Animal Classification using Convolutional Neural Network

Conference Paper · December 2023

4 authors:

Kimly Y.
Royal University of Phnom Penh

Malis Lany
Royal University of Phnom Penh

Soy Vitou
Royal University of Phnom Penh

Sokchea Kor
Royal University of Phnom Penh

Animal Classification using Convolutional Neural Network

Kimly Y[#1], Malis Lany[#2], Vitou Soy[#3], Sokchea Kor[*4]

[#]*Department of Information Technology Engineering, Royal University of Phnom Penh*

Russian Federation Boulevard, Toul Kork, Phnom Penh, Cambodia

[1]y.kimly.2821@rupp.edu.kh

[2]lany.malis.2821@rupp.edu.kh

[3]soy.vitou.2821@rupp.edu.kh

**Cambodia Academy of Digital Technology*

Bridge 2, National Road 6A, Sangkat Prek Leap, Khan Chroy Changva, Phnom Penh, Cambodia

[4]kor.sokchea@cadt.edu.kh

***Abstract*— Image classification is a process in computer vision that classifies images into related categories or labels based on specific rules. The ability to automatically and correctly classify images is a pivotal process in various real-world applications, spanning from animal recognition, face recognition, medical diagnosis, and self-driving cars to object detection. It is easy for humans to swiftly and accurately classify animals into different labels, while it is tough for machines. With the development of new hardware technologies and the availability of large dataset, machine learning subdiscipline known as deep learning has seen a tremendous rise in prominence over the last 13 years. Many deep learning-based classifiers have been proposed and developed to enable machines to mimic humans' ability to categorize the image. In this study, we propose the architecture of a deep learning network specifically designed for animal classification, known as a Convolutional Neural Network (CNN). To perform the experiment, the Kaggle Dogs vs. Cats public dataset was used to train the classifiers. In the dataset, each image is associated with corresponding class labels, enabling classifiers to generalize and make accurate predictions on new, unseen image between cat and dog. The performance and accuracy of each classifier were evaluated by four evaluation metrics including accuracy, precision, recall, and F1-score. In our experiment, CNN achieved with an accuracy of 96%. As a result, CNNs offer assistance across a wide range of applications within the field of animal conservation, proving valuable in addressing diverse needs and challenges in this domain.**

***Keywords*— Classification, Convolutional Neural Network, Forward Propagation, Back Propagation, FeedForward Fashion.**

I. INTRODUCTION

Image classification is an important technique in computer vision that is used to categorize image into labels. The core of image classification task is to build a model that learns from existing data to enable computers to make predictions about the classes of unseen data. It is a complicated and time-consuming task for computer system to deal with large image datasets. With the advancement of technology as well as the availability of dataset, deep learning emerges as a key technique in a wide range of real-world applications such as speech recognition, visual object recognition, and object detection. Deep learning enables computers to learn complex patterns in data at multiple levels of abstraction by using multiple processing layers [1]. Image classification with deep learning can be achieved using either supervised learning or unsupervised learning. Convolutional neural network (CNN) is a supervised learning type of machine learning in which the model is trained on a set of annotated data, where each input consists of a known output. These algorithms generate a function mapping inputs to the desired outputs. Convolutional neural network (CNN) is a domain of artificial neural networks that has become dominating on various computer vision challenges. CNN is designed to automatically and adaptively learn hierarchies of complex and relationship spatial features through backpropagation by updating their knowledge attempt to achieved high performance. As we know CNN is one of the most popular algorithms and widely used for image-related tasks, that’s why popular libraries like TensorFlow, Keras, PyTorch offer convenient ways to build efficient CNN models, but there is nothing worst to build CNN from scratch. The benefit of this implementation such as understand deeper level, easily to expand knowledge with others high algorithms, last but not least you have an ability to make your model for continual learning.

Our main objective of this paper is to develop a deep learning model that is capable to generalize and make accurate predictions on the new, unseen images between cat and dog. Nevertheless, to examine how well the models perform and how efficiently they operate by evaluating the accuracy and effectiveness of the classifier.

The rest of paper is structured as follow: in section 2 we will cover the literature review from previous work. In section 3 methodology is discussed in detail including the dataset and proposed architecture. In section 4 will discuss about the results of the model performance in our experiment. Finally, we devote a section to conclude the paper and future works.

