Pakistan Veterinary Journal

ISSN: 0253-8318 (PRINT), 2074-7764 (ONLINE)
DOI: 10.29261/pakvetj/2025.280

RESEARCH ARTICLE

Measurement of the Norberg Angle Using Artificial Intelligence in Diagnosing Canine Hip Dysplasia

M.K Yönez[1], U Alpman[1*], N.E Aslan[2], F.İ Bahar[3] and E. Alpman[4]

[1]Deparment of Surgery, Faculty of Veterinary Medicine, University of Erciyes, 38280, Kayseri, Turkey: [1]Deparment of Surgery, Faculty of Veterinary Medicine, University of Bozok, 66000, Yozgat, Turkey: [3]Faculty of Veterinary Medicine, University of Erciyes, 38280, Kayseri, Turkey: [4]Deparment of Mechanical Engineering, Faculty of Engineering, Marmara University, 34854, Istanbul, Turkey
*Corresponding author: umut_alpman@hotmail.com

ARTICLE HISTORY (25-775)

Received: August 04, 2025
Revised: September 15, 2025
Accepted: September 20, 2025
Published online: October 16, 2025

Key words:
Artificial
Canine
Hip Dysplasia
Intelligence
Norberg Angle

A B S T R A C T

Hip dysplasia in dogs is a developmental disorder caused by a lack of alignment between the acetabulum and femur. This disorder leads to an abnormal development of the hip joints between the caput femoris and the acetabulum. The study included Ventro-Dorsal pelvis radiographs of 2,306 dogs aged 12 months or older, without considering breed or sex. In the first part of the study, 2,306 radiographs in DICOM format were labelled both for the region encompassing the caput femoris and acetabulum and for four critical anatomical points (the centre of the right and left caput femoris and the dorsocranial projection of the right and left acetabulum). In the study, the performance of all configurations of the YOLOv8, YOLOv9, YOLOv10, YOLOv11, and YOLOv12 object detection algorithms was systematically evaluated in terms of precision, recall, mAP@50, and mAP50-95 metrics. According to the evaluation results, YOLOv11 achieved the best performance on the mAP50-95 metric with a value of 0.95397, outperforming the other configurations. This was followed by YOLOv8 (0.95245), YOLOv9 (0.95183), YOLOv10 (0.95005), and YOLOv12 (0.93832), respectively. In the mAP@50 analysis, the ranking was identified as YOLOv8 (0.99409), YOLOv9 (0.99386), YOLOv11 (0.99355), YOLOv12 (0.99345), and YOLOv10 (0.9933). This study concludes that artificial intelligence is a reliable alternative for diagnosing hip dysplasia in dogs. It has been found to be a more practical and accurate diagnostic method.

To Cite This Article: Yönez MK, Alpman U, Aslan NE, Bahar FI and Alpman E, 2025. Measurement of the norberg angle using artificial intelligence in diagnosing canine hip dysplasia. Pak Vet J, 45(4): 1930-1937. http://dx.doi.org/10.29261/pakvetj/2025.280

INTRODUCTION

Hip dysplasia is a developmental disorder caused by a discrepancy between the acetabulum and the femoral head (Schachner and Lopez, 2015). An abnormal position of the hip joint causes secondary osteoarthritis (Akula *et al.,* 2022). Genetic and environmental factors form the basis of hip dysplasia (Ginja *et al.,* 2015; Schachner and Lopez, 2015). Hip dysplasia tends to occur most frequently in large-breed dogs that grow fast (Loureiro *et al.,* 2024).

In hip dysplasia, the cycle of degenerative joint disease and bone remodeling continues. Cartilage degeneration, thickening of the joint capsule, stretching or rupture of the femoral head ligament, proliferation of the dorsal acetabular rim, thickening of the femoral neck, and local muscle atrophy are characteristic features of advanced hip dysplasia. These disorders cause mechanical stress during kinematic, static and dynamic activities of the joint, resulting in significant functional impairment among the various tissues and structures that comprise the hip (Pinna *et al.,* 2022). Although different systems are used in the evaluation of hip dysplasia, radiographic diagnosis of the acetabulum, femoral head and its position in the acetabulum, femoral neck, joint space and Norberg angle (NA) is taken into account and is the gold standard (Schachner and Lopez, 2015). Looseness, instability, and inflammation in the joint result in hip joint degeneration, characterized by the formation of osteophytes and exostoses. Hip dysplasia is scored, graded, and evaluated based on these findings (Mikkola *et al.,* 2019; Aghapour *et al.,* 2023). However, achieving standardization in the assessment of hip dysplasia is quite difficult. Examinations must be rigorous, well-defined, and more objective (Santana *et al.,* 2021).

In dogs, a Ventro-Dorsal (VD) hip radiograph is taken, and hip dysplasia is assessed using the Norberg-Olsson angle. During bilateral pelvic images, the tarsal joints are rotated 15° medially, ensuring dorsal alignment of the patellae. When measuring the Norberg-Olsson angle, the centres of both femoral heads are marked. The Norberg angle is the angle formed by the line between the centres of both femurs and the line drawn from the femoral centres of both hips to the cranial acetabular margin (Comhaire and Schoonjans, 2011). Studies have reported that the Norberg-Olsson angle is a reliable method for diagnosing hip dysplasia and provides a guide for interpreting hip radiographs (Ajadi *et al.,* 2018; Klever *et al.,* 2020).

With technological advancements, animal models, innovative treatment methods, device developments, and artificial intelligence-assisted imaging systems are being used in the veterinary field (Choudhary 2025a). Artificial intelligence (AI)-supported diagnostic systems improve clinical examination, diagnosis, and treatment by analyzing large data sets, increasing accuracy, and detecting subtle anomalies and anatomical points (Choudhary *et al.,* 2025b). As AI enables computer programs to analyze large imaging datasets and perform tasks related to veterinary diagnostic imaging, it has the potential to help meet the growing need for radiologic services through workflow, quality control, and image interpretation applications (Fitzke *et al.,* 2021).

