COMPUTER Society

ASCEE

Developing deep learning architecture for image classification using Convolutional Neural Network (CNN) algorithm in forest and field images

CrossMark

Meiga Isyatan Mardiyah [a,1,*], Tuti Purwaningsih [a,2]

[a] Departemen of Statistics, Universitas Islam Indonesia, Indonesia
[1] 16611087@students.uii.ac.id; [2] tuti.purwaningsih@uii.ac.id
* corresponding author

ARTICLE INFO

ABSTRACT

Indonesia is an agricultural country with a variety of natural resources such as agriculture and plantations. Agriculture and plantations in Indonesia are diverse, such as rice fields that can produce rice, soybeans, corn, tubers, and others. Meanwhile, plantations in Indonesia are like forests with timber products, bamboo, eucalyptus oil, rattan, and others. However, rice fields, which are examples of agriculture, and forests that are examples of plantations, have the same characteristics. It is not easy to distinguish when viewed using aerial photographs or photographs taken from a certain height. For recognizing with certainty the shape of rice fields and forests when viewed using aerial photographs, it is necessary to establish a model that can accurately recognize the shape of rice fields and forest forms. A model is to utilize computational science to take information from digital images to recognize objects automatically. One method of deep learning that is currently developing is a Convolutional Neural Network (CNN). The CNN method enters (input data) in the form of an image or image. This method has a particular layer called the convulsive layer wherein an input image layer (input image) will produce a pattern of several parts of the image, which will be easier to classify later. The convolution layer has the function of learning images to be more efficient to be implemented. Therefore, researchers want to utilize this CNN method to classify forests and rice fields to distinguish the characteristics of forests and rice fields. Based on the classification results obtained by testing the accuracy of 90%. It can be concluded that the CNN method can classify images of forests and rice fields correctly.

1.Introduction

Indonesia is an agricultural country with a variety of natural resources such as agriculture and plantations. According to the world food and agriculture organization, Indonesia has the

most productive paddy fields in Asia, with 4.9 million hectares of paddy fields. Many Indonesians work as farmers and planters because regions in Indonesia mostly have agricultural land and plantations [1]. One example of plantation types is forest. Based on data from the Ministry of Forestry and the Environment as of March 2019, the achievement of social forestry has now reached 2.56 million hectares (ha), consisting of village forests (1.28 million ha), community forests (245,593 ha), community plantations (331,993 ha), forestry partnerships (549,785 ha), and customary forests (28,286 ha) [2].

Agriculture and plantations in Indonesia [3] are diverse, such as rice fields that can produce rice, soybeans, corn, tubers, and others. Meanwhile, Indonesia's plantations are like forests with timber products, bamboo, eucalyptus oil, rattan, and others. However, rice fields, which are examples of agriculture, and forests that are examples of plantations, have the same characteristics. It is not easy to distinguish when viewed using aerial photographs or photographs taken from a certain height. For recognizing with certainty the shape of rice fields and forests when viewed using aerial photographs, it is necessary to establish a model that can accurately recognize the shape of rice fields and forest forms.

At this time, advances in information technology can not be avoided with the development of hardware in advancing computer performance and developing software that can resemble human intelligence (artificial intelligence). Computers nowadays can get things done more efficiently, more quickly, and in less time. One of the technologies that can facilitate humans is deep learning. Deep learning is a branch of machine learning based on artificial neural networks [4] that train or teach an action that is considered reasonable for humans. Deep learning can automatically classify images [5], text, and videos by making videos into images [6].

One method of deep learning that is currently developing is a Convolutional Neural Network (CNN). The CNN method enters (input data) in the form of an image or image. This method has a particular layer called the convulsive layer wherein an input image layer (input image) will produce a pattern of several parts of the image, which will be easier to classify later. The convolution layer has the function of learning images to be more efficient to be implemented. Therefore researchers used this CNN method to be able to classify forests and rice fields. In order to distinguish the characteristics of forests and rice fields.

