



INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

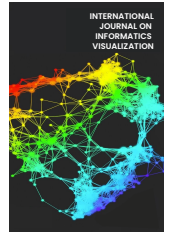


Image Prediction of Exact Science and Social Science Learning Content with Convolutional Neural Network

Mambang^{a,*}, Finki Dona Marleny^b

^a Department of Information Technology, Faculty of Science and Technology, Sari Mulia University, Banjarmasin, Indonesia

^b Department of Informatics, Faculty of Engineering, University of Muhammadiyah Banjarmasin, Indonesia

Corresponding author: *mambang@unism.ac.id

Abstract—Learning content can be identified through text, images, and videos. This study aims to predict the learning content contained on YouTube. The images used are images contained in the learning content of the exact sciences, such as mathematics, and social science fields, such as culture. Prediction of images on learning content is done by creating a model on CNN. The collection of datasets carried out on learning content is found on YouTube. The first assessment was performed with an RMSProp optimizer with a learning rate of 0.001, which is used for all optimizers. Several other optimizers were used in this experiment, such as Adam, Nadam, SGD, Adamax, Adadelta, Adagrad, and Ftrl. The CNN model used in the dataset training process tested the image with multiple optimizers and obtained high accuracy results on RMSprop, Adam, and Adamax. There are still many shortcomings in the experiments we conducted in this study, such as not using the momentum component. The momentum component is carried out to improve the speed and quality of neural networks. We can develop a CNN model using the momentum component to obtain good training results and accuracy in later studies. All optimizers contained in Keras and TensorFlow can be used as a comparison. This study concluded that images of learning content on YouTube could be modeled and classified. Further research can add image variables and a momentum component in the testing of CNN models.

Keywords— Image; exact and non-exact; learning content; CNN; deep learning.

Manuscript received 28 May 2022; revised 4 Jul. 2022; accepted 16 Aug. 2022. Date of publication 31 Dec. 2022.
International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

Learning content is something important in supporting the learning process. Learning content can be integrated on many digital platforms [1]. Easily accessible learning content has a positive impact on those who use it. Visual media is experiencing significant growth with various digital platforms [2]. Learning content can be identified through text, images, and also videos. Identification with images can be made with Deep Learning. Deep Learning provides an image classification process with the Keras module and TensorFlow. Artificial intelligence has shown excellent work in image and speech recognition [3]. Image object recognition with high-quality images can improve the performance of the created model [4]. Image recognition and processing are very well done with Convolutional Neural Networks [5]. Training and Testing are carried out after all datasets have been collected so that the process of validating drawings can be carried out [6].

Learning content is very important in supporting the learning process but is still not optimally utilized. Interactive digital content provides new opportunities and challenges for educators and students [7]. Using the latest technologies in many areas accelerates technology adaptation for users [8]. Identifying learning content in the exact field and non-exact fields in digital platforms such as YouTube needs to be researched. Deep learning needs to be optimized using new datasets to dig data into knowledge. Deep learning is developed and creates a framework for model formation, resulting in new knowledge [9]. Convolutional Neural Networks require a model of the framework architecture and the number of laps processed on the network. Convolutional Neural Networks need to be implemented on many images around us because they provide high accuracy values. The quantity and quality of the various image objects used for training impact the performance of Convolutional Neural Networks [10].

Several studies related to image prediction using a Convolutional Neural Network, such as in research by M.P. Islam et al. [11] used the tomato leaf image as an object to

