Journal of Theoretical and Applied Information Technology
15th March 2022. Vol.100. No 5
2022 Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FACIAL EXPRESSION RECOGNITION USING CNN FOR HEALTHCARE

[1]CHRISMORGAN SHINTARO, [2]WILLIAM, [3]FELIX NOVANDO, [4]AZANI CEMPAKA SARI
[1,2,3,4]School of Computer Science, Department of Bina Nusantara University, Jakarta, Indonesia
E-mail: [1] chrismorgan.shintaro@binus.ac.id, [2] william073@binus.ac.id, [3] felix.novando@binus.ac.id, [4] acsari@binus.edu

ABSTRACT

Human emotion is the way how human express anything from themselves. We can extract many informations from a human emotion alone like temper and health condition. Due to in Indonesia, people usually don't visit doctor just to seeing their own face condition. We can use a computer vision task to do so. This paper is aim to train a computer vision model that can recognize a human face emotion. Inspired by Convolutional Neural Networks (CNNs) that capable of doing image classification, we proposed our own designed multiple CNN classification model with different layer settings to do facial expression recognition. Next we applied the model on the FER2013 dataset. The best training accuracy achieved by the second model in 92.40% accuracy and the best validation accuracy achieved by third model in 67.43% accuracy.

Keywords: *Computer vision, CNN, Facial Expression Recognition, Health, Classification*

1. INTRODUCTION

Face reflects our health status. From face, we can sometimes tell someone's condition, from expressions to symptoms of a disease. There have been many studies that state that expression is closely related to our health condition. Starting from the expression of someone who in pain to someone who is energetic. Facial expressions are recognized based on the classification with various universal expressions such as anger, fear, surprise, sadness, disgust, and happiness. Based on this expression, the conditions and circumstances of a person can be extracted.

There are diseases that unconsciously affect expression such as neurological disorders that can occur in both children and adults. So that it can be identified based on certain characteristics which can then be used for assessment and the severity of this disorder. Facial images are captured via cameras and other devices, and then, the features are extracted to perform classification. Usually, facial expression processing is analyzed based on curves on the face such as eyebrows, eyes, lips, and nose. If the system can easily distinguish any variation in expression, these characteristics will eventually serve as disease-specific biological markers to aid in clinical diagnosis and evaluate the patient's therapeutic response [1]. Generally facial expression recognition includes three main stages, namely preprocessing, feature extraction, and classification.

In this topic, we will focus on facial expression recognition using Computer Vision and Deep Learning as well as some classification techniques to determine a person's expression from photographs. We will build a facial expression recognition model using python. For the training that will be carried out, we use the FER 2013 dataset.

2. LITERATURE REVIEW

Research on technology that focuses on expression recognition has been carried out [2]. Facial recognition technology is also present to be a popular topic at this time [3]. The use of deep learning methods in facial recognition technology also plays an important role in development. Convolutional Neural Network (CNN) is the method most often used in facial expression recognition [3].

Computer vision is a branch of Artificial Intelligence that trains computers to understand the objects they see. Computer vision itself has been widely used in many fields. Our goal is to implement computer vision to scan photos of a person's face

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

whether the person is sick or is suffering from a certain disease [4]. We looked for several sources to compare the algorithms and methods used in several experiments that have been carried out. A picture contains many meanings, as well as a face image that illustrates many details such as gender specifications, age, and the emotional state of the mind. Facial expressions play an important role in interactions in the community and are usually used in analyzing the nature of emotions [5]. Facial expression recognition can be implemented in everyday human activities and also for animals such as sheep to carry out disease evaluation through their facial expressions [6]. As for research that uses direct thermal cameras, such as that conducted by Peter Wei [7], in research on livestock. Facial expression recognition technology is widely implemented in the medical world to obtain health diagnoses quickly and reduce privacy exposure[8]. Applications in the medical world include the use of facial expression recognition for case studies of children with ASD which are carried out using CNN [9], prediction of depressive conditions using the DRR Depression Net network [10], analysis of pain using hybrids deep networks (CNN + RNN) [11], monitor neuropsychiatric disorders using CNN with a dataset from the Radboud Face Database (RaFD) [12], with SVM to study deficits in emotional expression and social cognition in neuropsychiatric disorders [13], and use facial expressions to construct interaction systems. Computer with humans with a case of users who have neuromuscular limitations [14]. Ghulam, et al [15] applied face expression recognition to improve health services in cities to monitor the patient's condition. Fei Wang, et al [16] also applied a facial expression recognition system through robots using CNN related to the 2013 FER dataset which aims to improve its performance in real life and use the digital healthcare (DHc) framework so that it can achieve substantial improvements for monitoring security and health care of parents. In addition, the application of facial expression recognition systems can also be applied for detecting driver’s fatigue [17].

