Paper

Convolutional Neural Networks for *C. Elegans* Muscle Age Classification Using Only Self-learned Features

Bartosz Czaplewski[1], Mariusz Dzwonkowski[1,2], and Damian Panas[3]

[1]*Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Department of Teleinformation Networks, Poland,*
[2]*Medical University of Gdańsk, Faculty of Health Sciences, Department of Radiology Informatics and Statistics, Poland,*
[3]*Institute of Animal Reproduction and Food Research of the Polish Academy of Sciences, Molecular Biology Laboratory, Poland*

https://doi.org/10.26636/jtit.2022.165322

Abstract — Nematodes *Caenorhabditis elegans* (*C. elegans*) have been used as model organisms in a wide variety of biological studies, especially those intended to obtain a better understanding of aging and age-associated diseases. This paper focuses on automating the analysis of *C. elegans* imagery to classify the muscle age of nematodes based on the known and well established IICBU dataset. Unlike many modern classification methods, the proposed approach relies on deep learning techniques, specifically on convolutional neural networks (CNNs), to solve the problem and achieve high classification accuracy by focusing on non-handcrafted self-learned features. Various networks known from the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) have been investigated and adapted for the purposes of the *C. elegans* muscle aging dataset by applying transfer learning and data augmentation techniques. The proposed approach of unfreezing different numbers of convolutional layers at the feature extraction stage and introducing different structures of newly trained fully connected layers at the classification stage, enable to better fine-tune the selected networks. The adjusted CNNs, as featured in this paper, have been compared with other state-of-art methods. In anti-aging drug research, the proposed CNNs would serve as a very fast and effective age determination method, thus leading to reductions in time and costs of laboratory research.

Keywords — biomedical imaging, C. elegans muscle aging, convolutional neural networks, deep learning, machine learning

1. Introduction

Aging is a complex process in which many changes take place simultaneously at every level of a biological organization - from organelles, cells, tissues, to the entire systemic environment. Naturally, aging and age-related diseases are studied primarily with human health and longevity in mind, but model organisms, e.g. *Caenorhabditis elegans* (*C. elegans*), may be successfully employed for those purposes as well [1]–[5]. *C. elegans*, as used in this study, is a free-living, 1 mm long nematode that feeds on bacteria. Its clear, age-dependent physiological changes, short lifespan of approx. 17–20 days, ease of maintenance in a laboratory setting and high genetic homology with humans make the worm a widely acknowledged system for aging research [6], [7].

C. elegans exhibits many age-associated changes [8], however deterioration of its muscle tissue draws particular attention. Since the nematode's body-wall muscle is analogous to human skeletal muscle, some disorders, especially those of particular concern to the public health, can be effectively studied with the use of this model. For example, due to its progressive loss of muscle mass, *C. elegans*, just like some elderly humans, suffers from sarcopenia. This condition has dire consequences for many activities, such as locomotion and ingestion, but surprisingly little is known regarding its cures and actual causes [9]–[11].

Extensive research has been performed on anti-aging drugs based on measuring the age of *C. elegans* and determining the effects that medication has at early stages of the aging process. The weakness of such studies has always been that the age of *C. elegans* was determined by human experts and, thus, the results were subjective and strongly dependent on the expert's experience. Gaining new insights in this domain is therefore potentially of great value, although succeeding advances might require increasingly sophisticated, state-of-the-art tools, such as convolutional neural networks (CNNs) [12]–[14]. This machine learning technique has recently shown some remarkable achievements and great potential in solving problems related to image classification, visual signal processing, biological and medical imaging, computer vision, and image recognition years [15]–[19]. In addition to being capable of outperforming medical diagnosis experts, machine learning-based tools, such as CNNs, may also significantly reduce the time and costs of such an evaluation. Hence, further explorations in this field are essential.

This paper shows an innovative solution to the problem of classifying *C. elegans* muscle aging using CNNs. The contributions of this paper are as follows.

– An innovative approach consisting in using a convolutional neural network for the purpose of *C. elegans* muscle aging classification was presented. New CNN models were obtained through fine-tuning at the feature selection stage,

and the introduction of new structures at the classification stage;

– A fully CNN-based solution was proposed which achieved the highest level of classification accuracy. The proposed solution relies solely on non-handcrafted, self-learned features, in contrast to the existing literature;
– The ability to use various networks known from the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) was investigated. New and different transfer learning strategies were applied. The effect that unfreezing of different numbers of convolutional layers during the transfer learning phase has on the overall classification accuracy was investigated. This step is usually ignored in other publications. The effect of different classifier structures was investigated and the fully connected layers were modified accordingly;
– Comparison of the proposed method with other methods used in tackling the same problem was conducted using the same or similar datasets.

The article is structured as follows. Related works are presented in Section 2. The proposed methodology is explained in Section 3. Section 4 describes experimental results, including the CNN structures, learning parameters, classification accuracy level obtained, and a comparison with other methods. Conclusions are given in Section 5.

2. Related Work

Image processing technologies are commonly incorporated into many studies to automate the measurement of age of *C. elegans*. A training-based approach relying on tracking the swimming motion of the worm was examined by Restif and Metaxas [20]. Johnston *et al.* devised a pattern recognition tool that assesses age using the pharynx pumping rate [21]. Machine learning techniques, more specifically a CNN trained for regression, that utilized the relationship between age and level of body bend, were proposed by Lin *et al.* in [22].

Many research programs described in the literature that investigate the topic rely on experiments with self-generated datasets and, therefore, preclude comparative analysis. However, some studies (as listed below) report their results based on data obtained from the IICBU Biological Image Repository, including, inter alia, a dataset concerned with *C. elegans* muscle aging.

An open-source utility classifier for biological image analysis called Wndchrm was proposed by Shamir *et al.* in [23], [24]. The classifier extracts image content descriptors (1025 image features including polynomial decompositions, high contrast features, pixel statistics, and textures) from the raw image. Feature extraction is also performed on image transforms, such as Fourier, wavelet and Chebyshev, as well as on compound transforms being different combinations of subsequent image transforms. The most informative features are selected and the feature vector of each image is used for classification and similarity measurement.

A BIOimage Classification and Annotation Tool (BIOCAT) was proposed by Zhou *et al. in* [25]. It allows automatic classification and annotation of entire 2- and 3-dimensional biological images or their specific regions of interest by utilizing pattern recognition algorithms. BIOCAT is equipped with approximately 20 built-in modularized algorithms of feature extractors that can be chained in a customizable way to form an appropriate model for image classification. The chains may be compared via BIOCAT to determine the most suitable model for the specific dataset.

Siji *et al.* enhanced the performance of a content-based biological image retrieval system by selecting discriminative feature sets from the set of canonical features [26]. Canonical features are extracted using Wndchrm [23] and are assigned to 4 separate feature sets. Each set is optimized using the principal component analysis (PCA) process and the Fisher score for the purpose of feature selection. The selected features are then used for training the support vector machine (SVM) and Bayesian classifiers.

Bioimage classification utilizing the SVM classifier was proposed by Song *et al.* [27]. In this approach, a high-dimensional multi-modal descriptor was introduced to combine multiple texture features. To enhance the discriminative power of the descriptors, a subcategory discriminant transform (SDT) was designed to transform the descriptors before performing SVM classification. Similarly to linear discriminant analysis (LDA), the descriptor transform obtained with the use of the SDT algorithm aims to minimize the within-class variation and maximize the between-class difference.

