

Article

Femoral Neck Thickness Index as an Indicator of Proximal Femur Bone Modeling

Pedro Franco-Gonçalo ^{1,2,3} , Ana Inês Pereira ^{1,4}, Cátia Loureiro ⁵, Sofia Alves-Pimenta ^{2,3,4} , Vítor Filipe ^{5,6} , Lio Gonçalves ^{5,6} , Bruno Colaço ^{2,3,4} , Pedro Leite ⁷, Fintan McEvoy ⁸ and Mário Ginja ^{1,2,3,*} 

- ¹ Department of Veterinary Science, University of Trás-os-Montes and Alto Douro, 5000-801 Vila Real, Portugal; pedrofranco@utad.pt (P.F.-G.); ines.pereira444@gmail.com (A.I.P.)
- ² Animal and Veterinary Research Centre (CECAV), University of Trás-os-Montes and Alto Douro, 5000-801 Vila Real, Portugal; salves@utad.pt (S.A.-P.); bcolaco@utad.pt (B.C.)
- ³ Associate Laboratory for Animal and Veterinary Sciences (AL4Animals), University of Trás-os-Montes and Alto Douro, 5000-801 Vila Real, Portugal
- ⁴ Department of Animal Science, University of Trás-os-Montes and Alto Douro, 5000-801 Vila Real, Portugal
- ⁵ Department of Engineering, University of Trás-os-Montes and Alto Douro, 5000-801 Vila Real, Portugal; catialoureiro1999@gmail.com (C.L.); vfilipe@utad.pt (V.F.); lgoncalv@utad.pt (L.G.)
- ⁶ Institute for Systems and Computer Engineering (INESC-TEC), Technology and Science, 4200-465 Porto, Portugal
- ⁷ Neadvance Machine Vision SA, 4705-002 Braga, Portugal; pleite@neadvance.com
- ⁸ Department of Veterinary Clinical Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, 1870 Copenhagen, Denmark; fme@sund.ku.dk
- * Correspondence: mginja@utad.pt

Simple Summary: Canine hip dysplasia development results in femoral neck modeling and an increase in thickness. The main objective of this work was to describe a femoral neck thickness index to quantify femoral neck width and to study its association with the degree of canine hip dysplasia using the *Fédération Cynologique Internationale* scoring scheme. A total of 53 dogs (106 hips) were randomly selected for this study. Two examiners performed femoral neck thickness index estimation to study intra- and inter-examiner reliability and agreement. Statistical analysis tests showed excellent agreement and reliability between the measurements of the two examiners and the examiners' sessions. All joints were scored in five categories by an experienced examiner according to the *Fédération Cynologique Internationale* criteria, and the results from examiner 1 were compared between these categories. The comparison of mean femoral neck thickness index between hip dysplasia categories using the analysis of variance test showed significant differences between groups. These results show that femoral neck thickness index is a parameter capable of evaluating proximal femur bone modeling and that it has the potential to enrich conventional canine hip dysplasia scoring criteria if incorporated into a computer-aided diagnosis software.



Citation: Franco-Gonçalo, P.; Pereira, A.I.; Loureiro, C.; Alves-Pimenta, S.; Filipe, V.; Gonçalves, L.; Colaço, B.; Leite, P.; McEvoy, F.; Ginja, M. Femoral Neck Thickness Index as an Indicator of Proximal Femur Bone Modeling. *Vet. Sci.* **2023**, *10*, 371. <https://doi.org/10.3390/vetsci10060371>

Academic Editor: Gerardo Fatone

Received: 21 April 2023

Revised: 18 May 2023

Accepted: 21 May 2023

Published: 24 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The alteration in the shape of the femoral neck is an important radiographic sign for scoring canine hip dysplasia (CHD). Previous studies have reported that the femoral neck thickness (FNT) is greater in dogs with hip joint dysplasia, becoming progressively thicker with disease severity. The main objective of this work was to describe a femoral neck thickness index (FNTi) to quantify FNT and to study its association with the degree of CHD using the *Fédération Cynologique Internationale* (FCI) scheme. A total of 53 dogs (106 hips) were randomly selected for this study. Two examiners performed FNTi estimation to study intra- and inter-examiner reliability and agreement. The paired *t*-test, the Bland-Altman plots, and the intraclass correlation coefficient showed excellent agreement and reliability between the measurements of the two examiners and the examiners' sessions. All joints were scored in five categories by an experienced examiner according to FCI criteria. The results from examiner 1 were compared between FCI categories. Hips that were assigned an FCI grade of A ($n = 19$), B ($n = 23$), C ($n = 24$), D ($n = 24$), and E ($n = 16$) had a mean $\pm$ standard deviation FNTi of 0.809 ± 0.024 , 0.835 ± 0.044 , 0.868 ± 0.022 , 0.903 ± 0.033 , and 0.923 ± 0.068 , respectively (ANOVA, $p < 0.05$). Therefore, these results show that FNTi is a parameter capable of evaluating proximal

femur bone modeling and that it has the potential to enrich conventional CHD scoring criteria if incorporated into a computer-aided diagnosis capable of detecting CHD.

Keywords: dog; canine hip dysplasia; femoral neck thickness index; bone modeling; osteoarthritis

1. Introduction

Canine hip dysplasia (CHD) is an inherited orthopedic disease predominant in large and giant dog breeds that causes lameness and disability. Phenotypic expression of CHD is influenced by genetic defects and environmental stresses that trigger hip joint laxity and incongruency, which often leads to bone modeling and progression to secondary osteoarthritis (OA) [1,2]. Molecular tests for the diagnosis of CHD have already been developed, but they still have not achieved acceptable diagnostic accuracy for the disease [3,4]. Radiography has remained the established imaging technology for diagnosing CHD, as it plays an important role in the selection of breeding stock with the aim of reducing the incidence of the disease in offspring [1,2]. In general, there is not a consistent relationship between clinical signs and radiographic joint changes [5,6]. The ventrodorsal hip extended (VDHE) view is recommended worldwide for CHD screening [7]. However, in young animals, a ventrodorsal hip stress view can also be used to evaluate the hip joint laxity [1,2]. Hip laxity is considered a main risk factor for CHD, but there are some important differences in the progression and final severity of CHD [5,6]. An image of good technical quality of the VDHE view requires radiographic images without pelvic tilting and with adequate femur extension and alignment [7]. However, in the last decades, despite the widespread use of CHD screening radiographs, the prevalence of the disease remains high in some breeds due to a number of factors. These include: variability between radiologists' assessments, which is due to different levels of expertise among radiologists; screening systems that are not yet sufficiently standardized and strict, allowing for some subjectivity; the late appearance of unequivocal pathognomonic radiographic signs, at an age stage in which sometimes the dog has already entered the breeding pool; and the absence of evaluation of potential early hip joint changes in this view [8–10]. Therefore, the current incidence of CHD highlights the need to introduce new parameters to improve diagnosis, such as the hip congruency index, which has the potential to confer greater objectivity to the assessment of hip congruency, consequently benefiting overall diagnostic and scoring accuracy [11].

