Convolutional Neural Network Face Recognition for Lecturer Attendance

Muhammad Rafi Muttaqin[1], Anshorulloh Nur Aziz[1], Dede Irmayanti[1] and Sumanto[2]
[1] *Informatics Engineering Study Program, Wastukancana College of Technology, Purwakarta, Indonesia*
[2] *Universitas Bina Sarana Informatika, Jakarta, Indonesia*

Keywords: Face Recognition, MobileNet V2, Attendance.

Abstract: Face recognition is a field of research that is widely used to solve various problems, but to apply face recognition requires high accuracy so that there are no errors in the system that applies face recognition. The purpose of this research is how to use one of the architectures of the Convolutional Neural Network (CNN), namely MobileNet v2 to perform the task of face recognition of STT Wastukancana lecturers. The data used is taken from the social media of each lecturer, data sharing is done with the K-Fold Cross Validation method. MobileNet v2 architecture will perform classification tasks using different hyperparameter values to find the best performance. From various patterns, the best accuracy is 85dropout of 0.3 to reduce overfitting. Data sharing using K-Fold Cross Validation provides results that improve accuracy. The addition of a dropout layer reduces overfitting of the model.

1 INTRODUCTION

A face is one way to recognize a person's identity. Humans can recognize someone's name from looking at their face, if they have known that person before. Many computer applications or systems that are made require a person's identity, and there are also many ways to recognize that identity. Attendance system is one of the examples. There are various ways used in an attendance system, one of the simplest is by signing on paper which is now used in the attendance system for lecturers at STT Wastukancana. To facilitate the attendance system, face recognition can be applied to replace the manual signature process on paper. Basically, face recognition is an image classification that is specialized for face classification only. Convolutional neural network (CNN) is the most suitable model used for image classification, because it has been specialized to separate and detect patterns in input images, thus making this approach useful in the field of face recognition(Farayola and Dureja, 2020). There are various CNN architectures such as AlexNet, GoogleNet, LeNet 5, or MobileNet. In this journal, the author will use the MobileNet v2 architecture, because this model was developed for efficiency and without sacrificing many resources (S. K. A. B. Singh, 2019). MobileNet is built using a deeply decoupled convolutional architecture for the development of a lightweight model(Howard, 2017). There was two versions of MobileNet, MobileNet v1 and MobileNet v2. The updates in MobileNet v2 are the addition of bottleneck layers and shortcut connections(Sandler et al., 8 12). Convolutional neural networks have been used in previous research for face recognition classification. Thirty-nine (39) classes were included in the dataset. Fully Connected Layer, pooling layer, and Convolutional layer without additional architecture were used for training and the accuracy obtained was 86.71 (Abhirawan et al., 2017).

Cross-Industry Standard Process for Data Mining or CRISP-DM is one of the datamining process models (datamining framework) which was originally (1996) built by 5 companies namely Integral Solutions Ltd (ISL), Teradata, Daimler AG, NCR Corporation and OHRA (Mauritsius and Binsar, 2020). CRISP-DM has the advantage over other models of a clear definition of the Business Understanding phase. This phase is not at all considered in detail in other Data Mining models(Chapman, 2020). Deep learning has been used in various areas such as computer vision, natural language processing, audio recognition, including face recognition. Deep learning is a multi−layer algorithm for extracting characteristics and identifying edges such as letters, numbers, faces, etc. (Farayola and Dureja, 2020).

Convolutional is a subset of deep neural networks

Muttaqin, M., Aziz, A., Irmayanti, D. and Sumanto, .
Convolutional Neural Network Face Recognition for Lecturer Attendance.
DOI: 10.5220/0012447800003848

In *Proceedings of the 3rd International Conference on Advanced Information Scientific Development (ICAISD 2023)*, pages 255-261
ISBN: 978-989-758-678-1

that have been introduced to evaluate visual images. Convolutional neural networks have been specialized to isolate and recognize patterns in visual inputs, thus applying this method to the field of face recognition (Farayola and Dureja, 2020). The problem that exists in the face recognition process is that there are differences in light intensity and also differences in poses in existing data (Zhao et al., 2003). In general, frameworks that process input face images through a feature extraction method and then the feature extraction is recognized by a classifier method for identification (Abhirawan et al., 2017).