II. Related Works

There are many articles and research work related to the area of image classification. According to the research paper [2], the author points out that the goal of supervised learning is to build a concise model of the distribution of class labels in term of predictor features, which enables the model to perform well on labelled training data. The paper gives an introduction of supervised machine learning techniques introducing the field of machine learning, which is an area of active research. The paper also summarizes the fundamental aspects and structure of these learning techniques.

In [3], the author presents a state-of-the-art image recognition system of deep image developed using end-to-end deep learning. The key components of their work are a custom-built supercomputer dedicated to deep learning. This system achieves excellent results on multiple challenging computer vision benchmarks using multi-scale high-resolution images, parallel algorithms, larger network models and data augmentation approaches. The authors achieve their proposed work by training on a dedicated supercomputer. In future work, they aim to investigate whether other approaches can yield the same results with less computational demand.

In [4], the authors work on the development of a classification software with the use of deep learning concepts for image processing. This software creates a model that recognizes bird species from input images, predicts the species, display the bird. The author conclude that the only drawback of their proposed work is that the accuracy of the identification depends entirely on the quality of the camera. This can lead to inaccuracies above a certain range for any camera. However, their work also will help upcoming researchers and engineers work on other areas, such as defect recognition in industries and image segmentation for various purposes in other fields.

[5] This previous study had demonstrated the effectiveness of Convolutional Neural Network in image Classification task by using Convolutional Neural Network to classify the cifar-10 public dataset which contains 3 classes--car, airplane and bird. Model was trained with 64 images per batch in 20 epochs to ensure its performance and accuracy. As a result, this model can classify the sample image accurately in 94.2%.

In paper [6], it had been worked on against image classification task by using several Convolutional Neural Network-Based Classifiers such as DenseNet, VGG16, VGG19, and InceptionV3 in purpose of enhancing the learning of CNNs model on camera-assisted fruit identification. To achieving the capable classification model, Kaggle Fruits 360 dataset which contains 4455 total images in 7 categories of fruits: Avocado, Banana, Cherry, Cocos, Kiwi, Mango, and Orange, are utilized in this experiment. Throughout the training, all of the mentioned Convolutional Neural Network-Based Models are successfully built to classify the image effectively with the highest accuracy, a DenseNet Model achieved 99.25% correctly.

As in [7], It addressed the challenge of animal recognition and identification in open environments using deep convolutional neural networks. The authors leverage dataset sourced from the Wildlife Spotter project, conducted by citizen scientists. Their approach involves employing state-of-the-art deep convolutional neural network architectures to handle two primary tasks: detection and identification. The dataset consists of both animal and non-animal images, with proportions of 67.74% and 32.26%, respectively. The model first determines whether an animal is present in the image and then proceeds to identify the species. The focus is on utilizing advanced CNN architectures for these tasks, specifically Lite AlexNet, VGG-16, and ResNet-50. The obtained results showcase a high accuracy of 96.4% for detecting images containing animals and 90.4% for identifying the three most common species.

This article [8] works on similar animal images classification between snub-nosed monkeys and normal monkeys by applying simple 2D CNN via python. The author uses the dataset downloaded from google images with 800 and 200 images for each classes respectively as train and test data. In the pre-processing step all images are set to 28x28 pixels, 3 RGB channels, and shuffle to create randomness. The used architecture consists of 3 convolutional layers that have its own max pooling layer and followed by fully connected layers. The model was trained with 100 epoch which achieved accuracy of 96.67%. Nevertheless, the author stated that the model also has some limitation which the presence of only two categories of the monkey which there are many spaces looking for improvements in the future.

In [9], The author proposed a novel deep convolutional neural network-based species recognition algorithm to address the challenging camera-trap imagery data for wild animal classification. Their work is the first attempt to the fully automatic computer vision-based species recognition the real camera-trap images. They collected and annotated dataset from a standard camera-trap dataset of 20 species common in North America. The author implemented two algorithms in the purpose of comparison between Bag-of-words (BOW) model and Deep Convolutional Neural Network (DCNN). In result, DCNN based species recognition outperforms the traditional BOW based with the accuracy of 38.315% on the overall species in the dataset. Author also mentioned that the learning capacity of the DCNN is very high and therefore the performance of the DCNN can be further improved if more training data are available.

III. Methodology

A. *Dataset*

In this experiment, the public dataset from Kaggle that is used to train the classifier consists of 8005 training images and 2023 testing images. Both training and testing set are divided into different folder with the exact label between cat and dog. In purpose to improving the performance of model and reducing both noise in public dataset, either computational complexity cost of training, each image is required to resize in the sample size as the table below. Table 1 show the technique that we use in data pre-processing of our implementation. The used dataset is coming with the different scales, to make it easier for model to learn we resized the dataset to a specific scaling to make sure that all the image in the dataset is in the same scaling of pixels.

