International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-7, Issue-6, June 2021
ISSN: 2395-3470
www.ijseas.com

Emotion Detection based on Facial Image Using Machine Learning Algorithm

Noor Bhagaskoro Liestantyo Hadhy[1], Aris Puji Widodo[2], Suryono Suryono[3]

[1]Magister of Information System, School of Postgraduate Studied, Diponegoro University, Semarang, Indonesia

[2]Depatemen of Computer Science, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia

[3]Doctoral Program of Information System, School of Postgraduate Studies, Diponegoro University, Semarang, Indonesia

Abstract

Emotions are an important issue in human life. Emotions can affect several aspects of human life such as decision making, level of aggressiveness, appetite, drug reactions taken and many related problems. Emotions need to be detected for medical and psychological purposes so that a person's mental state is in good condition and under control. Unfortunately, until now the emotion detection method is still manual and conventional through an expert. Another problem, the number of experts and the growth of emotional problems are not balanced so that the need for emotional diagnosis has not been met. In this study, a method of emotion detection with a computer information system is proposed. Systems and others where good and bad are strongly influenced by a person's positive or negative emotions. The method applied for emotion detection uses machine learning algorithms. In this study, training data was used in the form of the 2013 Facial Expression Recognition dataset (FER 2013). The emotion detection information system takes a sample image of a person being tested during screening. Tests were carried out on samples from various regions in Indonesia. From the tests conducted, the accuracy of emotion detection is above 70%.

Keyword:- Emotion, Decision, Medical, Expert, Diagnosis, Machine Learning Algorithm, Dataset, Accuracy

I INTRODUCTION

Face detection is the first step that must be done in facial analysis, including facial expression recognition [1]. Facial expressions are one of the visual clues that show a person's emotional state and human intentions. Emotions play an important role in human life and even affect physiological and psychological states [2]. Emotions can be known based on a person's micro-expressions and human micro-expressions are universal. Emotions can affect several aspects of human life such as decision making, level of aggressiveness [3],

appetite, and others where good and bad are strongly influenced by a person's positive or negative emotions. Currently, facial emotion recognition methods involving imaging technology are increasingly attracting attention [4]. Things are further complicated by the background and faces of the low resolution samples [5].

Face emotion detection aims to determine whether there are facial emotions or not in the image, and if so, where are the faces and the size of each face in the image [1]. Emotions can generally be divided into seven categories: anger, dislike, fear, happiness, sadness, surprise, and neutral [6]. The role of technology has brought changes in everyday life where technology has become an inseparable need. One of the emerging technologies, namely Computer Vision [7], which examines how computers work to process images such as images and videos that previously could only be done by human perception such as guessing someone's emotions from facial expressions [8]

Many studies have been conducted using computer vision techniques to identify emotions on the face [9]. In addition, the emergence of deep learning algorithms in computer vision is only able to carry out limited tasks, and requires codes that must be entered manually by programmers [10]. Many traditional methods that use learning are used for emotion classification, either based on handcrafted features [11] or based on deep neuron networks [12]. The existing classification method relies heavily on high-quality and large-scale data, especially deep learning methods [11]

Deep learning allows the process of reading objects to be easier [13]. One of the most important aspects of the success of this method is the availability of large amounts of training data [14]. Deep learning methods can be trained with very large data to learn features in representing data [15]. One of the most frequently used deep learning algorithms related to image classification is the Convolutional Neural Network [16]. The Deep Learning method based on Convolutional Neural Network is one of the most effective methods in analyzing image because its accuracy and performance is much better when compared to traditional methods [17].

In recent years, the popularity gained from deep learning methods has been increasing [18]. Convolutional Neural Networks and deep learning architectures have significant implications for diagnostic imaging [19]. Convolutional Neural Network is used to classify labeled data using the supervised learning method. The way the system works is by training using training data on target variables so that it can group data [16].

To obtain results from the right image processing method, an algorithm machine is needed that can classify images. Traditional Machine Learning methods have been able to identify

facial expression objects with a low level of accuracy, which is below 60% [20], MobileNetsV2 and NASNet models are Deep Learning models based on Convolutional Neural Networks that are most effective in analyzing images due to their much higher accuracy and performance. better than traditional methods [17].

MobileNetV2 is a Convolutional Neural Network architecture that attempts to work well on mobile devices. MobileNetV2 is based on an inverted residual structure where residual connections exist between bottleneck layers and enhances the performance of advanced mobile models across multiple tasks and benchmarks as well as across a spectrum of different model sizes. Unlike other deep learning models, MobileNetV2 provides high predictive accuracy without compromising computational and memory costs too much [21]. Meanwhile, Neural Architecture Search Network (NASNet) as a structure for the convolutional model. In its design, NASNet seeks to define high-performance building blocks in the categorization of a set of thumbnails [22].