Author at (Peryanto et al., 0 05) using K-Fold Cross Validation looking at data division can improve the accuracy of the model. KFold Cross Validation is a given data set divided into a number of K parts/folds where each fold is used as a test set at some point.

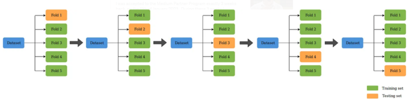


Figure 1: Fold Cross Validation (Krishni, 2018).

Data augmentation is a process in image data processing, augmentation is the process of changing or modifying images in such a way that the computer will detect that the changed image is a different image, but humans can still know that the changed image is the same image (Mahmud et al., 2019).

MobileNet is an efficient deep learning model that may be deployed on embedded devices or mobile devices such as smartphones without sacrificing a lot of resources(S. K. A. B. Singh, 2019). MobileNet is built using a depthseparable convolutional architecture to create lightweight models (Howard, 2017). There was two versions of MobileNet, MobileNet v1 and MobileNet v2. The updates in MobileNet v2 are the addition of bottleneck layers and shortcut connections (Sandler et al., 8 12).

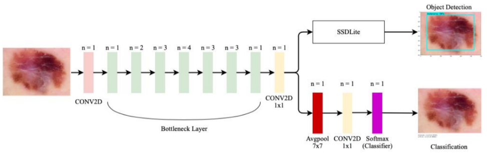


Figure 2: MobileNet v2 Architecture for object detection and classification (Wibowo et al., 0 07).

ReLU is the default activation in MobileNet v2's activation layer. ReLU is an activation function that was first introduced by H Sebastian Seung in 2000. The activation function serves to activate and deactivate neurons (Agarap, 2018). Specifically, ReLU6 is used in every layer except the last convolution layer. The equation for the ReLU6 activation function is shown in equation 1.

$$f(x) = min(max(0,x)6) \tag{1}$$

where $f(x)$ is the result of ReLU6 activation, and x is the value applied to be changed in the range (0,6). (Wibowo et al., 0 07).

1.1 Convolutional Layer

Convolution is considered to be a situation where a filter is applied to the input data (image) and gives the activation result. Also, it can be said to be a linear operation that involves multiplication performed between the set of weights and the input. These are the layers required for feature extraction from an input image (Farayola and Dureja, 2020).

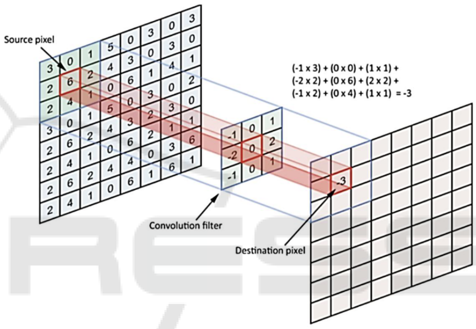

Source pixel
(-1 x 3) + (0 x 0) + (1 x 1) +
(-2 x 2) + (0 x 6) + (2 x 2) +
(-1 x 2) + (0 x 4) + (1 x 1) = -3
Convolution filter
Destination pixel

Figure 3: Visualization of convolution layer (Biswas and Islam, 2021).

1.2 Pooling Layer

This layer is usually and regularly used in CNNs to reduce the size of the input data to increase the computational speed of the network. It functions in each feature map independently. Hence, whenever a situation of excessive image input arises, the pool layer part will reduce the number of parameters. Also, pooling can be of different types. There are several types namely sum pooling, average pooling, and max pooling. Usually, the most commonly used is max pooling. Max pooling is a procces known as sample-based discretization.. It down-samples the input data, minimizing the dimensionality of the input and creating space for assumptions to be made regarding the sub-regions in which the features are located (Farayola and Dureja, 2020).