predict the performance of the Convolutional Neural Network. It shows that the Batch Normalization layer significantly improves network performance, achieving accuracy in the classification of more than 99.0% in the training process, 97.5% is validation and 98.0% results in the testing process. Wei et al. [12] use Convolutional Neural Networks to reconstruct hyperspectral image tomography, and the CNN used provides higher precision and shorter reconstruction times. Simone et al. [13] make predictions using CNN to predict the trajectory of pedestrians. The proposed CNN method introduces a new 2D convolution model. In 2022, Agustina et al. [14] performed HPV predictions in advanced oropharyngeal cancer with Convolutional Neural Network 2.5D. The results of their study were successfully applied and validated with two external cohorts in making predictions of HPV status in CT. Naseeb et al. [15] create a segmentation process on cotton using three Neural Network neural networks. Three neural networks are used, such as VGG16, InceptionV3, and ResNet34. Mingxia et al. [16] made detection and mapped the tree's crown with Convolutional Neural Network and Google Earth images, where the R-CNN Mask result has a detection value of 82.8%. Ananda et al. [17] create Transfer Learning classification of images in leaf disease in many plants using the VGG Convolutional Neural Network. The resulting classification of grape leaves affected by the disease with an accuracy of 98.40% and 95.71% on the leaves of tomatoes.

The use of Convolutional Neural Networks on many image objects has provided new knowledge, and this is what prompted us to predict images on learning content and find the best accuracy of several optimizers contained in TensorFlow and Keras, such as RMSProp, Adam, Nadam, SGD, Adamax, Adadelta, Adagrad, and Ftrl. From this instrument, of course, find a model and accuracy that matches the image dataset. Collecting image datasets is carried out by capturing the learning content. The learning content carried out by the retrieval process is categorized in the fields of the exact sciences and social sciences. The meaning of the exact field is mathematics learning content, and the non-exact is cultural learning content found on digital platforms such as Youtube. In addition to this experiment, the images we collect can be processed using sequential models on the Convolutional Neural Network that we will show in a series of methods, results, discussions, and conclusions.

II. MATERIAL AND METHOD

The data collection process and the stages of the process we carry out consist of the dataset's initiation and the model to be used. Then at the stage of model implementation using the dataset that has been created. Furthermore, the process of evaluating and validating the results has been done.

A. Dataset Collection

The collection of datasets that we do is obtained from learning content contained on digital platforms in this paper using the content contained on YouTube. The learning content is then used in both exact sciences and social sciences learning content.



Fig. 1 (a) Images of exact sciences learning content, (b) Images of social sciences learning content

The image above is a sample of the learning content image dataset on YouTube consisting of exact sciences learning content such as mathematics and social sciences learning content such as culture. All image datasets are processed with a Convolutional Neural Network model in the following process. CNN is very effectively used in processing image datasets[18]. In Convolutional Neural Networks, there is a maximum mean that serves to reduce the features that are extracted [19].

TABLE I
IMAGE DATA OF EXACT SCIENCES AND SOCIAL SCIENCES LEARNING CONTENT

Data Image	Image information used	
	Exact sciences image	Social sciences images
Dimensions	605 x 348	1098 x 549
Width	605 pixels	1098 pixels
Height	348 pixels	549 pixels
Bit depth	24	24

The table above shows information from later image data used in the test. The image used when collecting data is still not adjusted from its size. In loading the image on CNN, an ImageDataGenerator will be used to process the data before it is loaded. In the library Keras, ImageDataGenerator serves to transform data into the form of an array collection. There are several functions in ImageDataGenerator to vary images, such as horizontal_flip used to flip images horizontally.

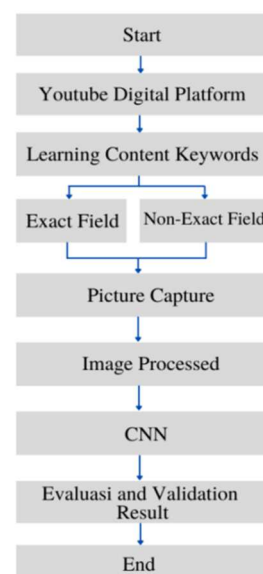


Fig. 2 Data collection flowchart used in this experiment

The rescale function is used to scale the image value so that when the image is uploaded, then the image is represented as a 3D array [20]. The validation split function is used to validate from training, and this validation split divides the data from the dataset as desired. To create a model, validation can use validation techniques. In machine learning, validation techniques often used are split validation and cross-validation. Simple validation can use the split validation technique, where the dataset is randomly divided into training and test data. Meanwhile, the cross-validation dataset is divided into many sets or subnets of training and validation data.

