



OPEN

Detecting medial patellar luxation with ensemble deep convolutional neural network based on a single rear view image of the hindlimb

Juho Jung^{1,2,4}, Geonwoo Park^{2,3,4} & Gwanghyeon Kim²✉

Medial patellar luxation (MPL) is a common orthopedic disease in dogs, which predisposes elderly and small-breed dogs. Unlike in humans, diagnosis in the early course of the disease is challenging because symptoms and joint-pain expression in canines are vague. Herein, we introduced a deep-learning system to diagnose MPL using a single rear-view hindlimb image. We believe that this is the first attempt to build a deep-learning system to diagnose MPL based on image analysis. Notably, 7689 images were collected from 2653 dogs in 30 private animal clinics between July 2021 and July 2022. Model performance was compared with ResNet50, VGG16, VGG19, Inception-V3, and veterinarian performance. For performance comparison, a professional veterinarian with >10 years of experience selected images of 25 normal dogs and 25 dogs with MPL. The proposed model showed the highest performance, with 92.5% accuracy, whereas human experts showed an average accuracy of 55.2%. Therefore, our model can diagnose MPL using only a single rear-view hindlimb image. Furthermore, to solve the image uncertainty caused by the input image noise, we used a one-class SVM and ensemble learning methods to ensure model robustness. Our study will help diagnose MPL in clinical settings using a single rear-view hindlimb image.

Patellar luxation (PL) is a common orthopedic condition in dogs that can cause lameness and secondary degenerative joint disease (DJD)^{1–3}. PL occurs in medial and lateral directions (MPL and LPL, respectively) and is graded for severity on a 4-point scale. MPL is more common than LPL, with over 90% of small-breed dogs diagnosed with MPL⁴. The incidence of MPL was reported to be 12 times higher in small dogs than in large dogs⁵. The breeds most affected by PL include Maltese, Poodle, and Pomeranian^{4,6}. Since PL is a progressive disease with a high potential to degrade the quality of life, it is of utmost importance to diagnose the disease early and take preventive measures³.

Physical examination and radiography are commonly used to diagnose PL, but these methods rely on the veterinarian's subjective evaluation⁷. The quadriceps angle, used to diagnose and grade PL, is measured between two lines connecting three anatomic sites: the line connecting the anterior superior iliac spine and midpoint of the patella and the line connecting the tibial tuber and midpoint of the patella⁸. This angle also represents the direction of the force of the quadriceps muscle⁹. The quadriceps angle is measured using diagnostic imaging methods, including radiography, computed tomography, and magnetic resonance imaging. Multiple studies have demonstrated a strong correlation between the quadriceps angle and PL, and the quadriceps angle is widely accepted as a reliable diagnostic index^{8–10}.

With advances in technology, canine PL has been successfully diagnosed using a region with convolutional neural networks (R-CNN) deep learning model and 2832 radiographic images of the canine stifle joint¹¹. However, this model requires radiographic images taken by veterinarians, which can only be obtained in veterinary clinics. In this study, we propose a deep-learning model that diagnoses PL using a single rear-view image of dogs without the need to visit veterinary clinics. To the best of our knowledge, this is the first deep-learning model to detect PL without diagnostic images. The performance of the model was compared with the diagnosis of veterinarians and gradient-weighted class activation mapping (Grad-CAM)¹² was applied to visualize the key areas with representative features that are utilized for the classification process.

¹Department of Applied Artificial Intelligence, Sungkyunkwan University, Seoul, Korea. ²Elevenliter Inc., Samuiwon Startup Center, 26 Kyunghee-Daero, Dongdaemun-Gu, Seoul, Korea. ³School of Computing, KAIST, Daejeon, Korea. ⁴These authors contributed equally: Juho Jung and Geonwoo Park. ✉email: official@elevenliter.com

Methods

Data collection and labeling

In this study, we enrolled 30 private animal clinics to collect images of dogs. From July 2021 to July 2022, the clinics collected 7689 rear-view images of the hindlimbs from 2563 dogs. The basic characteristics of all dogs are described in Table 1. There were three pictures of the hind limbs from three different angles for each dog. To label the dataset, veterinarians examined the dog's leg, touched the dogs' medial patella, and diagnosed the presence and severity of MPL. We provided guidelines to each clinic for high-quality images and allowed them to obtain hindlimb images more accurately. The guidelines contain the dogs' hip and leg locations, the distance from the camera, and brightness. In addition, information such as breed, age, weight, and presence of other diseases was also collected but not used in this study. In this study, no experiments were conducted on the animals, and hence, there is no legal issue of animal research ethics.

The data collectors captured pictures from three different angles for each dog. We enrolled two professional veterinarians who did not participate in the experiment to select one of the three pictures of each dog that showed the best MPL independently. The veterinarian excluded cases in which the MPL was not visible because of low image quality or excessive noise. Figure 1 shows images of normal and stages 1, 2, 3, and 4 MPL. The two veterinarians checked the quality of the images independently so they could double-check each other.