B. *Convolutional Neural Networks*

Convolutional Convolutional Neural Network is powerful algorithms for image classification because of its architecture, which consists of multiple layers of interconnected neurons in a feedforward manner. With Forward propagation, input data is transmitted through neural network to generate an output in the purpose of preserving the connection as in the feedforward fashion [10], which is calculated by :

$$z_j^l = \sum_{K=1}^{N^{l-1}} w_{jk}^l a_k^{l-1} + b_j^l$$

z_j^l is the weighted sum of inputs in layer L
N^{l-1} is the number of neurons in the previous layer
w_{jk}^l is the weight that connection between L1&L2
a_k^{l-1} is the activation of the k th neuron in layer $\mathrm{L}-1$
b_j^l is the bias term for the $\mathrm{j}-\mathrm{th}$ neuron in layer L

While Back propagation [11] optimizes the model by adjusting weights based on gradients of the loss function, minimizing the disparity between predicted and actual values which is done by using the chain rule of calculus, shown in Fig. 1.

Compute Loss Equation[11]:

$$E = \frac{1}{2}\sum_{K=1}^{N}(y_k - \hat{y}_k)^2$$

Backward Pass Equation[11]:

$$\frac{\partial E}{\partial w_{jk}^1} = -(y_k - \hat{y}_k) \cdot f'(a_k) \cdot h_j$$

$$\frac{\partial E}{\partial b_k^1} = -(y_k - \hat{y}_k) \cdot f'(a_k)$$

Update Weights Equation[11]:

$$w_{jk}^1 \leftarrow w_{jk}^1 - \alpha \frac{\partial E}{\partial w_{jk}^1}$$
$$b_j^1 \leftarrow b_j^1 - \alpha \frac{\partial E}{\partial b_j^1}$$

$\frac{\partial E}{\partial w_{jk}^1}$ is the residual to update knowledge
y_k is the actual output in the output layer.
$\hat{y}_k$ is the prediction for output layer.
$f'(a_k)$ is the derivative of activation function.
h_j is the output of the j-th neuron in the hidden.
α is the Hyperparameter called Learning rate.

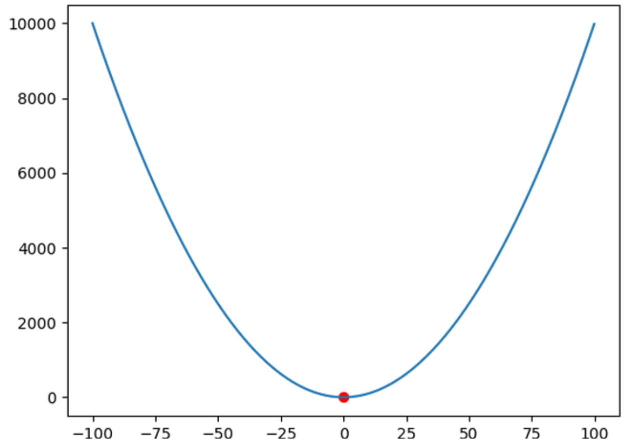

Fig. 1 Visualize Backpropagation Computing

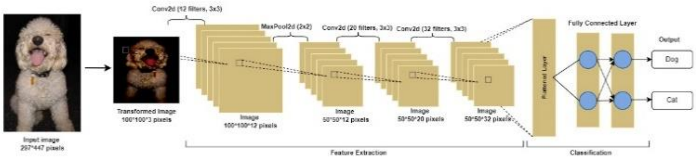

Fig. 2 Convolutional Neural Network Model

Fig. 2 shown CNN architecture with six layers implemented in two specials portions: feature extraction and classification portion. The input images are transformed to 100x100 pixels with three channels. The feature extraction portion consists of three convolutional layers with the same filter size (3x3 pixels), padding (1), and stride (1), but different numbers of filters, Con2d Image is shown in Fig. 3.

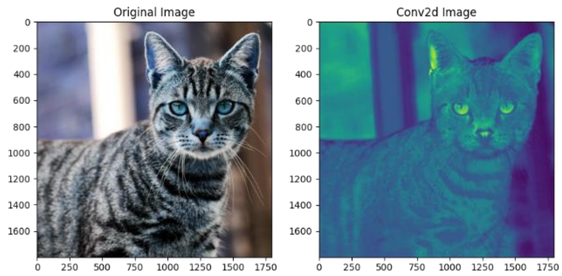

Fig. 3 Visualization of Convolutional Layer's Output Image

These layers play a crucial role in capturing spatial hierarchical features within an image through the application of convolutional operations and create feature maps that represent the height, width, and number of channels shown in Fig. 4.