The process and stages of collecting datasets begin with keywords related to learning content on the YouTube digital platform, such as learning mathematics. Mathematics belongs to the field of exact sciences. In the next stage, after the keyword process is used on the search menu on YouTube, the image capture process is carried out by taking a video on the

first index. Likewise, searching for images in non-exact fields is done in the same way as learning content in the exact field. After we capture the learning content on YouTube, the image dataset is divided into training and validation data (See Fig. 2).

B. Convolutional Neural Network

The algorithm in deep learning that functions in recognizing images and making predictions and classification is the Convolutional Neural Network. Convolutional Neural Networks are also used in detecting and recognizing objects in an image. The CNN architecture has the advantage of conducting image training that results in more detailed image predictions and classifications. Stages in the CNN architecture has several processes for conducting image dataset training.

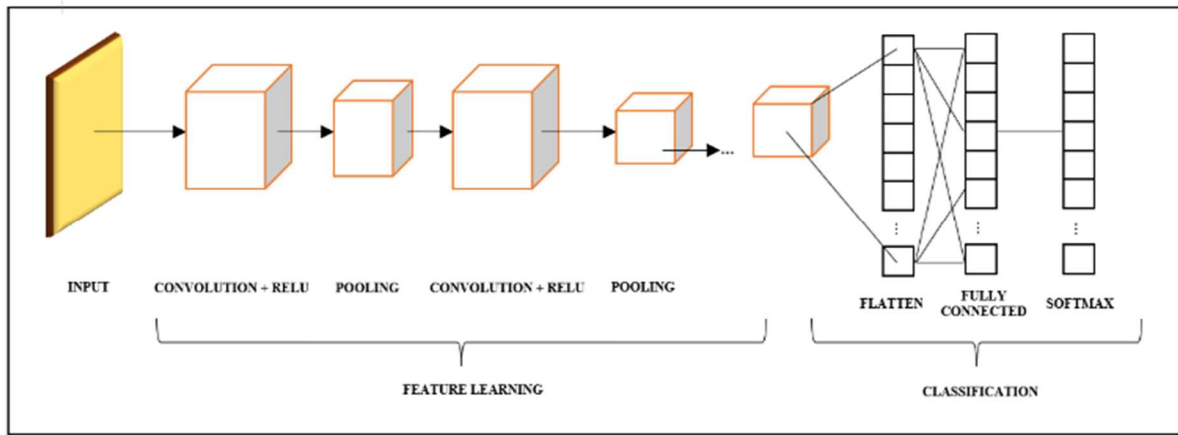


Fig. 3 Convolutional Neural Network Architecture

Each stage of the CNN architecture has a convolution process, a layer activation process, and a layer pooling process. Fig. 3 shows the first process in the CNN architecture, which is commonly referred to as convolution [21]. The first step in this process is to create a kernel at a specific size. The calculation in creating a processed kernel depends on the number of generated features and continues on the activation process. A prevalent activation function is ReLu. The next step is the pooling process after the activation process is carried out. At this stage of pooling, it can be processed and repeated so that later you get a feature map that can continue in the fully connected process and get a label or class output in the process carried out with the CNN architecture. The process operation on convolutions that can be used on CNN processes consists of:

$$s(t) = (x * t)(t) = \sum_{\alpha=-\infty}^{\infty} x(\alpha) * w(t - \alpha) \quad (1)$$

$S(t)$ results from convolution operations in this process, while x is input and w is a weight (kernel). The $s(t)$ function can produce a single output called the feature map. The first input x is an argument while the argument w as the second is a kernel or filter. Hyperparameters function in determining the output volume of each layer. Hyperparameters to make calculations of the number of activation neurons on the output we can show in the equation below:

$$(W - F + 2P) / (S + 1) \quad (2)$$

Based on the equation above, spatial size can be processed or calculated, thus obtaining the output volume with the hyperparameter used consisting of the size of the volume (W), filter (F), and shift or stride applied (S) as well as determining the amount of padding to be processed (P). In the equation above, W is a measure of the volume of the image, F is a filter, P is used to determine the value of the padding, and S is a value of the size of the shift or stride. Using stride in filtering the image input and zero padding will get a zero-value found around the image border area.

