Analysis of Drowsiness Detection Based on Images Using Convolutional Neural Network

Anwar Nasihin, Habibullah Akbar, Gerry Firmansyah, Budi Tjahjono

Universitas Esa Unggul, Jakarta, INDONESIA

E-mail: anwarnasihin725@gmail.com

| Submitted: August 04, 2023 | Revised: January 24, 2024 | Accepted: February 22, 2024 |

| Published: May 20, 2024 |

ABSTRACT

Drowsiness detection is crucial in maintaining the safety and alertness of individuals, especially in high-risk situations such as driving or operating heavy machinery. This research aims to develop a drowsiness detection system based on facial images using Convolutional Neural Network (CNN) with a focus on the AlexNet method and its comparison with ResNet. In this study, facial image data was collected from various conditions of drowsiness and normal conditions. Image preprocessing was performed to standardize the size and ensure consistent image quality. AlexNet and ResNet were implemented and trained using the image dataset to identify distinctive patterns that differentiate drowsy faces from faces in a normal state. The results of the experiments showed that the use of AlexNet and ResNet methods effectively detects drowsiness in facial images with high accuracy. However, there are performance differences between the two methods. ResNet demonstrated superior performance in certain conditions, while AlexNet showed advantages in other cases. This research contributes to the development of facial image-based drowsiness detection technology applicable in various fields, including smart vehicles and security systems. The comparison results between AlexNet and ResNet also provide valuable insights for selecting the most suitable CNN method for drowsiness detection applications based on facial images.

Keywords: drowsiness detection using CNN; AlexNet; ResNet; AlexNet; ResNet Comparison.

INTRODUCTION

In this era of globalization, the development of technology and information has dominated almost all aspects and fields of human life (Mowlana, 1997). These aspects that before being implemented with traditional methods slowly began to be replaced by the development of information technology (Rosa et al., 2013). The reason for the replacement of traditional methods is the advantages offered by information technology such as computerized systems that offer data processing and information reception in a faster time and higher accuracy (Zamzami et al., 2021). The majority of aspects that are replaced are techniques or ways of working on an activity that is now fixed using machines controlled by computer systems (Kittur et al., 2013).

The development of information technology is based on human thinking to be able to develop himself and human curiosity towards change (Mardhiyana & Sejati, 2016) (Asmoro & Mukti, 2019). This indirectly has an impact in the form of increasing needs in a short period of time. These increasing needs also directly increase the amount of burden and responsibility to every technology user and society at large (Karim et al., 2023).

Drowsiness is a common problem that we often encounter when a person's condition feels like sleeping (Ferber, 2006). There are many factors that can be found from this drowsiness, one of which is the fatigue factor (Rini et al., 2022). Fatigue can occur when we are working both in the office and at home due to the increasing number of these needs. The increase makes the work that must be done more and more increasing. So it tends to cause excessive fatigue or fatigue to every worker and every individual in life (Mangatta, 2016). According to some researchers, fatigue can significantly affect labor health and can reduce work productivity, where fatigue can contribute significantly to work accidents and fatigue can also affect health and body condition (Xing et al., 2020) (Techera et al., 2016) (Krueger, 1989).

pISSN 2302-4240
eISSN 2655-2086

Volume 13, Issue 2, June 2024, pp.378-388
DOI: http://dx.doi.org/10.32832/astonjadro.v13i12 http://ejournal.uika-bogor.ac.id/index.php/ASTONJADRO

From the statement above, a study will be proposed on the analysis of sleepiness detection through a picture, where the object in the picture will be used as a dataset to classify the level of sleepiness. The method to be used in this study is the Convolutional Neural Network method, as a convolution and kernel technique on image data samples.

Based on the above problems, you can take a thesis research entitled Analysis of Drowsiness Detection Based on Images Using the Convolutional Neural Network Method. While the formulation of the problem in this study is as follows: How to identify eye drowsiness accurately? How to calcify the identified object in terms of fatigue level? And the purpose of this study is to be able to find out things in a person's sleepiness from image processing of all postures, movement likeness and facial look. To Analyze the type of sleepiness in an image based on certain intervals.

Improving building infrastructure services has a significant impact on the activities of residents in the surrounding area. Good infrastructure not only improves the quality of life but also supports various social, economic and environmental aspects. Improved infrastructure such as smooth roads, safe sidewalks, and efficient public transportation makes it easier for citizens to access schools, hospitals, shopping centers, and workplaces. Improvements to buildings and public infrastructure such as street lighting, good drainage systems, and routine maintenance create a more comfortable and safe environment for residents (Nugraha IMGD et.al, 2023); (Devi NPKC, Trimarianto C, 2022); (Pratama SD et.al, 2024); (Sabariah I et.al, 2012).

