ISSN 0554-6397
UDK: 004.032.26
Original scientific paper
Received: 5th November 2020.

Daniel Štifanić
E-mail: dstifanic@riteh.hr
Jelena Musulin
E-mail: jmusulin@riteh.hr
Zlatan Car
E-mail: car@riteh.hr
University of Rijeka, Faculty of Engineering, Vukovarska 58, 51000 Rijeka, Croatia
Robert Čep
E-mail: robert.cep@vsb.cz

Use of Convolutional Neural Network for Fish Species Classification

Abstract

Fish population monitoring systems based on underwater video recording are becoming more popular nowadays, however, manual processing and analysis of such data can be time-consuming. Therefore, by utilizing machine learning algorithms, the data can be processed more efficiently. In this research, authors investigate the possibility of convolutional neural network (CNN) implementation for fish species classification. The dataset used in this research consists of four fish species (Plectroglyphidodon dickii, Chromis chrysura, Amphiprion clarkii, and Chaetodon lunulatus), which gives a total of 12859 fish images. For the aforementioned classification algorithm, different combinations of hyperparameters were examined as well as the impact of different activation functions on the classification performance. As a result, the best CNN classification performance was achieved when Identity activation function is applied to hidden layers, RMSprop is used as a solver with a learning rate of 0.001, and a learning rate decay of 1e-5. Accordingly, the proposed CNN model is capable of performing high-quality fish species classifications.

Keywords: Machine learning, convolutional neural network, fish species classification, underwater video

1. Introduction

Fish population monitoring is important, and it can be considered as a challenging task. Nowadays, many fish species are targeted by fisheries, for that reason is crucial to monitor their health and status as much as possible [1]. In underwater environments, manual methods usually require divers for the process of species classification and fish population estimation which is time-consuming [2]. In recent years fish monitoring approaches based on underwater video recording are becoming more popular due to affordable cost and high-quality video sequences. Therefore, the video-based fish monitoring approach is considered as non-invasive and effective according to Whitmarsh et al. (2017) [3]. Data collected in such a way can be processed and analyzed utilizing machine learning (ML) algorithms since they have been proven successful in image classification and pattern recognition problems [4, 5]. In addition, ML approaches also proven successful in solving complex problems in various fields of science, medicine, economy, maritime, and technology [6 – 16]. The algorithm that may provide satisfactory results in terms of fish species classification is convolutional neural network (CNN) [17]. In this algorithm within each convolutional layer, a convolution operation is performed by which unique feature maps are generated [18]. The data used in this research are publicly available and obtained from its authors' web page [2].

The aim of this research is to determine the CNN model performance that can be achieved for fish species classification as well as to examine the impact of different hyperparameter combination on the classification result. Implementation of the CNN model for fish species classification should help experts and reduce the time required for manual processing and analysis of image data.

Sajjad et al. (2019) proposed a novel CNN based multi-grade brain tumor classification system, and the results show that their proposed system has a better performance compared to existing methods [19]. Lorencin et al. (2019) present the CNN-based approach for marine object recognition and achieve good results [20]. Gao et al. (2018) present an object classification method based on CNN and image upsampling theory for vision and light detection and ranging fusion of autonomous vehicles. The experimental results show the effectiveness of the proposed method [21].

In order to train a CNN model, four clusters of fish species will be used since the original dataset with 23 clusters is highly unbalanced. These clusters are Plectroglyphidodon dickii, Chromis chrysura, Amphiprion clarkii, and Chaetodon lunulatus, which is resulting in a total number of 12859 fish images. The grid search algorithm will be performed to find the optimal combination of model hyperparameters. Afterward, the obtained results will be compared in order to examine the impact of different hyperparameter combinations on the performance measure.

2. Materials and Methods

In this section, brief descriptions of the dataset, convolutional neural network, and hyperparameter optimization algorithm are given along with a subsection where the confusion matrix as model performance measurement is presented.