C. Purposed Model

Building the model used in this experiment starts by loading the data by entering several libraries such as TensorFlow and Keras and then importing the image generator. In addition, there are also other libraries such as matplotlib, NumPy, and cv2. After the library components are inputted, an image load is displayed to display sample images from the entire dataset. Furthermore, the shaping process is carried out using the cv2 library to see the image information, and the initial image information consists of (545,1099,3). 545 is the width value, 1099 is the height, and 3 is the number of channels. Data preprocessing is done by creating an ImageDataGenerator (rescale= 1/255) for train and validation. The dataset is divided into 80% for training, 20% for validation. The training dataset and the validation dataset in

the directory are also done in this process, such as target size= (200,200), batch size = 3, and class mode binary.

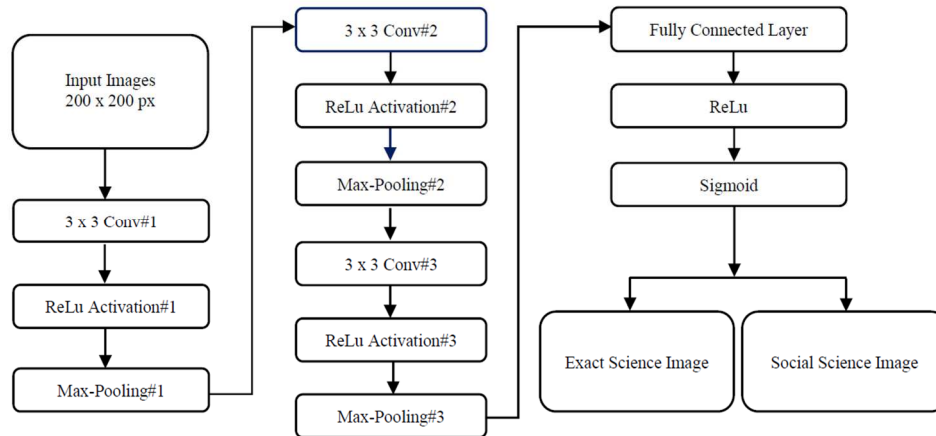


Fig. 4 CNN model used in this study

The classes contained in the dataset consist of 'exact science' : 0 and 'social science: 1. The next process performs a Fully Connected Layer with sigmoid activation. The next stage is compiling a model where for Loss using binary cross-entropy, optimizer used RMSprop, learning rate 0.001, and metrics used are accurate. The fit model on the training dataset uses steps_per_epoch five and epochs 30.

Fig. 4 indicates a summary model of the model created. After the model process is created, the next step is to evaluate the model. A CNN has dozens or hundreds of layers in which each layer learns to detect features of different types in an image. Convolutional Neural Network filters to each training image applied and processed through different resolutions. In contrast, the result or output on the convoluted image is used for input at the next layer.

III. RESULT AND DISCUSSION

The experiment process was carried out on the Convolutional Neural Network model created. The first test process was carried out with an RMSProp optimizer with a learning rate of 0.001. This learning rate was used for all optimizers. Learning rate is one of the training parameters to calculate the weight correction value during the training process. This learning rate value ranges from zero (0) to (1). The greater the learning rate, the faster the training process will run.

The greater the learning rate, the less network accuracy, but the opposite applies. If the learning rate is lower, the network accuracy will be greater or increase, with the training process consequences taking longer and longer. Several other optimizers were used in this experiment, such as Adam, Nadam, SGD, Adamax, Adadelata, Adagrad, and Ftrl.