RESEARCH METHODS

Steps of the Research Process

Problem Identification and Outcome Standardization

The problem in this study is an obstacle that will be faced when testing the method and *sample* data to be tested as research material. In this study, there is a major problem, namely the classification of images will be more processed if the dataset object contains other objects such as wearing glasses, wearing masks, other objects and others. Do not forget also from the results of problem identification, also accompanied by the determination of the final prediction results that will be obtained during testing for evaluation of testing in research. The result to be obtained is a prediction accuracy of 98% in the classification of fatigue levels based on images with deviation or miss levels, which is a maximum of 2%.

a. Research Data Collection
   Data or research data samples will be taken from public datasets from www.kaggle.com sites where there are several different images for trials in the matlab application
b. CNN Architecture Type Determination
   The study used the Convolutional Neural Network (CNN) method, so it is necessary to determine the architecture to be used. This affects the speed and accuracy of the classification results that will be generated in the CNN method and the model used in this model is alexnet.
c. Data Pre-processing
   The pre-processing data carried out is resize, crop, and grayscale as a standard for pre-processing models in the Convolutional Neural Network method using the AlexNet architecture for the application of the Convolutional Neural Network method. The image is taken from the public data set to carry out the pre-processing process to facilitate the classification process and also adjust to the chosen CNN architecture, namely AlexNet.

Data Training

Data training is an activity to provide a dataset set to the machine that will be used by algorithms against the dataset to perform learning on machine learning. Because this method uses deep learning in the form of the CNN method, the training data will be done by dividing some data from the dataset that has been obtained and collected as much as 70% for training data and 30% for testing data. The training data will first be carried out on the CNN method through a series of layers, namely input layer, convolutional layer, pooling layer, fully connected layer, and output layer and will be carried out again on the results of the CNN classification.

a. Data Testing

After obtaining learning results in the form of training, then use the remaining 30% of the data to test the results of the training that has been done.

b. Use of Convolutional Neural Network Method
Using the CNN method is a form of classification carried out in deep learning and produces a classification of images taken from public datasets In the image certainly produces features that can be used in classification.

Characteristics of the Final Research Results

The results or output of the Convolutional Neural Network (CNN) method are in the form of classifying a person's sleepiness from images or images through a series of processes in layers that focus on identifying open and closed eyes, and the width of mouth openness contained in the image when testing data is carried out, which provides a value in the form of a number as a result of measuring each of these three features. The results of measuring these features will determine the similarity label or likeness of various aspects to determine the level of fatigue of a person in the image.

CNN Alexnet architecture

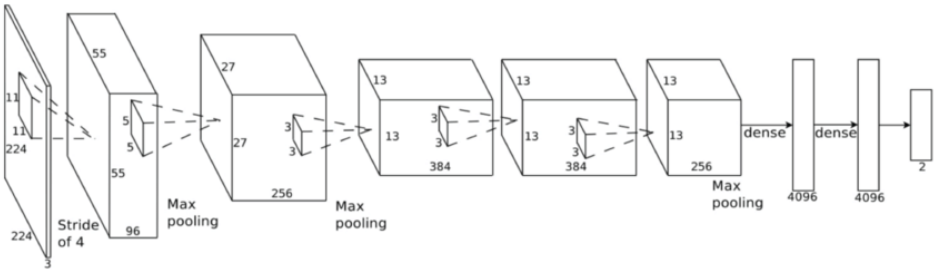

Figure 1. CNN Alexnet architecture

Alexnet is one of CNN's architects who won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition which is a large-scale image classification competition in 2012. AlexNet is a new breakthrough in deep learning by applying ConvsNet combined with Dropout Regularization techniques, the use of ReLu as an activation function and data augmentation.

CNN Resnet architecture

ResNet stands for "Residual Network" or "Residual Neural Network." ResNet is one of the most popular convolutional neural network.

(CNN) architectures in the field of image processing and image recognition.

The ResNet architecture was developed by Kaiming He and his colleagues in 2015. One of the problems that often arises in the use of neural networks in computer vision tasks is performance degradation as the network becomes deeper. That is, the deeper the network, the more difficult it is to achieve higher accuracy.

ResNet introduced the concept of a "skip connection" that allows information to skip multiple layers in a network. In ResNet, each layer has two branches: one is a direct path that contains convolution and activation operations, and the other is a connection shortcut that skips multiple layers.

Shortcut connections allow for easier flow of information from one layer to another, thus addressing performance degradation issues. Using connection shortcuts, ResNet can learn the difference or residual between the input and output of each layer block. Therefore, ResNet is able to overcome the challenge of building a deeper network with better accuracy.

ResNet has various variants, such as ResNet-50, ResNet-101, or ResNet-152, which indicate the number of layers or depth in the network. The deeper the network, the more complex the features that can be learned, but it also requires more computing resources.