Data preprocessing

We filtered 2563 rear views of the hindlimb images from 7689 total images. Because CNN-based models take 224×224 RGB images as input¹³, we down sampled all images to 224×224 RGB images while considering their original shape ratio. By doing so, we aimed to prevent deformation and preserve the intended shapes of the original photos within the downsized images. To prevent overfitting, in which the deep learning model was focused on specific weights during the training process, we applied data augmentation to all training images. Data augmentation¹⁴ is a representative method for ensuring generality in the training process of a deep learning model.

Specifically, data augmentation included the following processes: (i) random horizontal/vertical flips, (ii) random affine with no rotation but randomly translation to left and right in the range of [0.1, 0.1] pixels (iii) color jitter with brightness: 0.5, contrast: 0.5, saturation: 0.5, hue: 0.5. Data augmentation was used only in the training process, and only images without any transformations were used in the testing model.

Model architecture

The proposed model comprised (i) a dog-specific discriminator, (ii) a dog-specific encoder, and (iii) an ensemble classifier. Figure 2 shows the architecture of the proposed model.

	Dog, N	Image, n	Average age	Average weight (kg)	Most popular dog type (Top 3)	Previous PL* diagnosis	
						Yes	No
Total	2563	7689	5.52	5.59	1. Maltese (607)	117	2446
					2. Poodle (400)		
					3. Pomeranian (278)		
Normal	1072	3216	5.2	5.95	1. Maltese (221)	36	1036
					2. Poodle (206)		
					3. Pomeranian (147)		
Suspected ^a	732	2196	6.28	4.89	1. Maltese (154)	29	703
					2. Pomeranian (106)		
					3. Poodle (82)		
Stage 1 MPL	212	234	6.89	4.7	1. Maltese (72)	16	196
					2. Poodle (44)		
					3. Bichon (16)		
Stage 2 MPL	500	1500	5.32	4.67	1. Maltese (146)	31	469
					2. Poodle (67)		
					3. Pomeranian (59)		
Stage 3 MPL	44	132	6.85	4.35	1. Maltese (14)	4	40
					2. Pomeranian (9)		
					3. Chihuahua (7)		
Stage 4 MPL	2	6	9	6.5	1. Mix breed (1)	1	1
					2. Pekingese (1)		
Unknown ^b	1	3	1	6.7	1. Toy poodle (1)	0	1

Table 1. Basic characteristics of all dogs. ^aSuspected: Close to Stage 1 MPL but not yet. ^bUnknown: inability to distinguish between normal and MPL. MPL, Medial Patellar Luxation. PL*: Patellar Luxation.

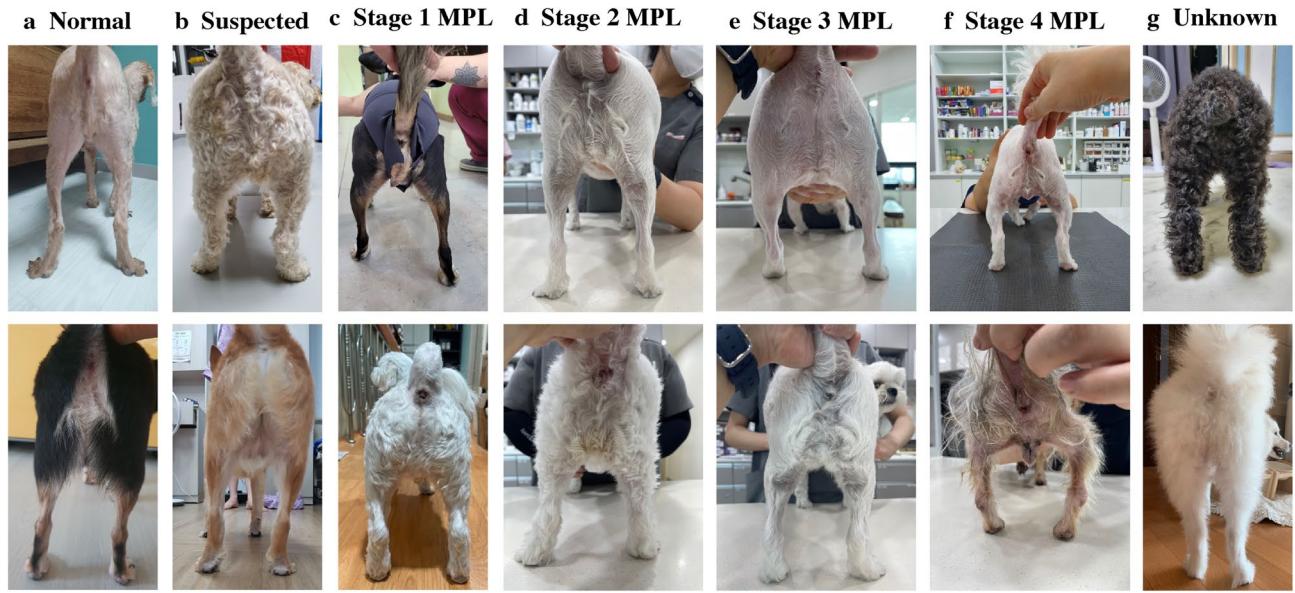


Figure 1. Representative cases of medial patellar luxation (MPL) for each Stage: Normal (a), MPL Suspected (b), Stage 1 MPL (c), Stage 2 MPL (d), Stage 3 MPL (e), Stage 4 MPL (f), Unknown (g). In MPL Suspected (B), compared to Normal (A) and Stage 1 MPL (C) Its bounded is unclear (2-Column).