2.1. Dataset description

The original dataset consists of 23 clusters, where each one of them is presented by a representative fish species. This data are obtained from a live video dataset, which resulted in 27370 verified fish images. Each fish species is manually labeled which makes this dataset ideal for image classification task. The original dataset is publicly available from its authors' web page [2]. Since the original dataset is highly imbalanced, only four clusters of fish species are used in order to create a dataset for this research. All four clusters have approximately the same number of images which makes the dataset nearly balanced. These clusters are: Plectroglyphidodon dickii, Chromis chrysura, Amphiprion clarkii, and Chaetodon lunulatus. The total number of fish images in created dataset is 12859, where 70% (9001) is used as a training set and the other 30% (3858) is used as a testing set. As a final step, these images are resized to 100 x 100 pixels. Resized images of fish species are shown in Figure 1.

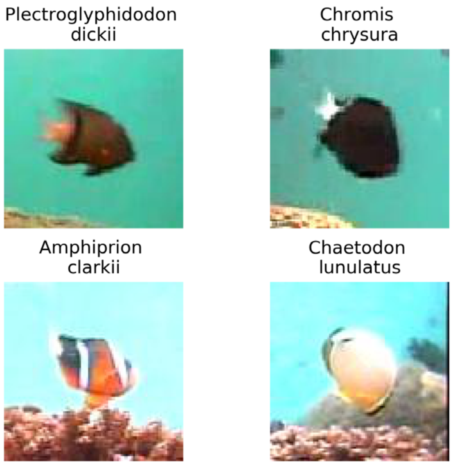

Figure 1 – Sample images from the dataset.
Each individual fish image represents one species.

2.2. Convolutional Neural Network

Convolutional neural network (CNN) is a class of feed-forward artificial neural network (ANN) commonly applied to analyzing images [22]. Characteristics like shared weights, local receptive fields and spatial sub-sampling ensure scale, distortion and shift-invariance of CNN models [23]. As the name suggests, this type of network uses a mathematical operation called convolution in order to generate a unique feature map for each convolutional layer, in other words, features are extracted from input images with convolutional layers [24]. Each convolution layer consists of a specific kernel function by which certain features can be extracted [25]. In complex neural network architectures between adjacent layers usually exists two types of connections; locally connected layer and fully-connected layer [26]. Utilizing local connection, the number of parameters that need to be calculated is reduced since each neuron within a layer can only receive a small group of pixels from the input image. Fully-connected layers are usually placed at the end of CNN and used to classify the image into a correct label [27]. The output from the previous layer must be flattened into a vector before is feed into a fully-connected layer. In addition to the aforementioned layers, pooling layers are also commonly used in CNNs for feature map dimensionality reduction and to achieve invariance to translation [28]. Pooling layer is typically applied after the convolutional layer and can calculate the maximum or average value of the feature map.

Since CNNs have been proved successful in areas of image recognition [29] and classification [30], it is appropriate choice when dealing with the problem of fish species classification. CNN architecture for fish species classification is visualized in Figure 2.

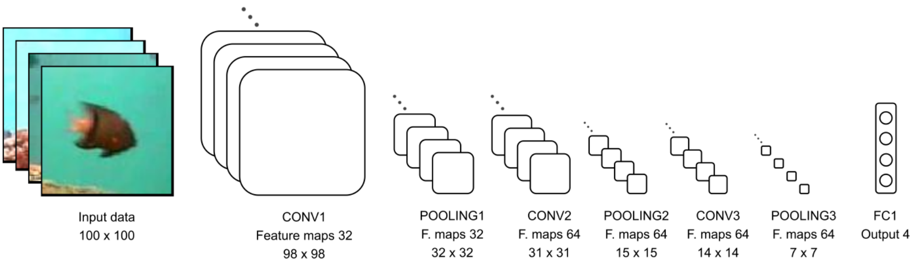

Figure 2 – CNN architecture for fish species classification (CONV1 – first convolutional layer with 32 feature maps, POOLING1 – first pooling layer with 32 feature maps, CONV2 – second convolutional layer with 64 feature maps, POOLING2 – second pooling layer with 64 feature maps, CONV3 – third convolutional layer with 64 feature maps, POOLING3 – third pooling layer with 64 feature maps, FC1 – output layer with one neuron for each class, four classes in total)