2.Method

3.1. Image

The image can be interpreted as a two-dimensional function, f (x, y), where x and y are spatial coordinates, and f (x, y) is a value at coordinates (x, y), which is often called intensity [7]. A digital image is an image of f (x, y) that has been digitalized or used as a digital system in terms of area coordinates and the intensity value. A digital image consists of several elements called picture elements or pixels.

Image is a representation (picture), similarity, or imitation of an object divided into two kinds, namely analog images, and digital images. Analog image is a continuous image such as images on television monitors, X-rays, and others. At the same time, digital images can be processed by computers [8].

3.2. Digital Imagery

Digital images represent images taken by machines in the form of an approach based on sampling and quantization [9]. Sampling states the size of the boxes arranged in rows and columns. In other words, sampling on the image states the pixel's size (point) in the image. Quantization states the value of the brightness level expressed in gray level values (grayscale) according to the number of binary bits used by the machine; in other words, quantization in the image stating the number of colors in the image.

Digital images are mapped into pixel elements in the form of two-dimensional matrices and grid shapes in a computer. Each pixel has a number that indicates the color channel [10], [11].

Numbers on each pixel are stored sequentially by a computer and are often subtracted for certain compression and processing purposes. A digital image can be represented by a matrix consisting of M columns N rows [12], [13], where the intersection between columns and rows is called pixels. (pixel = picture element), which is the smallest element of an image. Pixels have two parameters, namely coordinates and intensity or color. The value contained in the coordinates (x, y) is f (x, y), which is the intensity or color of the pixel at that point. Therefore, the image can be written into a matrix (1).

$$(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,M-1) \\ f(1,0) & f(1,1) & \cdots & f(1,M-1) \\ \vdots & \vdots & \vdots & \vdots \\ f(N-0,1) & f(N-1,1) & \cdots & f(N-1,M-1) \end{bmatrix} \tag{1}$$

Based on the formula above, an image f (x, y) can be written into a mathematical function as below (2)-(4)

$$0 \leq x \leq M-1 \tag{2}$$

$$0 \leq x\, N-1 \tag{3}$$

$$0 \leq f(x,y) \leq G-1 \tag{4}$$

Where M is the number of lines in the image, N is the number of columns in the image. The G is a grayscale scale value (gray). The magnitude of the values M, N, and G is usually the enhancements using the formulas: $M = 2m$; $N = 2n$; $G = 2k$.

Where the values of m, n, and k are positive, the interval (0, G) is called (grayscale). The value of G depends on the digitization process. Usually, gray (0 (zero) expresses black intensity, and 1 (one) indicates white intensity. For an 8-bit image [14], G's value is equal to 28 = 256 colors (gray degree).

3.3. Artificial Intelligence (AI)

Artificial Intelligence or Artificial Intelligence (AI) is a technique or method applied to imitate the intelligence possessed by living things to solve a problem. Artificial intelligence (artificial intelligence) is a study of how to make computers do things that are currently done better by humans. Artificial Intelligence (AI) functions to identify or model human thought processes and design machines so that they can mimic human behavior. Machines can act like humans with good knowledge and logical thinking skills. The function of Artificial Intelligence (AI) is to make robots that can resemble intelligence or even more than the intelligence possessed by humans.

3.4. Machine Learning

Arthur Samuel first defined the term machine learning in 1959. According to Arthur Samuel, machine learning is a computer science field that provides learning to computers to know things without an exact programmer [15]. In simple machine learning is to build an algorithm that allows computer programs to learn and do their work without any instructions from users. This algorithm works by making a model of input produce a decision based on existing data. Machine learning is directly related to computational statistics centered on making decisions based on computer usage. Some machine learning applications are image processing, search and recommendation engines, finance, speech understanding, and text analysis [16].

3.5. Deep Learning

Deep learning is one technique in machine learning that utilizes many layers [17] of non-linear information processing to perform feature extraction, pattern recognition, and classification [18].