Sekhar and Mohan proposed another SVM approach based on distance metric learning techniques for semi-supervised pattern classification [28]. The authors assigned label information to the selected group of images, creating a sparingly labeled training set. The optimal kernel gram matrix was determined by using pairwise similar/dissimilar constraints derived from the available examples in the labeled training set. Label information for the remaining unlabeled examples in the entire dataset is determined in an iterative self-training manner, utilizing SVM. The method uses a confidence measure in the decision-making process.

Nanni *et al.* proposed a method that combines a CNN with an SVM classifier [29] to obtain a structure capable of synthesizing a large number of different image classification tasks. Features extracted from the deep layers of the CNNs are mixed with more traditional hand-crafted features. The output of each layer is treated as a feature vector for which a dimensionality reduction method, such as PCA or discrete cosine transform (DCT), can be applied, depending on the size of the vector. All of the vectors are then processed by SVM to provide the final output. This approach was further improved in [30], where a combination of local features, dense sampling features, and deep learning features was proposed. Each descriptor was used to train a different SVM, which were then combined with the sum rule. In 2020, Nanni and Ghidoni proposed an alternative approach [31] that boosts the performance of trained CNNs by composing multiple CNNs into an

ensemble and combining scores based on the sum rule. The method, dedicated in general for bioimage datasets, features the combination of multiple descriptors based on different feature types, both of the learned and handcrafted variety.

In addition to the above-mentioned approaches relied upon in the classification of nematodes, there are also numerous novel works that demonstrate the feasibility of incorporating deep learning technology into radiology practice. Ahn *et al.* proposed a mammographic density estimation [32] process based on CNN trained with local and global statistics extracted from the image set, whereas Li *et al.* showed that pretraining may improve the effectiveness of deep learning-based tissue microstructure estimation [33]. This property is particularly evident with abundant, publicly available high-quality datasets, as is the case with the dMRI scans used by the authors.

In this paper, we propose a novel approach to classifying muscle aging in *C. elegans* with the use of a CNN. To the best of our knowledge, this paper is the first work to achieve classification accuracy at a level comparable with (or, in some cases, better than) state-of-the-art methods relying solely on different fine-tuned CNN models, without the need to introduce additional machine learning classifiers, such as linear a SVM (as proposed, for instance, in [29] and [30]), additional metadata or a description of the input images based on expert knowledge.

3. Proposed Methodology

The goal of this paper is to classify *C. elegans* muscle aging based on microscope imaging, using self-learned features only. CNNs are a perfect tool for the considered classification problem, due to their appropriable characteristics. For example, there is no need to design an analytical feature selection model which usually has to be based on expert knowledge. On the other hand, however, a CNN analyzes local spatial properties in the patches of meaningful pixels, which reflects the human interpretation of input data.

The experiments were conducted based on the *C. elegans* muscle aging dataset [34], [35] compiled by the Laboratory of Genetics at the U.S. National Institute on Aging. The dataset consists of images of muscles taken with a fluorescence microscope, with the said images representing various ages of nematodes. The individuals are genetically identical and live in a controlled laboratory environment. The images are saved in the TIFF format, in grayscale, with a resolution of 1600×1200 pixels and a depth of 16 bits. The size of the dataset is 237 observations. The images are unevenly divided into 4 classes: "day 1" (52 observations), "day 2" (48 observations), "day 4" (95 observations), and "day 8" (42 observations). Examples of such images are presented in Fig. 1.

In the proposed solution, the input has the form of a grayscale image of a *C. elegans* muscle with resolution of 1600×1200 pixels and a depth of 16 bits. Due to the structure of the CNNs used for transfer learning, the pixel values from a single-color channel are duplicated to create a 3-channel image. In general, the image can be of any resolution, as it is scaled down to a fixed size corresponding to the input layer of the CNN. The output is a decision about the age of the muscle shown in the image, i.e. the image is classified into one of the four age classes: "day 1", "day 2", "day 4", "day 8". The flowchart for the proposed method is shown in Fig. 2.

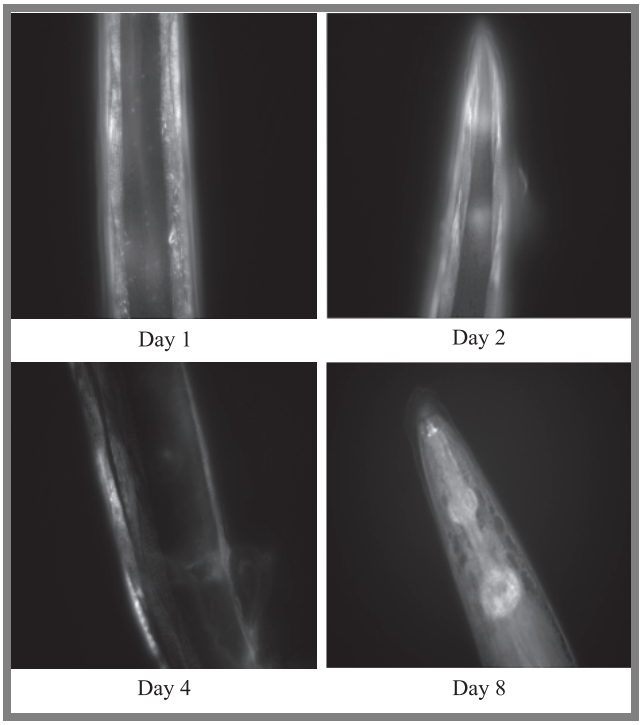

Fig. 1. Examples from the *C. elegans* muscle aging dataset [34], [35].

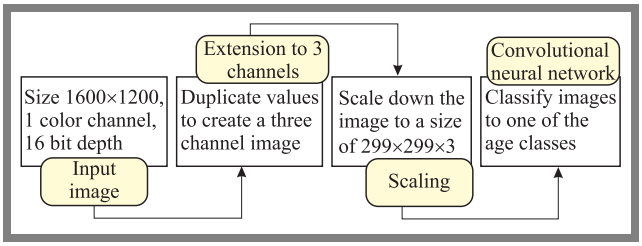

Fig. 2. Flowchart of the proposed approach.

The research presented in this paper can be divided into the following tasks. The first part consisted in selecting the best base network for transfer learning. Designing a new CNN for the classification problem under consideration from scratch would be very time-consuming and ultimately unnecessary. The literature has shown that it is possible to retrain an image classification CNN that has already been trained to extract powerful and informative features from natural images (usually based on ImageNet dataset) to become a classification CNN for a specific medical problem. Based on the results shown in the references [15], [22], three candidates for the base (starting point) network for transfer learning were selected: VGG16, InceptionV3, and InceptionResNetv2.

The second task was to select the best strategy for unfreezing the layers of the base network during transfer learning. The classical approach used in transfer learning is to freeze the weights and biases of the convolutional layers of the

base network, so that they are not retrained at all and only allow the layers responsible for classification to be retrained. During the research, it was found that this is not optimal. Instead, unfreezing a number of layers, both of the convolutional and fully connected variety, significantly increased the classification accuracy of the learned network.