According to Figure 1, all input images are rescaled to 100 x 100 pixels before applying the first CONV layer. CONV1 layer contains a kernel size of 3 x 3, while CONV2 and CONV3 layers contain a kernel size of 2 x 2. POOLING1 is a 3 x 3 maximal pooling layer, POOLING2 is a 2 x 2 maximal pooling layer and POOLING3 is a 2 x 2 average pooling layer. As input data propagates through CNN, feature size after each layer is decreasing which means that the kernel size directly affects the resulting size of the feature map. Feature maps sizes are CONV1 – 98 x 98, POOLING1 – 32 x 32, CONV2 – 31 x 31, POOLING2 – 15 x 15, CONV3 – 14 x 14, POOLING4 – 7 x 7. The last layer takes the flattened vector of POOLING4 feature map and calculates the final probabilities for each class. Activation function used in output layer is Softmax.

2.3. Hyperparameters of CNN

The influence of hyperparameters on CNN model performance is significant, therefore, it is necessary to implement the grid search algorithm, by which, an optimal combination of parameters can be determined. Such an approach requires a manually defined search-space of possible parameters [31]. After defining the search-space, the CNN model is trained and the performance result is stored for every possible parameter combination.

This research includes adjusting the following hyperparameters of the CNN model: activation function for convolutional layers, type of solver, learning rate, and learning rate decay. An overview of possible hyperparameters used in the model training process is shown in Table 1.

Table 1 – Combinations of hyperparameters used in CNN model training process

Hyperparameter	**Parameter value**
Activation function for convolutional layers	ELU, ReLU, Identity, Sigmoid, Tanh, Swish
Solver	SGD, Adam, RMSprop, Adadelta, Adamax
Learning rate	0.1, 0.01, 0.001, 0.0001, 0.00001
Learning rate decay	1e-5, 1e-6, 1e-7, 1e-8, 1e-9

2.4. Model Performance Evaluation

In this research, CNN models are evaluated by utilizing performance measurement called confusion matrix. This way, the classification performance of M-class problems can be visually interpreted by observing the relations between predicted values and the true ones [32]. For example, in the case of a two-class classification problem, four possible results can be derived: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). The confusion matrix in a two-class classification problem is shown in Table 2.

Table 2 – Example of confusion matrix in a two-class classification problem

		Predicted Class	
		Positive	Negative
Actual Class	Positive	TP	FP
	Negative	FN	TN

The elements in the confusion matrix are represented with n_{ij}, where i is considered as a row identifier, and j is considered as a column identifier. Moreover, TP represents correctly classified, FP represents incorrectly classified, FN represents incorrectly rejected, while TN is correctly rejected [33]. In the confusion matrix, the misclassified elements are located out of the main diagonal n_{ii} while correctly classified ones are placed in the main diagonal. Furthermore, the classification accuracy (ACC) can be easily derived from the confusion matrix as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

3. Results and discussion

The CNN model with architecture presented in Figure 2. was trained with different combinations of hyperparameters in order to achieve high-quality classification. Computations were performed using Intel i5-9400 processor along with 16 GB of RAM. Best result with the highest classification accuracy value is achieved when Identity activation function is applied to hidden layers, RMSprop is used as a solver with a learning rate of 0.001, and a learning rate decay of 1e-5. Moreover, a confusion matrix is used in order to visually interpret the classification performance of the CNN model as shown in Table 3.

Table 3 – Confusion matrix in a four-class classification problem (n: total number of images).