According to [19], deep learning is an approach to solving computer learning systems that use hierarchy. The concept of hierarchy makes computers able to learn complicated concepts by combining more straightforward concepts. If a graph is described how the concept is built on top of other concepts, this graph will be deep with many layers, and this is the reason referred to as deep learning (deep learning).

3.6. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is one of the algorithms from Deep Learning, the development of Multi-Layer Perceptron (MLP), designed to process data in grid form, one of which is a two-dimensional image, for example, images or sound [20]. CNN is included in the type of Deep Neural Network because of the high network depth and is widely applied to image data or images. Image classification can be used with MLP, but the MLP method is not appropriate because it does not store spatial information from image data and argues that each pixel is an independent feature (characteristic), resulting in unfavorable results.

Methodically, CNN is an architecture that can be trained and has several stages. Input (input) and output (output) of each stage are composed of several arrays called feature maps. Each stage consists of three layers, namely convolution, layer activation function, and layer pooling.

3.7. Confusion Matrix

The confusion matrix is a method used to evaluate decision tree classifications [21]. The confusion matrix is a table consisting of many rows of test data that are predicted to be true and incorrect by the classification model. This table is needed to measure the performance of a classification model [22].

A confusion matrix is a useful tool for analyzing how well a classifier/grouping can recognize tuples from different classes [23]. In making confusion matrix tables, four things must be known as in Table 1.

Table 1. Confusion Matrix

Class	Classified Positive	Classified Negative
Positive	TP (True Positive)	FN (False Negative)
Negative	FP (False Positive)	TN (True Negative)

Information:

1. True Positive (TP): the number of positive data sets that are classified as positive.

2. True Negative (TN): the number of negative data sets that are classified as negative.

3. False Positive (FP): the number of negative data sets classified as positive.

4. False Positive (FP): the number of positive data sets classified as negative.

3.8. Precision, Recall, and Accuracy

Precision is the accuracy between the information requested by the user and the system's answers. To calculate the value of precision using (5).

$$Precision = TP/(TP + FP) \tag{5}$$

The recall is the success rate of the system in finding back information. To calculate the recall value using (6).

$$Recall = TP/(TP + FN) \tag{6}$$

Accuracy is the level of closeness between the predicted value and the actual value. Accuracy can be interpreted in the form of truth or accuracy of errors. To calculate the Accuracy value, use (7).

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{7}$$

3. Results and Discussion

3.1. Dataset Scenario Training and Testing

Fig. 1 shows 100 image data/images that are two categories of forests and fields in the DataHutanSawah folder. Because naming image data/images using their respective category names, the image data have an automatic sequence by the name category's alphabet, i.e., the 1st category's location is the forest, and the 2nd category is the rice field.
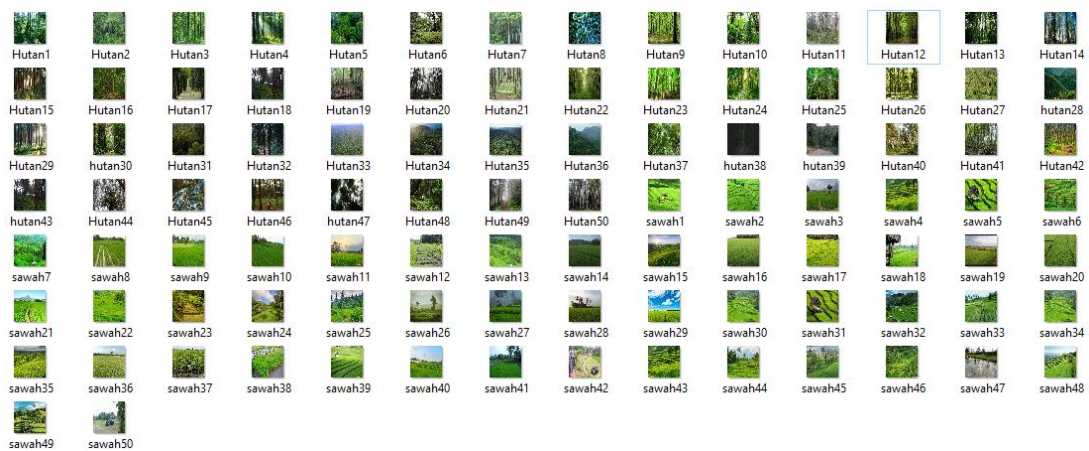