In practice, ResNet has proven to be very effective in a variety of computer vision tasks, including

pISSN 2302-4240
eISSN 2655-2086
Volume 13, Issue 2, June 2024, pp.378-388
DOI: http://dx.doi.org/10.32832/astonjadro.v13i12 http://ejournal.uika-bogor.ac.id/index.php/ASTONJADRO

image classification, object detection, segmentation, and more.

Convolutional Neural Network Process for Detection of Drowsiness in Matlab

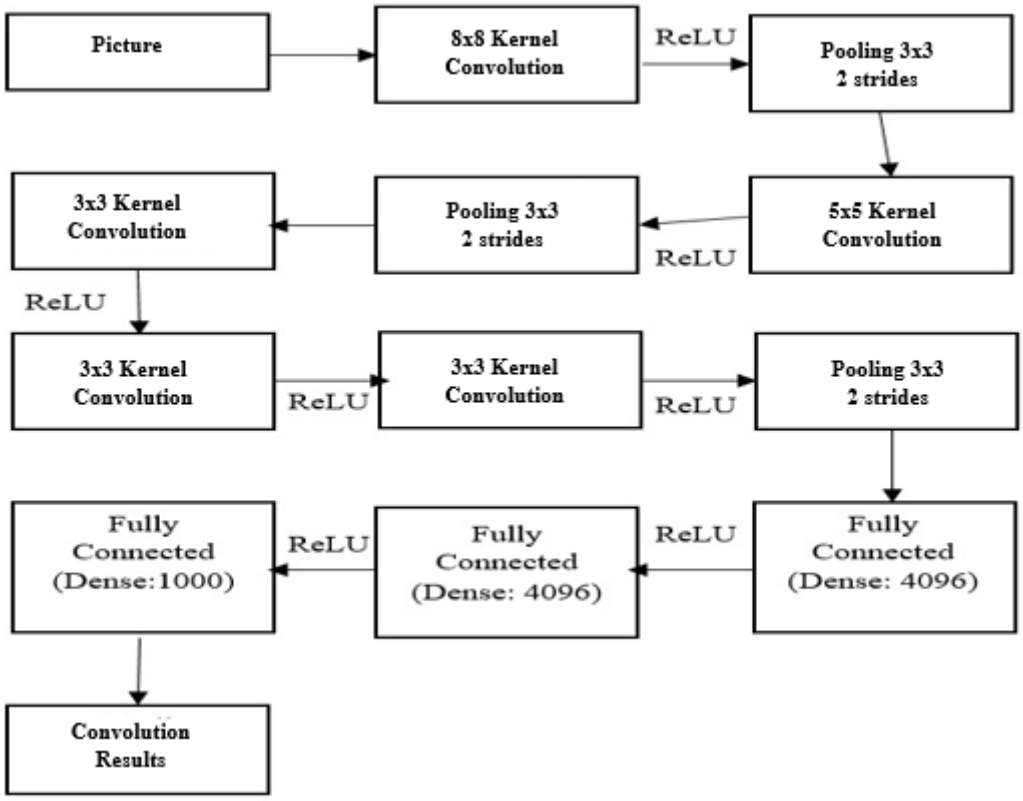

Figure 2. The image goes into the CNN via the input layer

Then an 8x8 kernel will be installed at the beginning of the image as the beginning of convolution. The first convolution will produce the first image filter and feature as a result of feature extraction. Then it will be continued on the ReLU activation function to do pooling in the form of max pooling at the pooling layer with two stride and 3x3 kernel pooling.

After that, the image will be processed again into a convolution layer with a smaller kernel size of 5x5 for and will go through the ReLU activation process again into the next pooling layer for more detailed feature extraction and image dimension reduction. After that, there will be a convolution of three layers with a kernel size of 3x3 where the ReLU function is used between the three layers and ushered to the 3x3 layer pooling with stride still numbering 2 after going through 3 convolution layers for clearer features in the image. Finally, it will be ushered to the fully connected layer for merging each node or the results of features obtained as many as 3 layers to get prediction and detection of sleepiness.

RESULT AND DISCUSSION

Experiments

To analyze the performance of our proposed method, we conducted a series of experiments, specifically, we discussed the loss function, training loss, accuracy and validation and the time required in training the model.