		CNN predicted label			
	n = 3858	**Plectroglyphidodon dickii**	**Chromis chrysura**	**Amphiprion clarkia**	**Chaetodon lunulatus**
True label	**Plectroglyphidodon dickii**	789	15	1	0
	Chromis chrysura	3	1075	0	0
	Amphiprion clarkia	1	0	1212	2
	Chaetodon lunulatus	0	0	1	759

The horizontal axis of the confusion matrix represents a predicted label while the vertical axis represents true label. According to Eq. (1), the overall classification accuracy value is 99.40 %, thereby the proposed CNN model is capable of performing high-quality fish species classification. From Table 2., it can be concluded that within Plectroglyphidodon dickii class, 16 out of 805 images are misclassified which is approximately 0.0199 %. CNN model also misclassified 3 out of 1078 (≈ 0.0028 %) images from Chromis chrysura class, 3 out of 1215 (≈ 0.0025 %) images from Amphiprion clarkia class, and 1 out of 760 (≈ 0.0013 %) from Chaetodon lunulatus class. The best performing model successfully classified 3835 out of 3858 images of fish species.

Moreover, the ten best combinations of hyperparameters that achieve high values of performance measures are shown in Table 4. Additionally, the computation time required to train the CNN models is also measured and included in the following table.

Table 4 – Experimental results obtained with the ten best combinations of hyperparameters

Number of hyperparameter combination	Activation function	Solver	Learning rate	Learning rate decay	Accuracy [%]	Computation time [s]
1.	Identity	RMSprop	0.001	1e-5	99.40	40.37
2.	ELU	RMSprop	0.001	1e-6	99.14	52.45
3.	ELU	Adamax	0.01	1e-8	98.89	52.26
4.	ReLU	Adamax	0.01	1e-5	98.81	45.74
5.	Tanh	Adam	0.001	1e-8	98.81	50.71
6.	ReLU	SGD	0.1	1e-7	98.50	45.95
7.	Identity	Adadelta	0.1	1e-6	98.39	41.58
8.	ReLU	Adam	0.01	1e-7	98.32	43.98
9.	Tanh	RMSprop	0.001	1e-7	98.19	51.57
10.	Identity	Adam	0.1	1e-6	97.62	41.16

CNN models that achieve a value of performance measure higher than 99 % use RMSprop as solver with the same learning rate of 0.001. Furthermore, when all results are summed up, it can be seen that all ten best combinations of hyperparameters achieved an accuracy value of 97.62 % or higher.

If the CNN models are compared in terms of computation time, the model with the highest accuracy also has the shortest computation time of 40.37 seconds. Additionally, CNN models where the Identity activation function is utilized for the design of convolutional layers required the shortest computation time. Graphical representation of computation time for each hyperparameter combination is shown in Figure 3.

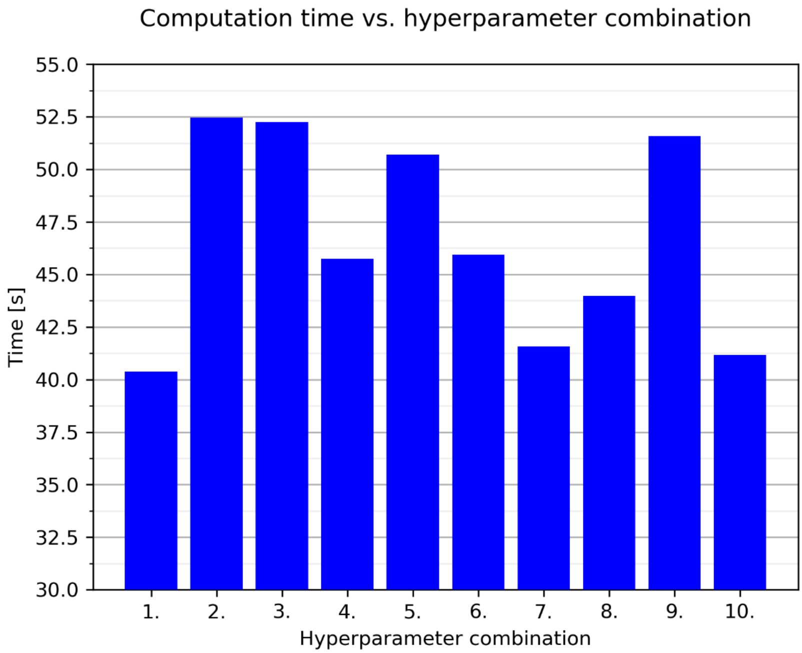

Figure 3 – Computation time versus the hyperparameter combination of the model

In order to examine the impact of activation functions on classification performance, six CNN models were trained with different activation functions in convolutional layers. The CNN architecture was the same as presented in Figure 2. while hyperparameters, except activation functions, are the ones that achieved the best classification performance (solver: RMSprop, learning rate: 0.001, learning rate decay: 1e-5). Classification performances of six CNN models with different activation functions in convolutional layers are compared in Figure 4.