Fig. 1. Forest and Field Dataset

The available dataset will then be determined by the amount of data divided into training and testing data. Researchers determine each category to use as much as 50 image data with 3 comparisons of training data scenarios namely 90%, 80%, and 70% (Table 2), for scenarios 70% of training data used is (50 x 70): 100 = 35 so for testing data 50 - 35 = 15, for the scenario of 80% of the training data used is (50 x 80): 100 = 40 so the testing data is 50 - 40 = 10, and for the scenario of 90% the training data used is (50 x 90): 100 = 45, then the testing data is 50 - 45 = 5 images.

Table 2. Data Sharing scenarios

Scenario 70%:30%		Scenario 80%:20%		Scenario 90%:10%	
Data Training	Data Testing	Data Training	Data Testing	Data Training	Data Testing
70	30	80	20	90	10

At this stage, the researcher will conduct several experiments to compare models using the Convolutional Neural Network (CNN) architecture to obtain the best classification results. The stages of obtaining the best CNN architecture can be seen from the parameters used. In this study, researchers compared the epoch value and compared the number of train data and test data. This architecture design is compared with the initial architecture with 80% data train parameter comparison, 20% test data, and 50 epoch.

For finding out the differences and find the best architecture, the results obtained from the CNN architecture design are explained in Comparing Epoch Values. The epoch values to be compared in this experiment are epoch 50, 65, and 80. Based on the training results, the comparison of the accuracy values for each epoch is obtained, shown in Table 3.

Table 3. Comparison of epoch values

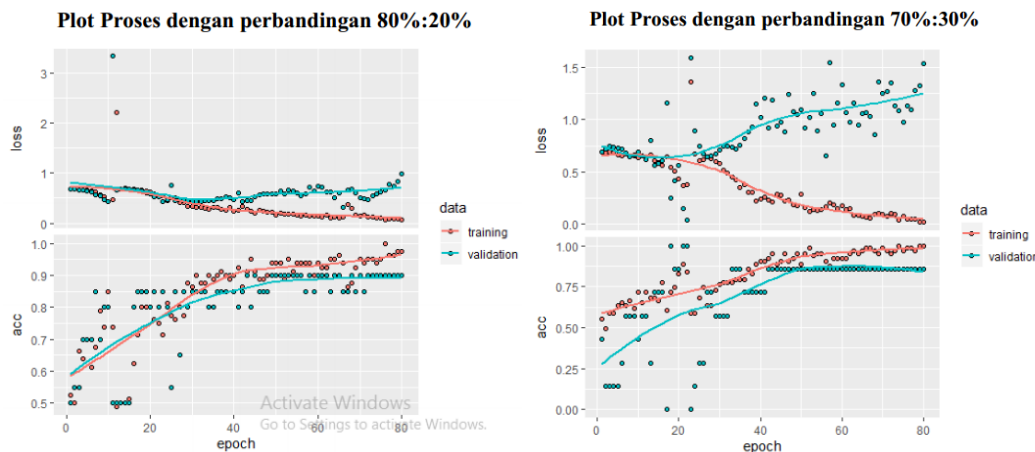| Data	Epoch 50 Accuracy	Epoch 65 Accuracy	Epoch 80 Accuracy
Training	0.7875	0.9125	0.9875
Loss Training	0.5315	0.2025	0.0313
Testing	0.75	0.9	0.9
Loss Testing	0.5631	1.0733	0.993

The following results obtained from the comparison of epoch 50, 65, and 80 show that the greater the epoch used, the better the accuracy results obtained. Many epochs are used to make the algorithm more trained and able to recognize patterns better. Many epochs make models even better.