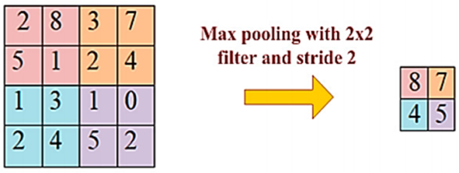


Figure 4: Visualization of max-pooling layer (Biswas and Islam, 2021).

1.3 Global Average Pooling

Global average pooling is to generate one feature map for each corresponding category of the classification task in the last convolution layer. Instead of adding fully connected layers on top of the feature maps, it then takes the average of each feature map, and the resulting vectors are fed directly into the softmax layer. One advantage of pooling global averages over fully connected layers is that it is more native to the convolution structure by enforcing the correspondence between feature maps and categories (Lin et al., 2014).

Another advantage is that there are no parameters to optimize in global average pooling so overfitting is avoided at this layer. In addition, global average pooling summarizes spatial information, so it is more robust to spatial translation of inputs(Lin et al., 2014).

1.4 Dropout

ropout is a process that prevents overfitting and also speeds up the learning process. Dropout refers to removing neurons that are either hidden or visible layers in the network. The neurons to be dropped will be chosen randomly (Abhirawan et al., 2017).

2 RESEARCH METHODS

The proposed architecture for performing face classification uses MobileNet v2. This section describes the data and architecture in more detail.

2.1 Business Understanding

The purpose of this research is to create a system that can recognize the face of a lecturer using a camera that can be used for a lecturer attendance system. The system will use artificial neural network, using convolutional neural network method for face recognition. The system will receive input in the form of a face image and will be processed on an artificial neural network, then will produce an output of the lecturer's name from the input face image.

2.2 Convolutional Layer

The data to be used is in the form of facial images of STT Wastukancana lecturers. The raw data needed is a photo of a lecturer with RGB colors that are visible on his face. The size of the required photo must be more than 224x224 pixels. The face part must be clearly visible, because the face part will be used.

Figure 5: Raw Data Example.

2.3 Convolutional Layer

For this face recognition research, photos of each STT Wastukancana lecturer will be taken from each social media that will be used. Where there are 10 labels which are the names of lecturers from each lecturer which can be seen in Table 1.

Table 1: Total Datasets.

No	Lecturer Name	Label	Total Data
1	Agus Sunandar	Agus	10
2	Syariful Alam	Nature	63
3	Chandra Dewi Lestari	Chandra	99
4	Dede Irmayanti	Dede	18
5	Irsan Jaelani	Irsan	20
6	Meriska Defriani	Meriska	109
7	Mochzen G. Resmi	Mochzen	55
8	M. Rafi Muttaqin	Raf	16
9	Rani Sri Wahyuni	Rani	166
10	Yusuf Muhyidin	Yusuf	335
	Total Data		891

Table 1 shows the names of the lecturers to be used, the labels to be used for each lecturer, the number of photos of each lecturer, and the total amount of data. The data obtained will be cropped, which is to cut part of the digital image so that only the necessary parts of the face are visible. Then the resize process is carried out, which changes pixels in Figure 6.

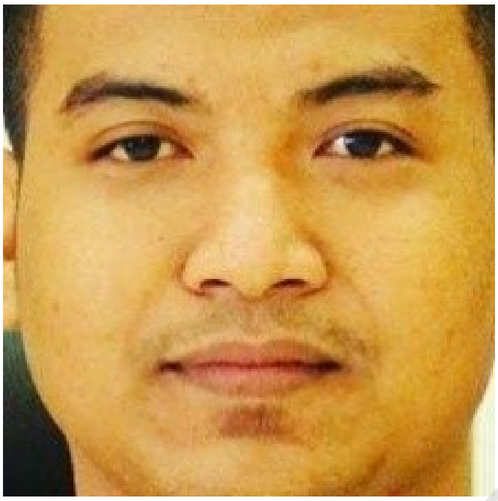


Figure 6: Example of Dataset photo after Cropping Process.