Worldwide, there are three main international entities for CHD scoring: the Fédération Cynologique Internationale (FCI), with implementation in the countries of continental Europe, the British Veterinary Association/The Kennel Club (BVA/KC), used mainly in the United Kingdom, and the Orthopedic Foundation for Animals (OFA), used in the United States of America [1,6,12]. All of these scoring systems place great emphasis on bone modeling and OA [12]. They also take a qualitative evaluation approach to these parameters, which leaves some margin for interpretation and error [8,13,14]. Early screening for CHD, commonly referred to as the PennHIP method, is based on distinct hip abnormalities. It assesses hip joint laxity by analyzing the femoral head separation from the acetabulum under stress by measuring the distraction index (separation distance divided by the radius of the femoral head) [1,2,6].

In CHD, early joint osteoarthritis and a consequent increase in joint fluid leads to incongruency, laxity of the soft tissue of the hip joint, and subluxation that results in abnormal stresses placed on the bony and soft tissue components of the joint. The joint capsule is attached to the margin of the acetabulum and around the femoral neck. The femoral neck is considered to be normal when its diameter narrows slightly directly below the head. The presence of biomechanical imbalance results in cartilage wear and tear and in the development of mechanosensitive pathways that drive proteases to initiate the mechanism of joint breakdown, subchondral and periosteal reaction, and new bone

production in the capsule attachment area and neighboring tissues, particularly around the junction between the head and neck [15,16]. One of the main signs of CHD is the widening of the appearance of the femoral neck on a craniocaudal radiograph view due to osteophyte development in conjunction with the flattening of the femoral head, the former becoming progressively thicker until it is indistinguishable from the head in severe stages of CHD due to bony proliferation (exostoses) [17–19]. The FCI proposes a 5-grade classification system to represent the severity of the disease: A (normal), B (near normal/transition), C (mild), D (moderate), and E (severe) [6,12,16,17]. Current recommendations in CHD FCI scoring are to include Norberg angle measurement, joint space evaluation, congruence, osteoarthritic signs, and all aspects of hip joint changes, commonly referred as Brass' method [17]. A previous study reports that the femoral neck thickness (FNT) is altered in hip joints classified as near normal (grade B) by the FCI system, adopting a slightly cylindrical shape. This morphological change becomes even more evident in hip joints classified as moderate grade (grade D) [17].

The main objective of this study was to create a new measurement method focused on alterations of the femoral neck as a means of determining proximal femoral changes associated with bone modeling and OA for CHD. For this purpose, we calculated the femoral neck thickness index (FNTi), an objective parameter that relates the minimal FNT to the diameter of the ipsilateral femoral head, and compared it to the different FCI grades. Our hypothesis was that there would be an association between FCI grades and FNTi, and that FNTi would increase with disease severity. To our knowledge, the FNTi's association with the FCI grades has not been previously studied and could be integrated into a classification system as a parameter for evaluating bone modeling of the proximal femur.

2. Materials and Methods

This was a retrospective study based on the evaluation of VDHE views that were randomly selected from the Veterinary Teaching Hospital of the University of Trás-os-Montes and Alto Douro database and from the Danish Kennel Club database, obtained between 2010 and 2023. Recorded data included breed, sex, and weight. The inclusion criteria were dogs older than 12, 15, or 18 months (according to FCI recommendations for medium, large, and giant breeds) and VDHE views with adequate technical quality in terms of image (good bone contrast and spatial resolution) and pelvis positioning (femurs parallel to each other, patellae centred between the femoral condyles, and pelvic symmetry) for CHD scoring. VDHE views showing radiographic signs compatible with other hip or hindlimb diseases, such as bone fractures, previous surgeries, neoplasia, and knee osteoarthritis, were excluded. All five FCI categories were similarly represented in the sample. Due to the observational nature of this study, owner consent and ethical committee approval were waived.

The sample included 53 dogs of 13 different breeds: Portuguese Pointers (13/53, 25%), Estrela Mountain dogs (12/53, 23%), Transmontano Mastiffs (8/53, 15%), German Shepherd dogs (5/53, 9%), Labrador Retrievers (3/53, 9%), Bernese Mountain dogs (2/53, 4%), Barbado da Terceira (1/53, 2%), French Bulldog (1/53, 2%), German Shorthaired Pointer (1/53, 2%), Old Danish Pointer (1/53, 2%), Portuguese Sheepdog (1/53, 2%), Portuguese Water dog (1/53, 2%), and Rottweiler (1/53, 2%). There were 19 (36%) males and 34 (64%) females. The age ranged from 12 to 166 months, and the mean $\pm$ standard deviation (SD) was 31.87 ± 29.07 months. The average body weight was 36.1 ± 14.1 kg.

2.1. Radiographic Measurements

The minimal FNT was measured by drawing a straight line, roughly perpendicular to the anatomical axis of the femoral neck, connecting the two closest points between the proximal and distal margins of the femoral neck in a VDHE view. The femoral head diameter was determined as a diameter of a circle outlining the margin of the femoral head (Figure 1). These measurements were performed by examiner 1 (E1) in two independent sessions

(S1 and S2) to evaluate repeatability and by examiner 2 (E2) to test the reproducibility using specific DICOM viewer and editor software (Dys4Vet version 2.0, accessed between 1 October and 31 December 2022). The FNTi was determined by dividing the FNT by the femoral head diameter.