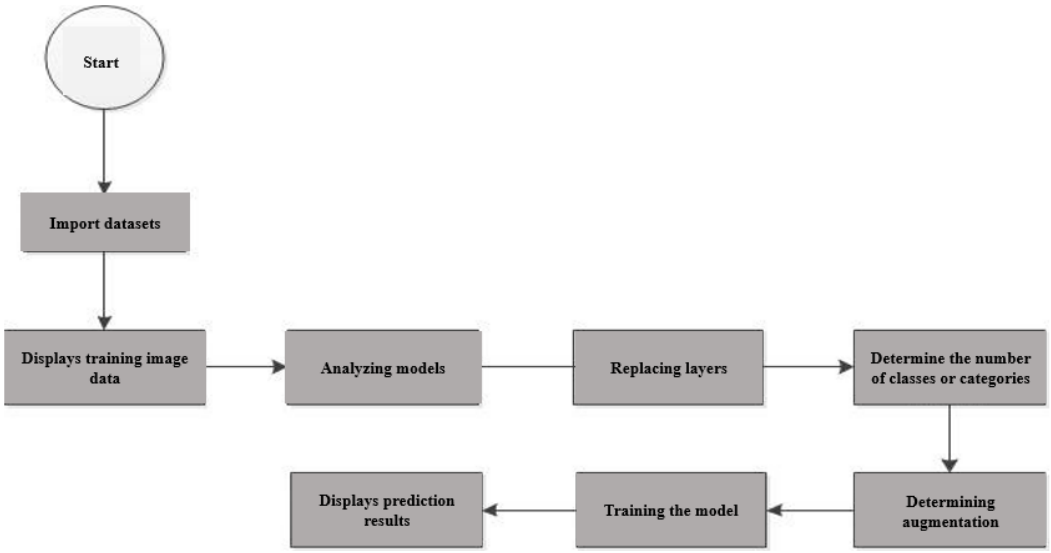

Figure 3. Flow Experiment

Like the plot described above, the experiment begins with the import of the dataset needed to classify the images.

At a later stage we downloaded the dataset which is a public dataset on the Kaggle site. https://www.kaggle.com/datasets/dheerajperumandla/drowsiness

Our dataset uses 2900 datasets consisting of multiple photos of multiple identities.

The number of data sets is divided into two parts, namely for training sets totaling 2030 and for validation sets totaling 870. In general, for data sharing there are 3 parts, namely:

a. Tarining sets are used to train the model i.e. calculate losses and adjust the weight of the model using gradient descent.

b. Validation sets are used to evaluate the model while training, adjust and select the best model

c. Test sets are used to compare different models or different types of modeling approaches and display the final accuracy of the model.

The process of changing layers during image processing can be done as follows:

Load model: First of all, you need to load the pretrained model. The model must have the appropriate architecture you need for image processing.

Replacing layers: Next, you need to change the layer you want to replace on the model. For example, if you want to replace the last layer on the model, then you can use the pop() function to delete the last layer and then add a new layer using the add() function.

Compile model: After changing the layer, you need to recompile the model so that the changes can be carried out properly. You can define optimizers, loss functions, and other evaluation metrics that suit your needs.

At this stage the system displays sample images for traning, images are taken randomly as many as 16 images with different categories. Below is an example of a picture to be trained.

pISSN 2302-4240
eISSN 2655-2086

Volume 13, Issue 2, June 2024, pp.378-388
DOI: http://dx.doi.org/10.32832/astonjadro.v13i12 http://ejournal.uika-bogor.ac.id/index.php/ASTONJADRO

Figure 4. Example of a picture to be trained

Sample dataset to detect drowsiness with four categories, open (open eyes), close (closed eyes), yawm (yawn), no yawm (not yawning) After taking the existing images in the training system conduct training of the selected model model using the alexnet model.

This architecture process includes the following steps and a brief explanation of the AlexNet algorithm:

a. Convolution: AlexNet has 5 convolution layers with filters measuring 11x11, 5x5, and 3x3. Each convolution layer is followed by max-pooling.

b. Batch Normalization: AlexNet implements batch normalization on initial layers to accelerate convergence.

c. Activation: After each convolution layer, AlexNet uses ReLU activation to increase the convergence speed.

d. Dropouts: AlexNet also uses dropouts on the last layers to prevent overfitting.

e. Fully Connected Layers: AlexNet has 3 fully connected layers with 4096 neurons each.

f. Output Layer: The output layer consists of 1000 neurons representing different classes in the ImageNet dataset.

g. Optimizer: AlexNet uses a stochastic gradient descent (SGD) optimizer to minimize function loss.

In determining the class, the author sets 2 calss categories, namely clasification yawn (yawn) and no yawn (not yawn)

Randomized data augmentations

Data augmentation in general is the process of processing datasets by experimenting with datasets in the form of taking pixels from images, with activities in the form of reversing images, changing the diagonal direction to see angles and others.

To train an AlexNet model, you need to follow these steps: Prepare datasets: This includes collecting and organizing your datasets into appropriate folders for use during training, validation, and testing. We may also consider adding variations to your data to increase the number of datasets and reduce overfitting.