Loureiro *et al.* (2024) developed a three-stage deep learning-based system that can automatically determine and measure the femoral neck thickness index. They noted that there was no significant difference between the system's measurements and those of the experts. Gomes *et al.* (2021) used the Inception-V3 transfer model to classify hip dysplasia. The data obtained indicated a 75% accuracy rate, and AI was reported as an alternative method for diagnosing hip dysplasia. In the literature, it has been emphasized that AI is an important diagnostic method in the field of radiology, that it is a strong alternative with the time savings and standardization it provides, and that comprehensive studies are necessary (LeCun *et al.,* 2015; Vinicki *et al.,* 2018; Hennessey *et al.,* 2022). In this study, we aimed to develop a deep learning-based application to improve and standardize the diagnosis of hip dysplasia in dogs by analyzing Norberg-Olsson measurements using the YOLO (You Only Look Once) and HrNET (High Resolution Network) algorithms.

MATERIALS AND METHODS

The study evaluated VD radiographs of 2,306 dogs presented to the Surgery Department of Erciyes University's Faculty of Veterinary Medicine between 2011 and 2025. Breed and gender were not considered in the evaluated radiographs. Only pelvic VD radiographs from dogs older than 12 months were included in the study. Images from the acetabulum to the distal femur were selected from DICOM (Digital Imaging and Communications in Medicine) format, showing no pathological lesions. Due to the retrospective and observational nature of the study, ethics committee approval and owner consent were not sought.

Labelling: First, the region encompassing the femoral head and acetabulum, as well as four critical anatomical landmarks (the centre of the right and left femoral head and the dorsocranial processes of the right and left acetabulum), was identified on 2,306 radiographs. The MKY application, a software we developed and made openly available on the GitHub website, was used for labelling. After labelling, the radiographic images were randomly divided into groups: 70% (1614) for training, 15% (346) for validation, and 15% (346) for testing. The training and test images were used to assess accuracy during training, while the images in the validation dataset were used to determine the overall map50-90 and map50 values, as well as to evaluate the extent to which the points deviated from the expert-marked points. All measurements were performed by MKY, UA, and NEA. The datasets were prepared separately for training two different models using the YOLO and HrNET architectures. Intraclass correlation coefficient (ICC) analysis was used to assess measurement consistency among experts, with values ranging from 0 to 1. Bland-Altman analysis was also performed to determine whether there were systematic differences among experts and to determine the limits within which 95% of measurement differences fell.

Data Set Preparation: To prepare the dataset for YOLO training, the coordinates of the region encompassing the caput femoris and acetabulum were determined in radiographic images, and the labelled coordinates were converted to YOLO format. To prepare the HrNET dataset, the region encompassing the caput femoris and acetabulum in each radiographic image was cut using Python software, and the coordinates of the labelled points were calculated. This ensured that the coordinate information was not distorted and that the HrNET model focused solely on the relevant anatomical region.

Study Design: After the labelling process, the overall algorithm flow was performed in two stages. In the first stage, the pelvic region was approximated using YOLO, and in the second stage, an HrNET algorithm was developed to calculate the NA. The aim was to determine the model that most successfully detected the pelvis among the YOLO models (YOLOv8-v12). Because successfully labelling the four different points that will determine the NA is critical for successful pelvic detection, the HrNET structure was combined with the Binary Heatmap method to compare the accuracy of the points and angles (Fig. 1).

Training: Testing was performed using Windows 11, 16GB of RAM, and an NVIDIA GeForce GTX 1650Ti GPU. Model training was conducted in Google Colab utilizing 80GB of RAM and an A100 GPU. PyTorch (2.0.1) and CUDA (11.8).

Training of YOLO: The YOLO architecture was chosen for pelvic region detection due to its single-pass detection approach, real-time processing capability, and proven effectiveness in medical applications (Ragap *et al.,* 2024).