Based on previous research, this study proposes a model for developing an information system with the application of deep learning by applying the MobileNetsV2 model [23] which is compared with NasNet [24] in detecting facial expressions. The evaluation of the MobileNetsV2 and NASNet models is expected to produce models with good performance and accuracy for detecting facial expressions, and can provide benefits for evaluating models related to facial expression data into something that can be useful for future research.

II MACHINE LEARNING TECHNOLOGY

Machine learning technology is part of artificial intelligence (AI). Machine learning consists of many layers of information processing stages in a hierarchical architecture that is utilized for unattended feature learning and for pattern analysis and classification. The essence of learning is to calculate features or hierarchical representations from observational data, where higher level features or factors are determined from lower level ones [25]. Computers are trained to use large data sets and then convert the image pixel values into an internal representation where the classifier can detect patterns in the input [16].

Machine learning algorithms perform learning that is represented in the form of a multilayer artificial neural network [26]. While representation learning itself is a method in machine learning to automatically extract/learn representations (features) from raw data. The raw data representation is then used for recognition or classification tasks. Some of the fundamental learning architectures are Convolutional Neural Network (CNN), Deep Belief Network

(DBN), Autoencoder (AE), and Recurrent Neural Network (RNN). Although an old idea, there are 3 important reasons for deep learning's current popularity: first, the discovery of new techniques (e.g., pretraining and dropout) and new activation functions (e.g. ReLU), secondly, huge data supply (big data), and the three processing chip capabilities that have drastically improved (eg GPU units) [27].

Convolutional Neural Network (CNN) is included in the type of deep learning because of the depth of the network. Deep learning is a branch of machine learning that can teach computers to do work like humans, including computers that can learn from the training process [8]. CNN has been successful in large-scale video processing and image recognition [28]. The application of CNN has been widely used by industries such as Amazon, Facebook, and Google [21]. On CNN each neuron is presented in 2-dimensional form, so this method is suitable for processing with image input [29].

CNN consists of two main stages, namely feature learning and classification. The feature learning stage consists of a convolution layer, ReLU (activation function) and pooling layer, while the classification stage consists of a flatten, fully-connected layer, and prediction. In each section of CNN there are two main processes, namely feed-forward and backpropagation [1]. CNN is one of the Deep learning models, which can be seen as an automatic extractor. The feature extractor contains the feature map layers and retrieves the distinguishing features of the raw image through three main layers: convolution layers, pooling layers, and the fully connected layer. The special operations are as follows [1]:

a. Convolution Layers: convolution operations are performed which is the main process on CNN. Convolution is a mathematical term which means applying a function to the output of another function repeatedly. When testing the presence of a feature on a new image, CNN will try all possible positions in the image. Filters are created to calculate feature matches across the entire image.

b. Pooling layers: a subsampling operation is performed, which is the process of reducing the size of an image data [17]. Pooling aims to reduce large images but still retain important information. One of the widely used approaches for CNN is max pooling, max pooling divides the image matrix into smaller parts and selects the largest value in them to be used when a new reduced image matrix is formed.

c. The fully connected layer: is a layer commonly used in Multilayer Perceptron (MLP) and aims to transform the data dimensions so that it can be classified linearly MLP with 2 fully-connected hidden layers shown in Figure 1. The output of this layer is an array with length of the number of classes that the model must select. In the fully connected layer, the

greatest weight from the previous layer will determine which features are most related to the class or label.

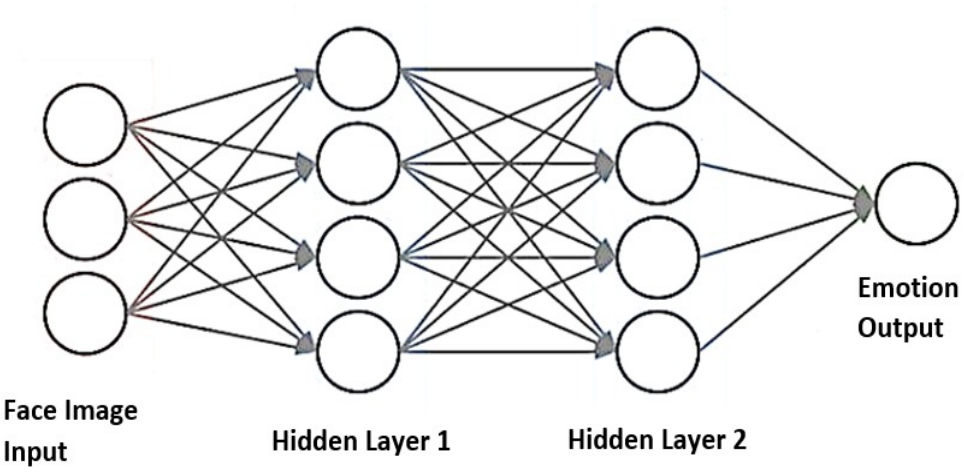


Figure 1. Simple Multilayer Perceptron 2 Hidden Layer

III BUILDING ANALYTICAL DATA FOR EMOTION DETECTION

The research describes the characteristics of the research object and the results of the research analysis (classification) will be presented. The implementation of this research procedure is carried out in several stages, including preparation, planning, training, testing, and conclusions from the results of the performance evaluation and the accuracy of the model for emotion detection.