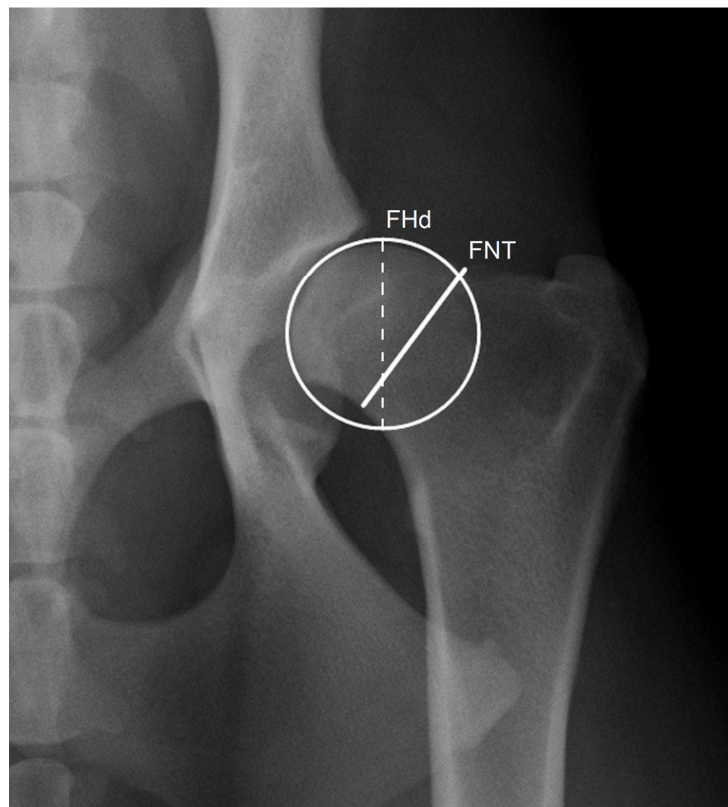


Figure 1. Detail of the left hip joint of a dog showing a circle delimiting the head of the femur with 31.5 mm diameter (FHd; dashed line) and a line joining the proximal and distal borders of the femoral neck to determine its thickness of 28.7 mm (FNT), corresponding to an index (FNTi) of 0.91.

In order to associate FNTi with the five FCI grades for CHD, the hip joints were scored using FCI criteria: grade A (normal hip—Norberg angle (NA) $> 105^\circ$ and excellent congruency); grade B (borderline or transitional hip joint—NA around 105° and mild incongruency); grade C (slight CHD—NA around 100° , centre of femoral head outside of dorsal acetabular margin, and moderate incongruency); grade D (moderate CHD—NA $> 90^\circ$, signs of osteoarthritis, and obvious incongruency); and grade E (severe CHD—NA $< 90^\circ$, signs of osteoarthritis, and severe incongruency) [12,17]. The NA was measured between a line joining the centre of the circle encompassing the femoral heads and another line connecting each centre of the femoral head with the ipsilateral, effective cranial acetabular rim [20]. Some radiographic parameters related to the femoral head centre position and the dorsal acetabular margin and joint space were also considered for final FCI joint scoring [17].

The E1 and E2 measurements were performed by I.P. and P.F., respectively, and the hip FCI scoring was performed by M.G. in a single-blind fashion for each examiner (i.e., E1 and E2 were unaware of each FNTi measurement and FCI scores, and the FCI scorer was unaware of FNTi measurements).

2.2. Statistical Analysis

SPSS computer software (SPSS Statistics for Windows Version 27.0: IBM Corp., Armonk, NY, USA) was used to perform the statistical analysis.

Parametric tests were used for statistical analysis [21]. The paired *t*-test was used for comparison of duplicate E1S1–E1S2 measurements to evaluate repeatability, as well as E1S1–E2 measurements to evaluate reproducibility [22]. The Bland–Altman analysis and the intraclass correlation coefficient (ICC) were used to investigate intra- and inter-examiner agreement and reliability, respectively. In the Bland–Altman method, the 95% limits of agreement (LA) were calculated as the mean difference (*d*) $\pm$ 1.96 standard deviation (SD) [22–25]. Measurements were considered in agreement when the 95% confidence interval (CI) of the mean differences included zero and equivalent when the 95% upper and lower LA were small (irrelevant difference) [23,24]. The ICC was considered as poor, moderate, good/acceptable, and excellent reliability when the lower limit of 95% CI was <0.50 , ≥ 0.50 – 0.75 , ≥ 0.75 – 0.90 , and ≥ 0.90 , respectively [25,26]. Cohen's *d* was used to measure the effect size when significant differences between measurements were registered: negligible < 0.20 , small ≥ 0.20 , medium ≥ 0.50 , and large ≥ 0.80 [27].

The comparison of FNTi values of E1S1 measurement between FCI categories was performed using the Welch's ANOVA, followed by the post hoc Games–Howell test. The null hypothesis was that there were no significant differences in the FNTi mean values between the FCI categories [28]. A *p*-value of <0.05 was considered statistically significant. The statistical analysis was performed considering each joint individually.

3. Results

Measurements were performed on 106 hip joints. The FNTi mean $\pm$ SD in E1S1 was 0.86 ± 0.06 ; in E1S2 it was 0.86 ± 0.06 ; and in E2 it was 0.87 ± 0.06 . The main statistical analysis results related to the intra-examiner (repeatability) ($p > 0.05$ in paired *t*-test; ICC = 0.94 [95% CI, 0.92–0.96]; $d \pm$ SD -0.001 ± 0.019 [95% LA, -0.038 , 0.036]) and inter-examiner (reproducibility) ($p < 0.05$ in paired *t*-test; ICC = 0.93 [95% CI, 0.90–0.95]; $d \pm$ SD -0.007 ± 0.021 [95% LA, -0.084 , 0.034]) measurements of FNTi are presented in detail in Table 1, Figures 2 and 3.

Table 1. Statistical analysis of the intra- and inter-examiner agreement and reliability of the FNTi in measuring 106 hips.