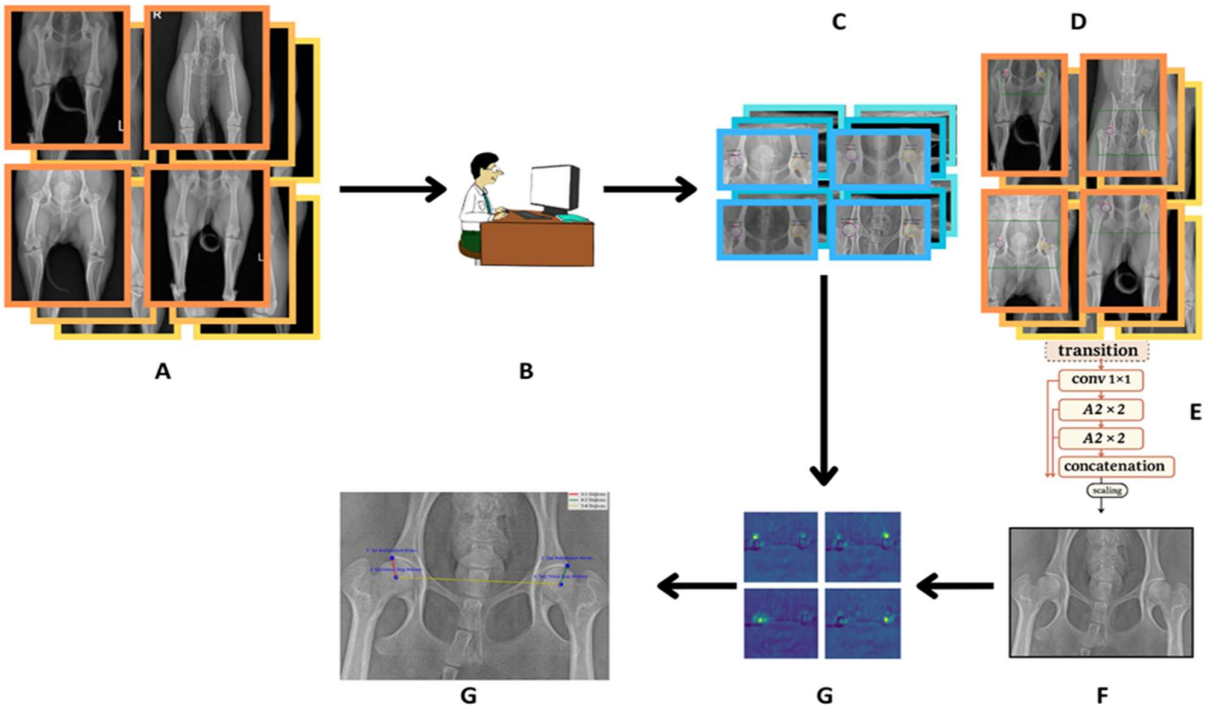

Fig 1: Flow diagram of the study.

All models were trained for 40 epochs to ensure adequate learning and adaptation. Batch size was set to 32, which balances memory usage and processing speed. Images were resized to 640x640 pixels. Model weights were updated using Stochastic Gradient Descent (SGD) optimization.

In the hyperparameter settings, the initial learning rate was 0.01, and the final learning rate was 0.001. Momentum was maintained at 0.937 across epochs, while weight reduction of 0.0005 prevented overfitting. The training process began with a 3-epoch warm-up phase using a momentum value of 0.8. The following parameters were used for data augmentation: hue shift (hsv_h=0.015), saturation increase (hsv_s=0.7), brightness change (hsv_v=0.4), horizontal flip (fliplr=0.5), and mosaic increase (mosaic=1.0). Additionally, for data set enrichment, images were subjected to ±3° random rotation, ±2° horizontal and vertical shearing, grayscale (hue value ±3°) for 4% of the images, saturation ±3°, brightness ±8°, and random noise addition up to 0.26% of the pixels. For geometric transformations, translation (translate=0.1) and scaling (scale=0.5) were used.

HRNet Training

HRNet Architecture: In this study, the HrNET-18 backbone architecture was chosen for the precise localization of four anatomical landmarks. This model was used to generate 128x128 heatmaps for each target point using a 512x512 input image. A 1x1 convolutional layer was added to the output of the HrNET-18 backbone, followed by a custom Binary Head Block containing Batch Normalization and ReLU activation functions. In the final layer, the 270-channel high-resolution feature map from the last layer of HrNET was converted to a 4-channel output (one channel for each anatomical landmark) using 1x1 convolution. Probability values in the range [0-1] were calculated for each pixel in the heatmaps using sigmoid activation following Zakareya *et al.* (2023).

Binary Heat Map: Instead of direct coordinate regression, heat maps were used to map anatomical landmarks. This approach enhanced localization accuracy by preserving spatial information. The Binary Heat Map function (Table 1) assigned a single positive pixel to each anatomical landmark. To increase the robustness of the model, noise was randomly assigned to neighbouring coordinates in proportion to the errors in the coordinates. This stochastic approach enabled controlled noise injection during the training process, strengthening the model's ability to detect artefacts that may occur during X-ray imaging.

Table 1: HRNet Binary Heat Map Parameters

Parameter	Value	Explanation
Heat Map Resolution	128x128	Pixel dimensions of the output Heat Map
Positive Pixel Value	1.0	Pixel value at anatomical point location
Negative Pixel Value	0.0	Value of background pixels
The Noise Injection Technique	Stochastic Rounding	Assigning a random neighbour coordinate proportional to the coordinate errors
Output Coordinate Calculation	Top-k (k=9)	Providing sub-pixel accuracy by taking the weighted average of the 9 pixels with the highest activation
Positive Class Weight	100	The weight coefficient given to positive pixels in the BCE loss function
Model Backbone	HRNet 18	-Network architecture that preserves high-resolution features
Feature Channels	270	Number of channels of HRNet-18 output tensor

Coordinate Extraction: A new approach was developed to obtain the final coordinates from the heat maps generated by the trained model. A colour scale ranging from white to red was used for Binary Heat Map point detection (Fig. 2B). The pixels with the highest activation were weighted, and coordinate extraction was performed at sub-pixel precision. Red regions were identified as indicators of high activation (Fig. 2C). The pixels with the highest k activation were evaluated, and their weighted averages were calculated using Softmax Scores as described by Liu (2025).

Loss Function: Implemented binary cross-entropy loss function for binary heat map, which addresses class imbalance using a positive weighting factor.

$$L(y,\hat{y}) = -\frac{1}{N}\sum_{i=1}^{N} \left[p \cdot y_i \cdot log\ log\ (\sigma(\hat{y}_i)) + (1-y_i) \cdot log\ log\ (1-\sigma(\hat{y}_i))\right]$$

Formula 1: Binary cross-entropy formula.

Here, $\boldsymbol{y_i}$ represents the true binary label, $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ notes the predicted logarithmic value, $\boldsymbol{\sigma}(\cdot)$ is the sigmoid activation function, and $\boldsymbol{p}$ refers to the positive class weighting factor. This approach, where the positive weight factor was set to 100, made the few anatomical point pixels in the heat map more evident than the background pixels. The AdamW algorithm was used for optimization, and a weight decay value of 0.01 was applied as the default. The model was trained for 34 epochs using a one-cycle policy with a maximum learning rate of 5e-4. Training was performed on an NVIDIA A100 GPU and lasted for 42 minutes. Data augmentation techniques were ±10° rotation, 5-15% scaling and ±10% brightness adjustments. Horizontal flip was intentionally disabled to preserve anatomical landmark symmetry.