Facial expressions are usually classified into several classes such as anger, fear, surprise, disgust, sadness, and happiness. Based on this expression, the status of a patient can be recognized [18]. Torki Altameem, Ayman Altameem introduces multi-modal visualization analysis to improve a slightly complicated process related to human-machine interaction [18]. they also used 3 CNN layers for classification, correlation, and detection of facial expressions.

Research conducted to improve the performance of algorithms and methods related to facial expression recognition such as using the Back Propagation Neural Network (BPNN) was conducted by Trieu H. Trinh [19] in image embedding training[20], Abir Fathallah, et al [21] also introduced the model. Visual Geometry Group is in order to fine tuning the CNN architecture which is made in order to improve its performance against various kinds of datasets. Jonitta Meryl C, et al [22] also developed the effectiveness of combining CNN and Radial Basis functions to have a better level of accuracy against the 2013 FER dataset.The region based method algorithm can be used to examine pixels in frames in epilepsy research to examine facial behavior of people with epilepsy. [23]. Application of region based algorithms is also discussed in the journal Victor Wiley and Thomas Lucas [24] on image processing. Tingting Zhao, et al [25] used the Facial Landmark model to perform processing on facial images. SVM is also used to detect facial spoofing [26]. Apart from its use as facial expression recognition, deep learning can also play a role in analyzing the severity of eye disease [27].

3. DATASET

Recognition 2013 (FER2013) dataset. The FER2013 has 28709 images in the training set and 7178 images in the testing set and each of the images has been assigned an emotion label of seven emotions: happy, sad, angry, fear, surprise, disgust, and neutral. Where images labeled happy emotions have the most data as you can see in Table I The FER2013 dataset was created by the google image search API with emotion-related keywords. The images in the dataset consist of posed and unposed headshots, which are rendered in grayscale mode at 48x48 pixels, as you can see the examples in the Figure 1.

	Surprise	Fear	Angry	Neutral	Sad	Disgust	Happy
Label	6	2	0	4	5	1	3
Train	3171	4097	3995	4965	4830	436	7215
Test	831	1024	958	1233	1247	111	1774

Table I: Table of Dataset FER2013

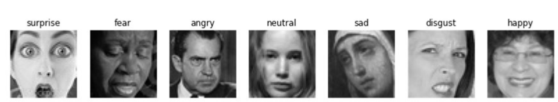


Figure 1: Emotion Samples

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. PROPOSED MODEL

In the current study, seven states of facial emotion are recognized by using a convolutional neural network. In general, Convolutional Neural Network (CNN) is one type of neural network that is usually used for data in the form of images. Convolutional Neural Network (CNN) itself has similarities to neural networks such as Multi Layer Perceptron (MLP) which consists of neurons that have weight, bias and activation functions and also similar to Neural Networks in general, CNN has several hidden layers of an input is a single vector, but in CNN each neuron is represented in two dimensions, unlike MLP where each neuron is only one dimension. So CNN can be called a further development of MLP because it uses a similar method with more dimensions.

According to CNN, The convolutional layer also consists of neurons arranged in such a way that they form a filter with length and height (pixels). CNN itself utilizes the convolution process by moving a filter of a certain size onto an image, so that the computer will get new representative information from the result of multiplying that part of the image with the filter used. If it is analogous to the features of the human face, the first layer is a reflection of strokes in different directions, in the second layer features such as the shape of the eyes, nose, and mouth begin to appear, this is because pooling is done from the first layer which is still in the form of scratches. In the third layer, a combination of the features of the eyes, nose, and mouth will be formed which will later be concluded with the face of a certain person.

CNN itself has the ability to reduce tuning time. CNNs are very effective in reducing the number of parameters without losing on the quality of models. Also CNNs are trained to identify the edges of objects in any image so that CNNs are really effective for image classification as the concept of dimensionality reduction suits the huge number of parameters in an image. Inpired by the CNN architecture from [28] [29], we developed 3 different models containing 3-6 layers of convolution, each with the names CNN 1, CNN 2, and CNN 3 in order.

