterinarian agreements (B-to-A and A-to-B) and model-to-veterinarian agreements (the remaining four columns from CNN-to-A to CNN-to-A?B).
Group B had the AR of 0.56 for the detected radiographic signs by group A, and vice versa of 0.54. CNN predicted with ARs of 0.79 and 0.78 for each detection of groups A and B and with ARs of 0.76 and 0.86 for all detection (A?B) and the consensus detections (A?B). Note that CNN prediction for the detections by veterinarians had higher ARs than groups A and B.

CONCLUSIONES / CONCLUSIONS

The disparity between CNN's AR and those of the two veterinary groups emphasizes the potential advantages of using DL. One might wonder how a computer model can outperform trained, experienced veterinarians in interpreting radiographs. Our interpretation is that the answer lies in the fundamental nature of DL. In other words, in this study, each radiograph with annotations contains the collective knowledge and judgment made by professionals. The aggregation of diverse knowledge enables the DL model to recognize patterns that might escape the human eye. Furthermore, unlike humans, DL models are unaffected by fatigue, cognitive biases, or other factors influencing human judgment. Thus, a well-trained model can consistently generate more valid diagnostic candidates than an individual or a group of individuals.

It is important to stress that these findings do not diminish the essential role of veterinary professionals. Instead, they demonstrate the potential of DL technologies to augment veterinary expertise and improve diagnostic outcomes. Therefore, incorporating DL techniques could help the field of veterinary radiology, serving as an efficient tool in enhancing the precision of diagnoses, thereby leading to improved patient outcomes.

BIBLIOGRAFÍA /BIBLIOGRAPHY

1. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton: ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 2012; 25.
2. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov et al: An image is worth 16x16 words: transformers for image recognition at scale. International Conference on Learning Representations 2021.
3. Ashish Vaswani, Noam Shazeer, Niki Parmar et al: Attention is all you need. Advances in Neural Information Processing Systems 2017; 30.
4. OpenAI: GPT-4 technical report. arXiv:2303.08774v3 2023.
5. Andre Esteva, Alexandre Robicquet, Bharath Ramsundar et al: A guide to deep learning in healthcare. Nature Medicine 2019; 25: 24-29.
6. Pranav Rajpurkar, Emma Chen, Oishi Banerjee, Eric J. Topol: AI in health and medicine. Nature Medicine 2022; 28: 31-38.
7. Nicolae Sapoval, Amirali Aghazadeh, Michael G. Nute et al: Current progress and open challenges for applying deep learning across the biosciences. Nature Communications 2022; 13: 1728.
8. Nahid Parvez Farazi, Bo Zou, Tanvir Ahamed, Limon Barua: Deep reinforcement learning in transportation research: a review. Transportation Research Interdisciplinary Perspectives 2021; 11: 100425.
9. Pauline Ezanno, Sébastien Picault, Gaël Beaunée et al: Research perspectives on animal health in the era of artificial intelligence. Veterinary Research 2021; 52: 40.
10. Parminder S. Basran, Ryan B. Appleby: The unmet potential of artificial intelligence in veterinary medicine. American Journal of Veterinary Research 2022; 83 (5): 385-392.
11. Pereira, Ana Inês, Pedro Franco-Gonçalo et al: Artificial intelligence in veterinary imaging: an overview. Veterinary Sciences 2023; 10 (5): 320.
12. Léo Dumortier, Florent Guépin, Marie-Laure Delignette-Muller, Caroline Boulocher, Thomas Grenier: Deep learning in veterinary medicine, an approach based on CNN to detect pulmonary abnormalities from lateral thoracic radiographs in cats. Scientific Reports 2022; 12: 11418.
13. Emilie Boissady, Alois De La Comble, Xiajuan Zhu, Jonathan Abbott, Hespel Adrien-Maxence: Comparison of a deep learning algorithm vs. humans for vertebral heart scale measurements in cats and dogs shows a high degree of agreement among readers. Frontiers in Veterinary Science 2021; 8: 764570.
14. Peyman Tahghighi, Ryan B. Appleby, Nicole Norena, Eranga Ukwatta, Amin Komeili: Machine learning can appropriately classify the collimation of ventrodorsal and dorsoventral thoracic radiographic images of dogs and cats. American Journal of Veterinary Research 2023 (online).
15. Abhishek Dutta, Andrew Zisserman: The VIA annotation software for images, audio and video. Proceedings of the ACM International Conference on Multimedia 2019; 2276-2279.
16. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun: Faster R-CNN: towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems 2015; 28.
17. Tsung-Yi Lin, Michael Maire, Serge Belongie et al: Microsoft COCO: common objects in context. European Conference on Computer Vision 2014; 740-755.