

Canine age classification using Deep Learning as a step toward preventive medicine in animals

Szymon Mazurek [1,2], Maciej Wielgosz [1,2], Jakub Caputa [1], Rafał Frączek [1,2], Michał Karwatowski [1,2], Jakub Grzeszczyk [1], Daria Łukasik [1], Anna Śmiech [3], Paweł Russek [1,2], Ernest Jamro [1,2], Agnieszka Dąbrowska-Boruch [1,2], Marcin Pietroń [1,2], Sebastian Koryciak [1,2], Kazimierz Wiatr [1,2]

1 ACC Cyfronet AGH, Nawojki 11, 30-950 Kraków, Poland

2 AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, Poland

3 University of Life Sciences, al. Akademicka 13, 20-950 Lublin, Poland

Email: {rafalfr, mkarwat, wielgosz, russek, adabrow, jamro, pietron, koryciak, wiatr}@agh.edu.pl

{s.mazurek, d.lukasik, j.grzeszczyk, j.caputa}@cyfronet.pl

anna.smiech@up.lublin.pl

Abstract—The main goal of this work was to implement a reliable machine learning algorithm that can classify a dog’s age given only a photograph of its face. The problem, which seems simple for humans, presents itself as very difficult for the machine learning algorithms due to differences in facial features among the dog population. As convolutional neural networks (CNNs) performed poorly in this problem, the authors took another approach of creating novel architecture consisting of a combination of CNN and vision transformer (ViT) and examining the age of the dogs separately for every breed. Authors achieved better results than those in initial works covering the problem.

I. INTRODUCTION

DELAYING the aging and preventing diseases associated with it is becoming increasingly prevalent in state-of-the-art medicine. Biological and chronological age comparison is one of the primary assessment tools to estimate the health status of a given subject. The trend is also prevalent in veterinary medicine as animals, especially domesticated ones such as dogs, play an essential role in modern societies.

Machine learning and its growing effectiveness and versatility provide valuable tools that can be used in medicine and anti-aging research to create simple and easy-to-use methods for assessing biological age.

The age assessment can help approximate an animal’s health status, allowing the animal owner to take preventive actions before a health condition is developed. The authors hope this work’s findings will also be useful for further research in human and animal preventive medicine.

The main goal of this work was to create a tool to assess a dog’s age, given only its picture. The age groups of the dogs were classified as follows:

- Young dogs, aged from 0 to 2 years,
- Adult dogs, aged from 2 to 6 years,
- Senior dogs, aged from 6 years and more.

The tool will be a part of the larger project aiming to detect animals’ health conditions and belongs to a realm of preventive medicine solutions. The system’s overall architecture is presented in Fig. 1.

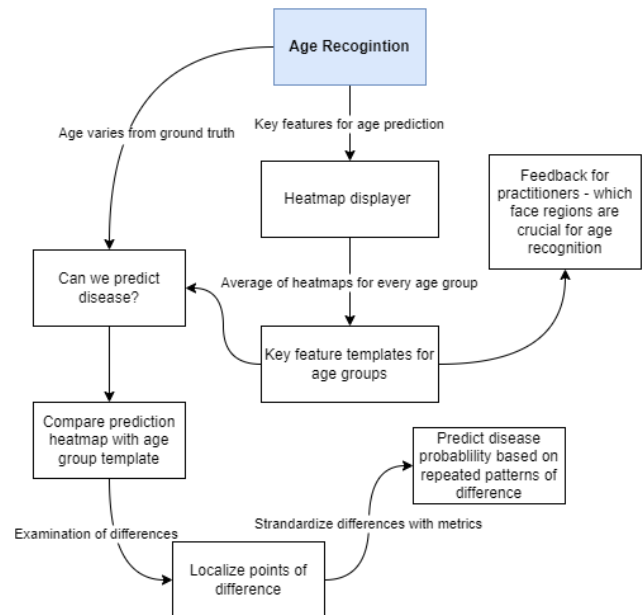


Fig. 1. A block diagram of the CyfroVet system for early disease diagnostics (a part of it presented in [11])

This work is part of CyfroVet, a research project run by ACC Cyfronet to utilize technology, especially AI, in veterinary medicine. The anti-aging and presented system is one of the topics covered by the project, alongside cancer cell detection and segmentation.

The summary of this paper’s contributions is as follows:

- new architecture for addressing the difficult task of dogs classification, which outperforms available state-of-the-art models,
- new dataset which enables effective training of the classification models,
- conceptualization of using age assessment in animals for disease prevention and their early diagnosis.

We make our code and the dataset publicly available.