Comparison		Paired <i>t</i> -Test <i>p</i> -Value	Cohen's <i>d</i>	ICC [95% CI]	<i>d</i> $\pm$ SD	95% LA	
						Lower Bound	Upper Bound
Intra-examiner	E1S1–E1S2	0.583	–	0.94 [0.92,0.96]	-0.001 ± 0.019	-0.038	0.036
Inter-examiner	E1S1–E2	<0.001	0.14	0.93 [0.90,0.95]	-0.007 ± 0.021	-0.048	0.034

Abbreviations: CI, confidence interval; *d*, mean difference; E, examiner; FNTi, femoral neck thickness index; ICC, intraclass correlation coefficient; LA, limits of agreement; S, session; SD, standard deviation.

A total of 19 (18%) hip joints were scored as FCI grade A, and the FNTi mean $\pm$ SD was 0.809 ± 0.024 ; 23 (21%) hip joints were scored as FCI grade B, and the FNTi mean $\pm$ SD was 0.835 ± 0.044 ; 24 (23%) hip joints were scored as FCI grade C, and the FNTi mean $\pm$ SD was 0.868 ± 0.022 ; 24 (23%) hip joints were scored as FCI grade D, and the FNTi mean $\pm$ SD was 0.903 ± 0.033 ; and 16 (15%) hip joints were scored as FCI grade E, and the FNTi mean $\pm$ SD was 0.923 ± 0.068 . Data were assessed for normality using the Shapiro–Wilk test ($p_A = 0.501$; $p_B = 0.275$; $p_C = 0.739$; $p_D = 0.983$; $p_E = 0.383$), and Levene's test indicated unequal variances among group samples ($p_{A,B,C,D,E} < 0.05$). Significant statistical mean differences using Welch's ANOVA followed by the post hoc Games–Howell test were verified in FCI categories with means marked with different letter superscripts ($p < 0.05$) (Table 2, Figure 4).

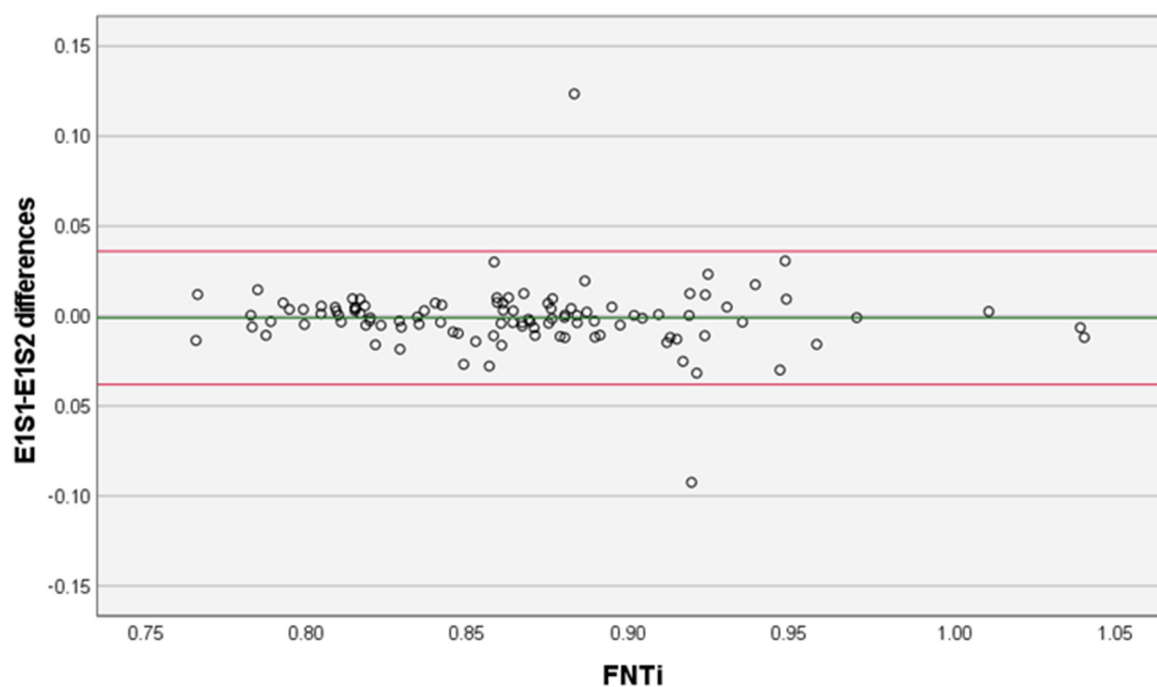


Figure 2. Examiner 1 differences between Sessions 1 and 2 (E1S1; E1S2). The green line represents the mean of the differences (-0.001) and the red lines represent the lower and upper 95% limits of agreement, -0.038 and 0.036 , respectively. FNTi—femoral neck thickness index.