Model architecture

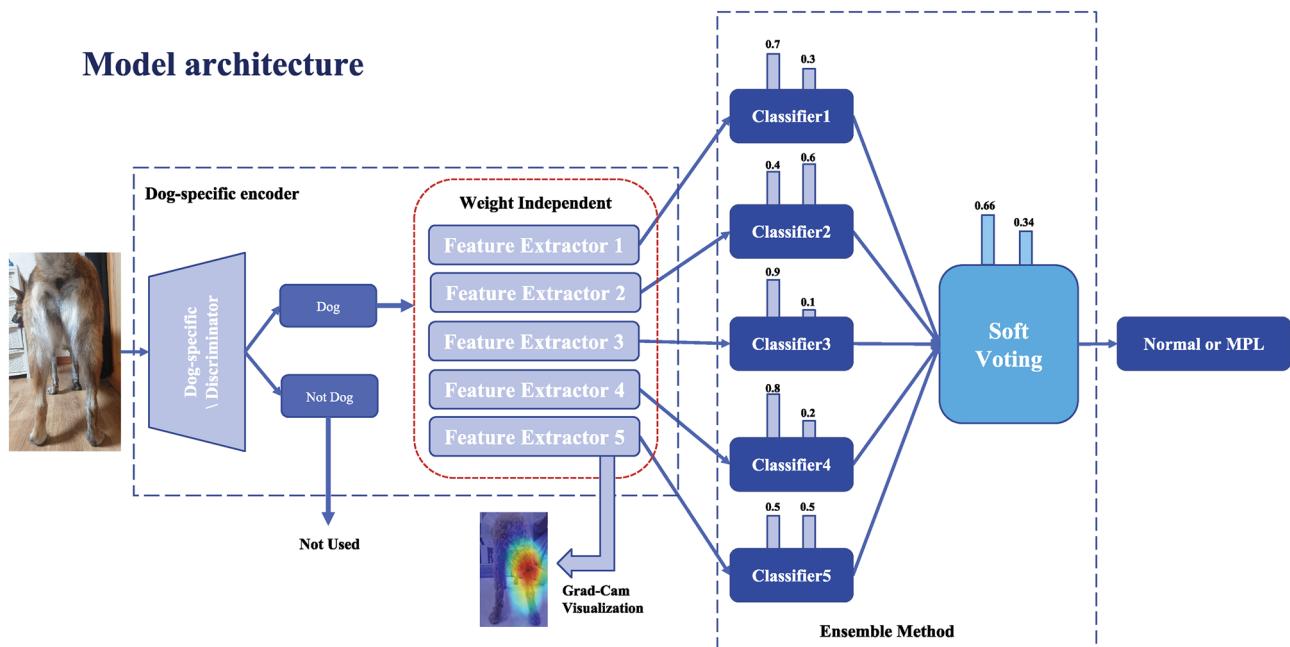


Figure 2. Architecture of the proposed Ensemble model based on ResNet18. Our proposed model consists of (i) a dog-specific discriminator, (ii) a dog-specific encoder (iii) an ensemble classifier. Specifically, a ResNet18-based dog-specific encoder consisted of an input layer, and 18 convolutional neural network layers with ReLU activation functions in five residual blocks with adaptive-pooling layers, an FC layer, and sigmoid. The final sigmoid activation function has been used to predict normal or Medial Patellar Luxation (MPL). (2-Column).

Since the input image may not be the rear-view image of the hindlimb or may be taken at an odd angle that can occur in single-image classification¹⁵, in the first stage, the dog-specific discriminator (Supplementary information 1) discriminates whether the input image is a dog's hindlimb. The dog-specific discriminator consists of a one-class support vector machine (SVM)¹⁶ placed in the front part of the model to perform a role in discriminating whether it is a dog. The one-class SVM discriminator process is as follows. First, representation features were extracted from an input image, and the dimension of features was reduced with the principal component analysis (PCA)¹⁷. PCA is one of the most widely used dimensionality reduction methods that can transform high-dimensional data into low-dimensional data while preserving the distribution of the original data as much as possible. Consequently, we reduced 50,176 raw features to 200 principal components for each image using

PCA. We then projected these 200 components to the vector space and found a decision line that could separate dogs and others. Moreover, by adjusting the threshold of the discriminators, our model selected high-quality dog images. Specifically, the discriminator filters out an image in which MPL cannot be distinguished owing to various noises and increases the accuracy and reliability of the model. This is enabled by the dog-specific discriminator, which identifies images as dogs. Images identified as dogs by this discriminator are then directed to the second stage, where they are processed by the feature extractor to extract image features.