Preprocess the data: This preprocess includes resizing the image to a standard size, normalizing the pixel values in the image, and other transformations needed so that your data can be used in the training process.

Define model parameters: The parameters you should specify include batch size, number of epochs, and learning rate.

Create and train a model: We need to structure the AlexNet model and do the training using the dataset already prepared. During the training process, the model will correct its weights and biases to produce more accurate predictions.

Model evaluation: After the training process is complete, you need to evaluate the performance of the model using datasets that were not used during the training process. This will help you find out how well your model is at making predictions on new data. After performing the above steps the system displays the following prediction results:

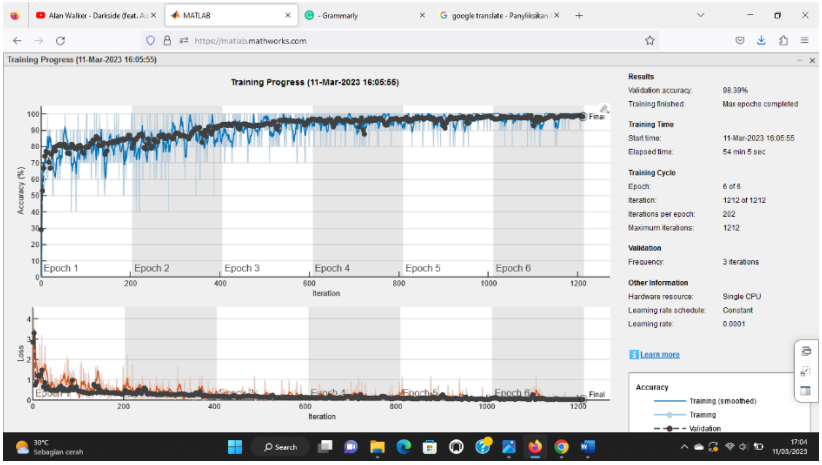

Figure 5. Results of accuracy and loss analysis in image detection training using alexnet

Based on the results of the training above, the Alexnet model has an accuracy of 98% with a time of 54 min 5 seconds with an epoch of 6 times out of 6 epochs with an iteration of 1212. Hardware resources for training images with Single CPU and Learning average 0.0001.

Previously we also conducted several experiments using alexnet with parameters that we randomly determined to get the best accuracy results with the best time as well, following the results of the trials we did.

Table 1. Here are the results of the graph of the experiment that has been carried out

Try to	Epoch	Learning Rate	Accuracy	Times
1	5	0.0001	96.55%	29min 39sec
2	4	0.001	49.89%	27min 56sec
3	5	0.00001	90.34%	77min 53sec
4	3	0.0001	87.82%	42min 3sec
5	3	0.0001	91.95%	18min 38sec

pISSN 2302-4240
eISSN 2655-2086

Volume 13, Issue 2, June 2024, pp.378-388
DOI: http://dx.doi.org/10.32832/astonjadro.v13i12 http://ejournal.uika-bogor.ac.id/index.php/ASTONJADRO