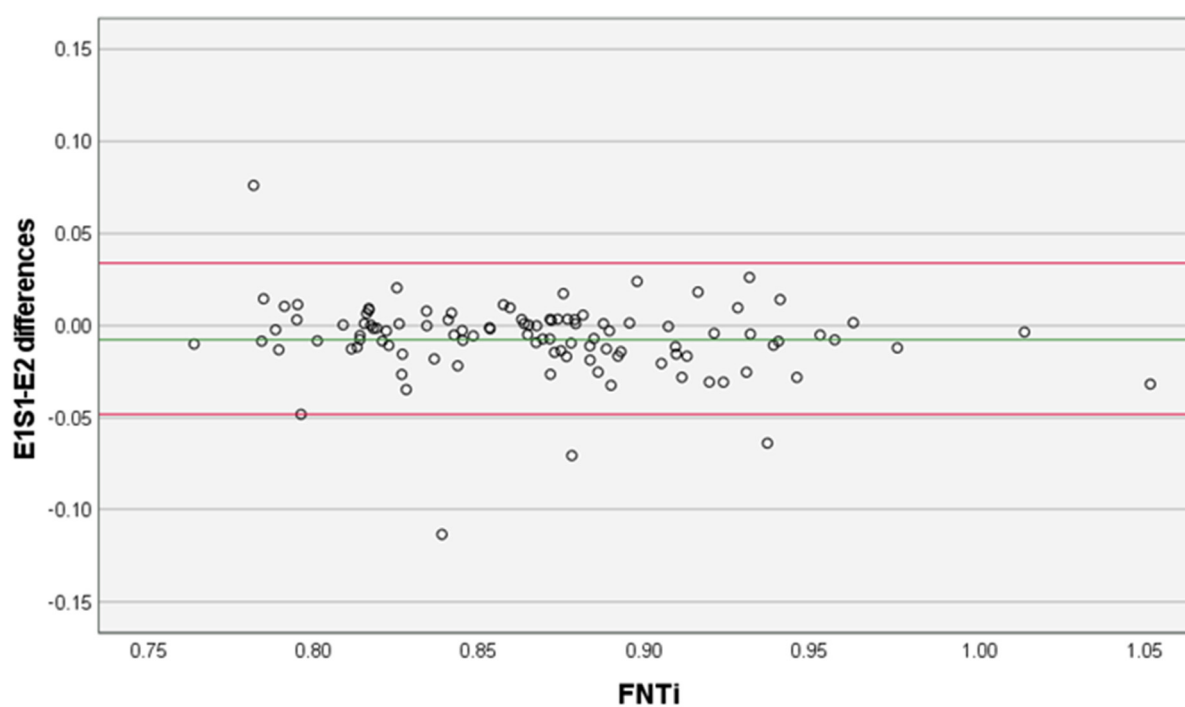


Figure 3. Examiner 1, Session 1 (E1S1) and Examiner 2 (E2) measurement differences. The green line represents the mean of the differences (-0.007) and the red lines represent the lower and upper 95% limits of agreement, -0.048 and 0.034 , respectively. FNTi—Femoral neck thickness index.

Table 2. Statistical descriptive analysis of the femoral neck thickness index by FCI categories.

FCI Categories	N	Mean *	SD	Mean 95% CI		Min	Max
				Lower Bound	Upper Bound		
A	19	0.809 ^a	0.024	0.797	0.823	0.771	0.864
B	23	0.835 ^{a,b}	0.044	0.812	0.859	0.757	0.952
C	24	0.868 ^{b,c}	0.022	0.856	0.880	0.831	0.918
D	24	0.903 ^{c,d}	0.033	0.886	0.921	0.841	0.969
E	16	0.923 ^d	0.068	0.887	0.959	0.781	1.036

* Means with different superscripts are statistically different ($p < 0.05$) in the post hoc Games–Howell test that followed Welch’s ANOVA. Abbreviations: CI, confidence interval; FCI, Fédération Cynologique Internationale; SD, standard deviation.

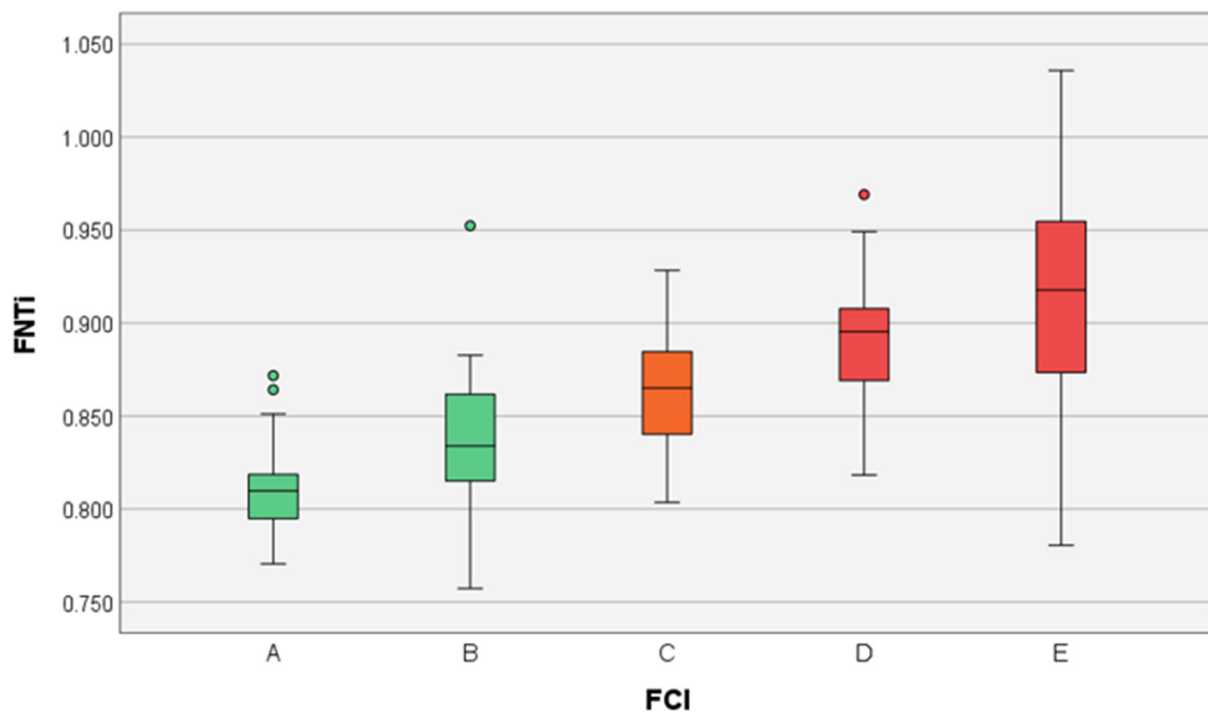


Figure 4. Box and whisker plot presenting the femoral neck thickness index (FNTi) classified in Fédération Cynologique Internationale (FCI) categories (A–E). The box plots show from bottom–up minimum data value, lower quartile value, median value, upper quartile value, maximum data value, and outliers (distant from the rest of the data).

4. Discussion

The objective of this study was to create a methodology that would allow a more objective radiographic evaluation of the changes that the proximal femur undergoes in CHD (namely, thickening of the neck), and to understand this relationship across the various FCI degrees of CHD in the VDHE view. The use of a parametric test in statistical analysis was based on the Central Limit Theorem, which states that in large sample sizes ($n > 30$), the distribution of standardized samples’ means tends to be normally distributed independently of the distribution of the population from where it originated [21]. The FCI scoring can be considered a dynamic system that has undergone regular updates over the years in order to harmonize CHD classifications in different countries and to improve the reliability of associating morphological alterations with the genetic profile of the animal. We highlight here the CHD panelist meetings of Dortmund 1991 and Copenhagen 2007 and 2022 [29].