II. MATERIALS AND METHODS

There are multiple methods for assessing canines age. Most of them involve a certain kind of precise examination. The few most popular are:

- Ocular lens examination [1]
- Teeth examination [2]
- Bone examination [3]

Authors have noticed an opportunity to use computer vision in this field, as it can recognize image features and patterns sometimes impossible to notice for humans with a bare eye examination. As mentioned before, a reliable age assessment based on appearance can be a straightforward tool to use for veterinary medicine practitioners and pet owners. Facial appearance differing vastly from the typical one for a given age group can be a sign of accelerated biological aging, which may be an indicator of the overall health status of an individual. Having preliminary warning signs, practitioners can take preventive steps and make a detailed diagnosis to detect early signs of disease before more severe symptoms develop. This early warning can be beneficial in treating potentially terminal conditions in their early stage. Quick detection and treatment of medical conditions such as cancer can significantly increase an individual's chances of survival.

A. Datasets

In this experiment, the authors initially decided to use the DogAge dataset containing dogs of various age groups created by the Tech4Animals research group [5]. This dataset was used in the Dog Age Challenge competition organized by the group mentioned earlier, which required participants to create a neural network algorithm that could assess the dog's age. It took place in 2019.

The expert dataset consisted of photographs classified into age groups by animal scientists. The photographs were high-quality, and the dog was directly facing the camera.

The Petfinder dataset consisted of photographs downloaded from *Petfinder.com*, a website for pet adoption [4]. Experts did not verify the classification of each age group. The images were often of poor quality, with dogs at various angles and distances.

Several preprocessing techniques were used for those datasets, such as cropping and data augmentation. They will be described in detail in the next section. The Petfinder dataset was cleared of irrelevant photos (i.e., dog's face not clearly visible, other dogs appearing).

After the initial experiments, the authors sought to expand the datasets, as the performance was not reaching satisfying results. The data was collected from the *Petfinder.com* [4], as the website provided the divisions of the dogs by age and breed. The photographs were downloaded and cleared, as many repeating adoption offers were on the website. Two approaches were taken, with the first dividing the problem into examining only one breed at a time and the second by expanding the mixed breeds datasets. Obtained photos were similar to those described in the previous case of the

Petfinder dataset. The datasets are described in Tab. I. *Expert dataset* and *Petfinder dataset* refer to datasets presented in DogAge Challenge. *Big dataset* refers to a dataset created by downloading new photos. Two datasets consisting only of dogs of single breeds are also presented in Tab. I. The dataset name describes the breeds. Given breeds were selected based on data availability - the number of photos and class balance.

III. EXPERIMENTS

The initial datasets presented many challenges. Firstly, they were greatly imbalanced, with the adult category containing more images than the other two combined. Secondly, the data quality was also an obstacle - the *Petfinder* set contained many invalid photos (i.e., dog's face not visible, humans present in the image) for the network training. Those problems were also present within the new datasets. The experiments were run on the Prometheus cluster, using one node with 2 Nvidia Tesla v100 GPUs and 36 CPU cores.

A. Image cropping algorithm

A two-way approach was introduced to address the problem of low image quality. Firstly, the *Petfinder* dataset was manually cleared, removing the inadequate pictures (e.g., a dog looking back or to the side, two dogs in the picture, dog with toys in its mouth). Secondly, two cropping algorithms were tested: YOLO (You Only Look Once)[6] and FaceDetector. After the evaluation, the latter was chosen. Both algorithms could correctly detect the dogs. However, FaceDetector focused only on the dog's face. Therefore, it presented itself not only as a cropping mechanism but also as a tool to clear the new datasets of irrelevant photos, as manual clearing is time inefficient. Results of both cropping algorithms are visible in Fig. 2. for FaceDetector and Fig. 3. for YOLO.



Fig. 2. An image cropped with FaceDetector



Fig. 3. An image cropped with YOLO

B. Metrics

Categorical accuracy (CA), F1-score (F1), and recall (RA) metrics were used to examine the performance of tested models. In the DogAgeChallenge announcement, [10] a modified version of mean average error (MAE), average accuracy (ACA), and average recall (aRA) is used. In this paper, the authors decided not to use mMAE since it was concluded that this metric would provide no meaningful information.