After the cropping process, the dataset photos that are too large are resized. An example of a resized photo can be seen in Figure 7.

Figure 7: Example of Dataset photo after Resize Process.

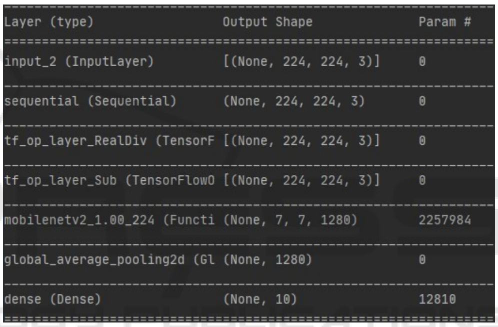
Then a dataframe will be created with 2 columns, namely filename and label (Table 2). The filename column will be filled with the file name of all images and the label column will be filled with the label of each image according to the image in the same row.

2.4 Modeling

The model built will use the MobileNet v2 architecture using the tensorflow framework. With several additional layers before MobileNet v2 including input layer, data augmentation, and preprocessing to scale image data between 0-255 to -1-1. Some additional layers after MobileNet v2 include Global Average Pooling and output layer.

Table 2: Example of Dataframe in Use.

No	File Name	Label
1	Agus001.jpg	Agus
2	Agus002.jpg	Agus
3	Agus003.jpg	Agus
4	Agus004.jpg	Agus
5	Agus005.jpg	Agus
...	...	...
887	Yusuf332.jpg	Yusuf
888	Yusuf333.jpg	Yusuf
889	Yusuf334.jpg	Yusuf
890	Yusuf335.jpg	Yusuf



```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         [(None, 224, 224, 3)]     0
_________________________________________________________________
sequential (Sequential)      (None, 224, 224, 3)       0
_________________________________________________________________
tf_op_layer_RealDiv (TensorF [(None, 224, 224, 3)]     0
_________________________________________________________________
tf_op_layer_Sub (TensorFlowO [(None, 224, 224, 3)]     0
_________________________________________________________________
mobilenetv2_1.00_224 (Functi (None, 7, 7, 1280)        2257984
_________________________________________________________________
global_average_pooling2d (Gl (None, 1280)              0
_________________________________________________________________
dense (Dense)                (None, 10)                12810
=================================================================
```

Figure 8: Design Model.

3 RESULT AND DISCUSSION

Python 3.7.9 with Tensorflow 2.3.1 framework was used in this study, which was conducted on NVIDIA GeForce GTX 1050 3GB GPU and AMD Ryzen 5 3550H laptop processor. The proposed architecture has been run on tensorflow with several different parameters. Once the architecture has been implemented, training on the model is done. In the training process, the dataset is divided into 2 parts, training data and validation data using k-fold cross validation with k = 5 to find the best data division. If overfitting occurs, dropout will be added to the model to reduce overfitting. Then to improve accuracy, training experiments will be conducted with a larger number of epochs. The training process will use Adam as model optimization, calculate loss with Crossentropy Loss, and calculate how often the prediction is correct by calculating the accuracy. Comparison of the results of the training process will be done by comparing parameter values. The parameters compared are the k-fold value, dropout, and number of epochs. Comparison of parameter values is done based on research from (Rokhana, 9 03)

3.1 Effect of K-Fold Value

The effect of sharing data with k-fold with the number of k=5 with the number of epochs of 50, resulting in 5 models that have different accuracies. The effect of k-fold value on model accuracy can be seen in Table 3.

Table 3: Effect of K-Fold Value on Accuracy Value and Loss Model.