In dysplastic hips, progressive bone modeling and osteophyte formation are induced by osteoarthritic pathways that result in subchondral and periosteal response with new bone production. Particularly around the junction between the head and neck, biomechanical stresses in the hip joint occur at a faster rate, and new bone and osteophytes are placed in some areas and reabsorbed in other areas of the femur and acetabulum [15]. These changes in the bone structure are concurrent with hip osteoarthritis [30]. Pinna et al. (2022) observed a significant increase in the thickness of the femoral neck in VDHE views of hips classified as grade B, a grade assigned to hips considered healthy and in which no osteoarthritic signs are visible [17]. In another study, Andronescu et al. (2015) studied the ratio between head volume and femoral neck volume on 3D computed tomography images of dogs at high risk of developing CHD (distraction index > 0.3) from 16 to 32 weeks of age and found a decrease in the ratio, even though the differences between values corresponding to OA severities were not statistically significant [18]. Therefore, the study of bone modeling and OA as two interrelated topics seems particularly important to us. We use the FNTi to get around the difficulty that exists in veterinary medicine to use absolute measures in anatomic measurements due to the different sizes of dogs. The relationship between measurements and the size of the femoral head is a strategy that has been successfully used previously for other purposes, such as the hip distraction index (mainly in young animals without severe bone changes) [1,6] and the hip congruency index [11]. In a previous study, the femoral neck width was related with the femoral neck length to create a widening index [31]. However, this methodology was not followed in this work because we think that the relationship between the length of the femoral neck and the size of the breed is much less studied and used than the diameter of the femoral head [6,11]. As such, there may be greater variability in the length of the femoral neck between breeds of similar sizes, which makes this parameter a less effective ratio measure. On the other hand, artificial intelligence is currently being introduced in digital image analysis [32,33], and since the identification of the femoral head seems to us to be an essential parameter in the application of artificial intelligence to the diagnosis of CHD, any index that resorts to its use can be more easily integrated in the near future for this purpose.

Bone modeling is admittedly one of the aspects of classification that is undervalued by the FCI criteria. In some iterations of the FCI criteria [34], morphological changes of the proximal femur are only explicitly mentioned in the most severe degree of dysplasia (grade D), pointing out the characteristic mushroom-like appearance that the femoral head assumes. Other changes related to bone modeling of the proximal femur have been overlooked and are presumably relegated to the topic of “osteoarthritic signs”, addressed with a yes/no question [34,35], which makes the evaluation less accurate. The BVA/KC uses a phenotypic evaluation criteria based on a points system, addressing changes in bone shape from minor modeling up to severe OA [19]. However, it is still a qualitative scoring system, so there is also some subjectivity in the analysis because it ends up being dependent on the level of experience of the examiner. Taking this into consideration, there is a perceived need for novel approaches that grant more objectivity to the assessment of bone modeling and, in addition, CHD scoring.

Our results showed that the methodology behind the FNTi had excellent reliability and agreement. Given the results of the paired *t*-tests and Bland-Altman plots with the mean differences near zero and the narrow 95% CI, there was no evidence of bias between the examiners' measurements, and they can be considered statistically similar. The intra-examiner ICC was 0.94, and the lower bound of the 95% CI was 0.92, which translates to excellent reliability. This indicates adequate repeatability and reproducibility of the described FNTi determination methodology. Significant mean differences between FNTi examiners' measurements were observed in the paired samples *t*-test. However, the recorded effect size was negligible ($d = 0.14$) and, considering the corresponding Bland-Altman plot (Figure 3), the 95% LA can be interpreted as clinically small because 95% of all calculated inter-examiner differences lie in the short range of -0.048 to 0.034 . Furthermore, the inter-examiner ICC shows excellent reliability. This shows that even though the two

examiners themselves produce somewhat different values, their measurements are clearly related and functionally consistent. Both examiners had a similar amount of practice time with the method beforehand, so no conclusions can be drawn based on experience. On the other hand, one would expect the mean differences to be greater among bigger FNTi mean values due to the subjectiveness of the annotation imposed by exostosis and/or osteophyte formation in the concave fossa, making the bone contours more intricate, which could have possibly caused disparities between sessions and examiners. Ultimately, by analyzing the two diagrams (Figures 2 and 3), we can see that the mean difference values do not increase or decrease in proportion to the mean FNTi values, thus concluding that there is no proportional bias.

The Welch's ANOVA of FNTi values in FCI categories revealed a statistically significant main effect ($p < 0.001$), indicating that not all FCI categories had the same mean FNTi value. Post hoc comparisons using the Games–Howell post hoc procedure were conducted to determine which pairs of the five categories' means differed significantly. The results indicate that different FCI categories have mean FNTi values that increase gradually with the degree of severity of the disease: grade A hips had a statistically significant lower mean FNTi value than grades C, D, and E ($p < 0.05$); grade B had significantly lower mean FNTi value than grades D and E ($p < 0.05$); and grade C had significantly lower mean FNTi value than grade E ($p < 0.05$). Therefore, the null hypothesis is rejected, which supports our assumption that FNTi changes with FCI CHD grades. This parameter can help distinguish between CHD grades, potentially improving FCI scoring criteria. However, there is some overlap in the FNTi values corresponding to the different FCI categories, specifically between adjacent categories. This is one of the reasons that an assessment solely dependent on this is ambiguous and, therefore, impractical. The “whiskers” of the box plot corresponding to category E stretch over a wider range of values than the other box plots, overlapping almost every other category. This can be explained by how long hip subluxation has been in place and by the difficulty of adequate delimitation of the femoral head for its diameter measurement in joints with very severe bone osteoarthritic deformations. In severe cases with long-term subluxation, the surface of the femoral head cartilage at the non-articulated margin of the head and neck becomes thickened due to a lack of contact with the opposite acetabular surface [15]. Hence, a long-term subluxated hip favors the appearance of the so-called mushroom head deformity, which inflates the diameter of the femoral head compared to the neck thickness, thus producing lower FNTi values pertaining to category E, values that lie in the lower “whisker” of the box plot. We strongly advocate that this parameter should never be used by itself to classify hips, but rather in complementarity with other parameters. On the other hand, given the recognized variability that exists in the progression of CHD, femoral neck bone modeling, and femoral head size, it is expected that more robust FNTi results will be obtained if a study is performed using only one breed in the sample [5].