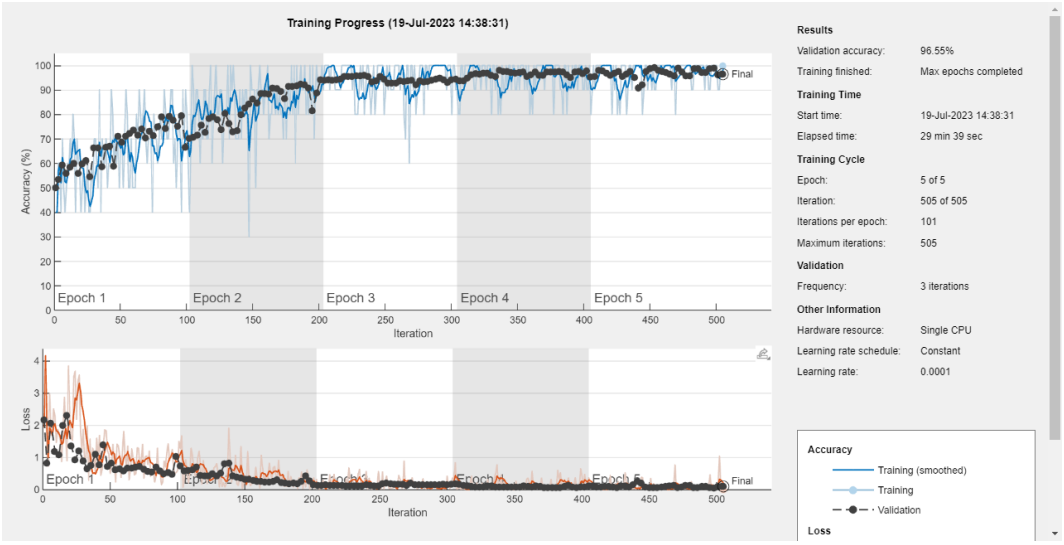

Figure 6. Results of the first experiment of the alexnet method

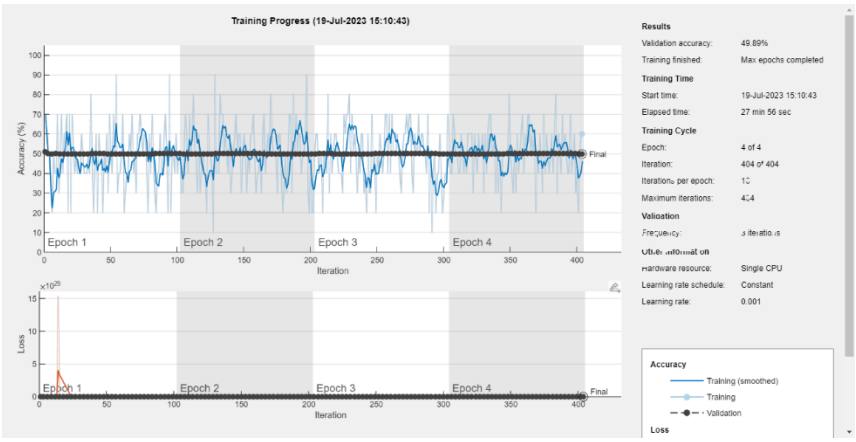

Figure 7. Results of the second experiment alexnet method

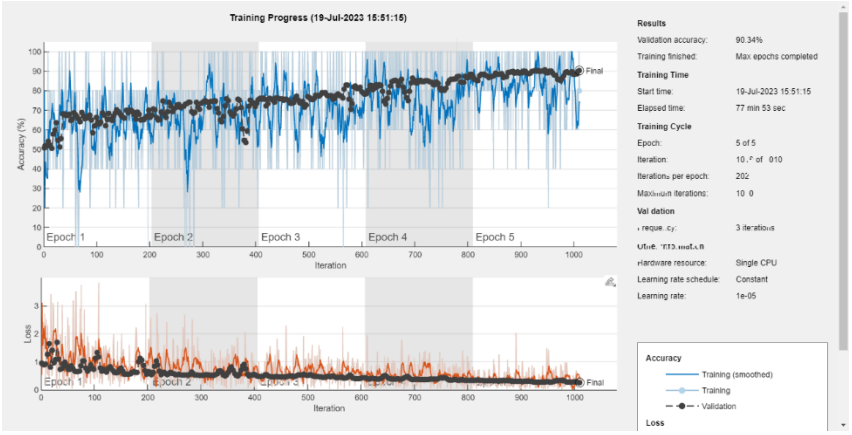

Figure 8. Results of the third alexnet method

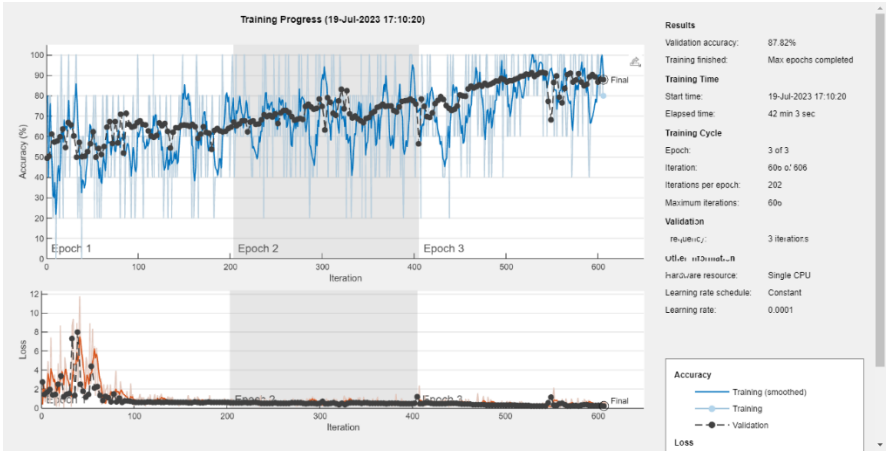

Figure 9. Results of the fourth trial of the Alexnet Method

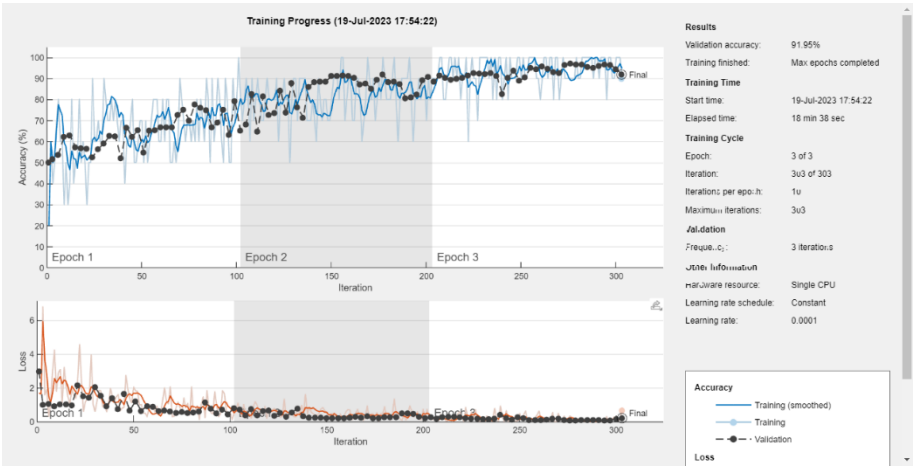

Figure 10. Results of the fifth experiment the alexnet method

Example of Prediction Results

The system will display images and labels of prediction results, like the example below, the system performs random image predictions in the dataset and displays images and labels of prediction results.

Figure 11. Prediction example results using alexnet model randomly

Comparison with resnet

pISSN 2302-4240
eISSN 2655-2086

Volume 13, Issue 2, June 2024, pp.378-388
DOI: http://dx.doi.org/10.32832/astonjadro.v13i12 http://ejournal.uika-bogor.ac.id/index.php/ASTONJADRO

Using the same dataset, Alexnet has an accuracy of 96.55% with a training time of 29 minutes while if using Resnet18 it has an accuracy of 99% with a time of 91 minutes. From the second result, alexnet comparison is better because with a faster training time of 62 minutes, it only has an accuracy difference of 3% from resnet, which is 98%.