After comparing the epoch values, the researcher compares the number of train data and test data 70%: 30%, 80%: 20%, and 90%: 10% with the previous optimal architecture, the results obtained after comparing the number of train data and test data presented in Table 4.

Table 4. Comparison of the number of train data and test data

Data	Epoch 50 Accuracy	Epoch 65 Accuracy	Epoch 80 Accuracy
Training	0.985	0.985	0.988
Loss Training	0.174	0.031	0.06
Testing	0.866	0.9	0.9
Loss Testing	1.374	0.993	0.781

Based on Table 4, it can be seen that the highest accuracy value is obtained by using a ratio of 90%: 10% with data train 90 and data test 10. Accuracy results obtained are equal to 0.988 or 98.8% for training accuracy and by 0.9 or 90% for accuracy testing. A comparison of the three results can be seen in the plot's process, as shown in Fig. 2.
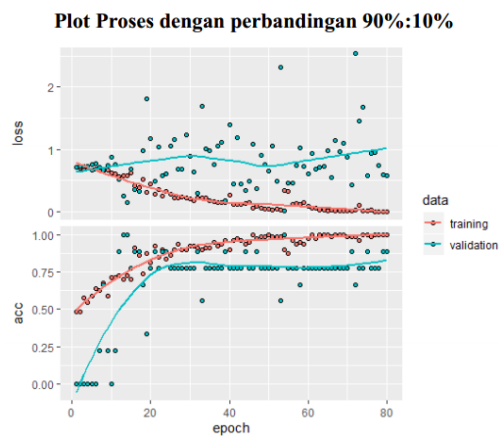
Fig. 2. Plot Process Comparison of 70%, 80%, and 90% of Data Train Scenarios
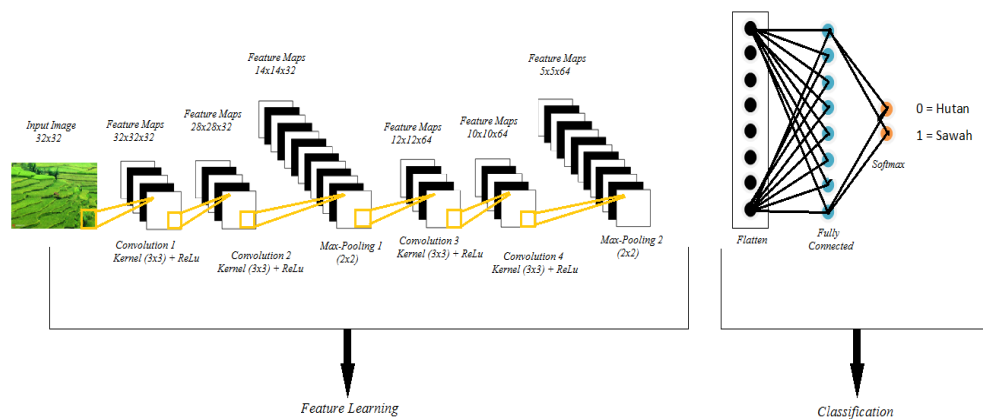
3.2. CNN Architecture

Fig. 3. CNN Architecture

After CNN architecture has been designed, as shown in Fig. 3. The best model is obtained, having a total number of parameters of 475,938 with the details in Table 5.