RESULTS

Pre-training heat map analysis: The binary heat map generated before training revealed that the model had not yet fully learned the anatomical structures. The activation regions were broad and scattered. It was observed that the pelvic joint positions were unclear, and high activation values did not provide a precise localization. Instead of critical anatomical landmarks such as the acetabulum and femoral head, low-precision estimates were distributed throughout a wide area of the pelvis. This indicated that the untrained model had not yet acquired the ability to determine the precise locations of anatomical structures.

Post-training heat map and landmark detection: A significant improvement was observed in the heat map after training. The model was now able to detect each anatomical landmark (left and right acetabulum, left and right femoral head) with high accuracy. The activation regions became sharp and focused, with the red regions indicating high confidence, matching the exact locations of the target anatomical structures. Background activation was almost eliminated, leaving a concentrated activation pattern at relevant points (Fig. 3).

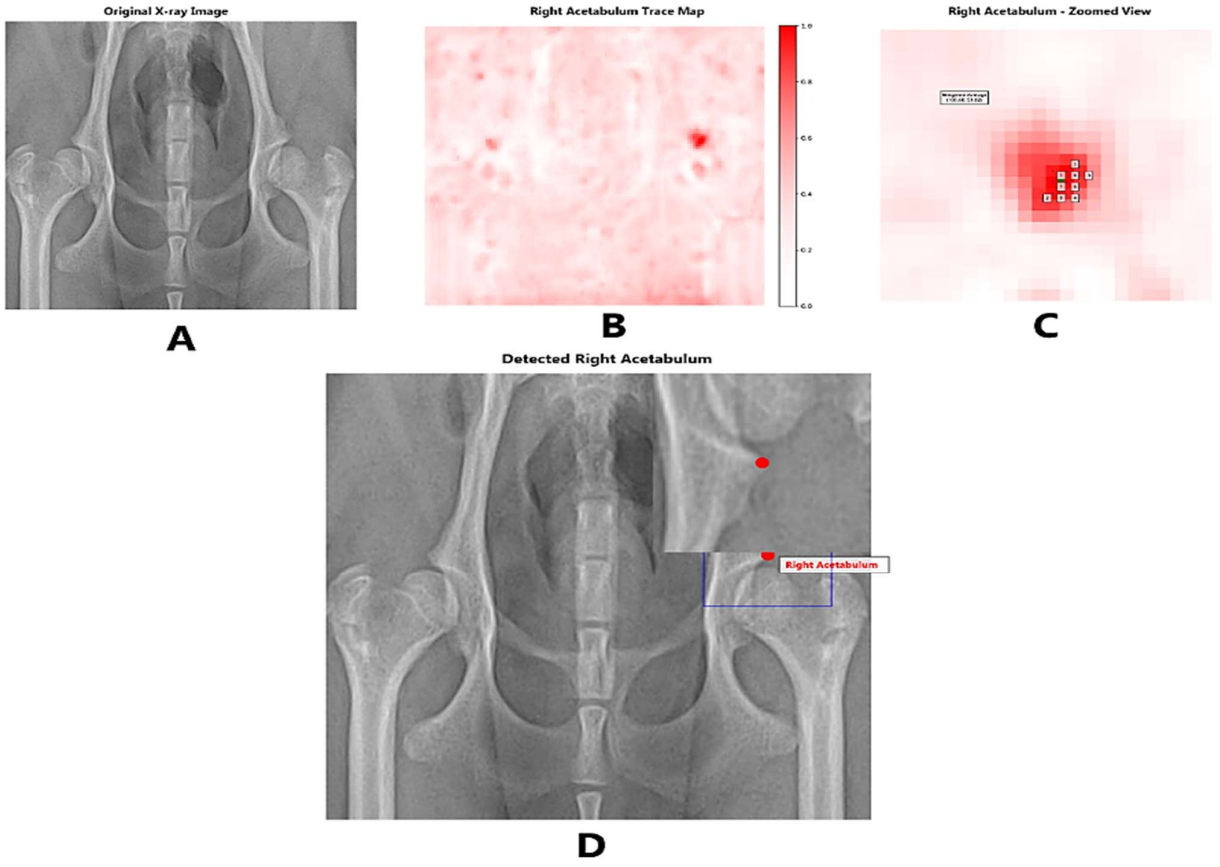

Fig. 2: Point detection on the pelvis radiograph A) Original VD radiograph B) Heat map generated for the right femur, C) Weighted average of the pixels with the highest activation, D) Marking of the detected points.