In the second stage of the architecture, we built a dog-specific encoder with a Resnet18-based encoder that diagnoses MPL using a single rear-view image of the hindlimb. As shown in Table 2, we chose resnet18¹⁸ among other well-known CNN models as the baseline model because its accuracy was the best. Interestingly, in the CNN model, the performance of the model improved as the layer became deep; however, when the layer became too deep, the performance degenerated. This phenomenon is called the gradient-vanishing problem¹⁹. ResNet is a well-known deep neural network that can solve the vanishing gradient problem through residual learning using skip connections. This advantage of resnet18, with few parameters, demonstrated the best performance in the monotonous dog hindlimb image with residual learning. As shown in Table 2, the accuracy of ResNet18 was higher than ResNet 50, which had a deeper layer depth. As a result, our dog-specific encoder comprised 18 CNN layers in five residual blocks, followed by a rectified linear unit (ReLU). In addition, we used an adaptive pooling layer²⁰ as the last pooling layer, which automatically adjusted the stride and kernel size to the input feature map. We applied the transfer learning method²¹ to reduce overfitting and to train the model faster. Specifically, we initialized 18 convolutional layers with pre-trained weights from the ImageNet dataset²². Because the rear-view image of the dog's hind limb has a simpler structure than the other images and has a similar shape, fewer layers and learning parameters performed well. Consequently, our proposed model has 11,177,538 trainable parameters.

Finally, we applied the ensemble method to the model's training process to overcome image uncertainty. As shown in Fig. 2 we built five weight-independent classifiers, and each classifier made independent predictions with the same input image. The model made the final decision by soft voting on these five predictions. Soft voting²³ selects the final predicted probability by averaging the prediction probabilities of the five classifiers for each class. Thus, the model can make fair decisions without bias or overfitting. It also supplements the need for more information on a single image.

Grad-CAM visualization

We used Grad-Cam¹² to visualize the region recognized as important in the image of the dog's hind limb. The Grad-cam extracts representative features from the last layer output of the model and shows their importance using a heatmap. Interestingly, the regions represented by this heatmap are those that the model recognized as important in decision-making. We calculated the activated part using the gradient of the feature map that came out through the CNN layer. The resulting heatmap highlights the area of the image used in the classification of the model.

Experiment setting

We performed fivefold cross-validation²⁴ to train and evaluate the model. We randomly divided the dataset into five folds, trained the model four times, and tested the model on the remaining fold. This is because even if we shuffled data and randomly split it into five folds, certain class data can be gathered in one fold due to statistical probability. We trained our model using an ensemble learning method for each of the five folds. The average of the five test results obtained through the five classification validation processes was evaluated as the final performance of the model. We divided the data by dogs such that the same dog was not duplicated in the training and test data. To compare our model with other CNN models, we performed the same five-fold cross-validation for these CNN models with the same five data folds. All models, including our proposed model, were trained using a batch size of 32, epochs of 200, utilizing the Adam optimizer with a learning rate of 0.001.

To compare our model with that of five veterinarians, we selected 25 images of normal dogs and 25 images of dogs with MPL that were not used for training and testing the model. The data collector selected 50 images, and a professional veterinarian who did not participate in the experiment verified the image.

Statistical analyses

Cohen's kappa coefficient was used to rate the agreement level between the two veterinarians. We used a well-known Python library, Scikit-learn, to perform statistical analysis.

	Accuracy	Precision	Recall	F1-Score
Model	0.925	0.930	0.940	0.929
ResNet18	0.910	0.905	0.905	0.910
ResNet50	0.885	0.897	0.900	0.885
VGG16	0.870	0.830	0.810	0.870
VGG19	0.860	0.875	0.880	0.860
Inception_V3	0.905	0.900	0.900	0.905

Table 2. Performance comparison between the proposed model and CNN model.

Calculation

Performance of the proposed model was evaluated using the following metrics: accuracy, recall (also called sensitivity), F1 score, and precision (also called positive predictive value [PPV]). These metrics were calculated as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{F1 score} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \quad (4)$$

where TP is the total number of images of dogs correctly diagnosed with MPL, FP is the total number of images of normal dogs incorrectly categorized as having MPL, FN is the total number of images of MPL that were improperly identified as normal, and TN is the total number of images of normal dogs that were misdiagnosed.

Results

We contacted 30 private animal clinics and collected 7689 images of 2563 dogs from three different angles for each dog. The average age of the dogs was 5.52 years, and the average weight was 5.59 kg. Table 1 shows the characteristics of all dogs. We hired an experienced veterinarian to select and verify the best-quality image among the three pictures for each dog. Moreover, low-quality images were excluded. Consequently, 1072 normal dogs and 758 dogs with MPL were used to train and test the model.

Model performance

Table 2 lists the accuracy of the proposed model and well-known CNN model. To compare the performance of the models, we utilized 1630 images from 972 normal dogs and 658 dogs with MPL for training and utilized 200 images from 100 normal dogs and 100 MPL dogs for testing the models. Our model, based on ResNet18, showed a 92.5% accuracy in detecting MPL. Its performance was better than that of VGG16 (87%), VGG19 (86%), ResNet50 (88.5%), and Inception-V3 (90.5%). Our model showed the highest accuracy among the well-known CNN-based models.