A. Data Set Preparation

At the preparatory stage, namely determining the research topic, conducting theoretical studies, collecting information from journals, books, and other studies related to the research theme, and determining the model. Collecting data related to the research to be carried out.

The first stage is preprocessing for the training process, namely by automating face detection, in this case by detecting the face area and normalizing the image size. Furthermore, histogram equalization is carried out to expand the contrast of the image, as well as masking to cover the corners of the image so as to reduce variations that arise in these parts.

The research data used is a dataset from The Facial Expression Recognition 2013 (FER-2013) which was introduced at the 2013 International Conference on Machine Learning (ICML) [30]. The dataset for the training process uses the Facial Expression Recognition

(FER) dataset with 35,888 total grayscale image data and the size of 64 x 64 pixels in 7 categories of emotions that have been classified, namely anger, disgust, fear, joy, sadness, surprise and neutral. The data is already in the form of pixel values for each image in the form of an array with .csv format, so it can speed up the training process.

The FER-2013 data set contains 28,711 face grayscale images measuring 48x48 consisting of 6 different types of emotions. The data has been labeled and classified into 6 classes with an index between 0 to 5 as shown in Table 1. Data analysis was carried out with the aim of obtaining information about the data to be used for deep learning modeling. The labeling performed is shown in Table 1.

At this stage it is used to ensure data integrity so that when the dataset is processed at a later stage it does not cause problems in the training process. Class identification is carried out with the aim of knowing the class and the number of members of each existing class. Data integrity is needed to ensure the learning process runs as expected.

Table 1 Emotion class at FER2013

Label	Emotion Type	Total
0	Angry	4593
1	Disgusting	547
2	Afraid	5121
3	Happy	8989
4	Very Sad	6000
5	Surprised	4002
6	Neutral / Sad	6000

B. Training Stage

At this stage, the CNN training process consists of training using the MobileNetV2 model and training using the NASNets model. This stage requires input in the form of dictionary training obtained from the output of the data planning stage. Dictionary training is normalized at the data normalization stage, then the MobileNetV2 model and the NASNets model are trained at the CNN stage.

C. Validation Stage

Part of the data from the dataset is set aside to validate the results of the training carried out, in this process validation is carried out at each iteration during the training process.

D. Model Testing and Analysis Phase

Tahap ini dilakukan proses pengujian dengan mencoba memasukan gambar wajah untuk dilakukan prediksi. Metode *face detection* CNN diperlukan agar gambar yang akan dilakukan prediksi merupakan hanya gambar wajah. Hasil dari *face detection* CNN berupa hanya gambar wajah tersebut, maka dilakukan prediksi menggunakan model yang telah dilatih sebelumnya. Hasil akurasi klasifikasi ekspresi wajah dengan ekstraksi ciri menggunakan model *MobileNetV2*, pelatihan menggunakan model *NASNets* dan CNN.

VI RESULTS AND DISCUSSION

The emotion detection information system is made by computer programming with a data interface system as shown in Figure 2. Some examples of detected facial images are shown in Figure 3. Training on the mobilenet v2 model uses image input with a size of 224x224 pixels. The emotion prediction performance is quite fast under the 1 second per picture time range. This system is suitable to be selected in the emotion detection information system. The training process uses a previously created model, the model is taken from the Tensorflow Hub which is a repository of machine learning models.