TABLE I
SUMMARY OF THE DATASETS

Dataset name	Total number of photos	Number of adult group photos	Number of senior group photos	Number of young group photos
Expert dataset	1088	370	495	223
Petfinder dataset	22573	13016	1898	7659
Big dataset	49010	26021	5523	17464
Big dataset (balanced)	16964	5999	5517	5448
Chihuahua	3933	1898	875	1160
Pitbull terrier (balanced)	3195	1128	996	1071

C. Augmentation, class weights and data balancing

To cope with a class imbalance within the datasets, the authors tried to introduce data augmentation via random flip, rotation, brightness, and contrast changes. However, augmentation harmed models' performances, lowering CA on validation sets on average by 5% when augmenting by flip and rotation and 10% when augmenting via contrast and brightness changes. Those tests were run on a balanced Pitbull Terrier dataset, excluding the influences of class imbalance being magnified by augmentation. Authors suspect that augmentation via brightness and contrast shifts may negatively influence the results by skewing some features crucial for age assessment (for example, fur color). Data augmentation failed to solve the class imbalance problem, so the authors used class weights during training. Weights were chosen as 1, 5, and 2 for Adult, Senior and Young classes. Only the Expert dataset, showing different data distribution, was assigned weights 2, 1, and 2 for the same class order. Also, class balancing via removal of the photos from most numerous classes was tested. The number of removed photos was chosen so that all classes had roughly the same number of photos as the least numerous ones. In the case of balanced datasets, no class weights were applied in the final experiments, as this approach did not perform well for balanced datasets (once again reducing CA and F1).

D. Initial network architecture and experiments

Initial trials were conducted using the transfer learning protocol. The backbone of the model consisted of CNN pre-trained on ImageNet dataset with a classifier consisting of fully connected layers. During the experiments, the EfficientNetB7 [7] was chosen as the final backbone for the model, as the results of different commonly used CNN architectures were performing on a very similar level. The architecture of the initial model is based on the EfficientNetB7 backbone, followed by global average pooling and two fully connected layers.

The initial model was tested on the available datasets. During the experiments, the random data splits were as follows:

- training split - 80% of the dataset
- validation split - 10% of the dataset
- test split - 10% of the dataset

The network was tested on both single and mixed-breed datasets. It was trained using AdamW [8] optimizer with a learning rate parameter equal to 0.0001 and the weight decay

TABLE II
ARCHITECTURE OF THE NETWORK WITH ViT AS CLASSIFIER

Layer number	Layer name	Parameters
1	EfficientNetB7	Weights from pretraining on Imagenet
2	Patch Encoding 2D	Embedding dimension of 64
3	Layer normalization	-
4	Multi-Head Attention	6 attention heads
5	Skip connection	Adding outputs from layers 2 and 4
6	MLP layer	-
7	Skip connection	Adding outputs from layer 5 and 6
8	Layer normalization	-
9	Flattening layer	-
10	Dropout layer	50% dropout
11	MLP layer	dropout increased to 50%
12	Dense layer	Softmax activation

parameter of the same value. Different configurations of the classifier with the usage of regularization, dropout layers, more dense layers, and different values of units within dense layers did not significantly influence the model's performance. The results of experiments with the initial network are visible in Tab. III and Tab. IV.

E. Transformer-based classifier as a method to improve the classification results

During the experiments, the CNN model obtained better results than those found in the literature [10]. However, they still were not satisfying, so the authors decided to modify the architecture, adding a small version of ViT as a classifier after the backbone CNN block. ViT implements an attention-based model known in the field of NLP. The decoder block is omitted in the computer vision version compared to the original implementation. The model uses an attention mechanism to learn relations between different parts of the input image. Dosovitskiy et al. [9] showed that ViT can outperform current state-of-the-art CNN architectures in image classification. As it requires large datasets, the authors decided to extract features from the data using pre-trained CNN. Architecture is presented in Tab. II.

The model was again trained using available datasets, with the same data split as for experiments with the initial network. Results are visible in Tab. III and Tab. IV.

TABLE III

FINAL RESULTS OF THE INITIAL AND MODIFIED NETWORKS TRAINED ON THE DESCRIBED DATASETS.

Model	Initial network		Network with ViT as classifier	
	CA	RA	CA	RA
Expert dataset	0.49	0.44	0.51	0.49
Petfinder dataset	0.49	0.23	0.55	0.50
Big dataset	0.44	0.23	0.46	0.29
Big dataset (balanced)	0.53	0.25	0.54	0.45
Chihuahua	0.45	0.35	0.48	0.45
Pitbull Terrier (balanced)	0.50	0.22	0.50	0.44

TABLE IV

FINAL F1 SCORE OF THE INITIAL AND MODIFIED NETWORKS TRAINED ON THE DESCRIBED DATASETS FOR SPECIFIC CLASSES (A - ADULT, S - SENIOR, Y - YOUNG).