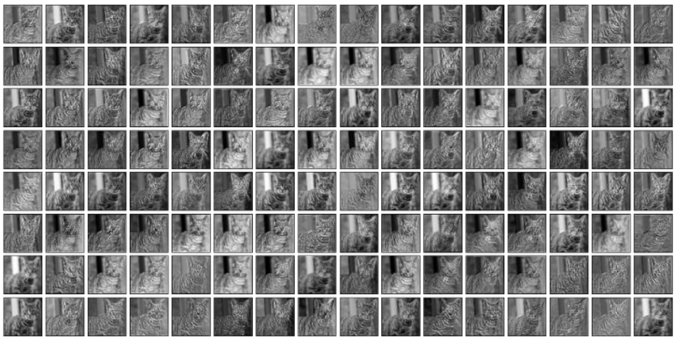

Fig. 4 Visualization a batch of convolutional output

[12] One max pooling layer with a filter size of 2x2 pixels, is a down sampling operation used between the first two convolutional layers to reduce the spatial dimensions of an input feature map while retaining its most important information. It is a form of spatial subsampling that helps in reducing the computational complexity of the network and mitigating the risk of overfitting as shown in Fig. 5.

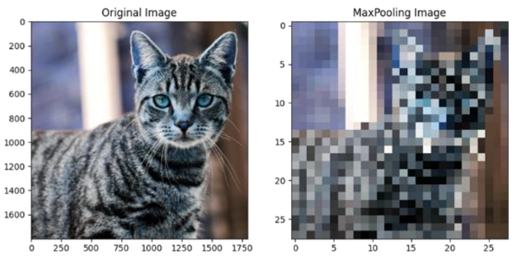

Fig. 5 Visualization of MaxPooling Image

After each convolutional layer, an activation function is a crucial component which serving as a mathematical operation applied to the output of each neuron in order to introduce non-linearity into the network, allowing it to learn complex patterns in data, and Rectified Linear Units (ReLU) is a commonly used, shown in Fig. 6 and Fig. 7.

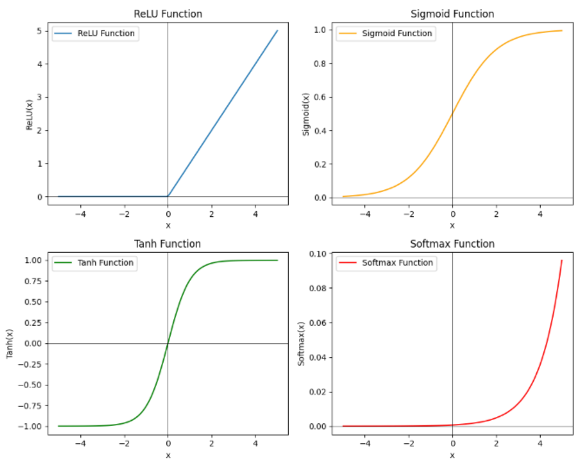

Fig. 6 Some commonly used activation function [13]

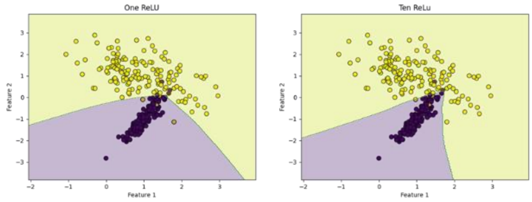

Fig. 7 Visualization of ReLU for clustering data

In classification portion, a flattened layer takes the input tensor, reshapes it into a 2D tensor, and pass it to fully connected layers. Fully Connected layer uses the mathematics function to operate on flattened vector data and classify the image. The images are then initialized into the matching category.

IV. RESULT AND ANALYSIS

A. *Confusion Matrix*

A confusion matrix provides insight the accuracy of predictions comparing to the actual values. It summarizes the number of correct and incorrect predictions, which is useful to understand the detail of model's performance.

According to[10], there are 4 important terms:

- **True Positive (TP)**: The outcome where a model accurately classifies the positive class.
- **True Negative (TN)**: The outcome where a model accurately classifies the negative class.
- **False Positive (FP)**: The outcome where a model inaccurately classifies the positive class.
- **False Negative (FN)**: The outcome where a model inaccurately classifies the negative class.