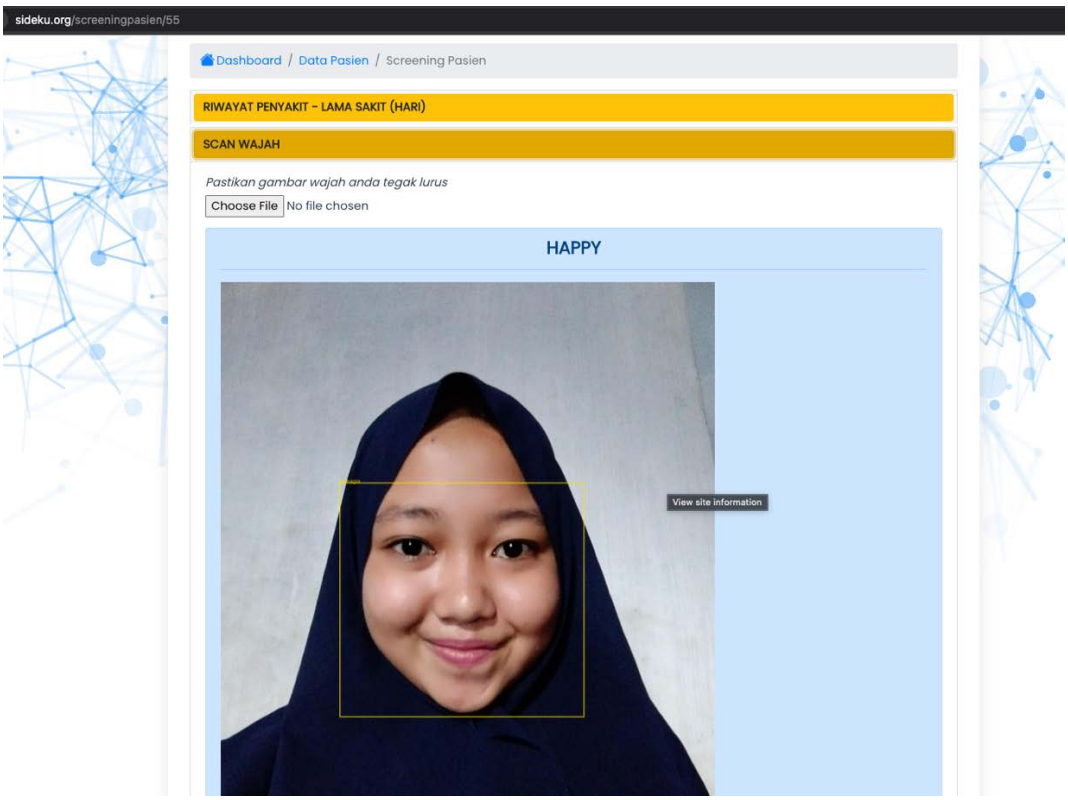


Figure 2. Information system for scanning face detection with mobilenet v2

Figure 3. An example of an Indonesian face image from the sample tested

In this study, 22,970 images were used for the training process. The training process is continued with the back propagation scheme that we have set the iteration when creating the training script. This process is repeated to obtain the smallest loss. After the dataset and model are prepared, a training process is needed to get accurate weights so that they can predict emotions from facial images as shown in Figure 4.

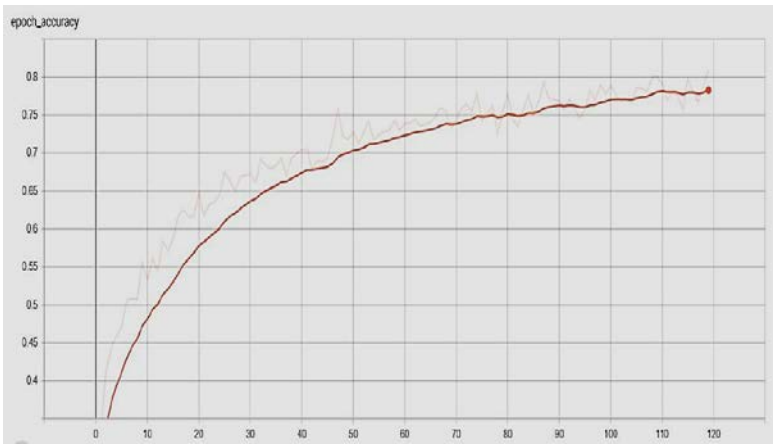


Figure 4. Accuracy in the mobile net v2 model process 224x224 pixel input

In Figure 5, it can be seen that the accuracy of using mobile net v2 with input of 224x224 pixels obtained the highest accuracy of 0.78 from a scale of 0 -1.