The third task was to investigate how the base network's architecture could be modified to achieve even better results. Since the part of the network corresponding to feature selection should not be changed, as this would mean the rejection of all of the benefits from transfer learning, the main focus was on changes in the part responsible for the classification. Several new classification stage models were considered and the results achieved by the best of them are presented below.

The fourth part was to compare the proposed CNN with other prominent solutions available in the literature. The accuracy of the classification was a comparative criterion.

4. Experimental Results

4.1. Network Structures

The authors decided to rely on the transfer learning technique in their research, as it is much faster and easier than training a network from scratch. It involves the use of an already pre-trained network for general image classification and, relying on its ability to recognize powerful and informative features from an image, training it for a new, much more specific classification task. For this purpose, three CNNs were selected: VGG16 [36], InceptionV3 [37], and InceptionResNetV2 [38]. In this paper, they are referred to as base networks. All three networks had been pretrained on the ImageNet dataset [39] which contains over a million images of natural objects, and are capable of classifying the said images into one of 1000 labels, such as goldfish, coffee mug, pencil, etc. These networks have already learned an effective set of image features, but the goal is to fine-tune the base network, so it can learn features specific to the new dataset and the new set of labels. A comparison of the base networks is given in Tab. 1.

Tab. 1. Comparison of the base networks for transfer learning.

Network	Depth (layers)	No. of parameters	Image size
VGG16	16	$144 \cdot 10^6$	$224 \times 224 \times 3$
InceptionV3	48	$23.9 \cdot 10^6$	$299 \times 299 \times 3$
InceptionResNetV2	164	$55.0 \cdot 10^6$	$299 \times 299 \times 3$

The easiest approach to transfer learning is as follows. One should adopt the architecture of the base network, set the initial pre-trained values of weights and biases, remove the last layer of the classifier and replace it with a new, fully connected layer with the number of neurons equal to the new number of classes, freeze all convolution layers (set the learning rate factors to 0), speed up learning for the fully connected layers (set the learning rate factors to e.g. 10), and then start retraining using the new dataset. In this way, the deep network designer assumes that the starting network has already acquired adequate knowledge for the feature selection stage, suitable for solving the new problem, so the learned values in the convolutional layers would not be changed. A designer could make this decision if the new dataset for the new problem is very similar to the original dataset for which the underlying network was trained. Changing the last fully connected layer at the classification stage is the minimum condition that need to be fulfilled to teach the network to solve a new problem. Alternatively, instead of freezing the layers completely, one can just slow down their learning process, i.e. set the learning rate factor to 1, while setting a very low initial learning rate in the training options (e.g. 0.0001). In this way, the developer allows for minor changes to the feature selection stage.

Unfortunately, such an approach produces results that are not satisfactory for the purpose of our research. This is mainly because the dataset differs significantly from the ImageNet dataset. Unfortunately, such a straightforward approach can be found throughout the literature. Therefore, fine-tuning is essential. In this research, two approaches to fine-tuning were applied: firstly, we unfroze a number of the convolutional layers at the feature extraction stage, and secondly, we introduced different structures of newly trained, fully connected layers at the classification stage.

When selecting layers to unfreeze, one should not select any of the initial layers. The initial convolutional layers are used to identify simple, general patterns in an image, e.g. edges. Even if the new dataset used for transfer learning to solve the new problem is significantly different from the original dataset used for the base network, we still want the network to be able to recognize such elementary features in the image. Moreover, allowing the re-learning of the initial layers will render the knowledge in the subsequent layers useless due to the alternation of the output from the initial layers. On the other hand, deeper convolutional layers are used to identify complex features specific to a given dataset, based on previously identified simple patterns. In the case concerned, both the images and

Tab. 2. General information concerning network structures used.

	CNN#1	CNN#2	CNN#3
Depth	17	49	165
Base CNN	VGG16	InceptionV3	Inception-ResNetV2
Input size	$224 \times 224 \times 3$	$299 \times 299 \times 3$	$299 \times 299 \times 3$
Frozen conv. layers	12	89	238
Unfrozen conv. layers	1	9	6
Fully connected layers	4	2	2
No. of parameters	150 527 244	30 211 936	60 648 676

Tab. 3. Detailed network structures in relation to base networks.

	CNN#1	CNN#2	CNN#3
Frozen feature extraction part	– Convolutional layers from "conv1_1" to "conv5_2": • number and sizes of filters as in VGG16, • bias and weight learn rate factors set to 1, • pre-trained biases and weights initialized – ReLU activation and max pooling layers as in VGG16	– Convolutional layers from "conv2d_1" to "conv2d_89": • number and sizes of filters as in InceptionV3, • bias and weight learn rate factors set to 0, • pre-trained biases and weights initialized – Batch normalization, ReLU activation and max pooling layers as in InceptionV3	– Convolutional layers from "conv2d_1" to "conv2d_199", from "block17_1_conv" to "block17_20_conv", from "block35_1_conv" to "block35_10_conv" from "block8_1_conv" to "block8_9_conv": • number and sizes of filters as in InceptionResNetV2, • bias learn rate factor set to 0, weight learn rate factor set to 1, • pre-trained biases and weights initialized – Batch normalization, ReLU activation, max pooling, addition, scaling, and depth concatenation layers as in InceptionResNetV2
Unfrozen feature extraction part	– Convolutional layer "conv5_3": • number and sizes of filters as in VGG16, • bias and weight learn rate factors set to 10, • biases initialized as zeros and weights initialized randomly – ReLU activation and max pooling layer as in VGG16	– Convolutional layers from "conv2d_86" to "conv2d_94": • number and sizes of filters as in InceptionV3, • bias and weight learn rate factors set to 10, • pre-trained biases and weights initialized. – Batch normalization, ReLU activation and max pooling layer as in InceptionV3	– Convolutional layers from "conv2d_200" to "conv2d_203", "block8_10_conv", and "conv_7b" layers: • number and sizes of filters as in InceptionResNetV2, • bias and weight learn rate factors set to 10, • pre-trained biases and weights initialized. – Batch normalization, ReLU activation and max pooling layer as in InceptionResNetV2
Classification part	– Fully connected layers "fc6" and "fc7": • 4096 neurons, • bias and weight learn rate factors set to 1, • pre-trained biases and weights initialized – ReLU activation and 0.5 dropout layer – Fully connected layers "new_fc8": • 4096 neurons, • bias and weight learn rate factors set to 10, • biases initialized as zeros and weights initialized randomly – ReLU activation and 0.5 dropout layer – Fully connected layer "new_fc9": • 4 neurons, • bias and weight learn rate factors set to 10, • biases initialized as zeros and weights initialized randomly – Softmax and classification layer	– Global average pooling layer – Fully connected layer "new_fc1": • 4096 neurons, • bias and weight learn rate factors set to 10, • biases initialized as zeros and weights initialized randomly – ReLU activation and 0.5 dropout layer – Fully connected layer "new_fc2": • 4 neurons, • bias and weight learn rate factors set to 10, • biases initialized as zeros and weights initialized randomly – Softmax and classification layer	– Global average pooling layer – Fully connected layer "new_fc1": • 4096 neurons, • bias and weight learn rate factors set to 10, • biases initialized as zeros and weights initialized randomly – ReLU activation and 0.5 dropout layer – Fully connected layer "new_fc2": • 4 neurons, • bias and weight learn rate factors set to 10, • biases initialized as zeros and weights initialized randomly – Softmax and classification layer

their associated labels are fundamentally different from the images and labels in ImageNet, so we expect that the network will learn to recognize completely different complex features. This means that many deep convolutional layers should be unfrozen and, furthermore, they should be able to change their values rapidly during the training process. This means that if we want to unfreeze convolutional layers (set the learning rate factors to 10, for instance) and allow them to be re-trained, we have to do it from the end of the network, starting with the layers immediately preceding the fully connected part. In this study, the number of unfrozen convolutional layers ranged from 1 to 9, depending on the network.