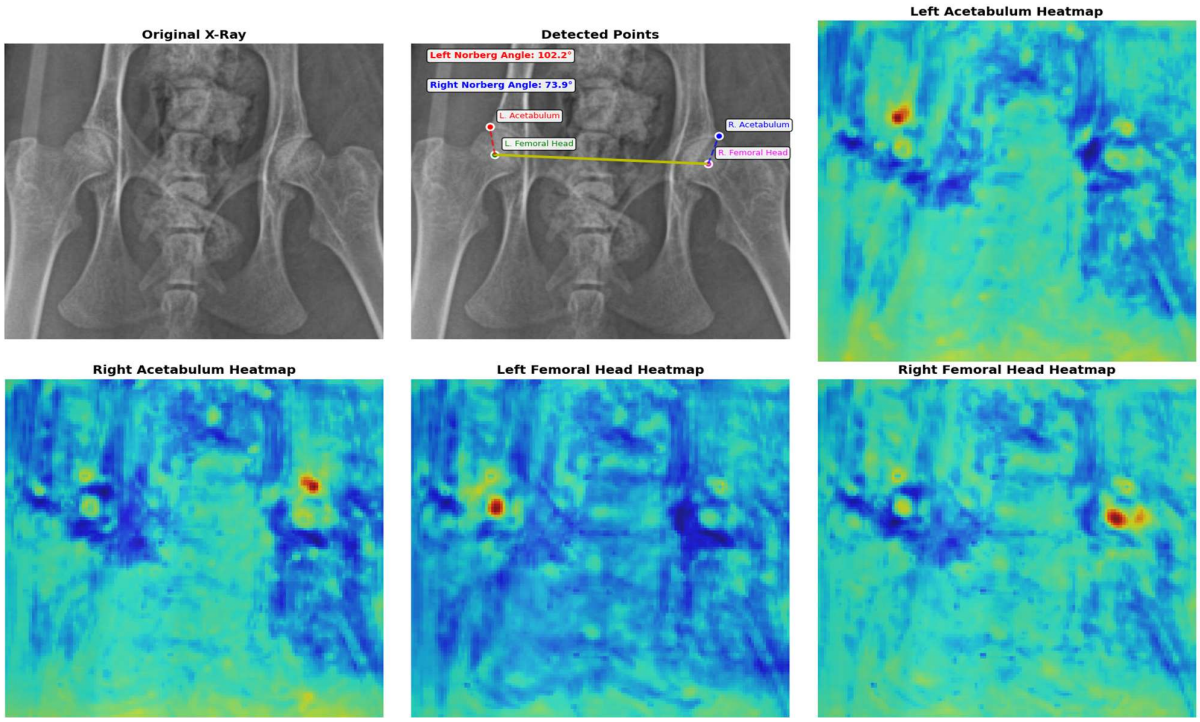

Fig. 3: Binary heat map after training.

Comparative evaluation of YOLO models: In the study, the performance of all configurations of the YOLOv8, YOLOv9, YOLOv10, YOLOv11, and YOLOv12 object detection algorithms was systematically evaluated in terms of precision, recall, mAP@50, and mAP50-95 metrics. According to the evaluation results, YOLOv11 achieved the best performance on the mAP50-95 metric with a value of 0.95397, outperforming the other configurations. This was followed by YOLOv8 (0.95245), YOLOv9 (0.95183), YOLOv10 (0.95005), and YOLOv12 (0.93832), respectively. In the mAP@50 analysis, the ranking was identified as YOLOv8 (0.99409), YOLOv9 (0.99386), YOLOv11 (0.99355), YOLOv12 (0.99345), and YOLOv10 (0.9933).

When examining the precision metric, the YOLOv11 and YOLOv12 configurations demonstrated superior accuracy performance with a value of 0.99174, followed closely by YOLOv8 and YOLOv9 at 0.99173. The YOLOv10 model reached a value of 0.99172. In terms of the recall metric, all models achieved an equal level of success with a notable identical score of 0.99174. Regarding computational efficiency, YOLOv8 exhibited the fastest performance with a training time of 40.4 minutes, followed by YOLOv11 (44.4 minutes), YOLOv9 (48.8 minutes), YOLOv10 (50.3 minutes), and YOLOv12 (56.1 minutes).

An analysis of the training processes of the models revealed that YOLOv9 had reached its optimal mAP50-95 value at an early stage, specifically at the 22nd epoch. At the same time, YOLOv10 and YOLOv12 achieved the same level at the 38th epoch. YOLOv8 reached its peak accuracy at the 31st epoch, and YOLOv11 at the 37th epoch. These findings highlight the correlation between model complexity and convergence speed, particularly demonstrating YOLOv9's rapid optimization capability. Additionally, the data indicated that despite its more complex architecture, YOLOv12 offered a lower validation performance-efficiency balance compared to the other models (Table 2).

Table 2: YOLO (You Only Look Once) Results

Metrics	YOLOv8	YOLOv9	YOLOv10	YOLOv11	YOLOv12
mAP50	0.99409	0.99386	0.9933	0.99355	0.99345
mAP50-95	0.95245	0.95183	0.95005	0.95397	0.93832
Precision	0.99173	0.99173	0.99172	0.99174	0.99174
Recall	0.99174	0.99174	0.99174	0.99174	0.99174
Training Duration(min)	40.4	48.8	50.3	44.4	56.1

HrNET-based landmark detection: Detection was performed using the HrNET model, which was trained by cropping 346 validation images from the YOLO v11 model. Thanks to the Binary Heat Map and high-resolution network structure, the model's average pixel error was determined to be 4.2866 pixels. The resulting RMSE (mean pixel error) value was 4.7851 pixels, indicating a consistent error distribution. The results of the detection process revealed that detecting the caput femoris was relatively more difficult than detecting the acetabulum (Table 3).

Table 3: Average pixel errors obtained for each anatomical point

Left Acetabulum	3.9561	pixel
Right Acetabulum	3.9699	pixel
Left Caput Femoris	4.7139	pixel
Right Caput Femoris	4.5066	pixel

Training process and loss analysis: The model showed a sudden decrease in loss values within the first three epochs (from 0.307212 to 0.008288). This indicated that our proposed binary heat map approach and HrNET architecture were able to learn anatomical landmarks quickly and effectively. The lowest validation loss value was recorded at 0.003585 in the 15th epoch. This epoch was found to be the step with the highest generalization ability of the model.