Table 5. Comparison of epoch values

Data	Epoch 50	Epoch 65
conv2d (Conv2D)	(None, 30, 30, 32)	896
conv2d_1 (Conv2D)	(None, 28, 28, 32)	9248
max_pooling2d (MaxPooling2D)	(None, 14, 14, 32)	0
dropout (Dropout)	(None, 14, 14, 32)	0
conv2d_2 (Conv2D)	(None, 12, 12, 64)	18496
conv2d_3 (Conv2D)	(None, 10, 10, 64)	36928
max_pooling2d_1 (MaxPooling2D)	(None, 5, 5, 64)	0
dropout_1 (Dropout)	(None, 5, 5, 64)	0
flatten (Flatten)	(None, 1600)	0
dense (Dense)	(None, 256)	409856
dropout_2 (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 2)	514
Total params : 475.938		
Trainable params : 475.938		
Non-trainable params : 0		

3.3. Classification Results

After getting the best CNN architecture, then proceed with the results of the classification. The classification process uses the test data as ten images, with each category as many as five images. The best classification results table from the testing data is presented in Table 6.

Table 6. Classification results of the best models with data testing (test)

Prediction	Actual	
	Forest	Field
Forest	5	1
Field	0	4

Based on Table 6, the classification results are obtained using testing data. The number of forest images predicted as forest images with five images means the forest image classification is correct. Furthermore, the number of paddy imagery predicted as paddy imagery is four images, and there is an unpredictable missing data as paddy imagery of 1 image. Calculation of classification results and accuracy values can be calculated using the (7). The testing accuracy has 90% of the result of testing.

$$Testing\ Accuracy = \frac{Total\ Prediction\ Testing}{Total\ Testing} \qquad (8)$$

Based on the (8) calculation results obtained accuracy values based on the classification results with the best CNN architecture using 90% testing data, the results of the classification of predicted forest images according to the actual forest images are only five images and 4 fields images for paddy images.

4. Conclusion

Based on the results and discussion related to the researcher's analysis, it can be concluded that CNN in the forest and field image classification process can be done by finding the best model architecture. This architecture compares two parameters: the number of epochs and the scenario for comparison of train and test datasets. The classification results have an accuracy level of 90%, so the CNN method with these parameters can be used to classify forest and field images.

References

[1] Indonesia.go.id, "Sawah Berbentuk Jaring Laba-laba di Flores," *Portal Informasi Indonesia*, 2019. Available: https://indonesia.go.id/ragam/pariwisata/pariwisata/sawah-berbentuk-jaring-laba-laba-di-flores.

[2] Indonesia.go.id, "Usaha Berbasis Kayu Tetap Jalan, Hutan Pun Tetap Lestari," *Portal Informasi Indonesia*, 2019. Available: https://www.indonesia.go.id/narasi/indonesia-dalam-angka/ekonomi/usaha-berbasis-kayu-tetap-jalan-hutan-pun-tetap-lestari.

[3] Z. Hidayah, *A Guide to Tribes in Indonesia*. Singapore: Springer Singapore, 2020, doi: 10.1007/978-981-15-1835-5.

[4] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017, doi: 10.1109/MSP.2017.2743240.

[5] M. Z. Alom *et al.*, "A state-of-the-art survey on deep learning theory and architectures," *Electronics*, vol. 8, no. 3, p. 292, Mar. 2019, doi: 10.3390/electronics8030292.

[6] S. Pouyanfar *et al.*, "A survey on deep learning: algorithms, techniques, and applications," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–36, Jan. 2019, doi: 10.1145/3234150.

[7] S. González Izard, J. A. Juanes Méndez, P. Ruisoto Palomera, and F. J. García-Peñalvo, "Applications of Virtual and Augmented Reality in Biomedical Imaging," *J. Med. Syst.*, vol. 43, no. 4, p. 102, Apr.

2019, doi: 10.1007/s10916-019-1239-z.

[8] U. Yuandari, "Improving the quality of digital images using the image averaging method," *Int. J. Informatics Comput. Sci.*, vol. 4, no. 1, pp. 5–11, 2020, doi: 10.30865/ijics.v4i1.1982.

[9] Kwang In Kim, Keechul Jung, and Jin Hyung Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1631–1639, Dec. 2003, doi: 10.1109/TPAMI.2003.1251157.