K-Fold	Accuracy (Training)	Loss (Training)	Accuracy (Validation)	Loss (Validation)
1	0,94	0,24	0,85	0,4
2	0,94	0,24	0,76	0,80
3	0,94	0,23	0,77	0,74
4	0,93	0,25	0,80	0,58
5	0,96	0,19	0,83	0,70

Based on Table 3, it can be seen that in the training data, the best accuracy is in the fifth fold of 0.96, and in the validation data is in the first fold of 0.85. If the first fold and the fifth fold are compared in the amount of overfitting, the first fold is better against overfitting than the fifth fold, with a distance of 0.09 in the first fold and 0.13 in the fifth fold. And when viewed at the loss value, the first fold has a smaller loss value in the validation data of 0.47 while the fifth fold is 0.70. Therefore, from the results of data division using k-fold cross validation, the first fold is considered the best.

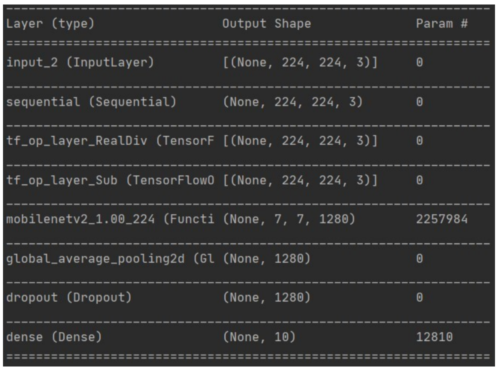
3.2 Effect of Dropout Rate

Based on the results of the effect of data division using k-fold cross validation, the first fold is the best. However, there is a little overfitting in the model. To reduce overfitting in the model, we will retrain the first fold by adding dropouts to the model which can be seen in Figure 9.

The number of dropouts is set from the smallest value between 0 to 1 to reduce overfitting. The dropout value starts from 0.1. If the overfitting value is still large, the dropout value will be added little by little by 0.1, until the training results on the model have no overfitting or have the smallest possible overfitting value. Training results on accuracy and loss can be seen in Table 4.

3.3 Effect of Number of Epochs

With the aim of increasing the amount of accuracy in the model, retraining is carried out with a larger number of epochs, namely 100 epochs. The results of training on accuracy and loss can be seen in Table 5.



```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         [(None, 224, 224, 3)]     0
_________________________________________________________________
sequential (Sequential)      (None, 224, 224, 3)       0
_________________________________________________________________
tf_op_layer_RealDiv (TensorF [(None, 224, 224, 3)]     0
_________________________________________________________________
tf_op_layer_Sub (TensorFlowO [(None, 224, 224, 3)]     0
_________________________________________________________________
mobilenetv2_1.00_224 (Functi (None, 7, 7, 1280)        2257984
_________________________________________________________________
global_average_pooling2d (Gl (None, 1280)              0
_________________________________________________________________
dropout (Dropout)            (None, 1280)              0
_________________________________________________________________
dense (Dense)                (None, 10)                12810
=================================================================
```

Figure 9: Model Design with Dropout Added.

Table 4: Accuracy and Loss Value With the Addition of Dropout.

Dropout	Accuracy (Training)	Loss (Training)	Accuracy (Validation)	Loss (Validation)
1	0,94	0,24	0,85	0,47
0,1	0,92	0,26	0,87	0,48
0,2	0,91	0,34	0,84	0,52
0,3	0,85	0,42	0,85	0,55

Table 5: Comparison of Accuracy and Loss Model Values With 50 and 100 Epochs.

Epoch	Accuracy (Training)	Loss (Training)	Accuracy (Validation)	Loss (Validation)
50	0,85	0,42	0,85	0,55
100	0,92	0,27	0,85	0,52

In Table 5, it can be seen that there is no improvement in the accuracy of the validation data. In the training data, there is an increase of 0.07. However, the increase in training data causes overfitting in the model. Therefore, adding epochs to 100 does not improve the accuracy of the model. The graph of the results of training models with 50 epochs can be seen in Figure 10 and for training models with 100 epochs can be seen in Figure 10 and 11.