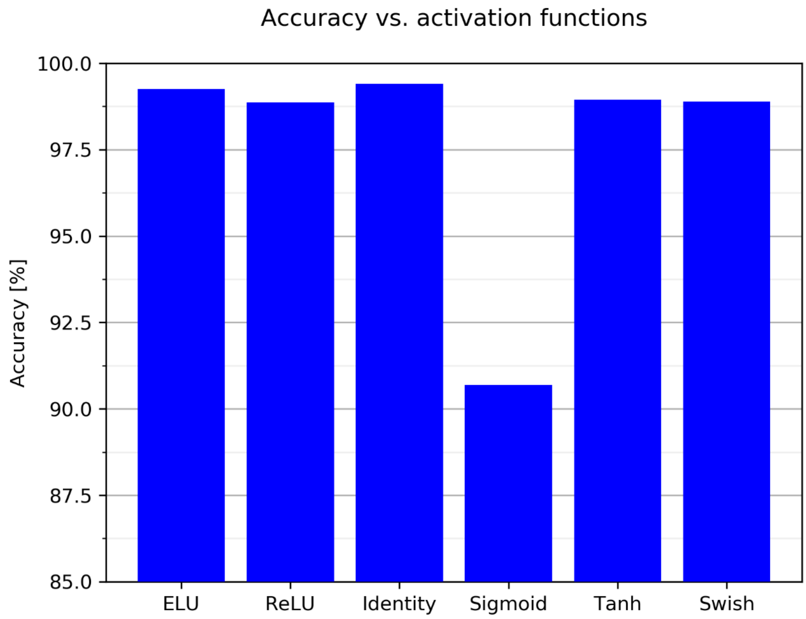

Figure 4 – Accuracy value versus the activation functions for CNN model

It can be concluded that for the aforementioned CNN model configuration, the lowest accuracy value of 90.69 % is achieved if Sigmoid activation function is utilized for the design of convolutional layers. Furthermore, in the case of Tanh, Swish, and ReLU activation functions, the performance measure is almost the same with the values of 98.94 % (Tanh), 98.89 % (Swish), and 98.86 % (ReLU). Accuracy value of 99.25 % is achieved with ELU activation function, while the highest value of 99.40 % is achieved with Identity activation function. From the obtained results, it can be seen that CNN models with the appropriate selection of hyperparameters are capable to perform fish species classification from images of underwater environments with high values of performance measures.

Sun et al. (2016) propose a framework to explicitly learn the discriminative features from low-resolution underwater images and obtained results are promising for fish recognition on underwater image datasets [34]. Deep and Dash (2019) demonstrate the effectiveness of the proposed hybrid Convolutional Neural Network framework by performing fish species classifications. Experimental results show the ability of the proposed framework to perform high-quality classification with high values of evaluation measures [35]. Salman et al. (2016) demonstrate the use of a Convolutional Neural Network in a hierarchical feature combination setup and achieve a correct classification rate of more than 90 percent [36].

4. Conclusion

In this research, CNN was utilized in order to perform a classification of fish species from images obtained with an underwater camera. Obtained results show that it is possible to achieve a high-quality classification with CNNs, thereby, such approach can be used to perform fish species recognition. However, the selection of CNN architecture and hyperparameter combination is crucial for achieving a high value of performance measure. With the implementation of this algorithm, the time needed for manual processing and analysis of image data can be significantly reduced.

Future work includes developing an advanced automatic system for fish species recognition and detection as well as to train the existing models on diverse data in order to build more robust systems.