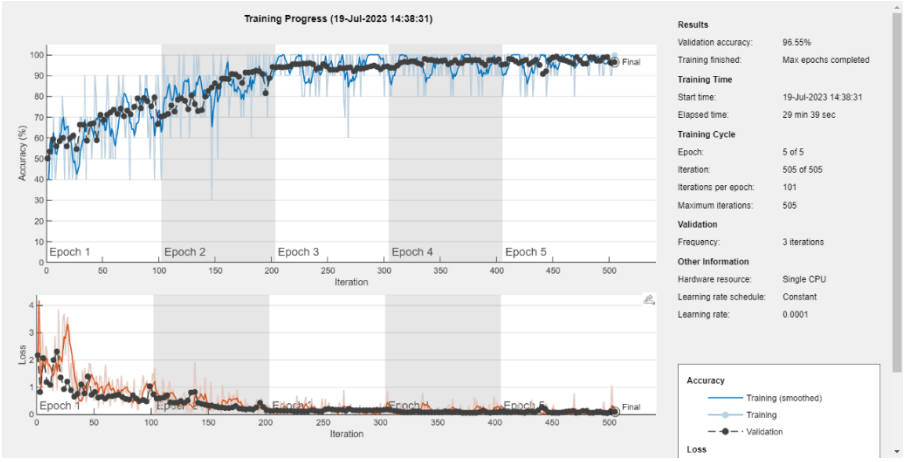

Figure 12. Image training results using Alexnet

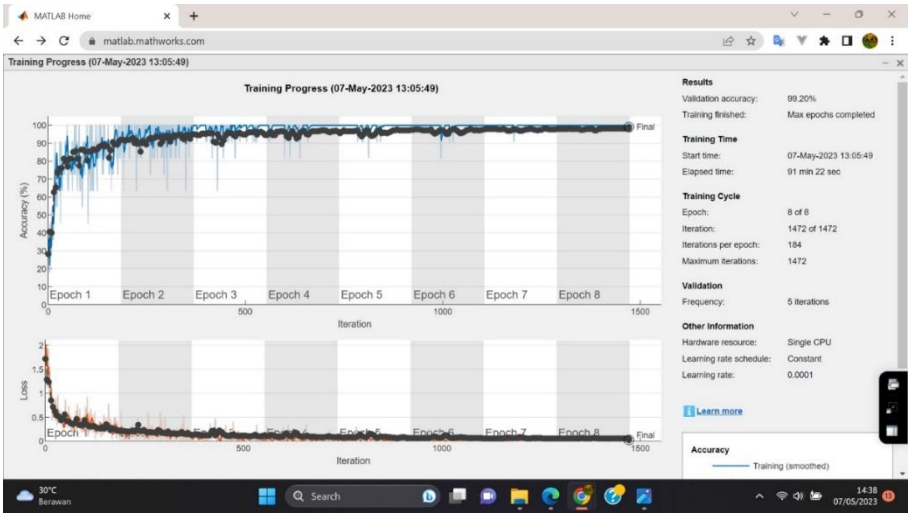

Figure 13. Training Results using Resnet18

CONCLUSION

We have used the alexnet model to detect sleepiness by dividing into 2 categories, namely yawn and no yawn with a fairly good accuracy of 96.55% with not too long time, in that process the data was divided into 70% training and 30% for testing. In addition, we apply augmentation to each image that loads images randomly so that each image is dynamically loaded, which allows training in many variations in loading images, allowing for better accuracy. We have compared the alexnet model with the resnet model on alexnet provides a smaller accuracy of 3% than the resnet model, but on alexnet only requires a training duration of 29 minutes while on resnet it reaches 91 minutes. In the restnet model, the resulting accuracy is 99% with a duration of minutes.

REFERENCES

Abdurohman, D., Putra, H. M., & Nurdin, I. (2020). Tinjauan Fiqih Muamalah Terhadap Jual Beli Online. *Ecopreneur*, *1*, 35–48.

Asmoro, B. P., & Mukti, F. D. (2019). Peningkatan rasa ingin tahu ilmu pengetahuan alam melalui

model contextual teaching and learning pada siswa kelas va sekolah dasar negeri karangroto 02. *Abdau: Jurnal Pendidikan Madrasah Ibtidaiyah*, *2*(1), 104–128.

Ferber, R. (2006). *Solve your child's sleep problems: new, revised*. Simon and Schuster.

Karim, H. A., Lis Lesmini, S. H., Sunarta, D. A., SH, M. E., Suparman, A., SI, S., Kom, M., Yunus, A. I., Khasanah, S. P., & Kom, M. (2023). *Manajemen transportasi*. Cendikia Mulia Mandiri.

Kittur, A., Nickerson, J. V, Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., & Horton, J. (2013). The future of crowd work. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 1301–1318.

Krueger, G. P. (1989). Sustained work, fatigue, sleep loss and performance: A review of the issues. *Work & Stress*, *3*(2), 129–141.

Mangatta, B. H. (2016). Strategi Adaptasi Tukang Becak Dalam Kehidupan Sosial Ekonomi (Studi Kasus Tukang Becak di Kelurahan Bontobiraeng Kecamatan Mamajang Kota Makassar). *HOLISTIK, Journal of Social and Culture*.

Mardhiyana, D., & Sejati, E. O. W. (2016). Mengembangkan kemampuan berpikir kreatif dan rasa ingin tahu melalui model pembelajaran berbasis masalah. *PRISMA, Prosiding Seminar Nasional Matematika*, 672–688.

Mowlana, H. (1997). Global information and world communication: New frontiers in international relations. *Global Information and World Communication*, 1–288.

Rini, W. N. E., Aurora, W. I. D., & Kusmawan, D. (2022). Pendekatan Structural Equation Modeling Kelelahan Kerja pada Pekerja Perkebunan Kelapa Sawit. *Window of Health: Jurnal Kesehatan*, 633–641.

Rosa, J., Teixeira, C., & Pinto, J. S. (2013). Risk factors in e-justice information systems. *Government Information Quarterly*, *30*(3), 241–256.

Techera, U., Hallowell, M., Stambaugh, N., & Littlejohn, R. (2016). Causes and consequences of occupational fatigue. *Journal of Occupational and Environmental Medicine*, *58*(10), 961–973.

Xing, X., Zhong, B., Luo, H., Rose, T., Li, J., & Antwi-Afari, M. F. (2020). Effects of physical fatigue on the induction of mental fatigue of construction workers: A pilot study based on a neurophysiological approach. *Automation in Construction*, *120*, 103381.

Zamzami, F., Nusa, N. D., & Faiz, I. A. (2021). *Sistem Informasi Akuntansi*. Ugm Press.

IMGD Nugraha, SA Paturusi, DGAD Putra. (2023). Poverty effect and population density towards the quality of fishermen's settlement environment in Kusamba. ASTONJADRO 12 (2), 360-367.

NPKC Devi, C Trimarianto. (2022). COMPATIBILITY OF WASTE MANAGEMENT PLACES AT TPS3R KSM BAYU SUCI, SUB-DISTRICT MARGA. ASTONJADRO 11 (2), 326-332.

SD Pratama, P Hadi, Hamkah, Mansyur, P Haerul, H Arman, IA Arifuddin, I Andi, S.Syaiful. (2024). MANAJEMEN PROYEK INFRASTRUKTUR. Tohar Media. Makassar, 1-153.

I Sabariah, S Syaiful, NI Hayati. (2012). ANALISIS METODE NETWORK PLANNING DAN S-CURVE PROYEK KONSTRUKSI DI BOGOR. ASTONJADRO 1 (1), 28-34.