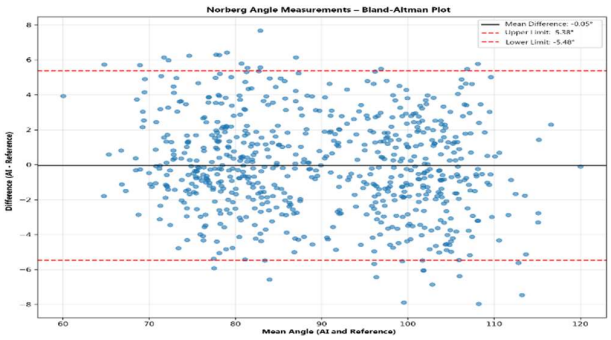

Fig. 4: Bland-Altman table.

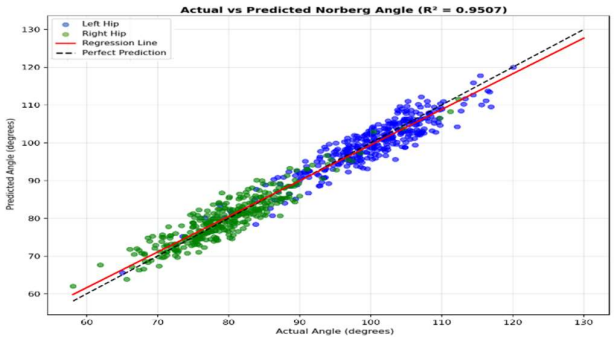

Fig. 5: Actual and predicted curve.

Validation of AI-estimated NA: The Bland-Altman plot (Fig. 4) confirmed the agreement between the artificial intelligence-estimated and manually measured NA. The mean difference was determined to be -0.05° (95% CI: -5.48° to 5.38°). This minimal deviation indicated that the AI algorithm did not over or underestimate the NA. The narrow limits of agreement (approximately ±5.4°) indicated high measurement precision across the entire range of observed angles (60°-120°). 95.1% of all measurement differences were observed to fall within the limits of agreement. The distribution of differences was symmetrical around the mean difference line, but a small number of outliers were found to lie beyond the limits of agreement. These outliers (n=17, 2.5% of all measurements) were observed at the extremes of the measurement range of the high (>110°) or low (<70°) NA.

Correlation between AI and manual NA measurements: A strong positive correlation (R^2=0.9507) was observed between the estimated and manually measured NA values. Regression analysis revealed consistent performance across the clinically relevant NA range (60°-130°). Right and left NA measurements (blue-green, respectively) showed similar distribution patterns along the line. (Fig. 5).

DISCUSSION

Current radiographic devices require manual labelling and verification of anatomical points by veterinarians for measuring the NA. To enhance the efficiency of this process and make it smoother and faster, this study aims to develop an AI application that can estimate the NA from radiographic images.

Artificial intelligence technologies are effective in processing large data sets and identifying complex patterns, which is critical for improving diagnostic accuracy and understanding disease mechanisms (Jacob *et al.,* 2022; Bouchemla *et al.,* 2023; Akinsulie, 2024). Without AI, many patterns may not be discovered because traditional methods struggle to analyze the large amounts of data generated in veterinary practice (Appleby and Basran, 2022; Kim *et al.,* 2022). By processing various data sources such as laboratory results and diagnostic images, it provides a better understanding of diseases and reduces the possibility of misdiagnosis (Albadrani, 2024). It has been observed that artificial intelligence in veterinary medicine is very effective in diagnostic imaging. Joslyn and Alexander (2022) reported that in research environments, it is possible to detect some radiographic abnormalities in veterinary radiology with AI. In the present study, radiographs of dogs were evaluated and analyzed to diagnose hip dysplasia with the support of artificial intelligence.

Lassalle *et al.* (2024) reported in their study evaluating hip dysplasia with AI that manual measurements can lead to errors and variability among observers. Consequently, they found that the use of AI has the potential to increase the reproducibility of measurements as part of comprehensive pelvic evaluations and thus improve diagnostic accuracy. Hansen *et al.* (2025) reported in their study that Bland-Altman plots showed good agreement between human and AI measurements when NA values were greater than 80°. Franco-Gonçalo *et al.* (2025) developed an AI algorithm that measures the distance between the center of the femoral head and the dorsal acetabular rim. This developed system has enabled more transparent and interpretable results in the assessment of hip dysplasia. In our study, the results obtained from the HrNET algorithm we used are consistent with those of Lassalle *et al.* (2024). The results obtained by using the YOLOv11 and HrNET algorithms together are consistent with the study by Hansen *et al.* (2025) in terms of minimizing inter-observer variability. Franco-Gonçalo *et al.* (2025) show parallels in terms of obtaining interpretable results in the evaluation of hip dysplasia.

In the evaluation of radiographs, factors such as the experience level and training, fatigue, lack of attention, or intense working tempo of the observer are important factors affecting the reliability of the evaluation. This subjectivity leads to inconsistencies between individual assessors (Pesapane *et al.,* 2024). Early diagnosis of hip dysplasia is critical for the breeding and genetic development of dog breeds. In the present study, NAs were measured using YOLO V8-9-10-11-12 and HrNET spine architecture in VD pelvis radiographs in dogs. Thus, by presenting an objective approach to evaluating hip dysplasia, an AI algorithm was developed that remains consistent regardless of the evaluator, provides fast and standardized results, and enables the evaluation of a large number of radiographs in a very short time period.

The primary difference between this study and other studies in the literature is that two distinct AI architectures are used in conjunction with each other. In this way, the shortcomings of the two algorithms are largely overcome. In the studies in the literature, either the YOLO algorithm alone (Den *et al.,* 2023) or the Convolutional Neural Network (CNN) algorithm alone (Boufenar *et al.,* 2023) is used. Within the scope of the presented study, six different methods were compared using YOLO v8-9-10-11-12 and HrNET architectures together.