Performance comparison with veterinarians

Table 3 shows the five veterinarians and the accuracy of the proposed model. The five professional veterinarians we recruited had an accuracy of 50–60% in diagnosing MPL by looking at only a single image. The average accuracy of these veterinarians was 55.2%, whereas that of our model was 90%. To conduct a fair experiment between our model and veterinarians, we selected 25 images of normal dogs and 25 images of dogs with the best quality for the comparison experiments, which were not used for model training. Therefore, the higher accuracy of the proposed model, even on 50 images that both veterinarians and the model had never seen, confirmed that our model had a higher accuracy in diagnosing MPL than experienced veterinarians.

We analyzed the cases in which our model was correct and those in which it was not. Figure 3 shows the confusion matrix of our model and that of the five veterinarians. The average specificity and sensitivity for veterinarians were 71.2% and 39.2%, respectively. As shown in Fig. 3b, the high specificity and low sensitivity indicate that veterinarians diagnosed most cases as normal dogs. Conversely, as shown in Fig. 1a, our model never misclassified MPL as normal but only misdiagnosed normal dogs as having MPL.

Figure 4 shows five cases where our model made an incorrect prediction. Although the real answer was normal in all five cases, the model predicted all of them as MPL. As shown in Fig. 4e, the curvature of the legs is not

	Accuracy	Sensitivity	Specificity	Precision	F1-score
Model	0.90	1.00	0.80	0.83	0.91
V1	0.50	0.28	0.72	0.50	0.36
V2	0.52	0.44	0.60	0.52	0.48
V3	0.58	0.44	0.72	0.51	0.61
V4	0.56	0.36	0.76	0.60	0.45
V5	0.60	0.44	0.76	0.65	0.52
V average ^a	0.552	0.392	0.712	0.556	0.484

Table 3. Performance comparison between the proposed model and human experts. ^aV average denotes the average of five veterinarians. V, veterinarian.

Confusion Matrix

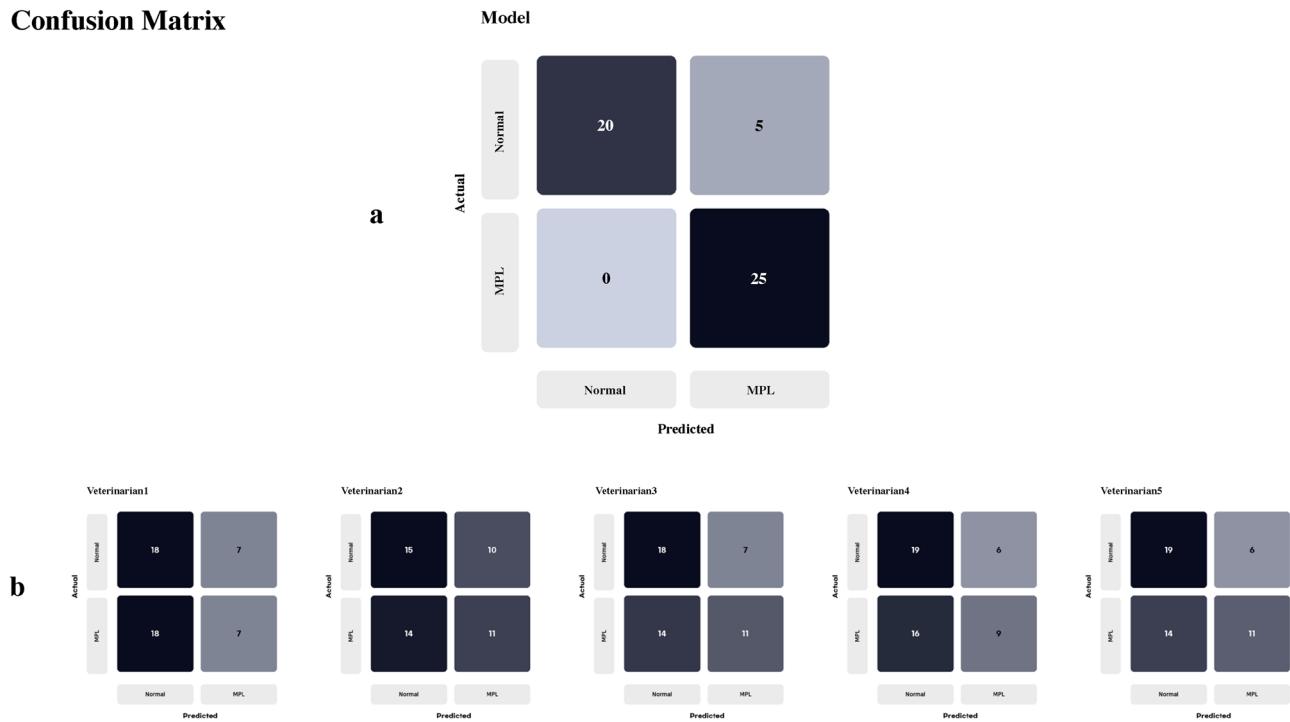


Figure 3. Comparison confusion matrix between the proposed model and the veterinarians. (a) Proposed model's confusion matrix. (b) Five veterinarians' confusion matrix. V1, 2, 3, 4, 5 denotes 5–10 years experienced Veterinarians. (2-Column).