In terms of the fully connected part of the network, the starting point was always the classification stage model established for the considered base network. Furthermore, the CNNs were also tested using other classification stage models, e.g. the

five models proposed in [15] were tested. Modifications of fully connected layers at the classification stage are motivated by the fact that the set of labels is completely different from that of the core network, both in meaning and in number. However, it has been noticed that the best results are usually obtained by extending the fully connected part with just one new fully connected layer with a large number of neurons immediately before the last classification layer, which has to be also a new layer with the number of neurons corresponding to the new number of classes.

As the last step in selecting the network structures, the influence of other network parameters was tested, e.g. activation layers were changed from ReLU layers to clipped ReLU layers, leaky ReLU layers or tanh layers, the training parameters were changed, etc. The impact of these changes was negligible and is, therefore, not discussed in this paper.

For all three base networks, hundreds of training sessions were run with various numbers of unfrozen layers as well as with various classifier models. In Tabs. 2 and 3, and throughout the paper, only the best obtained network structures for each of the base networks are presented.

4.2. Training Options

All of the CNNs were trained via supervised learning. For estimating performance of the trained models, the 5-fold cross-validation procedure was used. The dataset was randomly split in such a way that 60% of it was used for the training set, 20% for the validation set, and 20% for the test set. The training data was shuffled before each epoch, and the validation data was shuffled before each validation. Both shuffling and data augmentation were applied to reduce the variance in results and ensure that the model is general and not overfitted.

To avoid overfitting, CNNs rely heavily on the large datasets available. Overfitting occurs when the CNN learns to model the training data too perfectly, such that the learned model is characterized by a high variance for the test data. In analyzing medical images, big datasets are usually not available. With such a small dataset, the risk of overfitting is very high. In such a case, data augmentation [40] is essential to prevent overfitting by introducing some random geometrical transformations into the training data for each epoch. In this way, one randomly augmented version of each image is used during each training epoch. In the course of the research, the data augmentation settings were based on the data augmentation models presented in [15], which deals with a similar classification problem. The applied data augmentation consists of the following geometric transformations: random rotation in the 0–30 range, random uniform scaling with the 1–1.05 factor, random horizontal and vertical translations by 0 to 10 pixels, random horizontal and vertical reflections with a probability of 0.5 for each direction.

The goal of shuffling is to ensure that training set and validation sets are representative of the overall distribution of the data. This is important when using the minibatch gradient descent algorithm in which the minibatch must be representative of the overall dataset. Shuffling after each epoch helps avoid the risk of bad batches even further. This causes each image in the training set to have an independent impact on the model without being biased by the previous images.

All of the training options are presented in Tab. 4. They were chosen empirically in order to perform precise training in a reasonable time frame.

Tab. 4. Available training options.

Training parameter	Value
Mini-batch options	
Max epochs	30
Mini-batch size	10
Shuffle	Every epoch
Validation options	
Validation frequency	16
Validation patience	Infinity
Solver options	
Solving algorithm	Stochastic gradient descent with momentum (SGDM)
Momentum	0.9
Initial learn rate	0.0001
Learn rate schedule	None
Learn rate drop factor	0.1
Learn rate drop period	10
L2 regularization	0.0001

4.3. Classification Accuracy

The effectiveness of CNNs can be measured and compared using the following parameters: classification accuracy (the average accuracy in the case of k-fold cross validation), prediction rates for particular pairs of classes (confusion matrix), training time, and decision time. The most important thing when implementing a CNN-based solution is the classification accuracy, and more specifically, low false prediction rates and high true prediction rates. Table 5 shows the obtained classification accuracy. The prediction rates are stored as confusion matrices in Tabs. 6, 7, and 8. The confusion matrix for a given CNN shows the overall classification efficiency for each class. The correct classification of each class is rep-

Tab. 5. Classification accuracy for the proposed networks.

Network	Accuracy	Decision time	Training time
CNN#1	0.8174	$\sim$ 0.09 s	$\sim$ 5–7 min
CNN#2	0.9783	$\sim$ 0.13 s	$\sim$ 10–12 min
CNN#3	0.9304	$\sim$ 0.15 s	$\sim$ 15–17 min

Tab. 6. Confusion matrix for CNN#1.

True class	Predicted class				True positive	False negative
	day 1	day 2	day 4	day 8		
Day 1	0.715	0.093	0.136	0.056	0.715	0.285
Day 2	0.080	0.900	0.000	0.020	0.900	0.100
Day 4	0.046	0.000	0.874	0.080	0.874	0.126
Day 8	0.025	0.047	0.189	0.739	0.739	0.261

Tab. 7. Confusion matrix for CNN#2.

True class	Predicted class				True positive	False negative
	day 1	day 2	day 4	day 8		
Day 1	1	0	0	0	1	0
Day 2	0	1	0	0	1	0
Day 4	0.024	0.011	0.954	0.011	0.954	0.046
Day 8	0	0.022	0	0.978	0.978	0.022

Tab. 8. Confusion matrix for CNN#3.

True class	Predicted class				True positive	False negative
	day 1	day 2	day 4	day 8		
Day 1	0.904	0.018	0.078	0	0.904	0.096
Day 2	0.020	0.920	0.040	0.02	0.920	0.080
Day 4	0.023	0	0.965	0.012	0.965	0.035
Day 8	0.025	0.047	0.022	0.906	0.906	0.094

resented by the diagonal cells, while elements beyond the diagonal indicate a misclassification. The last two columns contain the true positive rate and the false negative rate for each true class, respectively.

Figure 3 presents receiver operating characteristic (ROC) curves, being the plots of the true positive rate against the false positive rate for a given class. The ROC curve for the best classifier is near the upper left corner of the graph. On the other hand, a diagonal line across the whole plot would represent random guessing. The classifier's overall quality is measured by the area under the ROC curve (AUC). The larger the area under the plot, the better the classification performance. It can be seen that the ROCs for CNN#2 are close to the ROCs of the ideal classifier.

It should be noted that the training time and the number of learnable parameters have very little significance in this research. Furthermore, the training time depends strongly on the hardware setup and the dataset used. In the case of decision time, the differences for the different considered CNN architectures are negligible. The application under consideration is not a real-time solution, so all of the results are very satisfactory. The training time and decision time are shown in Tab. 5.

4.4. Proposed Network

The experimental results show that CNN#2 is the best of the presented CNNs for solving the given problem. The overall classification accuracy of this network is equal to 0.9783. The accuracy for the "day 1" class is 1, for the "day 2" class is 1, for the "day 4" class is 0.954, and for the "day 8" class is 0.978. The learning time was about 12 minutes and the decision time was approximately 0.13 s. CNN#2 has a depth of 165, consists of 244 convolutional layers and 2 fully connected layers, and the number of learnable parameters is 60,648,676. The proposed network is shown in Fig. 4.