AI is used in various ways to detect hip dysplasia (Boufenar *et al.,* 2023; Den *et al.,* 2023). Methods based on the detection of a specific anatomical landmark are more frequently used in AI. Classification of hip dysplasia is performed using acetabular index measurements (Sha *et al.,* 2023). In the presented study, NA was determined by measuring the angle between the centre of the femoral head and the dorsal rim of the acetabulum. The use of AI to identify anatomical landmarks on pelvic radiographs is similar to the study by Sha *et al.* (2023).

McEvoy *et al.* (2013) reported that they used the partial least squares discriminant analysis (PLS-DA) model to identify pelvic joints. They reported a classification error of 6.7%. In another study, the researchers scored hip dysplasia using the YOLOv3 Tiny NN application. They emphasized that it demonstrated 53% sensitivity and 92% specificity in the analysis of radiographs without hip dysplasia and radiographs showing dysplastic features (McEvoy *et al.* 2021). Similarly, Gomes *et al.* (2021) investigated hip dysplasia diagnosis on VD radiographs using an InceptionV3-based NN model. They reported that the model achieved 75% accuracy and that its test performance was adequate. The YOLO (8-9-10-11-12) algorithm was used in the presented study. The analysis of these algorithms compared the metrics of precision, recall, mAP@50, and mAP50-95. As a result, the YOLOv11 and YOLOv12 configurations showed superior accuracy performance (0.99174), with a mean difference of -0.05° (95% CI: -5.48° to 5.38°). These results demonstrate a higher accuracy rate compared to the results obtained in the studies of McEvoy *et al.* (2013), McEvoy *et al.* (2021), and Gomes *et al.* (2021).

Boufenar *et al.* (2023) conducted radiographic analyses using a CNN model in their study on dogs with hip dysplasia. They noted that the 507 radiographic images used were limited and that no expert support was received during the study. Despite this, the best-performing VGG16 model achieved 98.32% accuracy, 98.35% recall, and 98.44% sensitivity. Consequently, they emphasized the need to investigate larger and more complex models. Sezer *et al.* (2020) demonstrated that pelvic ultrasound images can be classified and segmented using CNN. They found 97.9% accuracy, 96.17% sensitivity (95% CI: 92.85%-98.23%), and 98.02% specificity (95% CI: 96.39-99.05%) in their classification. In our study, radiographic images from 2,306 dogs were evaluated. The YOLOv11 model yielded the best results in terms of sensitivity (0.99174), recall (0.99174), and accuracy (95%). In light of the obtained data, it was concluded that larger data sets and more complex models should be investigated. Our observations align with the study conducted by Boufenar *et al.* (2023) and Sezer *et al.* (2020) in terms of limitations and results obtained.

In their study, Loureiro *et al.* (2024) developed a model capable of measuring the femoral neck thickness index in the AI diagnosis of hip dysplasia in dogs. In their study, they found that the majority of the data points in the Bland-Altman plot fell between -0.047 and 0.057. They also determined a 95% confidence interval. In our study, the mean difference between the artificial intelligence and manual measurements in the Bland-Altman plot was -0.05° (95% CI: -5.48° to 5.38°). This minimal deviation demonstrated that the AI was consistent in measuring the NA. Although the methodology used in the presented study differed from that of Loureiro *et al.* (2024), the confidence interval and the data obtained from the Bland-Altman plot were found to be consistent. The data obtained improved the performance of NA measurement models on pelvic radiographs. Fine-tuning the hyperparameters of the developed model is expected to reduce error rates and improve the generalization aptitude.

Conclusions: In conclusion, this study proves that AI can be used safely in the evaluation of hip dysplasia in dogs by calculating the NA. Although the canine pelvic radiography dataset used in this study is sufficient for the analysis, it is thought that using more data may increase the accuracy of the system. The dataset should include images of dogs of different breeds, ages, and sexes. For this, datasets covering a wider group of patients and obtained from different geographies can be used. Thus, prospective, multicentre, and large-sample clinical studies are needed to evaluate the use of AI in the diagnosis of early hip dysplasia.

Conflict of interest: The Authors declare that there is no conflict of interest.

Author's contribution: MKY, UA, NEA and FİB, conceived and designed the review, MKY, UA and NEA corrected and edited the final manuscript, MKY and EA conceptualized, critically revised the manuscript for important intellectual contents and approved the final version.

REFERENCES

Aghapour M, Bockstahler B, Kneissl S, *et al.,* 2023. Radiographic diagnosis of hip laxity in Rottweilers: interobserver agreement at eight- and twelve-months of age. Animals 13(2):231.

Ajadi RA, Sanni JL and Sobayo EL, 2018. Radiographic coxofemoral measurements in Boerboel dogs. Folia Vet 62(4):66-73.

Akinsulie OC, 2024. The potential application of artificial intelligence in veterinary clinical practice and biomedical research. Front Vet Sci 11:1347550.

Akula S, Akula P and Kamati N, 2022. Detection and classification of canine hip dysplasia according to FCI grading system using 3D CNNs. In: Proceedings of the First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR). IEEE, Glasgow, Scotland, pp:1-6.

Albadrani B, 2024. Artificial intelligence in veterinary care: A review of applications for animal health. Egypt J Vet Sci 55(6):1725-36.

Appleby RB and Basran PS, 2022. Artificial intelligence in veterinary medicine. J Am Vet Med Assoc 260(8):819-24.

Bouchemla F, Akchurin SV, Akchurina IV, *et al.,* 2023. Artificial intelligence feasibility in veterinary medicine: A systematic review. Vet World 16(10):2143-149.