Acknowledgments

This research has been (partly) supported by the CEEPUS network CIII-HR-0108, European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS), project CEKOM under the grant KK.01.2.2.03.0004, CEI project "COVIDAi" (305.6019-20) and University of Rijeka scientific grant uniri-tehnic-18-275-1447

References

1. Boom, B. J., Huang, P. X., Beyan, C., Spampinato, C., Palazzo, S., He, J., ... & Chen-Burger, Y. H. (2012). Long-term underwater camera surveillance for monitoring and analysis of fish populations. *VAIB12*.
2. Boom, B. J., Huang, P. X., He, J., & Fisher, R. B. (2012, November). Supporting ground-truth annotation of image datasets using clustering. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)* (pp. 1542-1545). IEEE.
3. Whitmarsh, S. K., Fairweather, P. G., & Huveneers, C. (2017). What is Big BRUVver up to? Methods and uses of baited underwater video. *Reviews in Fish Biology and Fisheries*, *27*(1), 53-73.
4. Zhang, M., Li, W., & Du, Q. (2018). Diverse region-based CNN for hyperspectral image classification. *IEEE Transactions on Image Processing*, *27*(6), 2623-2634.
5. Chowdhury, A., Kautz, E., Yener, B., & Lewis, D. (2016). Image driven machine learning methods for microstructure recognition. *Computational Materials Science*, *123*, 176-187.
6. Lorencin, I., Anđelić, N., Baressi Šegota, S., Štifanić, D., Musulin, J., Mrzljak, V., Markova-Car, E., Car, Z. (2020). Dataset Size-based Approach in Design of Artificial Neural Network for Breast Cancer Diagnosis. *World of Health, 3, 13-19.*
7. Car, Z., Baressi Šegota, S., Anđelić, N., Lorencin, I., & Mrzljak, V. (2020). Modeling the Spread of COVID-19 Infection Using a Multilayer Perceptron. *Computational and Mathematical Methods in Medicine*, *2020*.
8. Kieffer, B., Babaie, M., Kalra, S., & Tizhoosh, H. R. (2017, November). Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks. In *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)* (pp. 1-6). IEEE.
9. Musulin, J., Smolić, K., & Car, Z. (2020). Bladder Cancer Recognition Using SIFT and SURF Features.