Since our study sample was sourced from two different databases and included a variety of breeds, the results can be more easily extrapolated to general canine populations at risk of CHD. Nonetheless, as a limitation of this study, it is important to note that some breeds prone to CHD are clearly over-represented (Portuguese breeds); others, on the contrary, are under-represented (Labrador Retriever and Rottweiler). As such, in the future, there is a need to conduct more studies on a wider range of breeds and patients.

5. Conclusions

This study describes a methodology that allows for the evaluation of bone modeling of the proximal femur in the VDHE view, which can be used in the future with confidence as a criterion for CHD scoring. The FNTi shows adequate intra- and inter-examiner measurement agreement and reliability. Mean FNTi values are gradually higher in the different FCI categories, with statistically significant differences. The FNTi method shows potential to make CHD classification more objective if incorporated as a scoring criterion,

and it is also promising in view of its application through a computer-aided diagnosis capable of autonomously identifying and classifying CHD.

Author Contributions: M.G. and P.F.-G. contributed to conception and design of this study. B.C., F.M., P.F.-G., P.L., S.A.-P. and M.G. organized the database. M.G., P.F.-G., A.I.P. and C.L. defined the methodology. B.C., C.L., F.M., L.G., A.I.P., V.F. and M.G. performed validation and data analysis. P.F.-G. wrote the first draft of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financed by project Dys4Vet (POCI-01-0247-FEDER-046914), co-financed by the European Regional Development Fund (ERDF) through COMPETE2020, the Operational Programme for Competitiveness and Internationalisation (OPCI).

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the observational and retrospective nature of this study.

Informed Consent Statement: This study did not involve humans.

Data Availability Statement: Not applicable.

Acknowledgments: The authors acknowledge the UTAD Veterinary Teaching Hospital and Danish Kennel Club for allowing access to images from their data archive. The authors are also grateful for all the conditions made available by FCT—Portuguese Foundation for Science and Technology, under the projects UIDP/00772/2020, LA/P/0059/2020 and Scientific Employment Stimulus—Institutional Call—CEECINST/00127/2018 UTAD.

Conflicts of Interest: Author P.L. was employed by Neadvance Machine Vision SA. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. Bartolome N., Segarra S., Artieda M. et al. (2015) A genetic predictive model for canine hip dysplasia: integration of genomewide association study (GWAS) and candidate gene approaches. *PLoS ONE* 10, e0122558.

References

- Smith, G.K.; Paster, E.R.; Powers, M.Y.; Lawler, D.F.; Biery, D.N.; Shofer, F.S.; McKelvie, P.J.; Kealy, R.D. Lifelong Diet Restriction and Radiographic Evidence of Osteoarthritis of the Hip Joint in Dogs. *J. Am. Vet. Med. Assoc.* **2006**, *229*, 690–693. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ginja, M.M.; Gonzalo-Orden, J.M.; Melo-Pinto, P.; Bulas-Cruz, J.; Orden, M.A.; San Roman, F.; Llorens-Pena, M.P.; Ferreira, A.J. Early hip laxity examination in predicting moderate and severe hip dysplasia in Estrela mountain dog. *J. Small Anim. Pract.* **2008**, *49*, 641–646. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bartolomé, N.; Segarra, S.; Artieda, M.; Francino, O.; Sánchez, E.; Szczypiorska, M.; Casellas, J.; Tejedor, D.; Cerdeira, J.; Martínez, A.; et al. A genetic predictive model for canine hip dysplasia: Integration of Genome Wide Association Study (GWAS) and candidate gene approaches. *PLoS ONE* **2015**, *10*, e0122558. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bruun, C.S.; Bank, A.; Ström, A.; Proschowsky, H.F.; Fredholm, M. Validation of DNA test for hip dysplasia failed in Danish Labrador Retrievers. *Anim. Genet.* **2020**, *51*, 617–619. [\[CrossRef\]](#)
- Lopez, M.; Schachner, E. Diagnosis, Prevention, and Management of Canine Hip Dysplasia: A Review. *Vet. Med. Res. Rep.* **2015**, *6*, 181–192. [\[CrossRef\]](#)
- Ginja, M.M.D.; Silvestre, A.M.; Gonzalo-Orden, J.M.; Ferreira, A.J.A. Diagnosis, Genetic Control and Preventive Management of Canine Hip Dysplasia: A Review. *Vet. J.* **2010**, *184*, 269–276. [\[CrossRef\]](#)
- Martins, J.; Colaço, B.; Alves-Pimenta, S.; Gonzalo Orden, J.M.; Ferreira, A.J.; Ginja, M.M. Effect of the Dog positioning on X-ray Table on Hip Dysplasia Parameter Evaluation. *Vet. Comp. Orthop. Traumatol.* **2019**, *32*, 376–382. [\[CrossRef\]](#)
- Coopman, F.; Verhoeven, G.; Paepe, D.; Van Bree, H.; Duchateau, L.; Saunders, J.H. Inter-Observer Agreement on the Radiographic Assessment of Canine Hip Dysplasia. *Vlaams Diergeneesk. Tijdschr.* **2007**, *76*, 417–422.
- Genevois, J.-P.; Cachon, T.; Fau, D.; Carozzo, C.; Viguiet, E.; Collard, F.; Remy, D. Canine Hip Dysplasia Radiographic Screening. Prevalence of Rotation of the Pelvis along Its Length Axis in 7012 Conventional Hip Extended Radiographs. *Vet. Comp. Orthop. Traumatol.* **2007**, *20*, 296–298.
- Powers, M.Y.; Biery, D.N.; Lawler, D.F.; Evans, R.H.; Shofer, F.S.; Mayhew, P.; Gregor, T.P.; Kealy, R.D.; Smith, G.K. Use of the Caudolateral Curvilinear Osteophyte as an Early Marker for Future Development of Osteoarthritis Associated with Hip Dysplasia in Dogs. *J. Am. Vet. Med. Assoc.* **2004**, *225*, 233–237. [\[CrossRef\]](#)
- Franco-Gonçalo, P.; Moreira da Silva, D.; Leite, P.; Alves-Pimenta, S.; Colaço, B.; Ferreira, M.; Gonçalves, L.; Filipe, V.; McEvoy, F.; Ginja, M. Acetabular Coverage Area Occupied by the Femoral Head as an Indicator of Hip Congruency. *Animals* **2022**, *12*, 2201. [\[CrossRef\]](#)