The following process is to find the best results on the training dataset image of the exact and non-exact learning content and get the best accuracy results from the CNN model process used. In the first test, the accuracy results of the training dataset using the RMSprop optimizer and a learning rate of 0.001 (See Figs. 5 and 6).

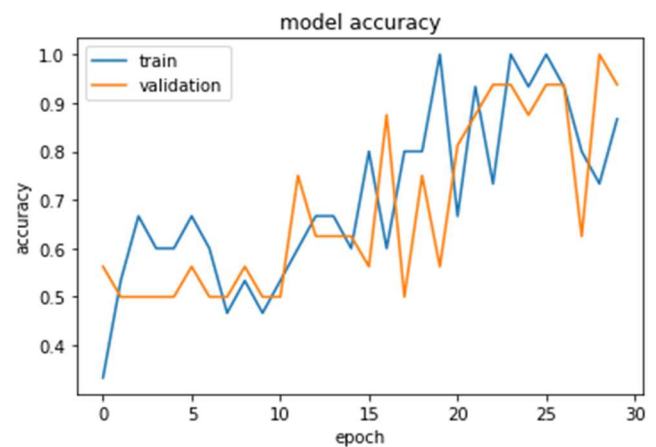


Fig. 5 Model accuracy on RMSprop optimizer

Of the 30 Epochs used in the RMSprop optimizer, the accuracy result was 0.8630, and the validation accuracy was 0.937.

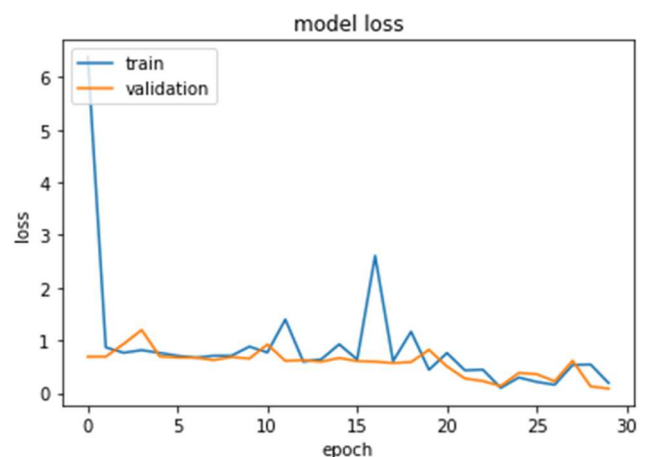


Fig. 6 Model loss optimizer RMSprop

Of the 30 Epochs used in the RMSprop optimizer, the Loss obtained was 0.1911, and the validation loss: was 0.0833. The table below shows the results of testing datasets of exact and non-exact learning content images with several optimizers.

TABLE II
TEST RESULTS WITH MULTIPLE OPTIMIZERS

Optimizer	Acc	Val Acc	Loss	Val Loss
RMSProp	0.8630	0.937	0.1911	0.0833
Adam	1.0000	1.000	0.0088	0.0245
Nadam	0.7185	0.500	8.5652	0.6598
SGD	0.5472	0.625	0.6655	0.6671
Adamax	1.0000	1.000	0.1102	0.0590
Adadelata	0.3963	0.562	0.6940	0.6814
Adagrad	0.6259	0.625	0.6673	0.6404
Ftrl	0.4519	0.500	0.6932	0.6931

In Testing with Nadam, SGD, Adadelata, Adagrad, and Ftrl optimizers, the testing process found prediction inaccuracies because there were gaps in the training datasets processed based on the accuracy and loss results. This can happen because all five optimizers require a momentum component. The momentum component is performed to improve the speed and quality of neural networks. The CNN model used in this experiment was without creating a momentum component in the compile process. The results of the predictions using the model made we show in Fig. 7 below:

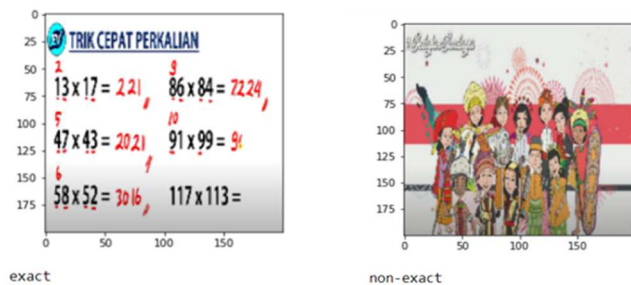


Fig. 7 Predicted results from the CNN model with RMSprop, Adam, Adamax

Fig. 7 shows the prediction results from the RMSprop optimizer, Adam, Adamax. The tests conducted with the three optimizers found accuracy with high values on the Adam and Adamax optimizers.

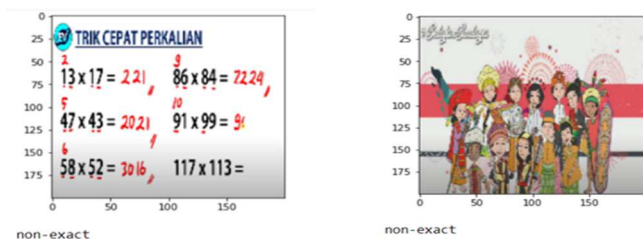


Fig. 8 Predicted results from CNN models with Nadam, SGD, Adedelata, Adagrad, and Ftrl

Fig. 8 above shows the prediction results from the Nadam, SGD, Adedelata, Adagrad, and Ftrl optimizers. Richard et al. [5] proposed the Convolutional Neural Network in their research to create a Cataract classification on Fundus Images by comparing four CNN architectures when applying the Adam optimizer with a learning rate of 0.001.

Their study had an equation of learning level of 0.001 with the experiment we were working on. M.P. Islam et al. [11] create a series of parallel convolution neural networks to predict tomato leaf disease by selecting an Adam optimizer to optimize tissue weights. Their research also combined Adam with several optimizers, such as Adagrad and RMSprop.

Simone et al. [13] used Adam's optimizer to build a Convolutional Neural Network model to predict pedestrian trajectories. We found the best results in Adam's optimizer from several test results with different optimizers. Several previous research studies and the results of our experiments concluded that making predictions on image objects is highly recommended using the Adam optimizer with a learning rate of 0.001. Figs. 9 and 10 below show the accuracy of the Adam optimizer that we are working on.