The anticipated and actual categorization of a classifier are displayed in a confusion matrix of size n × n (n rows and n columns), where n is the number of distinct classes. The following formulas are used to determine True Positive (TP), True Negative (TN), False Negative (FN), and False Positive (FP) for n × n matrices[10].

$$TP_i = a_{ii},$$

$$FP_i = \sum_{j=1, j\neq i}^{n} a_{ji},$$

$$FN_i = \sum_{j=1, j\neq i}^{n} a_{ij},$$

$$TN_i = \sum_{j=1, j\neq i}^{n} \sum_{k=1, k\neq i}^{n} a_{jk}.$$

Convolutional Neural Network: Fig. 8 shows the confusion matrix where the CNN model correctly classified 894 cat images, and 568 dog images. There were 117 FP and 444 FP where the model was incorrectly predicted cat as dog and dog as cat respectively. There is a noticeable that you may see the false positive gap between the two classes which is due to the dog images in the training data have more noise than the cat images. Despite the rescaling the image to 100*100 pixels we are still unable to ensure that all the noise in the image is totally eliminated that influenced to achieved this outcome. Generally, to be more efficient, the input images shall be resized to relatively small spatial resolution (eg. 224*224), and both training and inference are carried out at this resolution [14].

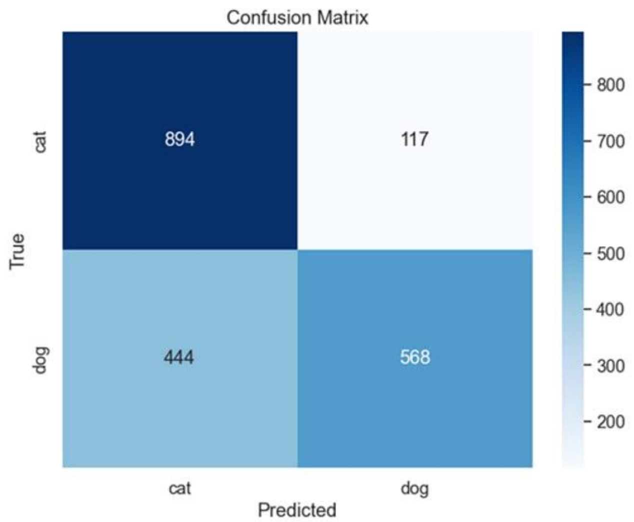

Fig. 8 Convolutional Neural Network Confusion Matrix

B. Accuracy

Accuracy is the value of correct classification achieve by the machine learning model that already trained. It's to evaluate the performance and the prediction accurate of the model. It's defined by the number of correct predictions divided by the total number of all classes' predictions[10].

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ all\ Classes'\ Prediction}$$

C. Precision

Precision is simply the ratio of correct positive predictions out of all positive predictions made, or the accuracy of minority class predictions. It is calculated as the number of true positives divided by the total number of true positives and false positives[10].

$$Pr\,e\,cision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

D. Recall

Recall measures the effectiveness of a classification model in identifying all relevant instances from a dataset. It is the ratio of the number of true positive instances to the sum of true positive and false negative instances[10].

$$Recall = \frac{True\ Positive}{True\ Positive\ + False\ Negative}$$

E. F1-Score

F1 Score is the Harmonic Mean between precision and recall. It tells you how precise your classifier is (how many instances it classifies correctly), as well as how robust it is (it does not miss a significant number of instances). The greater the F1 Score, the better is the performance of our model [10].

$$F1 - Score = 2 \cdot \frac{Pr\,e\,cision\ \cdot Recall}{Pr\,e\,cision + Recall}$$

F. Algorithms Evaluation Table

The performance of the model has been increasing and balancing from Epoch shown in Fig. 9.