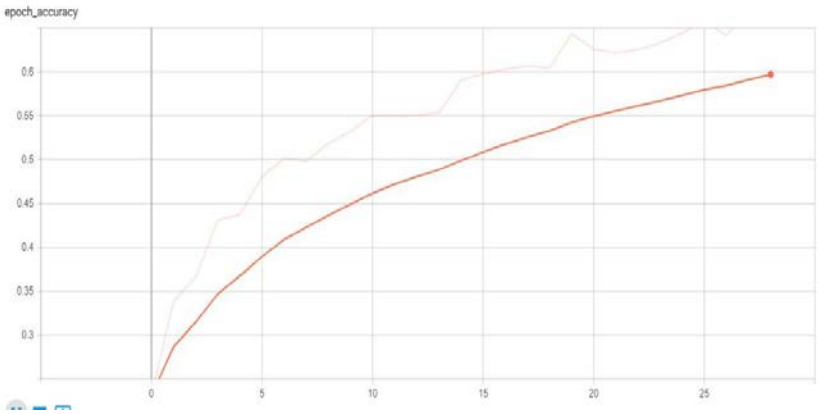
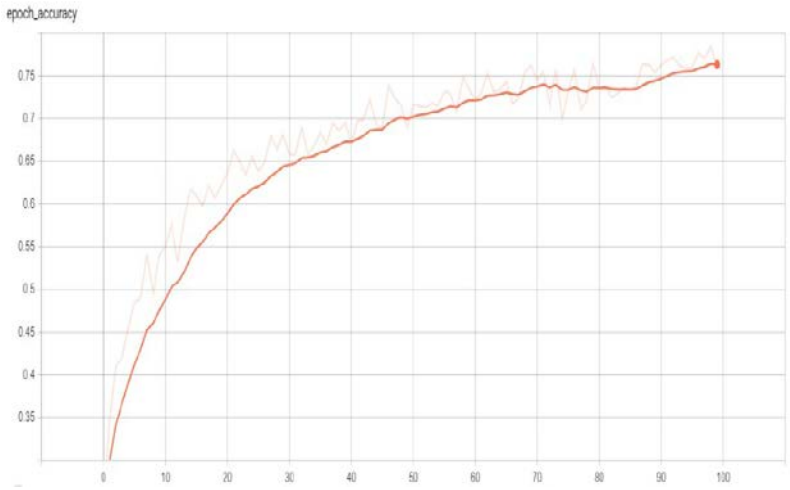


Figure 5. Accuracy of mobile net v2 model 128x128 pixel input

In Figure 6, it can be seen that the accuracy of using mobile net v2 with input 128x128 pixels obtained the highest accuracy of 0.6 from a scale of 0 -1.



Gambar 6 Akurasi model mobile net v2 96x96 pixel input

In Figure 7 it can be seen that the accuracy of using mobile net v2 with input of 96x96 pixels obtained the highest accuracy of 0.76 from a scale of 0 -1.

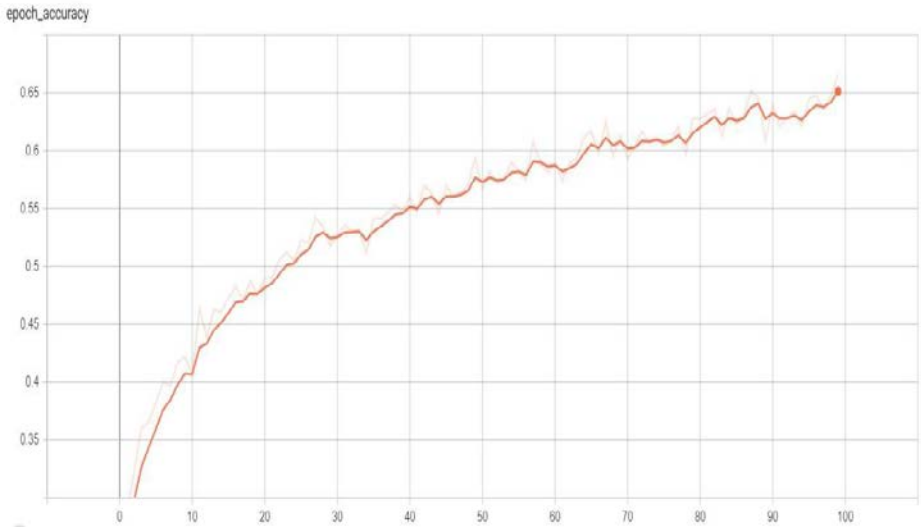


Figure 7. Accuracy of nasnet model 224x224 pixel input

In Figure 8 it can be seen that the accuracy of using NASNET with input of 224x224 pixels obtained the highest accuracy of 0.65 from a scale of 0 -1, the comparison results of the entire model being trained can be seen in Figure 8.