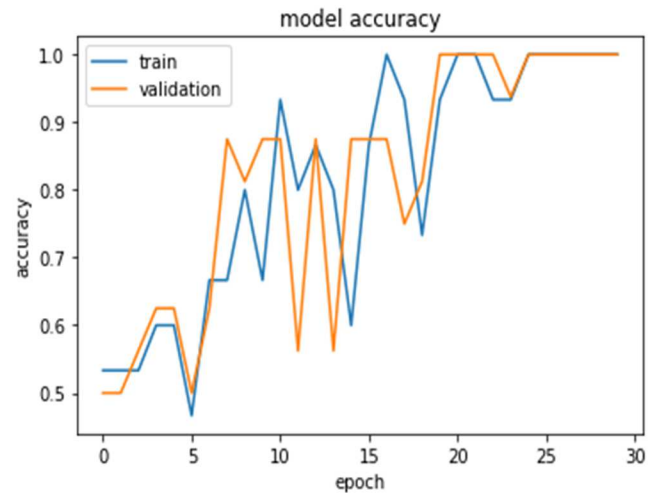


Fig. 9 Model accuracy on Adam optimizer

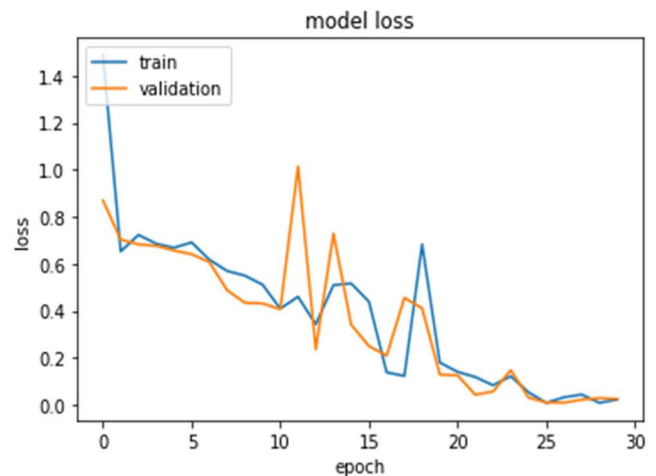


Fig. 10 Model loss on Adam optimizer

The Deep Learning approach can overall detect objects' features based on convolutional neural networks [22]. The Convolutional Neural Network model that we created using the captured dataset on learning content on YouTube can be processed with Adam's optimizer and has the best accuracy and Loss. In the experiment that we did in this study, there are still many shortcomings, such as not using the momentum component. The momentum component is performed to improve the speed and quality of neural networks. All optimizers contained in Keras and TensorFlow can be used as a comparison. Testing on the image dataset training can be done with any image input. In subsequent studies, we can develop a CNN model using the momentum component to get good training results and a high level of accuracy. Answering the question of the purpose of this study, whether capturing

images on learning content can be processed with the CNN model, it turns out that this can be done with the foundation is the experiment that we are working on and supported by previous research theories.

IV. CONCLUSION

Learning content is something important in supporting the learning process. Learning content can be integrated across many digital platforms, such as YouTube. Searching for learning materials using keywords such as learning mathematics on YouTube is still done by typing these keywords in the search menu. The CNN model used in the dataset training process tested the image with multiple optimizers and obtained high accuracy results on RMSprop, Adam, and Adamax. The CNN Prediction Model with Deep Learning in the future is expected to be able to directly predict learning content contained in digital platforms so that without typing in the search menu related to the material they want to learn, users can already be given display notifications on these contents related to the field of learning content in the exact sciences and social sciences. The experiments we worked on in this study showed that the CNN model could predict images of learning content contained on YouTube. Artificial Intelligence will certainly have an even greater positive impact on the lives of mankind.

ACKNOWLEDGMENT

The Information Technology Department of Sari Mulia University and the Informatics Department of the Muhammadiyah University of Banjarmasin support this research.