Model	Initial network			Network with ViT as classifier		
	A	S	Y	A	S	Y
Expert dataset	0.47	0.57	0.35	0.33	0.64	0.41
Petfinder dataset	0.47	0.40	0.55	0.57	0.42	0.56
Big dataset	0.28	0.42	0.56	0.32	0.45	0.57
Big dataset (balanced)	0.45	0.61	0.53	0.49	0.60	0.55
Chihuahua	0.27	0.54	0.55	0.49	0.51	0.43
Pitbull Terrier (balanced)	0.43	0.60	0.48	0.55	0.40	0.50

F. Results

The results obtained during the experiments are presented in this section compared to the previous experiments. In the comparison, classification accuracy (CA) and recall (RA) were used to compare the obtained results with the ones conducted previously during the DogAge Challenge [10]. During DogAge Challenge, authors showed that using Squeezenet and Inception v3 with dense layer classifiers resulted in the CA and RA of 32% for the Squeezenet model and 34% on both metrics for Inception v3. Comparing these values with the ones seen in Tab. III, which describes the ones obtained with solutions proposed by the authors trained on available datasets, we can see that significant improvement is made. The networks can reach both higher CA and RA. The CNN + ViT architecture improved CA and RA for every dataset. The single breed approach was similarly effective compared to the mixed breed approach. Examining Tab. IV, it is visible that CNN + ViT does not necessarily improve F1 for given classes but instead results in more even per-class F1. It can also be concluded that the class balancing and weighting approach can be a helpful tool to cope with class imbalances for this problem, as balanced datasets' results are comparable to those obtained on unbalanced datasets with class weights applied during network training.

IV. DISCUSSION

Several challenges affect the experiment and results presented in this work. The most important of them are as follows:

- limited access to high-quality data,

- lack of the effective data specification methods, which makes both augmentation and preparation of the validation dataset challenging,
- it is tough to deal with multiple breeds using a single model,
- granularity of the age classification is tightly related to an ability to label images precisely.

Other research projects in this domain may also consider these remarks to reach a high performance of ML models.

V. CONCLUSIONS AND FUTURE WORK

The problem of dog age classification using neural networks turned out to be more demanding than expected in the beginning. Experiments showed that state-of-the-art CNN models provided insufficient accuracy while using datasets with multiple dog breeds. The transformer-based architecture introduced in the paper improved the performance results, but there is still much space for progress in future research.

VI. DATA AND CODE AVAILABILITY

The code and datasets as well as additional results available: github.com/SzymonMazurekAGH/Age_recognition_Cyfrovet

REFERENCES

- [1] Abood S. Estimating Age in Dogs and Cats Using Ocular Lens Examination. *Compendium on Continuing Education for the Practising Veterinarian - North American Edition*. (2000).
- [2] Roccaro, M. & Peli, A. A determination in dog puppies by teeth examination: legal, health and welfare implications, review of the literature and practical considerations. *Veterinaria Italiana* 56. pp. 149-162 (2020). DOI: 10.12834/VetIt.1876.9968.2
- [3] Sutton, L., Byrd, J. & Brooks, J. Age Determination in Dogs and Cats. *Veterinary Forensic Pathology, Volume 2*. pp.151-163 (2018). DOI: <https://doi.org/10.1007/978-3-319-67175-8>
- [4] Petfinder.com, a website for pet adoption. <https://www.petfinder.com/>, date accessed 30.06.2021
- [5] Tech4Animals, *DogAge Challenge*, <http://132.75.251.84:3000/~tech4animals/dogchallenge/>, date accessed: 30.06.2021
- [6] Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition*. (2016). DOI : 10.1109/CVPR.2016.91
- [7] Tan, M. & Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proceedings of the 36th International Conference on Machine Learning, ICML 2019*. pp. 6105-6114 (2019).
- [8] Loshchilov I. & Hutter F. Decoupled Weight Decay Regularization. *International Conference on Learning Representations 2017*. (2017).
- [9] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. & Houlsby, N. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. (2020). DOI: <https://doi.org/10.48550/arXiv.2010.11929>
- [10] Zamansky, A., Sinitca, A., Kaplun, D., Dutra, L. & Young, R., Automatic Estimation of Dog Age: The DogAge Dataset and Challenge. *Artificial Neural Networks And Machine Learning - ICANN 2019: Image Processing*. pp. 421-426 (2019). DOI: 10.1007/978-3-030-30508-6
- [11] Caputa, J., Łukasik, D., Wielgosz, M., Karwatowski, M., Frączek, R., Russek, P. & Wiatr, K. Fast Pre-Diagnosis of Neoplastic Changes in Cytology Images Using Machine Learning. *Applied Sciences*. **11** (2021). DOI: <https://doi.org/10.3390/app11167181>