	a	b	c	d	e
GT	Normal	Normal	Normal	Normal	Normal
Model	MPL	MPL	MPL	MPL	MPL
V1	Normal	Normal	MPL	Normal	Normal
V2	Normal	Normal	MPL	Normal	Normal
V3	Normal	MPL	Normal	Normal	MPL
V4	Normal	Normal	Normal	Normal	Normal
V5	Normal	Normal	MPL	Normal	Normal

Figure 4. Comparison of medial patellar luxation (MPL) detection between the proposed model and human experts. The figure displays five instances where the proposed model exhibited incorrect detection results. The ground-truth (GT) annotations and the five veterinarians' (V1–V5) expertise were used for benchmarking and comparison with human detection performance. (2-Column).

observed clearly because of the dog's fur. Additionally, as shown in Fig. 4a, b, and d, the dog's legs seemed bent even though the case is normal. These noises caused our model to predict incorrectly.

The Kappa coefficients between the first and fifth veterinarians and the third and fourth veterinarians were 0.39, 0.50, demonstrating moderate agreement ($P < 0.001$) in diagnosing MPL. In contrast, the Kappa coefficients between the third veterinarian and the model and the fifth veterinarian and the model were 0.17 and 0.14, respectively, demonstrating slight agreement ($P < 0.001$) in diagnosing MPL.

Visualization of model's decision-making process with Grad-Cam

Figure 5 shows a feature-based heat map provided by the gradient-weighted class activation mapping (Grad-CAM)¹¹. The area highlighted in the Grad-Cam image is the medial patella. When MPL occurs, the corresponding part is bent¹². This heat map confirms that our proposed model diagnoses MPL while observing the actual medial patellar location.

Discussion

We aimed to build a deep-learning model that accurately diagnoses MPL using a single rear-view image of the hind limb. Many previous studies have focused on improving the performance of deep neural networks by drawing object outlines or using segmentation algorithms²⁵. A representative study improved object detection accuracy by proposing an accurate edge detector using richer convolutional features (RCF) of images²⁶. Therefore, by using a segmentation algorithm, the performance of the proposed model will also be more accurate. However, our study is the first attempt to accurately discriminate the presence or absence of MPL, even with a single picture taken with a smartphone. Our model also showed higher accuracy than other well-known CNN models and professional veterinarians without using a segmentation algorithm. Following the experiment, we conducted a survey of veterinarians and found that less-experienced professional veterinarians who have never diagnosed MPL by examining a single image have lower accuracy compared to the proposed model.

There are two major image problems when developing a model for diagnosing MPL with a single rear view of the image. First, the input image could not be an image of a dog or of low quality; second, a single image could not contain enough information to detect the MPL. These problems are referred to as image uncertainty²⁷. To build a stable model against image uncertainty, we added a special dog-specific discriminator to the front of our deep-learning system and trained our model using the ensemble method. The dog-specific discriminator distinguishes dogs from backgrounds by tuning the threshold; thus, the model can focus on dogs specifically. To take a picture of a dog's hind limb, the person must hold the dog. Thus, the person or background is inevitably included in the image as a noise. Therefore, the discriminator automatically discriminates dogs from others; hence, only trainable images are fed into the model's input. Moreover, we used the ensemble method²⁸ to overcome the lack of information in a single image. The ensemble method yields more accurate and stable predictions by generating multiple classifiers and combining these predictions. A single model's prediction can easily overfit; however, the ensemble method can prevent overfitting by combining the independent predictions of several classifiers²⁹. We trained the proposed model using the ensemble method and developed a robust model that can overcome the uncertainty of a single image.