10. Lorencin, I., Anđelić, N., Španjol, J., & Car, Z. (2020). Using multi-layer perceptron with Laplacian edge detector for bladder cancer diagnosis. *Artificial Intelligence in Medicine*, *102*, 101746.
11. Štifanić, D., Musulin, J., Miočević, A., Baressi Šegota, S., Šubić, R., & Car, Z. (2020). Impact of COVID-19 on Forecasting Stock Prices: An Integration of Stationary Wavelet Transform and Bidirectional Long Short-Term Memory. *Complexity*, *2020*.
12. Ren, R., Wu, D. D., & Liu, T. (2018). Forecasting stock market movement direction using sentiment analysis and support vector machine. *IEEE Systems Journal*, *13*(1), 760-770.
13. Medica-Viola, V., Baressi Šegota, S., Mrzljak, V., & Štifanić, D. (2020). Comparison of conventional and heat balance based energy analyses of steam turbine. *Pomorstvo*, *34*(1), 74-85.
14. Baressi Šegota, S., Anđelić, N., Lorencin, I., Saga, M., & Car, Z. (2020). Path planning optimization of six-degree-of-freedom robotic manipulators using evolutionary algorithms. *International Journal of Advanced Robotic Systems*, *17*(2), 1729881420908076.
15. Mrzljak, V., Blecich, P., Anđelić, N., & Lorencin, I. (2019). Energy and exergy analyses of forced draft fan for marine steam propulsion system during load change. *Journal of Marine Science and Engineering*, *7*(11), 381.
16. Baressi Šegota, S., Lorencin, I., Musulin, J., Štifanić, D., & Car, Z. (2020). Frigate Speed Estimation Using CODLAG Propulsion System Parameters and Multilayer Perceptron. *NAŠE MORE: znanstveni časopis za more i pomorstvo*, *67*(2), 117-125.
17. Li, X., Shang, M., Qin, H., & Chen, L. (2015, October). Fast accurate fish detection and recognition of underwater images with fast r-cnn. In *OCEANS 2015-MTS/IEEE Washington* (pp. 1-5). IEEE.
18. Francis, M., & Deisy, C. (2019, March). Disease Detection and Classification in Agricultural Plants Using Convolutional Neural Networks—A Visual Understanding. In *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)* (pp. 1063-1068). IEEE.
19. Sajjad, M., Khan, S., Muhammad, K., Wu, W., Ullah, A., & Baik, S. W. (2019). Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *Journal of computational science*, *30*, 174-182.
20. Lorencin, I., Anđelić, N., Mrzljak, V., & Car, Z. (2019). Marine objects recognition using convolutional neural networks. *NAŠE MORE: znanstveno-stručni časopis za more i pomorstvo*, *66*(3), 112-119.
21. Gao, H., Cheng, B., Wang, J., Li, K., Zhao, J., & Li, D. (2018). Object classification using CNN-based fusion of vision and LIDAR in autonomous vehicle environment. *IEEE Transactions on Industrial Informatics*, *14*(9), 4224-4231.
22. Agarap, A. F. (2017). An architecture combining convolutional neural network (CNN) and support vector machine (SVM) for image classification. *arXiv preprint arXiv:1712.03541*.
23. Liu, Y., Cui, H. Y., Kuang, Z., & Li, G. Q. (2017). Ship detection and classification on optical remote sensing images using deep learning. In *ITM Web of Conferences* (Vol. 12, p. 05012). EDP Sciences.
24. El Amin, A. M., Liu, Q., & Wang, Y. (2016, July). Convolutional neural network features based change detection in satellite images. In *First International Workshop on Pattern Recognition* (Vol. 10011, p. 100110W). International Society for Optics and Photonics.
25. Rahmouni, N., Nozick, V., Yamagishi, J., & Echizen, I. (2017, December). Distinguishing computer graphics from natural images using convolution neural networks. In *2017 IEEE Workshop on Information Forensics and Security (WIFS)* (pp. 1-6). IEEE.
26. Liu, K., Kang, G., Zhang, N., & Hou, B. (2018). Breast cancer classification based on fully-connected layer first convolutional neural networks. *IEEE Access*, *6*, 23722-23732.
27. Basha, S. S., Dubey, S. R., Pulabaigari, V., & Mukherjee, S. (2020). Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing*, *378*, 112-119.
28. Chang, J., Zhang, L., Gu, N., Zhang, X., Ye, M., Yin, R., & Meng, Q. (2019). A mix-pooling CNN architecture with FCRF for brain tumor segmentation. *Journal of Visual Communication and Image Representation*, *58*, 316-322.
29. Qassim, H., Verma, A., & Feinzimer, D. (2018, January). Compressed residual-VGG16 CNN model for big data places image recognition. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 169-175). IEEE.

30. Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., & Xu, W. (2016). Cnn-rnn: A unified framework for multi-label image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2285-2294).
31. Pontes, F. J., Amorim, G. F., Balestrassi, P. P., Paiva, A. P., & Ferreira, J. R. (2016). Design of experiments and focused grid search for neural network parameter optimization. *Neurocomputing*, *186*, 22-34.
32. Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, *91*, 216-231.
33. Zeng, G. (2020). On the confusion matrix in credit scoring and its analytical properties. *Communications in Statistics-Theory and Methods*, *49*(9), 2080-2093.
34. Sun, X., Shi, J., Dong, J., & Wang, X. (2016, October). Fish recognition from low-resolution underwater images. In *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 471-476). IEEE.
35. Deep, B. V., & Dash, R. (2019, March). Underwater Fish Species Recognition Using Deep Learning Techniques. In *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)* (pp. 665-669). IEEE.
36. Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., & Harvey, E. (2016). Fish species classification in unconstrained underwater environments based on deep learning. *Limnology and Oceanography: Methods*, *14*(9), 570-585.