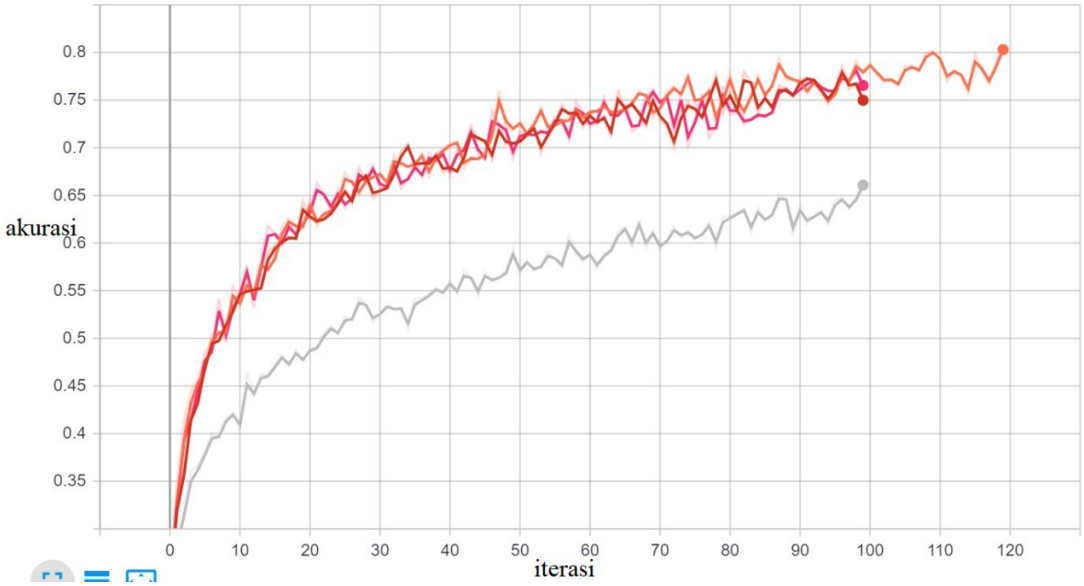


Figure 8 Accuracy of mobilenetv2 and nasnet training models

The validation process uses two different models. The results of the validation are as shown in Figure 9 and Figure 10. In Figure 9 it can be seen that the accuracy using mobile net v2 with input 224x224 pixels obtained a final validation accuracy of 0.43 from a scale of 0 -1