Deep learning techniques, in particular convolutional neural networks, are characterized by very high predictive power, but are not easily interpretable by humans. This is sometimes pointed out as a weakness of CNN compared to methods based on expert knowledge. Interpreting a non-linear classifier is important to gain trust into the prediction and to identify potential data selection mistakes. In order to solve this challenge, one can use the so-called importance maps or heatmaps.

Figure 5 shows exemplary heatmaps, calculated using the Grad-CAM (gradient-weighted class activation mapping) algorithm, which correspond to the sample data from the dataset presented in Fig. 1. These importance maps visualize the regions of the analyzed image that activated, the most, the outputs from the last convolution layer while calculating classification scores for specific class labels. Areas on a map with a large value are those that impact the network score for that class the most. There exists a direct analogy to magnetic resonance imaging (MRI) of the human brain, in which different regions activate for different stimulations. If the pixels on a heatmap are red, it means that the last convolutional layer has been strongly activated at a given location. The network considers this feature to be important and valid for a given class label. In this way, one can see that the network

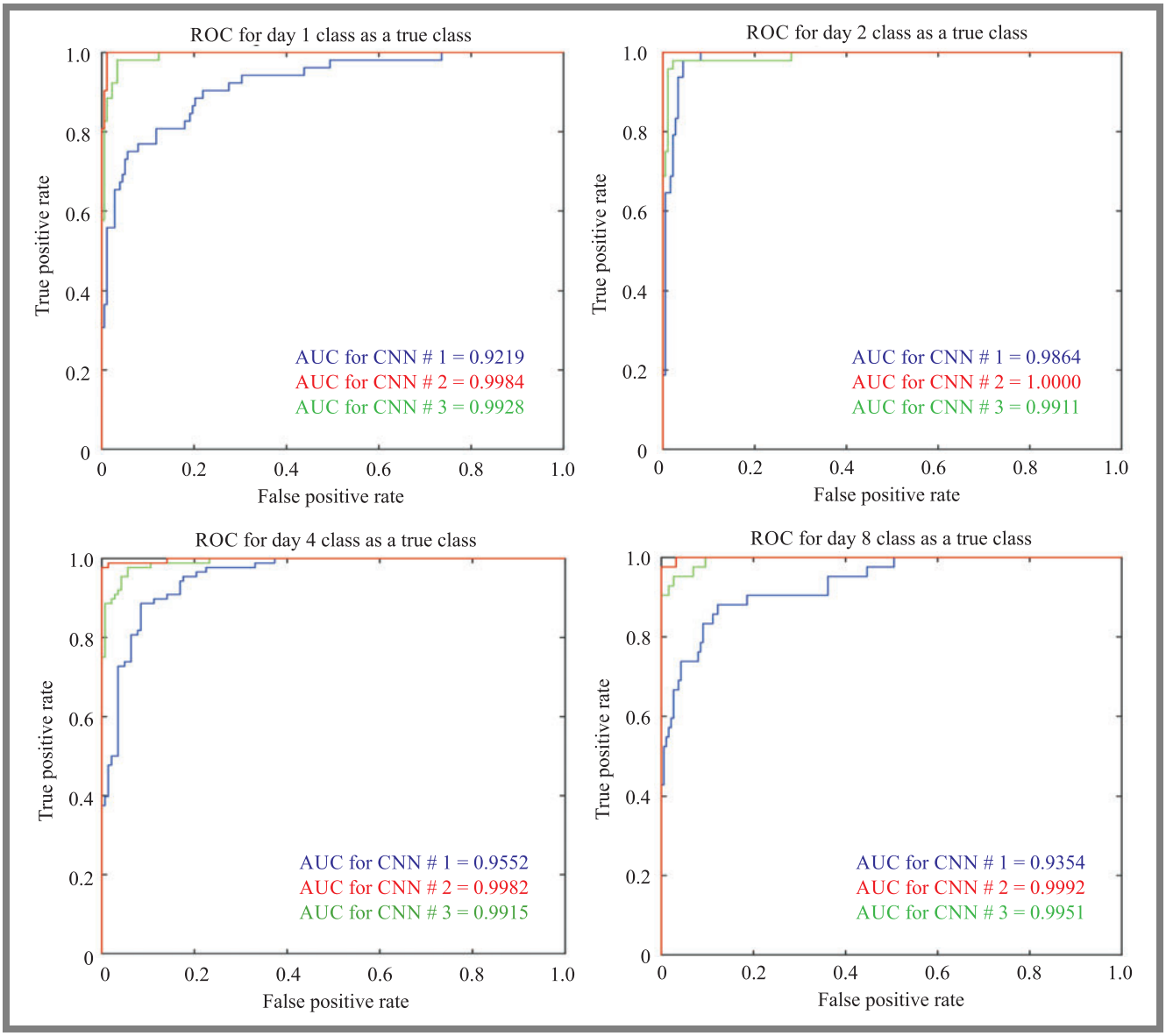

Fig. 3. ROC curves for the considered CNNs.

focuses on what is essential for distinguishing the age of *C. elegans*. By analyzing individual heatmaps, we can see with the naked eye what parts of the data are recognized and taken into consideration by the CNN.

As expected, the network extensively utilizes information from the rim of the nematode's body. This is the very region where the progressively deteriorating body-wall muscles are located. However, quite interestingly, the network also accommodates morphological changes in the pharynx, which is evident in Fig. 5, day 8, but also noticeable on day 2. The CNN appears to include even more regions (day 4). Unfortunately, the organs and tissues from such regions (excluding body-wall muscles) are difficult to determine.

4.5. Comparison of Methods

The three CNNs featured in this paper have been compared with other state-of-the-art methods in terms of classification performance. The methods used for the comparison were tested on the same dataset as the proposed method. The results are presented in Tab. 9. It can be seen that the proposed CNN#2 offers nearly the highest accuracy currently available in the literature. However, when comparing the methods, one should take into account that the proposed CNN#2 comes with all of the advantages of a fully self-learned CNN-based solution without the handcrafted feature selection stage requiring expert knowledge.