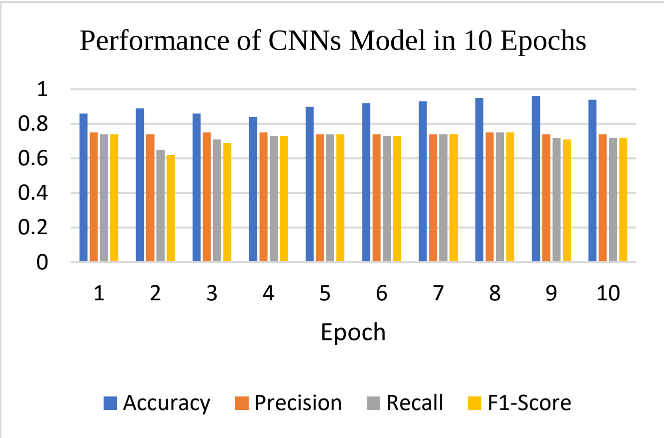

Fig. 9 Model performance in 10 epochs

V. CONCLUSIONS AND FUTURE WORKS

There are many ways to perform image classification, but one of the most popular and effective methods is using deep learning. Deep learning network are composed of layers of artificial neurons that can learn complex patterns from data by feeding images as input, it can learn to extract features and recognize objects in the images. Overall, the objective in this paper was to analyze the effectiveness and efficiency of models by assessing the performance and accuracy of the classifier. We do the evaluation of the technique based on the four-evaluation metrics on their performance including accuracy, precision, recall, and F1-score. For the result, the highest accuracy our proposed architecture could achieved is 96%. Despite that the complexity ratings are relative and depend on various factors such as the size of the dataset, the dimensionality of the data, model hyperparameters, available computational resources and specific implementation details chosen for each algorithm. It's important to consider these factors when selecting an algorithm for a particular classification task. However, in this study there are also many problems that could affected to the learning ability and accuracy of the model because of the challenging in the dataset. The used dataset contains various problems such as the occlusion and illumination of the image, the edited image, and some image that has several objects in the frame. So, in the future work, we are aiming to try implement the algorithms with more of the challenge in the dataset that could be helpful in pursue further research on the development of the model architecture.

ACKNOWLEDGMENT

We would like to express our gratitude and thanks to our supervisor Mr. Sokchea Kor, in his way guiding us to explore and doing research throughout this whole project and the author of every research paper and media as we can be able to extract the information as references and resources.

REFERENCES

[1] LeCun, Y., Bengio, Y., & Hinton, G. Deep learning. nature, 521(7553), 436-444. 2015.

[2] Nasteski, V. An overview of the supervised machine learning methods. Horizons. b, 4, pages 51-62, 2017.

[3] Wu, R., Yan, S., Shan, Y., Dang, Q., and Sun, G. Deep image: Scaling up image recognition. arXiv preprint arXiv:1501.02876, 7(8), 4. 2015.

[4] Rai, B. K., Sharma, S., Kumar, G., and Kishor, K. Recognition of Different Bird Category Using Image Processing. International Journal of Online & Biomedical Engineering, 18(7). 2022.

[5] Sharma, A., & Phonsa, G. (2021, April). Image classification using CNN. In Proceedings of the International Conference on Innovative Computing & Communication (ICICC), 2021.

[6] Tripathi, M. (2021). Analysis of convolutional neural network based image classification techniques. Journal of Innovative Image Processing (JIIP), 3(02), 100-117, 2021.

[7] Nguyen, H., Maclagan, S. J., Nguyen, T. D., Nguyen, T., Flemons, P., Andrews, K., ... & Phung, D. (2017, October). Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring. In 2017 IEEE international conference on data science and advanced Analytics (DSAA) (pp. 40-49). IEEE.

[8] Zeng, P. (2021, December). Research on similar animal classification based on CNN algorithm. In Journal of Physics: Conference Series (Vol. 2132, No. 1, p. 012001). IOP Publishing.

[9] Chen, G., Han, T. X., He, Z., Kays, R., & Forrester, T. (2014, October). Deep convolutional neural network based species recognition for wild animal monitoring. In 2014 IEEE international conference on image processing (ICIP) (pp. 858-862). IEEE.

[10] Shad, H. S., Rizvee, M. M., Roza, N. T., Hoq, S. M., Monirujjaman Khan, M., Singh, A., ... and Bourouis, S. (2021). Comparative analysis of deepfake image detection method using convolutional neural network. *Computational Intelligence and Neuroscience*, *2021*.

[11] Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., & Hinton, G. (2020). Backpropagation and the brain. *Nature Reviews Neuroscience*, *21*(6), 335-346.

[12] Shankar, V., Kumar, V., Devagade, U., Karanth, V., & Rohitaksha, K. (2020). Heart disease prediction using CNN algorithm. *SN Computer Science*, *1*(3), 170.

[13] Sharma, S., Sharma, S., & Athaiya, A. (2017). Activation functions in neural networks. Towards Data Sci, 6(12), 310-316.

[14] Talebi, H., & Milanfar, P. Learning to resize images for computer vision tasks. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 497-506). 2021.