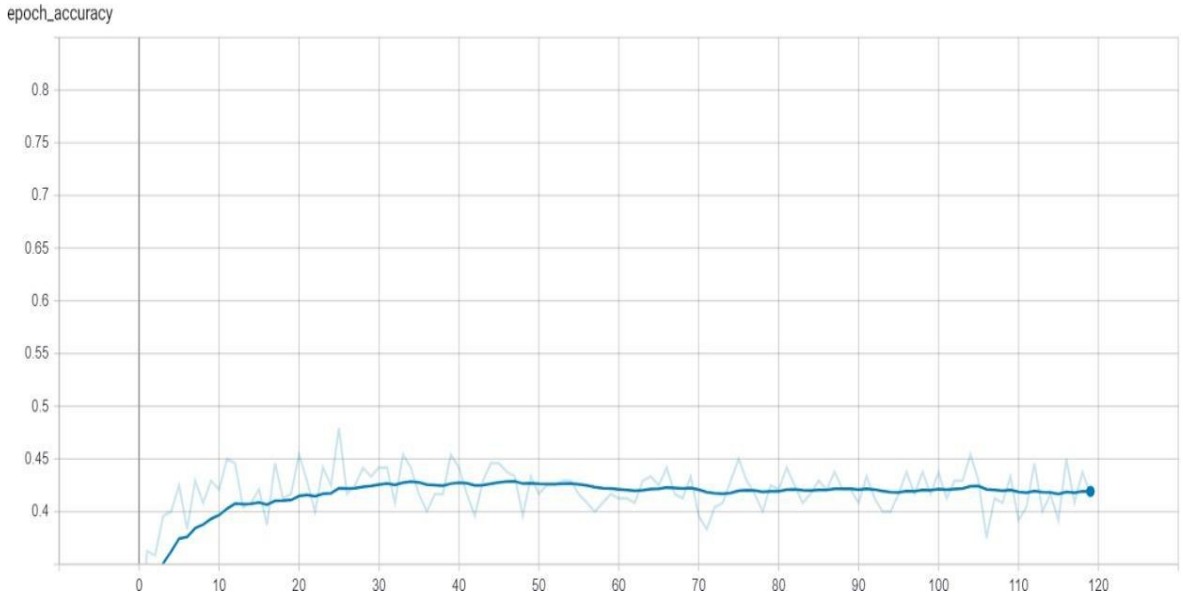


Figure 9. Validation test Mobilenet v2 224x224 pixel input

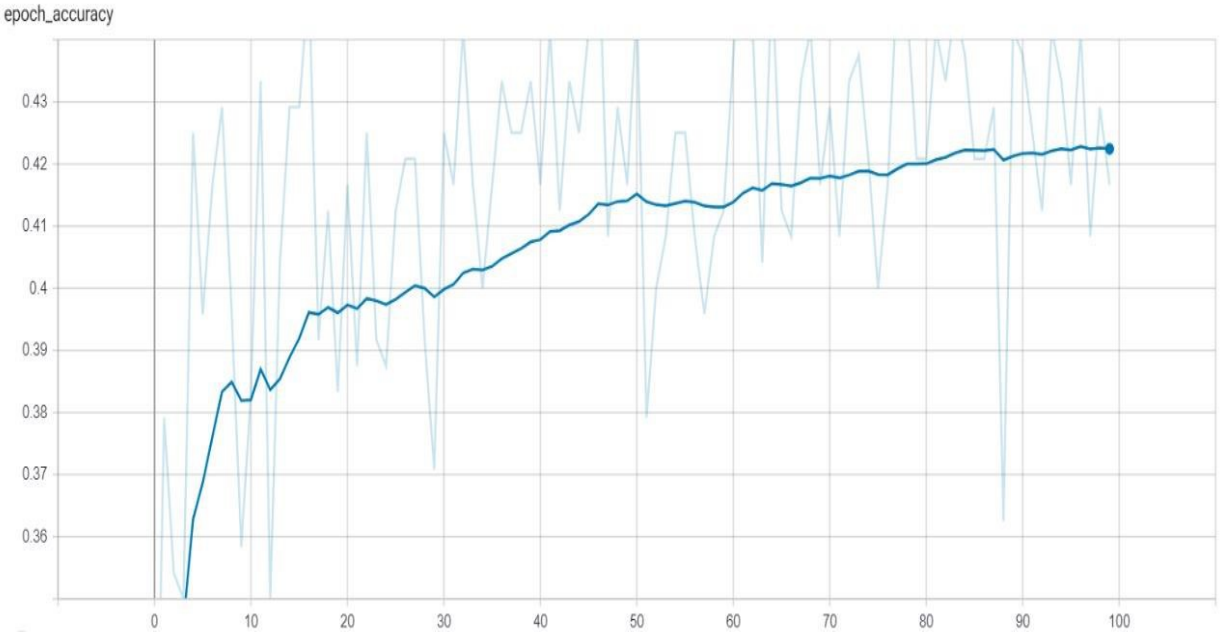


Figure 10. Validation test Mobilenet v2 128x128 pixel input

In this study, the process of testing the execution time of the information system was also carried out by carrying out various variations of the model and the number of itera. The test results are shown in Table 2.

Table 2. The iteration time test results for the training process

No	Type	Iteration	Time/Iteration	total time	Prediction Time
1	mobilenet v2 224x224	100	10,6 menute	17,6 hour	27,6 ms
2	mobeilenet v2 128x128	100	4,35 menute	7,25 hour	11,3 ms
3	mobilenet v2 96x96	100	3.36 menute	5,6 hour	8,7 ms
4	Nasnet 224 x 224	100	17,6 menute	29 hour	45 ms

VI CONCLUSION

The emotion detection information system using the mobileNet v2 machine learning model is able to detect emotions through images taken spontaneously in real time even though the detection device has low resources, predictions are made using a mobileNet model that has been trained using international standard datasets. Facial images are classified based on emotions with training results that show accuracy above 70%. From the results of the tests carried out, validated directly by the expert by taking a sample of 11 images of the detection results, the results were more than 90% valid. Some suggestions that can be given to the development and implementation of an emotion detection information system so that the detection results are more accurate are to develop a training dataset using images of the faces of Indonesian people because the users of the emotion detection information system are native Indonesians, besides that the development of an emotion detection information system can be developed. with video in real time because the mobileNet v2 model can detect in real time.

VII REFERENCES

[1] Yang, W., Zhang, D., dan Fu, Y., 2016, Research of a Diagonal Recurrent Neural Network and Artificial Neural Networks, Dalam *2016 International Symposium on Computer, Consumer and Control (IS3C)* (hlm. 374–377). Xi'an, China: IEEE.

[2] Salama, E. S., El-Khoribi, R. A., Shoman, M. E., dan Wahby Shalaby, M. A., 2020, A 3D-convolutional neural network framework with ensemble learning techniques for multi-modal emotion recognition, *Egyptian Informatics Journal*, S1110866520301389.

[3] Handasah, R. R., 2018, Pengaruh Kematangan Emosi Terhadap Agresivitas Dimediasi Oleh Kontrol Diri Pada Siswa Sma Negeri Di Kota Malang, 18.

[4] Kalchbrenner, N., Grefenstette, E., dan Blunsom, P., 2014, A Convolutional Neural Network for Modelling Sentences, *ArXiv:1404.2188 [Cs]*.

[5] Jain, D. K., Shamsolmoali, P., dan Sehdev, P., 2019, Extended deep neural network for facial emotion recognition, *Pattern Recognition Letters 120*, 69–74.

[6] Li, S., dan Deng, W., 2020, Deep Facial Expression Recognition: A Survey, *IEEE Transactions on Affective Computing*, 1–1.

[7] Menghani, P., Barthwal, S., dan Bansal, S., 2016, An extreme helping hand for handicap people: Using computer vision, Dalam *2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE)* (hlm. 1–5). Jaipur, India: IEEE.

[8] Ravi, R., Yadhukrishna, S. V., dan prithviraj, R., 2020, A Face Expression Recognition Using CNN & LBP, Dalam *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)* (hlm. 684–689). Erode, India: IEEE.

[9] Bantupalli, K., dan Xie, Y., 2018, American Sign Language Recognition using Deep Learning and Computer Vision, Dalam *2018 IEEE International Conference on Big Data (Big Data)* (hlm. 4896–4899). Seattle, WA, USA: IEEE.