The only method that has achieved a similar level of accuracy was [27], but their direct comparison is very difficult here, because it represents a completely different approach to solving the problem. The method applied in [27] uses separate feature selection and classification tools. Feature selection is based on a very sophisticated, hand-crafted, high-dimensional descriptor set, which is then assisted by a specialized SDT algorithm. Afterwards, classification is performed via SVM. On the other hand, the methodology proposed in this paper

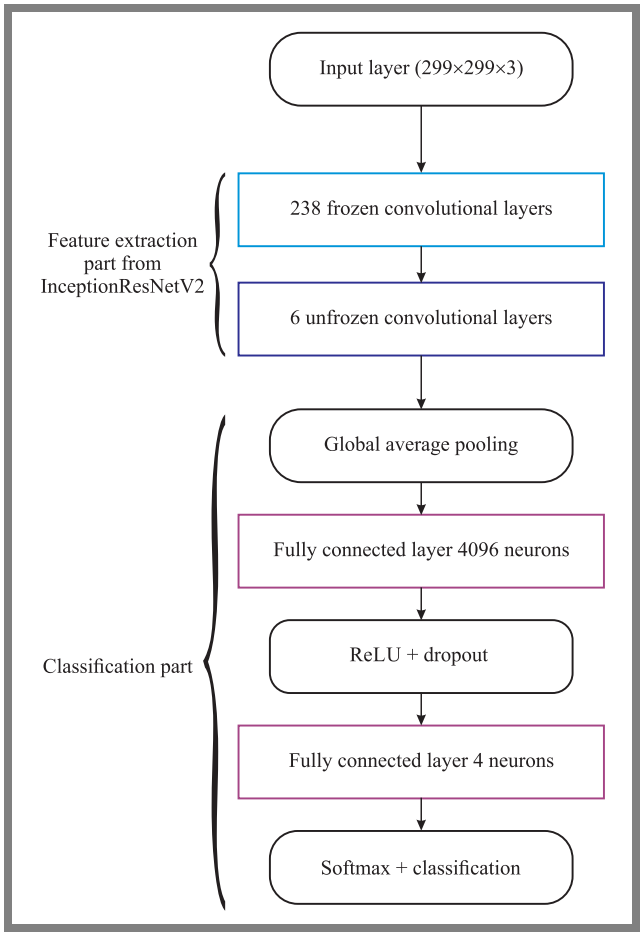

Fig. 4. The proposed CNN for *C. elegans* muscle aging classification.

Tab. 9. Classification accuracy for different state-of-the-art methods.

Method	Accuracy
Song *et al.* [27]	0.979
CNN#2	0.9783
CNN#3	0.9304
Nanni *et al.* [30]	0.9375
Nanni *et al.* [29]	0.9292
Zhou *et al.* [25]	0.896
Lin *et al.* [22]	0.8478
CNN#1	0.8174
Mohan and Sekhar [28]	0.75
Siji *et al.* [26]	0.75
Shamir *et al.* [35]	0.53

offers the same level of effectiveness by adopting a much simpler strategy. Firstly, feature extraction and classification are implemented using one coherent tool (CNN). Secondly, the CNN-based solution is more open should a need to add more classes or extend the dataset arise. Thirdly, the method presented in this paper uses date augmentation to protect against overfitting, which is ignored in [27]. Fourthly, the CNN-based solution is free from SVM flaws, such as high algorithmic complexity, extensive memory requirements, sensitivity to

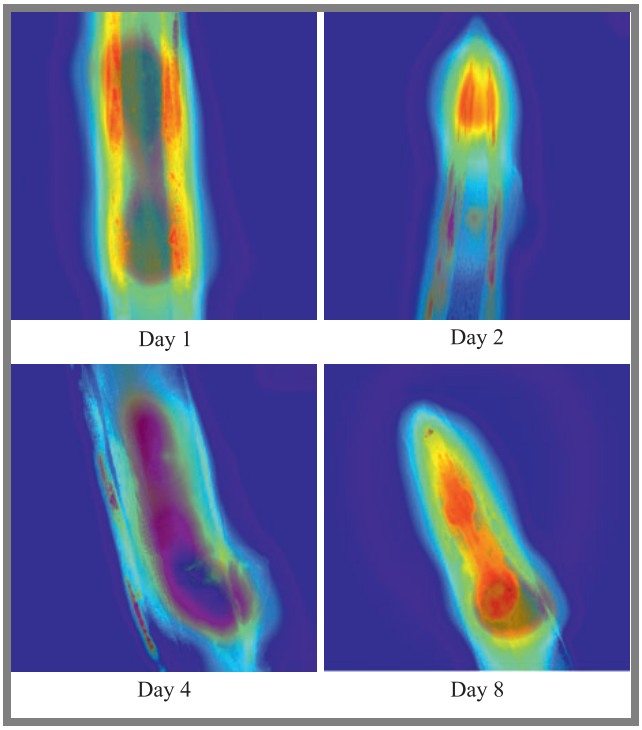

Fig. 5. Examples of heatmaps obtained with the use of CNN#2, in which the network accommodates not only body-wall muscles, but also pharynx, and regions that cannot be unambiguously determined.

outliers, lack of prediction scores (classification probabilities), very time-consuming training, unsuitability for large datasets, underperformance when the dataset is noisy, and underperformance if the number of features for each data point exceeds the number of training data samples. Ultimately, the biggest advantage of the proposed method over the method presented in [27] is that there is no need to develop a hand-crafted feature selection stage.

The next two methods from [29] and [30] produce slightly worse results. In both cases, CNNs are only used as a supplement to handcrafted features. Feature selection consists of local features, dense sampling features, and deep learning features which are then classified using SVM. Firstly, unfortunately, this still requires specialized expert knowledge to create a set of handcrafted features. Secondly, the CNNs presented in [29] and [30] have not been used in an optimal way, because the authors have not unfrozen (allowed re-learning) a single convolution layer and are using the deep learning features learned for a completely different, unrelated classification problem, completely relying on adaptation by SVM. As a result, the presented methodology for using CNNs, even when boosted by handcrafted features, does not provide as good results as the CNN#2 network presented in this paper. Thirdly, these methods also deal with the drawbacks of SVM outlined in the previous paragraph. In this context, the SVM acts as a bottleneck that blocks the useful classification properties of the CNN.

It should be noted that the method presented in [25], although less accurate than CNN#2, is intended for analyzing 3D images. In order for the proposed methodology to be able to

operate on 3D images, additional studies involving extrapolation to three dimensions would be necessary.

The last compared work that requires an additional comment is presented in [22]. It is a method that calculates *C. elegans* muscle age by regression. It is very difficult to fairly compare the regression method to the classification method, e.g. the results of the regression method can be interpreted as classification results with a class resolution of one day (i.e. many classes differing by one day). However, in [22], it is necessary to add some metadata or a description to the image, which implies additional work that needs to be supervised by a human. For the method proposed in this paper, only an image is entered as input, with no additional metadata or description.

5. Conclusion

This paper presents the results of research on *C. elegans* muscle age classification using CNNs with self-learned features only. The problem under consideration is very difficult for humans to solve. Therefore, machine learning-based solutions are being sought. In anti-aging drug research, the proposed CNN would be a very fast and effective age determination method, which would lead to reductions in time and costs.

The challenge in these studies was the small dataset, which made it difficult to train the CNNs effectively and posed the risk of overfitting. Transfer learning and data augmentation techniques were used to solve this problem. VGG16, InceptionV3, and InceptionResNetV2 networks were selected as the base networks for transfer learning. The impact of unfreezing successive convolutional layers on the classification accuracy was investigated, as well as the impact of extending the classifier model with additional fully connected layers. The overall classification accuracy of the proposed network is equal to 0.9783. The proposed network has a depth of 165, consists of 244 convolutional layers and 2 fully connected layers, and the number of learnable parameters is 60,648,676. The proposed network was compared with other solutions described in the literature.

The achieved results prove that CNNs can be successfully used for the classification of *C. elegans* muscle aging without handcrafted features. Moreover, given the fact that CNNs are free from the disadvantages of analytical feature selection and other machine learning methods, the results achieved are improvements in the research problem under consideration and are of great importance for anti-aging studies.