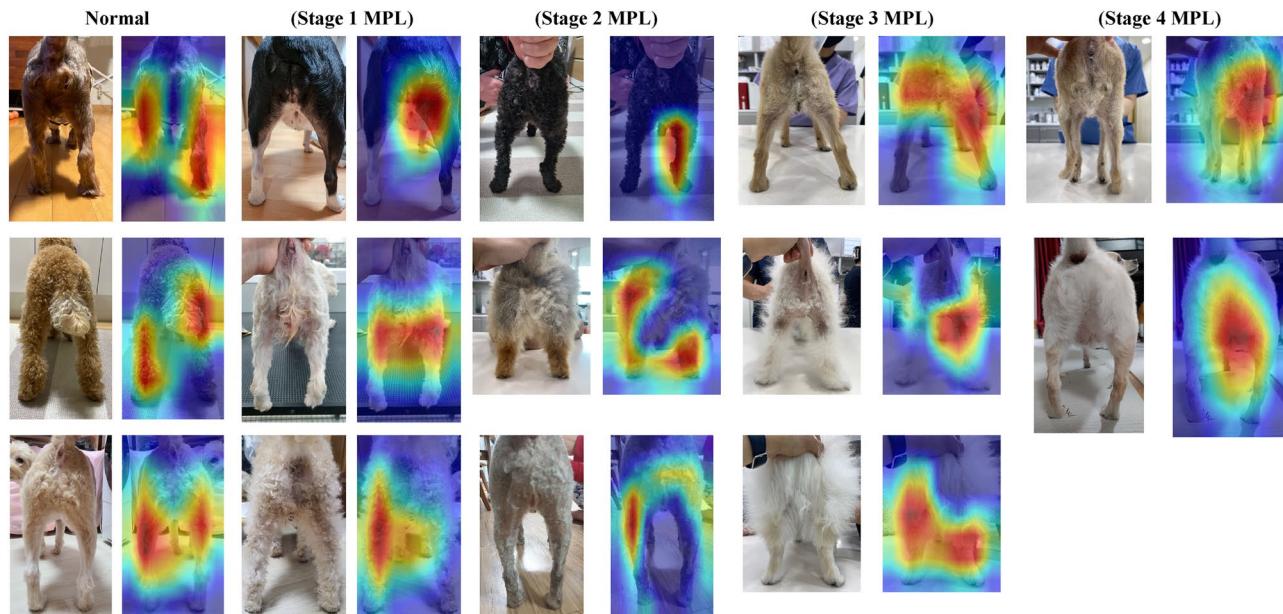


Figure 5. Heat maps for the proposed model's predictions by gradient weighted class activation mapping (Grad-CAM). Grad-cam calculates the representative region in the image while the training process. Grad-cam shows important parts of the model's decision-making process with a heat map. (2-Column).

Our study had several limitations. First, when diagnosing MPL, only a single image taken under strict standards was considered as the input of the model. We provided guidelines to the data collectors and let them take rear-view pictures of dog hindlimbs under the same conditions. The guidelines are as follows: when taking photos of the dog's hindlimb, one person should hold the dog, while another should take a horizontal picture of the dog's buttock, with the dog's buttocks, paws, kneecaps, and medial patella being visible. However, in the real world, people take photos of the dog's hindlimb from various angles, and the severity of MPL may differ depending on the angles and sights. Hence, diagnosing MPL using only a single image taken in the real world is limited. However, because our study used only images taken under specified conditions, the MPL was well revealed using only a single image. Second, we used only a single image formed in front of the hindlimb without considering all images taken from various angles. Thus, in our study, information on the MPL from multiple angles was not considered. When veterinarians diagnose MPL, they also examine the extent to which the patella is bent from various angles. The model would be more accurate if trained using multiple images taken from different angles. Thus, in future studies, we can develop a multi-image input model based on 3D-CNN³⁰, multi-instance model³¹, or attention model³² to increase the accuracy of MPL diagnosis by considering all images taken from various angles and creating a robust model. Third, noise exists in the image. Dogs have very different hair shapes depending on the breed. Because of this hair, there were complex cases for the model to determine the presence or absence of MPL. Figure 4 shows the incorrect decisions of the model for classifying hairy dogs. Based on a study of estimating the shape of the face by finding the key points of the face with a deep neural network³³, this problem can be solved by drawing the dog's outline or capturing the key points of the joints and calculating the bent patella angles. Lastly, MPL occurs more often in smaller and older dogs³⁴, and the model diagnoses MPL using only one rear-view image of the hindlimbs without considering other dog information. Figure 3 confirms that veterinarians cannot diagnose PL using just one hind leg image for diagnosing MPL. Moreover, we interviewed veterinarians after the experiments, and all the veterinarians answered that it was challenging to diagnose MPL by looking at only a single image because veterinarians had never diagnosed MPL by looking at only one image. Thus, as a follow-up study, we are trying to create a model for diagnosing patella dislocation considering the hind leg image and the dog's age, weight, sex, and breed. We expect the proposed model to be more robust by providing additional information as model input.

In summary, we created a deep learning system that diagnoses MPL with 90% accuracy using only a single image of the hind limb. We attached a dog-specific discriminator to the front of our model to discriminate the dog regardless of any image so that the model could distinguish whether the input image was a dog or not. In addition, to verify the robustness, we applied an ensemble learning method to solve various problems in diagnosing patellar dislocation using only a single image. The proposed model can support veterinarians in diagnosing patellar luxation.

Data availability

The data are not available for public access due to ethical restrictions and veterinary hospital privacy concerns, but are available from the corresponding author on reasonable request.