An example of face recognition using the created model can be seen in Fig. 12. In Fig. 12, the red square line shows the face area that will be used for classification. The photo used is a photo of one of the lecturers labeled Yusuf. The classification results using the model that has been trained show the output is Yusuf, with a confidence value of 99.82%. So that the classification results are declared correct. The results of the model have been trained showing the highest accuracy value of 85 valitation data. The best model configuration uses input layer, data augmentation layer, image preprocessing to scale image data between 0-255 to -1-1, MobileNet v2, Global Average Pooling layer for output layer ten-class classification. The initial dataset is resized to 224x224 pixels. The division between training data and validation

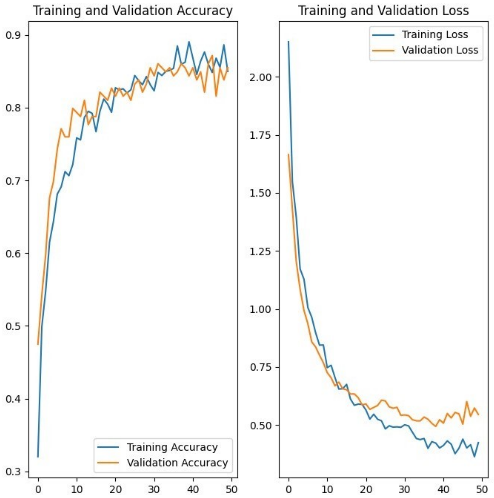


Figure 10: Graph of model training results with 50 epochs.

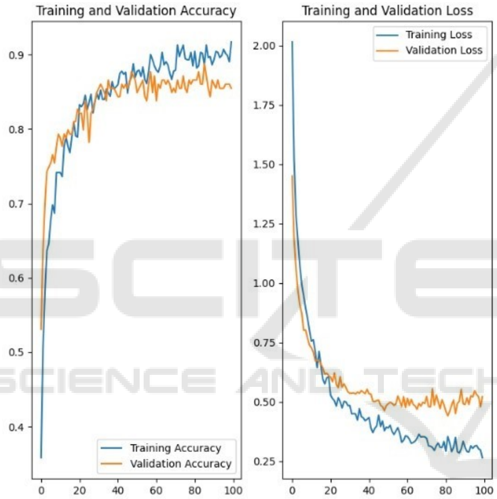


Figure 11: Graph of model training results with 100 epochs.

data uses K-Fold Cross Validation with the number k = 5, the final result shows the first fold is the division that gives the highest accuracy value. The process of training the model starts the image will be entered in the input layer with a size of 224x224x3. After going through the input layer, data augmentation will be carried out to train the model to recognize images from various points of view. Then the image data will be converted into a scale of -1 to 1 to match MobileNet v2. Next, the data will

go through the MobileNet v2 architecture for the training process, and will go through the global average pooling layer with a layer dropout of 0.3, the results of which will be classified into 10 classes. This model was trained and validated using 891 lecturer photo data. The accuracy obtained is 8550 epochs.

The addition of a dropout layer of 0.3 in the global average layer is enough to reduce the overfitting of the training results to make the accuracy of the training data and validation data the same at 85the number of epochs was not effective enough because the accuracy on the validation data did not change but only on the training data. However, it should be noted that the data used in this study only used 10 lecturers, not all lecturers. For application to the lecturer attendance system, it is necessary to use data from all lecturers.