[10] S. Saifullah and V. A. Permadi, "Comparison of Egg Fertility Identification based on GLCM Feature Extraction using Backpropagation and K-means Clustering Algorithms," in *Proceeding - 2019 5th International Conference on Science in Information Technology: Embracing Industry 4.0: Towards Innovation in Cyber Physical System, ICSITech 2019*, Oct. 2019, pp. 140–145, doi: 10.1109/ICSITech46713.2019.8987496.

[11] Sunardi, A. Yudhana, and S. Saifullah, "Identification of Egg Fertility Using Gray Level Co-Occurrence Matrix and Backpropagation," *Adv. Sci. Lett.*, vol. 24, no. 12, pp. 9151–9156, 2018, doi: 10.1166/asl.2018.12115.

[12] A. Yudhana, Sunardi, and S. Saifullah, "Segmentation comparing eggs watermarking image and original image," *Bull. Electr. Eng. Informatics*, vol. 6, no. 1, pp. 47–53, 2017, doi: 10.11591/eei.v6i1.595.

[13] Sunardi, A. Yudhana, and S. Saifullah, "Identity analysis of egg based on digital and thermal imaging: Image processing and counting object concept," *Int. J. Electr. Comput. Eng.*, vol. 7, no. 1, pp. 200–208, 2017, doi: 10.11591/ijece.v7i1.12718.

[14] B. Goyal, A. Dogra, S. Agrawal, and B. S. Sohi, "Two-dimensional gray scale image denoising via morphological operations in NSST domain & bitonic filtering," *Futur. Gener. Comput. Syst.*, vol. 82, pp. 158–175, May 2018, doi: 10.1016/j.future.2017.12.034.

[15] M. Schuld, I. Sinayskiy, and F. Petruccione, "An introduction to quantum machine learning," *Contemp. Phys.*, vol. 56, no. 2, pp. 172–185, Apr. 2015, doi: 10.1080/00107514.2014.964942.

[16] Y. Fauziah, S. Saifullah, and A. S. Aribowo, "Design Text Mining for Anxiety Detection using Machine Learning based-on Social Media Data during COVID-19 pandemic," in *Proceeding of LPPM UPN "Veteran" Yogyakarta Conference Series 2020–Engineering and Science Series*, 2020, vol. 1, no. 1, pp. 253–261, doi: 10.31098/ess.v1i1.117.

[17] S. Loussaief and A. Abdelkrim, "Deep learning vs. bag of features in machine learning for image classification," in *2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET)*, Mar. 2018, pp. 6–10, doi: 10.1109/ASET.2018.8379825.

[18] L. Deng, "Deep Learning: Methods and Applications," *Found. Trends® Signal Process.*, vol. 7, no. 3–4, pp. 197–387, 2014, doi: 10.1561/2000000039.

[19] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. Available: https://www.deeplearningbook.org.

[20] S. Sunardi, A. Mahfurdz, and S. Saifullah, "Green turtle and fish identification based on acoustic target strength," *Int. J. Adv. Intell. Informatics*, vol. 4, no. 1, p. 53, Mar. 2018, doi: 10.26555/ijain.v4i1.147.

[21] M. Tayefi *et al.*, "hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm," *Comput. Methods Programs Biomed.*, vol. 141, pp. 105–109, Apr. 2017, doi: 10.1016/j.cmpb.2017.02.001.

[22] A. Tharwat, "Classification assessment methods," *Appl. Comput. Informatics*, vol. ahead-of-p, no. ahead-of-print, Aug. 2020, doi: 10.1016/j.aci.2018.08.003.

[23] J. Han, M. Kamber, and J. Pei, "Classification: basic concepts," in *Data Mining*, M. Kaufmann, Ed. 2012, pp. 327–391. Available: http://hanj.cs.illinois.edu/cs412/bk3/08.pdf.