Our study shows that during the early stages of ageing, *C. elegans* exhibits morphological alterations not only in the muscle tissue but also in other parts of the nematode's body, e.g. its pharynx. The CNNs proposed by us are capable of detecting such changes with a high degree of accuracy, therefore providing reliable information regarding potentially new sources of ageing.

Future research may include the following. Retraining the network on an extended dataset with more classes, e.g. each class could represent a separate day from a 2-week span. Integrating the proposed CNN with region-based active contour models for medical image segmentation [41]. Application of the proposed CNN to solve other classification problems in diagnostics [34], [35].

Authors' Contributions

Conceptualization: Damian Panas,
Data curation: Damian Panas,
Formal analysis: Mariusz Dzwonkowski,
Acquisition of funding: Bartosz Czaplewski and Mariusz Dzwonkowski,
Investigation: Bartosz Czaplewski, Mariusz Dzwonkowski and Damian Panas.
Methodology: Bartosz Czaplewski and Mariusz Dzwonkowski,
Project administration: Bartosz Czaplewski;
Resources: Damian Panas,
Software: Bartosz Czaplewski.
Supervision: Bartosz Czaplewski, Validation, Bartosz Czaplewski; Visualization: Bartosz Czaplewski,
Writing – original draft: Bartosz Czaplewski, Mariusz Dzwonkowski and Damian Panas. Writing – review and editing: Bartosz Czaplewski and Mariusz Dzwonkowski.

Funding

This work was supported by a ministerial research subsidy awarded to Gdańsk University of Technology.

References

[1] A.G. Alexander, *et al.*, "Use of *Caenorhabditis elegans* as a model to study Alzheimer's disease and other neurodegenerative diseases", *Front. Genet.*, vol. 5, 2014 (DOI: 10.3389/fgene.2014.00279).

[2] E. Moreno-Arriola, *et al.*, "*Caenorhabditis elegans*: A useful model for studying metabolic disorders in which oxidative stress is a contributing factor", *Oxid. Med. Cell. Longev.*, vol. 2014, 2014 (DOI: 10.1155/2014/705253).

[3] A.E. Apostolakou, *et al.*, "Exploring the conservation of Alzheimer-related pathways between H. sapiens and *C. elegans*: a network alignment approach", *Sci. Rep.*, vol. 11, no. 1, 2021 (DOI: 10.1038/s41598-021-83892-9).

[4] K. Podshivalova, *et al.*, "How a mutation that slows aging can also disproportionately extend end-of-life decrepitude", *Cell Rep.*, vol. 19, no. 3, pp. 441–450, 2017 (DOI: 10.1016/j.celrep.2017.03.062).

[5] M.M. Abdelsamea, "A semi-automated system based on level sets and invariant spatial interrelation shape features for *Caenorhabditis elegans* phenotypes", *J. Vis. Commun. Image Represent.*, vol. 41, pp. 314–323, 2016 (DOI: 10.1016/j.jvcir.2016.10.011).

[6] H.A. Tissenbaum, "Using *C. elegans* for aging research", *Invertebr. Reprod. Dev.*, vol. 59, no. sup1, pp. 59–63, 2015 (DOI: 10.1080/07924259.2014.940470).

[7] S. Zhang, F. Li, T. Zhou, G. Wang, and Z. Li, "*Caenorhabditis elegans* as a useful model for studying aging mutations", *Front. Endocrinol. (Lausanne)*, vol. 11, 2020 (DOI: 10.3389/fendo.2020.554994).

[8] J.J. Collins *et al.*, "The measurement and analysis of age-related changes in *Caenorhabditis elegans*", *WormBook: The Online Review of C. elegans Biology. Pasadena, CA, WormBook*, 2005–2018 (URL: https://www.ncbi.nlm.nih.gov/books/NBK116075/).

[9] D.S. Wilkinson, R.C. Taylor, and A. Dillin, "Analysis of aging in *Caenorhabditis elegans*", *Methods Cell Biol.*, vol. 107, pp. 353–381, 2012 (DOI: 10.1016/B978-0-12-394620-1.00012-6).

[10] L.A. Herndon, *et al.*, "Stochastic and genetic factors influence tissue-specific decline in ageing *C. elegans*", *Nature*, vol. 419, no. 6909, pp. 808–814, 2002 (DOI: 10.1038/nature01135).

[11] H.G. Son, O. Altintas, E.J.E. Kim, S. Kwon, and S-J.V. Lee, "Age-dependent changes and biomarkers of aging in *Caenorhabditis elegans*", *Aging Cell*, vol. 18, no. 2, 2019 (DOI: 10.1111/acel.12853).

[12] G.J. Lithgow, "The Future of Worm Ageing", *Ageing: Lessons from C. elegans. Healthy Ageing and Longevity, A. Olsen, M. Gill, Eds., Switzerland, Cham: Springer International Publishing, A. Olsen, M. Gill (eds) Ageing: Lessons from C. elegans. Healthy Ageing and Longevity. Springer, Cham.*, pp. 431–435 2017 (DOI: 10.1007/978-3-319-44703-2_19).

[13] A. Olsen and M.S. Gill, "Introduction in Ageing: Lessons from *C. elegans*. Healthy Ageing and Longevity, *A. Olsen, M. Gill, Eds., Switzerland, Cham: Springer International Publishing*, pp. 1–7, 2017 (DOI: 10.1007/978-3-319-44703-2_1).

[14] D. Gems and L. Partridge, "Genetics of longevity in model organisms: debates and paradigm shifts", *Annu. Rev. Physiol.*, vol. 75, pp. 621–644, 2013 (DOI: 10.1146/annurev-physiol-030212-183712).

[15] S.S. Yadav and S.M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis", *J. Big Data*, vol. 6, no. 133, 2019 (DOI: 10.1186/s40537-019-0276-2).

[16] B. Czaplewski and M. Dzwonkowski, "A novel approach exploiting properties of convolutional neural networks for vessel movement anomaly detection and classification", *ISA Transactions*, vol. 119, pp. 1–16, 2022 (DOI: 10.1016/j.isatra.2021.02.030).

[17] B. Czaplewski, "An improved convolutional neural network for steganalysis in the scenario of reuse of the stego-key", *I. Tetko, V. Kůrková, P. Karpov, F. Theis (eds) Artificial Neural Networks and Machine Learning – ICANN 2019, Lecture Notes in Computer Science*, vol. 11729, Springer, Cham. 2019 (DOI: 10.1007/978-3-030-30508-6_7).

[18] J. Zhang, K. Shao, and X. Luo, "Small sample image recognition using improved Convolutional Neural Network", *J. Vis. Commun. Image Represent.*, vol. 55, pp. 640–647, 2018 (DOI: 10.1016/j.jvcir.2018.07.011).

[19] M. Claro, *et al.*, "An hybrid feature space from texture information and transfer learning for glaucoma classification", *J. Vis. Commun. Image Represent.*, vol. 64, 102597, 2019 (DOI: 10.1016/j.jvcir.2019.102597).

[20] C. Restif and D. Metaxas, "Tracking the Swimming Motions of *C. elegans* Worms with Applications in Aging Studies", *Med. Image Comput. Comput. Assist. Interv.*, D. Metaxas *et al.* Eds., Germany, Berlin, Heidelberg: Springer, vol. 5241, pp. 35–42, 2008 (DOI: 10.1007/978-3-540-85988-8_5).