REFERENCES

- [1] P. D. Rebeca, A. Fern, and M. C. Rodr, "Integrating micro-learning content in traditional e-learning platforms," *Multimed. Tools Appl.*, vol. 80, pp. 3121–3151, 2020.
- [2] B. McCartney, B. Devereux, and J. Martinez-del-rincon, "A zero-shot deep metric learning approach to Brain – Computer Interfaces for image retrieval," *Knowledge-Based Syst.*, vol. 246, 2022.
- [3] Y. Matsuo, Y. Lecun, M. Sahani, D. Precup, and D. Silver, "Deep learning , reinforcement learning , and world models," *Neural Networks*, vol. 152, pp. 267–275, 2022.
- [4] A. Eko, M. Hazmi, C. Mandiri, Y. Azhar, and F. Bimantoro, "Classification of Diabetic Retinopathy Disease Using Convolutional Neural Network," *Int. J. INFORMATICS Vis.*, vol. 6, no. March, pp. 12–18, 2022.
- [5] R. Bina *et al.*, "Cataract Classification Based on Fundus Images Using Convolutional Neural Network," *Int. J. INFORMATICS Vis.*, vol. 6, no. March, pp. 33–38, 2022.
- [6] S. Aulia and D. Rahmat, "Brain Tumor Identification Based on VGG-16 Architecture and CLAHE Method," *Int. J. INFORMATICS Vis.*, vol. 6, no. March, pp. 96–102, 2022.
- [7] X. Xia and W. Qi, "Artificial Intelligence Temporal tracking and early warning of multi semantic features of learning behavior," *Comput. Educ. Artif. Intell.*, vol. 3, no. August 2021, p. 100045, 2022.
- [8] S. Minn, "AI-assisted knowledge assessment techniques for adaptive learning environments," *Comput. Educ. Artif. Intell.*, vol. 3, no. July 2021, p. 100050, 2022.
- [9] S. Liu *et al.*, "Tool path planning of consecutive free-form sheet metal stamping with deep learning," *J. Mater. Process. Tech.*, vol. 303, no. February, p. 117530, 2022.
- [10] P. M. Blok, G. Kootstra, H. Elchaoui, B. Diallo, F. K. Van Evert, and E. J. Van Henten, "Active learning with MaskAL reduces annotation effort for training Mask R-CNN on a broccoli dataset with visually similar classes," *Comput. Electron. Agric.*, vol. 197, no. December 2021, p. 106917, 2022.
- [11] M. P. Islam, K. Hatou, T. Aihara, S. Seno, S. Kirino, and S. Okamoto, "Performance prediction of tomato leaf disease by a series of parallel convolutional neural networks," *Smart Agric. Technol.*, vol. 2, no. March, p. 100054, 2022.
- [12] W. Huang, M. Svanborg, M. Juul, and M. Toudal, "The application of convolutional neural networks for tomographic reconstruction of hyperspectral images," *Displays*, vol. 74, no. January, p. 102218, 2022.
- [13] S. Zamboni, Z. Tilahun, S. Girdzijauskas, C. Norén, and L. Dal, "Pedestrian trajectory prediction with convolutional neural networks," *Pattern Recognit.*, vol. 121, p. 108252, 2022.
- [14] A. La *et al.*, "A 2 . 5D convolutional neural network for HPV prediction in advanced oropharyngeal cancer," *Comput. Biol. Med.*, vol. 142, p. 105215, 2022.
- [15] N. Singh, V. K. Tewari, P. K. Biswas, L. K. Dhruw, C. M. Pareek, and H. D. Singh, "Semantic segmentation of in-field cotton bolls from the sky using deep convolutional neural networks," *Smart Agric. Technol.*, vol. 2, no. March, p. 100045, 2022.
- [16] M. Yang *et al.*, "Detecting and mapping tree crowns based on convolutional neural network and Google Earth images," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 108, no. March, p. 102764, 2022.
- [17] A. S. Paymode and V. B. Malode, "Transfer Learning for Multi-Crop Leaf Disease Image Classification using Convolutional Neural Network VGG," *Artif. Intell. Agric.*, vol. 6, pp. 1–11, 2022.
- [18] H. Trang, N. Thanh, and D. Hwang, "Convolutional attention neural network over graph structures for improving the performance of aspect-level sentiment analysis," *Inf. Sci. (Ny)*, vol. 589, pp. 416–439, 2022.
- [19] I. Rodriguez-martinez, J. Lafuente, and R. H. N. Santiago, "Replacing pooling functions in Convolutional Neural Networks by linear combinations of increasing functions," *Neural Networks*, vol. 152, pp. 380–393, 2022.
- [20] H. Min, T. Ko, I. Young, and J. Myong, "Asbestosis diagnosis algorithm combining the lung segmentation method and deep learning model in computed tomography image," *Int. J. Med. Inform.*, vol. 158, p. 104667, 2022.
- [21] F. Chen and J. Yeu, "Assessing the effects of convolutional neural network architectural factors on model performance for remote sensing image classification: An in-depth investigation," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 112, p. 102865, 2022.
- [22] L. Ruo, H. Kamaludin, N. Zuraidin, M. Safar, N. Wahid, and N. Abdullah, "Intelligence Eye for Blinds and Visually Impaired by Using Region- Based Convolutional Neural Network (R-CNN)," *Int. J. INFORMATICS Vis.*, vol. 5, no. December, pp. 409–414, 2021.