[10] Nara, M., Mukesh, B. R., Padala, P., dan Kinnal, B., 2019, Performance Evaluation of Deep Learning frameworks on Computer Vision problems, Dalam *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)* (hlm. 670–674). Tirunelveli, India: IEEE.

[11] Li, H., dan Xu, H., 2020, Deep reinforcement learning for robust emotional classification in facial expression recognition, *Knowledge-Based Systems 204*, 106172.

[12] Simonyan, K., dan Zisserman, A., 2015, Very Deep Convolutional Networks for Large-Scale Image Recognition, *ArXiv:1409.1556 [Cs]*.

[13] Nishani, E., dan Cico, B., 2017, Computer vision approaches based on deep learning and neural networks: Deep neural networks for video analysis of human pose

estimation, Dalam *2017 6th Mediterranean Conference on Embedded Computing (MECO)* (hlm. 1–4). Bar, Montenegro: IEEE.

[14] Parkhi, O. M., Vedaldi, A., dan Zisserman, A., 2015, Deep Face Recognition, Dalam *Procedings of the British Machine Vision Conference 2015* (hlm. 41.1-41.12). Swansea: British Machine Vision Association.

[15] Trigueros, D. S., Meng, L., dan Hartnett, M., 2018, Face Recognition: From Traditional to Deep Learning Methods, *ArXiv:1811.00116 [Cs]*.

[16] Azizah, L. M., Umayah, S. F., Riyadi, S., Damarjati, C., dan Utama, N. A., 2017, Deep learning implementation using convolutional neural network in mangosteen surface defect detection, Dalam *2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)* (hlm.242–246). Penang: IEEE.

[17] Cotter, S. F., 2020, MobiExpressNet: A Deep Learning Network for Face Expression Recognition on Smart Phones, Dalam *2020 IEEE International Conference on Consumer Electronics (ICCE)* (hlm. 1–4). Las Vegas, NV, USA: IEEE.

[18] Goularas, D., dan Kamis, S., 2019, Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data, Dalam *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep- ML)* (hlm. 12–17). Istanbul, Turkey: IEEE.

[19] Banerjee, I., Ling, Y., Chen, M. C., Hasan, S. A., Langlotz, C. P., Moradzadeh, N., Chapman, B., Amrhein, T., Mong, D., Rubin, D. L., Farri, O., dan Lungren, M. P., 2019, Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification, *Artificial Intelligence in Medicine 97*, 79–88.

[20] Tingxuan Zhang., 2020, Face Expression Recognition Based on Deep Learning, Journal of Physics: Conference Series.

[21] Lum, K. Y., Goh, Y. H., dan Lee, Y. B., 2020, American Sign Language Recognition Based on MobileNetV2, *Advances in Science, Technology and Engineering Systems Journal 5* (*6*), 481–488.

[22] Martinez, F., Martínez, F., dan Jacinto, E., 2020, Performance Evaluation of the NASNet Convolutional Network in the Automatic Identification of COVID-19, *International Journal on Advanced Science, Engineering and Information Technology 10* (*2*), 662.

[23] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., dan Chen, L.-C., 2018, MobileNetV2: Inverted Residuals and Linear Bottlenecks, Dalam *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (hlm. 4510–4520). Salt Lake City, UT: IEEE.

[24] Zoph, B., Vasudevan, V., Shlens, J., dan Le, Q. V., 2018, Learning Transferable Architectures for Scalable Image Recognition, *ArXiv:1707.07012 [Cs, Stat]*.

[25] Deng, L., 2014, A tutorial survey of architectures, algorithms, and applications for deep learning, *APSIPA Transactions on Signal and Information Processing 3*, e2.

[26] LeCun, Y., Bengio, Y., dan Hinton, G., 2015, Deep learning, *Nature 521* (*7553*), 436–444.

[27] Gultom, Y., Arymurthy, A. M., dan Masikome, R. J., 2018, Batik Classification using Deep Convolutional Network Transfer Learning, *Jurnal Ilmu Komputer Dan Informasi 11* (*2*), 59.

[28] Pigou, L., Dieleman, S., Kindermans, P.-J., dan Schrauwen, B., 2015, Sign Language Recognition Using Convolutional Neural Networks, Dalam L.

[29] Maggiori, E., Tarabalka, Y., Charpiat, G., dan Alliez, P., 2017, Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification, *IEEE Transactions on Geoscience and Remote Sensing 55* (*2*), 645–657.

[30] Goodfellow, I. J., Erhan, D., Luc Carrier, P., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., Ionescu, R., Popescu, M., Grozea, C., Bergstra, J., Xie, J., Romaszko, L., Xu, B., Chuang, Z., dan Bengio, Y., 2015, Challenges in representation learning: A report on three machine learning contests, *Neural Networks 64*, 59–63.