12. Flückiger, M. Scoring Radiographs for Canine Hip Dysplasia—The Big Three Organizations in the World. *Eur. J. Companion Anim. Pract.* **2007**, *17*, 135–140.
13. Geissbühler, U.; Drazovic, S.; Lang, J.; Howard, J. Inter-Rater Agreement in Radiographic Canine Hip Dysplasia Evaluation. *Vet. Rec.* **2017**, *180*, 357. [\[CrossRef\]](#)
14. Fortrie, R.R.; Verhoeven, G.; Broeckx, B.; Duchateau, L.; Janssens, L.; Samoy, Y.; Schreurs, E.; Saunders, J.; van Bree, H.; Vandekerckhove, P.; et al. Intra- and Interobserver Agreement on Radiographic Phenotype in the Diagnosis of Canine Hip Dysplasia. *Vet. Surg.* **2015**, *44*, 467–473. [\[CrossRef\]](#)
15. Riser, W.H.; Rhodes, W.H.; Newton, C.D. Hip Dysplasia. In *Textbook of Small Animal Orthopedics*; U. Penn CAL Small Animal Orthopedics; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 1982.
16. Smith, G.K. New Paradigms for Hip Dysplasia Prevention and Control, Performance and Ethics of CHD Screening. In *12th ESVOT Congress Proceedings*; Vezzoni, A., Schramme, M., Eds.; LMU: Munich, Germany, 2004; pp. 125–131.
17. Pinna, S.; Tassani, C.; Antonino, A.; Vezzoni, A. Prevalence of Primary Radiographic Signs of Hip Dysplasia in Dogs. *Animals* **2022**, *12*, 2788. [\[CrossRef\]](#)
18. Andronescu, A.A.; Kelly, L.; Kearney, M.T.; Lopez, M.J. Associations between Early Radiographic and Computed Tomographic Measures and Canine Hip Joint Osteoarthritis at Maturity. *Am. J. Vet. Res.* **2015**, *76*, 19–27. [\[CrossRef\]](#)
19. Dennis, R. Interpretation and Use of BVA/KC Hip Scores in Dogs. *InPractice* **2012**, *34*, 178–194. [\[CrossRef\]](#)
20. Gaspar, A.R.; Hayes, G.; Ginja, C.; Ginja, M.M.; Todhunter, R.J. The Norberg Angle Is Not an Accurate Predictor of Canine Hip Conformation Based on the Distraction Index and the Dorsolateral Subluxation Score. *Prev. Vet. Med.* **2016**, *135*, 47–52. [\[CrossRef\]](#)
21. Kwak, S.G.; Kim, J.H. Cornerstone of Modern Statistics. *Korean J. Anesthesiol.* **2017**, *70*, 144–156. [\[CrossRef\]](#)
22. Watson, P.F.; Petrie, A. Method Agreement Analysis: A Review of Correct Methodology. *Theriogenology* **2010**, *73*, 1167–1179. [\[CrossRef\]](#)
23. Ludbrook, J. Confidence in Altman-Bland Plots: A Critical Review of the Method of Differences. *Clin. Exp. Pharmacol. Physiol.* **2010**, *37*, 143–149. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Bland, J.M.; Altman, D.G. Measuring Agreement in Method Comparison Studies. *Stat. Methods Med. Res.* **1999**, *8*, 135–160. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Lee, J.; Koh, D.; Ong, C.N. Statistical Evaluation of Agreement between Two Methods for Measuring a Quantitative Variable. *Comput. Biol. Med.* **1989**, *19*, 61–70. [\[CrossRef\]](#)
27. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates, Inc.: London, UK, 1988.
28. Marôco, J. *Análise Estatística Com o SPSS Statistics*, 7th ed.; Gráfica Manuel Barbosa & Filhos: Pêro Pinheiro, Portugal, 2018.
29. Hedhammar, Å. Swedish Experiences From 60 Years of Screening and Breeding Programs for Hip Dysplasia—Research, Success, and Challenges. *Front. Vet. Sci.* **2020**, *27*, 228. [\[CrossRef\]](#)
30. Edinger, D.T.; Hayashi, K.; Hongyu, Y.; Markel, M.D.; Manley, P.A. Histomorphometric Analysis of the Proximal Portion of the Femur in Dogs with Moderate Osteoarthritis. *Am. J. Vet. Res.* **2005**, *66*, 150–155. [\[CrossRef\]](#)
31. Martins, J.; Ferreira, A.J.; Ginja, M.M. Morphometric assessment of the hip joint in the Estrela Mountain Dog breed. *Vet. Comp. Orthop. Traumatol.* **2012**, *25*, 202–210. [\[CrossRef\]](#)
32. McEvoy, F.J.; Proschowsky, H.F.; Müller, A.V.; Moorman, L.; Bender-Koch, J.; Svalastoga, E.L.; Frellsen, J.; Nielsen, D.H. Deep Transfer Learning Can Be Used for the Detection of Hip Joints in Pelvis Radiographs and the Classification of Their Hip Dysplasia Status. *Vet. Radiol. Ultrasound.* **2021**, *62*, 387–393. [\[CrossRef\]](#)
33. Banzato, T.; Wodzinski, M.; Tauceri, F.; Donà, C.; Scavazza, F.; Müller, H.; Zotti, A. An AI-Based Algorithm for the Automatic Classification of Thoracic Radiographs in Cats. *Front. Vet. Sci.* **2021**, *8*, 731936. [\[CrossRef\]](#)
34. Morgan, J.K.; Stephens, M. *Radiographic Diagnosis and Control of Canine Hip Dysplasia*; Iowa State University Digital Press: Ames, IA, USA, 1985.
35. FCI Hip Dysplasia Classification. In Proceedings of the FCI/DKK Hip Dysplasia Conference, Copenhagen, Denmark, 24–25 May 2022; p. 11.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.