[21] J. Johnston, W.B. Iser, D.K. Chow, I.G. Goldberg, and C.A. Wolkow, "Quantitative image analysis reveals distinct structural transitions during aging in *Caenorhabditis elegans* tissues", *PLoS One*, vol. 3, no. 7, 2008 (DOI: 10.1371/journal.pone.0002821).

[22] J-L. Lin, W.-L. Kuo, Y-H. Huang, T-L. Jong, A-L. Hsu, and W-H. Hsu, "Using convolutional neural networks to measure the physiological age of *Caenorhabditis elegans*", *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 18, no. 6, pp. 2724–2732, 2021 (DOI: 10.1109/TCBB.2020.2971992).

[23] L. Shamir, N. Orlov, D.M. Eckley, T. Macura, J. Johnston, and I.G. Goldberg, "Wndchrm – an open source utility for biological image analysis", *Source Code Biol. Med.*, vol. 3, no. 13, 2008 (DOI: 10.1186/1751-0473-3-13).

[24] N. Orlov, L. Shamir, T. Macura, J. Johnston, D.M. Eckley, and I.G. Goldberg, "WND-CHARM: Multi-purpose image classification using compound image transforms", *Pattern Recognit. Lett.*, vol. 29, no. 11, pp. 1684–1693, 2008 (DOI: 10.1016/j.patrec.2008.04.013).

[25] J. Zhou, S. Lamichhane, G. Sterne, B. Ye, and H. Peng, "BIOCAT: a pattern recognition platform for customizable biological image classification and annotation", *BMC Bioinformatics*, vol. 14, no. 291, 2013 (DOI: 10.1186/1471-2105-14-291).

[26] K.K. Siji, B.S. Mathew, R. Chandran, B.S. Shajeemohan, and K.S. Shanthini, "Feature selection, optimization and performance analysis of classifiers for biological images", *2014 IEEE National Conference on Communication, Signal Processing and Networking (NCCSN)*, pp. 1–5, 2014 (DOI: 10.1109/NCCSN.2014.7001150).

[27] Y. Song, W. Cai, H. Huang, D. Feng, Y. Wang, and M. Chen, "Bioimage classification with subcategory discriminant transform of high dimensional visual descriptors", *BMC Bioinformatics*, vol. 17, no. 465, 2016 (DOI: 10.1186/s12859-016-1318-9).

[28] B.S. Shajee Mohan and C.C. Sekhar, "Distance metric learnt kernel based SVMs for semi-supervised pattern classification", *2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR)*, pp. 1–6, 2017 (DOI: 10.1109/ICAPR.2017.8592956).

[29] L. Nanni, S. Ghidoni, and S. Brahnam, "Handcrafted vs. non-handcrafted features for computer vision classification", *Pattern Recognit.*, vol. 71, pp. 158–172, 2017 (DOI: 10.1016/j.patcog.2017.05.025).

[30] L. Nanni, S. Brahnam, S. Ghidoni, and A. Lumini, "Bioimage classification with handcrafted and learned features", *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 16, no. 3, pp. 874–885, 2019 (DOI: 10.1109/TCBB.2018.2821127).

[31] L. Nanni, S. Ghidoni, and S. Brahnam, "Ensemble of convolutional neural networks for bioimage classification", *Appl. Comput. Inform.*, vol. 17, no. 1, 2020 (DOI: 10.1016/j.aci.2018.06.002).

[32] C.K. Ahn, C. Heo, and H. Jin, "A novel deep learning-based approach to high accuracy breast density estimation in digital mammography", *Med. Imaging: Computer-Aided Diagnosis*, vol. 10134, SPIE, 2017 (DOI: 10.1117/12.2254264).

[33] X. Li, Y. Quin, Z. Liu, "Pretraining improves deep learning based tissue microstructure estimation", *Computational Diffusion MRI, Springer, Cham*, 173–185, 2021 (DOI: 10.1007/978-3-030-73018-5_14).

[34] –, *IICBU dataset website*, 2008 (URL: https://ome.irp.nia.nih.gov/iicbu2008/).

[35] L. Shamir, N. Orlov, D.M. Eckley, T. Macura, and I.G. Goldberg, "IICBU 2008: a proposed benchmark suite for biological image analysis", *Med. Biol. Eng. Comput.*, vol. 46, no. 9, pp. 943–947, 2008 (DOI: 10.1007/s11517-008-0380-5).

[36] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", 2014 (DOI: 10.48550/arXiv.1409.1556).

[37] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826, 2016 (DOI: 10.1109/CVPR.2016.308).

[38] C. Szegedy, S. Ioffe, V. Vanhoucke, and A.A. Alemi., "Inception-v4, Inception-ResNet and the impact of residual connections on learning", *Proc. Conf. AAAI Artif. Intell.*, vol. 31, no. 1, pp. 4278–4284, 2017 (DOI: 10.1609/aaai.v31i1.11231).

[39] –, *ImageNet – image database*, 2021 (URL: http://www.image-net.org).

[40] C. Shorten and T.M. Khoshgoftaar, "A survey on image data augmentation for deep learning", *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019 (DOI: 10.1186/s40537-019-0197-0).

[41] A. Pratondo, C-K. Chui, and S-H. Ong, "Integrating machine learning with region-based active contour models in medical image segmentation", *Journal of Visual Communication and Image Representation*, vol. 43, pp. 1–9, 2017 (DOI: 10.1016/j.jvcir.2016.11.019).

Bartosz Czaplewski received the M.Sc. and Ph.D. degrees from the Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, in 2011 and 2015, respectively. Currently, he works as an assistant professor at Gdańsk University of Technology, Poland. His research interests include cybersecurity, cryptography, digital fingerprinting, image processing, machine learning, and deep learning.

https://orcid.org/0000-0001-7904-5567

E-mail: bartosz.czaplewski@eti.pg.edu.pl

Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Department of Teleinformation Networks, Poland

Mariusz Dzwonkowski received the M.Sc. and Ph.D. degrees from the Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, in 2011 and 2017, respectively. He is an assistant professor with Medical University of Gdańsk, Poland and with Gdańsk University of Technology, Poland. His research interests include cryptography, hypercomplex algebra, information coding and reversible data hiding.

https://orcid.org/0000-0003-3580-7448

E-mail: mard@gumed.edu.pl

Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Department of Teleinformation Networks, Poland

Medical University of Gdańsk, Faculty of Health Sciences, Department of Radiology Informatics and Statistics, Poland

Damian Panas is a bioinformatician at the Molecular Biology Laboratory at the Institute of Animal Reproduction and Food Research, Polish Academy of Sciences. He received his M.Sc. degree from the Faculty of Applied Physics and Mathematics, Gdańsk University of Technology, Poland, and his Ph.D degree from the Faculty of Health Sciences, Medical University of Gdańsk, Poland. His areas of interest include mathematical modeling, statistics, machine learning, computational biology, and multi-omic data analysis.

https://orcid.org/0000-0003-2181-2414

E-mail: d.panas@pan.olsztyn.pl

Institute of Animal Reproduction and Food Research of the Polish Academy of Sciences, Molecular Biology Laboratory, Poland