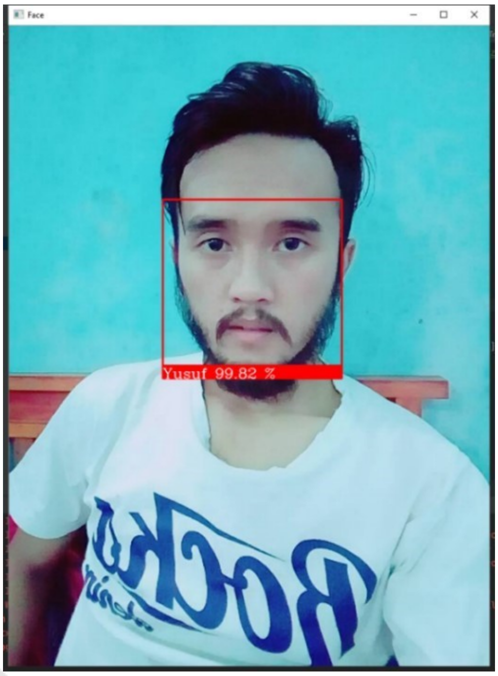


Figure 12: Example of Face Recognition Using a Model That Has Been Created.

4 CONCLUSIONS

The main topic of this research is to create a deep learning model that can recognize lecturers' faces to be applied to the lecturer attendance system at STT Wastukancana. This research uses photos of 10 lecturers taken from social media with a total amount of 891 data. The architecture model used is MobileNet v2 with the best parameter configuration resulting in an accuracy of 85In this model, the MobileNet v2 architecture is the main layer that carries out the training process, and classification is carried out through the global average pooling layer. The best results use data sharing with k-fold cross validation in the first fold with k = 5. The dropout layer added to the global average pooling layer of 0.3 is enough to reduce overfitting on training results so that the accuracy value of training and validation data becomes the same at 85%.

REFERENCES

Abhirawan, H., Jondri, and Arifianto, A. (2017). Face recognition using convolutional neural networks (cnn. *e-Proceeding Eng*, 4(3):4907-4916.

Agarap, A.F.M. (2018). Deep Learning using Rectified Linear Units (ReLU).

Biswas, A. and Islam, M. (2021). An efficient cnn model for automated digital handwritten digit classification. *J. Inf. Syst. Eng. Bus. Intell*, 7(1):42.

Chapman (2020). CRISP-DM ready for Machine Learning Project.

Farayola, M. and Dureja, A. (2020). A Proposed Framework: Face Recognition With Deep Learning.

Howard, A. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *Jounal Apr*. Accessed: Mar. 05, 2017

Krishni (2018). Evaluating a Machine Learning model can... 'by Krishni — Data Driven Investor — Medium. *Data Driven Investor*.

Lin, M., Chen, Q., and Yan, S. (2014). Network in network. In *2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc*, pages 1-10.

Mahmud, K., Adiwijaya, and Faraby, S. (2019). Multi-class Image Classification Using Convolutional Neural Network, e-Proceeding Eng. vol. 6, pag. 2127-2136.

Mauritsius, T. and Binsar, F. (2020). *Cross-Industry Standard Process for Data Mining (CRISP-DM*. MMSI BINUS University.

Peryanto, A., Yudhana, A., and Umar, R. (2020-05). Image classification using convolutional neural network and k fold cross validation. *J . Appl. Informatics Comput*, 4(1):45-51,.

Rokhana, R. (2019-03). Convolutional neural network for femur fracture detection in b-mode ultrasonic image. *J. Nas. Tech. Electro and Technol. Inf*, 8(1):59.

S. K. A. B. Singh, D. T. (2019). Shunt connection: An intelligent skipping of contiguous blocks for optimizing mobilenet-v2. vol. 118, pag.192-203.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. (2018-12). Mobilenetv2: Inverted residuals and linear bottlenecks. *Proc.IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*, pages 4510-4520.

Wibowo, A., Hartanto, C., and Wirawan, P. (2020-07). Android skin cancer detection and classification based on mobilenet v2 model. *J. Adv. Intell. Informatics*, 6(2):135-148.

Zhao, W., Chellappa, R., Phillips, P., and Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys*.