Boufenar C, Logovi TEMN, Samir D, *et al.,* 2023. Computer-aided diagnosis of canine hip dysplasia using deep learning in a novel X-ray dataset. Comput Methods Biomech Biomed Eng Imaging Vis 11(7).

Choudhary OP, 2025a. Animal models for surgeries and implants: a vital tool in medical research and development. Ann Med Surg 87(7):4090-95.

Choudhary OP, Infant SS, As V, *et al.,* 2025b. Exploring the potential and limitations of artificial intelligence in animal anatomy. Ann Anat 258:152366.

Comhaire FH and Schoonjans FA, 2011. Canine hip dysplasia: significance of the Norberg angle for healthy breeding. J Small Anim Pract 52:536-42.

Den H, Ito J and Kokaze A, 2023. Diagnostic accuracy of a deep learning model using YOLOv5 for detecting developmental dysplasia of the hip on radiographs. Sci Rep 13(1):6693.

Fitzke M, Stack C, Dourson A, *et al.,* 2021. Rapidread: global deployment of radiology AI for veterinary teleradiology. arXiv 2111.08165.

Franco-Gonçalo P, Leite P, Alves-Pimenta S, *et al.,* 2025. A Computer-Aided Approach to Canine Hip Dysplasia Assessment: Measuring Femoral Head–Acetabulum Distance with Deep Learning. Appl Sci 15(9):5087.

Ginja M, Gaspar A and Ginja C, 2015. Emerging insights into the genetic basis of canine hip dysplasia. Vet Med Res Rep 6:193-202.

Gomes D, Alves-Pimenta S, Ginja M, *et al.,* 2021. Predicting canine hip dysplasia in X-ray images using deep learning. In: Optimization, Learning Algorithms and Applications (Pereira AI, ed). Springer, Berlin, Germany, pp:393-400.

Hansen GC, Yao Y, Fischetti AJ, *et al.,* 2025. Next-generation machine learning model to measure the Norberg angle on canine hip radiographs increases accuracy and time to completion. Am J Vet Res 86(9):ajvr.25.02.0044.

Hennessey E, DiFazio M, Hennessey R, *et al.,* 2022. Artificial intelligence in veterinary diagnostic imaging: A literature review. Vet Radiol Ultrasound 63:851-70.

Jacob N, Kumar BAAS and Padodara RJ, 2022. A glimpse into artificial intelligence in animal physiology and allied sciences. Anim Reprod Update 2(1):72-81.

Joslyn S and Alexander K, 2022. Evaluating artificial intelligence algorithms for use in veterinary radiology. Vet Radiol Ultrasound 63(S1):871-79.

Kim E, Fischetti AJ, Sreetharan PS, *et al.,* 2022. Comparison of artificial intelligence to the veterinary radiologist's diagnosis of canine cardiogenic pulmonary edema. Vet Radiol Ultrasound 63(3):292-97.

Klever J, Brühschwein A, Wagner S, *et al.,* 2020. Comparison of Norberg angle and distraction index for hip laxity in dogs. Vet Comp Orthop Traumatol 33(4):274-78.

Lassalle L, Regnard NE, Durteste M, *et al.,* 2025. Validation of AI-driven measurements for hip morphology assessment. Eur J Radiol 183:111911.

LeCun Y, Bengio Y and Hinton G, 2015. Deep learning. Nature 521(7553):436-44.

Liu C, Wang W, Lian J, *et al.,* 2025. Lesion classification and diabetic retinopathy grading via vision transformer. Front Public Health 12:1442114.

Loureiro C, Gonçalves L, Leite P, *et al.,* 2024. Deep learning-based automated assessment of canine hip dysplasia. Multimed Tools Appl 1-17.

McEvoy F and Amigo J, 2013. Using machine learning to classify image features from canine pelvic radiographs. Vet Radiol Ultrasound 54(2):122-26.

McEvoy F, Proschowsky H, Müller A, *et al.,* 2021. Deep transfer learning for hip joint detection and dysplasia classification in pelvis radiographs. Vet Radiol Ultrasound 62(4):387-93.

Mikkola L, Holopainen S, Pessa-Morikawa T, *et al.,* 2019. Genetic dissection of canine hip dysplasia phenotypes and osteoarthritis reveals three novel loci. BMC Genomics 20:1027.

Pesapane F, Gnocchi G, Quarrella C, *et al.,* 2024. Errors in radiology: a standard review. J Clin Med 13(15):4306.

Pinna S, Tassani C, Antonino A, *et al.,* 2022. Prevalence of primary radiographic signs of hip dysplasia in dogs. Animals 12(20):2788.

Ragap MG, Abdulkadır SJ, Muneer A, *et al.,* 2024. A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023). IEEE Access 12:57815-57836.

Santana A, Alves-Pimenta S, Martins J, *et al.,* 2021. Imaging diagnosis of canine hip dysplasia without human exposure to ionizing radiation. Vet J 276:105745.

Schachner E and Lopez M, 2015. Diagnosis, prevention, and management of canine hip dysplasia: A review. Vet Med Res Rep 6:181-192.

Sezer A and Sezer HB, 2020. Deep CNN-based automatic classification of neonatal hip ultrasound images with speckle noise reduction. Ultrasound Med Biol 46(3):735-49.

Sha J, Huang L, Chen Y, *et al.,* 2023. Clinical thought-based software for diagnosing developmental dysplasia of the hip in pediatric radiographs. Front Pediatr 11:1080194.

Vinicki K, Ferrari P, Belic M, *et al.,* 2018. Using CNNs to determine reticulocyte percentage in cats. Clin Orthop Relat Res arXiv 1803.04873.

Zakareya S, Izadkhah H and Karimpour J, 2023. A new deep-learning model for breast cancer diagnosis from medical images. Diagnostics (Basel) 13(11):1944.