Received: 15 May 2023; Accepted: 29 September 2023

Published online: 10 October 2023

References

- Hulse, D. A. Pathophysiology and management of medial patellar luxation in the dog. *Vet. Med. Small Anim. Clin.* **76**, 43–51 (1981).
- Ness, M. G., Abercromby, R. H., May, C., Turner, B. M. & Carmichael, S. A survey of orthopaedic conditions in small animal veterinary practice in Britain. *Vet. Comp. Orthop. Traumatol.* **09**, 43–52 (1996).
- Johnson, A. L. *et al.* Comparison of trochlear block recession and trochlear wedge recession for canine patellar luxation using a cadaver model. *Vet. Surg.* **30**, 140–150 (2001).
- Alam, M. R. *et al.* Frequency and distribution of patellar luxation in dogs. 134 cases (2000 to 2005). *Vet. Comp. Orthop. Traumatol.* **20**, 59–64 (2007).
- Priester, W. A. Sex, size, and breed as risk factors in canine patellar luxation. *J. Am. Vet. Med. Assoc.* **4**, 633–636 (1972).
- O'Neill, D. G., Meeson, R. L., Sheridan, A., Church, D. B. & Brodbelt, D. C. The epidemiology of patellar luxation in dogs attending primary-care veterinary practices in England. *Canine Genet. Epidemiol.* **3**, 4 (2016).
- L'Epplattenier, H. & Montavon, P. Patellar luxation in dogs and cats: Management and prevention. *Compendium* **24**, 292–300 (2002).
- Kaiser, S. *et al.* Magnetic resonance measurements of the deviation of the angle of force generated by contraction of the quadriceps muscle in dogs with congenital patellar luxation. *Vet. Surg.* **30**, 552–558 (2001).
- Pinna, S. & Romagnoli, N. Radiographic measurement of the quadriceps angle in dogs. *PLOS ONE* **12**, e0185833 (2017).
- Towle, H. A. *et al.* Pre- and postoperative radiographic and computed tomographic evaluation of dogs with medial patellar luxation. *Vet. Surg.* **34**, 265–272 (2005).
- Shim, H. *et al.* Deep learning-based diagnosis of stifle joint diseases in dogs. *Vet. Radiol. Ultrasound* **64**, 113–122 (2023).
- Selvaraju, R. R. *et al.*, Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proc. of the IEEE International Conference on Computer Vision*, 618–626 (Venice, Italy, 2017)
- O'Shea, K. & Nash, R. An introduction to convolutional neural networks. Preprint at [arXiv:1511.08458](https://arxiv.org/abs/1511.08458) (2015).
- Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**, 1–48 (2019).
- Tian, Y., Zhong, Z., Ordonez, V., Kaiser, G. & Ray, B. Testing dnn image classifiers for confusion & bias errors. In: *Proc. of the ACM/IEEE 42nd International Conference on Software Engineering*, 304–305 (Seoul, South Korea, 2020).
- Manevitz, L. M. & Yousef, M. One-class SVMs for document classification. *J. Mach. Learn. Res.* **2**, 139–154 (2001).
- Jolliffe, I. T. & Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. A Math. Phys. Eng. Sci.* **374**, 20150202 (2016).
- He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (Las Vegas, NV, USA, 2016).
- Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. Preprint at [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014).

20. Liu, S., Qi, L., Qin, H., Shi, J. & Jia, J. Path aggregation network for instance segmentation. In: *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 8759–8768 (Salt Lake City, UT, USA, 2018).
21. Pan, S. J. & Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**, 1345–1359 (2010).
22. Deng, J. *et al.*, Imagenet: A large-scale hierarchical image database. In: *Proc. of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255 (Miami, FL, USA, 2009)
23. Wang, H., Yang, Y., Wang, H. & Chen, D. Soft-voting Clustering Ensemble. In *Multiple Classifier Systems MCS* (eds Zhou, Z. H. *et al.*) (Springer, 2013).
24. Payam, R., Tang, L. & Liu, H. Cross-Validation. In *Encyclopedia of Database Systems* (eds Liu, L. & Özsu, M. T.) (Springer, 2009).
25. Minaee, S. *et al.* Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 3523–3542 (2022).
26. Liu, Y. *et al.* Richer convolutional features for edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 1939–1946 (2019).
27. Wilson, R. & Granlund, G. H. The uncertainty principle in image processing. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 758–767 (1984).
28. Sagi, O. & Rokach, L. Ensemble learning: A survey. *WIREs Data Mining Knowl. Discov.* **8**, e1249 (2018).
29. Zhou, Z. H. *Ensemble Learning* (Springer, 2021).
30. Ji, S., Yang, M. & Yu, K. 3D convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 221–231 (2013).
31. Ramon, J. & De Raedt, L., Multi instance neural networks. In: *Proc. of the ICML-2000 workshop on attribute-value and relational learning*, 53–60 (Stanford, CA, 2000).
32. Vaswani, A. *et al.* Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30** (2017).
33. Sun, Y., Wang, X. & Tang, X. Deep convolutional network cascade for facial point detection. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 3476–3483 (Portland or, USA, 2013).
34. Di Dona, F., Della Valle, G. & Fatone, G. Patellar luxation in dogs. *Vet. Med. (Auckl)* **9**, 23–32 (2018).

Author contributions

J.J.: Conceptualization, Methodology, Software, Investigation, Resources, Writing-Original Draft, Supervision, Figures & Tables G.P.: Software, Formal analysis, Writing-Review & Editing, Visualization, Data Collection G.K.: Conceptualization, Validation, Data Collection & Curation, Writing-Review & Editing. All authors reviewed and modified the manuscript prior to submission.

Funding

This research did not receive any specific grant from funding agencies in the government, commercial, or not-for-profit sectors.

Competing interests

The authors declare the following financial interests/personal relationships, which may be considered as potential competing interests: Currently, Ju-ho Jung and Gun-woo Park work as freelancers in the company (Elevenliter, Seoul, Korea), and receive salary; however, their official affiliation is Sungkyunkwan University. Gwanghyeon Kim is the CEO of Elevenliter. No additional funding from the company or other sources was received for the writing of this thesis or the conduction of this project.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-43872-7>.

Correspondence and requests for materials should be addressed to G.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com