A. *Model Architecture*

The 3 CNN architectures are outlined in Table II. The only difference among the 3 model archictecture is the number of convolutional layers, filter and kernel regulizer. The convolutional layer works like image filters, and is aimed at learning different features. The more convolutional layers, the more detailed features this CNN could learn.

a) Input Layer: We use input layer with input shape 48 x 48 x 1 to receive input from raw images that have been preprocessed.

b) Convolution Layer: Convolutional layer is a layer that do a convolving process of a filter with determined size to input image. The layer do dot product to each pixel of the image. The result of the layer is an activation map. We use 2D convolution layers throughout all 3 CNN model with 3×3 kernel size. Padding value is same, and the convolution stride is fixed to 1, so the resolution of the image could be maintained after each convolution.

c) Non-linearity Layer: Non-linearity layer is followed after every convolution layer. We use Rectified Linear Units (ReLUs, $f(x) = \max(0, x)$) activation function. The advantage of using ReLUs is it can alleviate the gradient vanish problem. Nevertheless, when $x \leq 0$, the ReLU function doesn't active, it filters out the negative responses and obliterate some gradient information.

d) Regularizers Layer: Regularizers allow to apply penalties on layer parameters or layer activity during optimization. These penalties are summed into the loss function that the network optimizes.

e) Batch Normalization Layer: Batch normalization applies a transformation that maintains the mean output close to 0 and the output standard deviation close to 1.

f) Pooling Layer: Pooling Layer is a layer that used to reduce input spatially by using down-sampling operation. Our pooling layer using max pooling strategy with a 2×2 window. The pooling stride is none.

g) Drop Out Layer: Dropout layer randomly sets input units to 0 with a frequency of rate at each step during training time, which helps prevent overfitting. We set first drop out layer with 0.75 keep probability or we drop 25% of the nodes and the second drop out layer we drop 50% nodes.

h) Flatten Layer: Flatten layer reshape the multiple feature dimensions to one dimension.

i) Fully Connected / Dense Layer: Fully connected layer can also be treated as convolution layer with a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1×1 kernel size. We stack three fully connected layers at the end of the network.

j) Output Layer: We use Softmax classification which is commonly used in CNN architecture.

$$Li = \sum_{j \neq y_i} \max(0, f(x_i, W)_j - f(x_i, W)_{y_i} + \Delta)$$

B. Overall architecture

The model consists of three different model that will be run separately can see in Figure 2. All models are started from input layer who takes image inputs. Then, process it through convolution layers with kernel size 3 x 3, padding same, and relu activation. After that, normalization layers, max pooling layers with pool size 2 x 2, and dropout layers. After the features are extracted, we pass it through fully connected layers with relu activation function. For the last fully connected layers, we choose 7 neurons because we will have 7 output classes (Surprise, Fear, Angry, Neutral, Sad, Disgust, Happy) with softmax activation layers. For the second and third model, we use kernel_regularizer that is Ridge Regression in the middle convolution layers.

C. Training detail

a. *Data preprocessing and augmentation*: In this study, we first resize all images to 48x48 pixels. Due to the low resolution given so that the training process into the model becomes faster. For preprocessing data, we use scaling by dividing each pixel by 255 (maximum light in RGB color pixel) and for data augmentation, we randomly zoom the original images, and flip the images horizontally. Data augmentation itself is a technique used to prevent overfitting and can also improve model performance and performance.

Table II: 3 Model CNN

<table>
<tr><th>CNN 1</th><th>CNN 2</th><th>CNN 3</th></tr>
<tr><td>Input</td><td>Input</td><td>Input</td></tr>
<tr><td rowspan="2">Conv2D-16</td><td>Conv2D-32</td><td>Conv2D-32</td></tr>
<tr><td>Conv2D-64</td><td>Conv2D-64</td></tr>
<tr><td>Batch Norm</td><td>Batch Norm</td><td>Batch Norm</td></tr>
<tr><td rowspan="2">MaxPool</td><td>MaxPool</td><td>MaxPool</td></tr>
<tr><td>Drop Out 0.25</td><td>Drop Out 0.25</td></tr>
<tr><td rowspan="2">Conv2D-32</td><td>Conv2D-128</td><td>Conv2D-128</td></tr>
<tr><td>Conv2D-256</td><td>Conv2D-256</td></tr>
<tr><td>Batch Norm</td><td>Batch Norm</td><td>Batch Norm</td></tr>
<tr><td rowspan="2">MaxPool</td><td>MaxPool</td><td>MaxPool</td></tr>
<tr><td rowspan="6">Drop Out 0.25</td><td>Drop Out 0.25</td></tr>
<tr><td rowspan="2">Conv2D-48</td><td>Conv2D-512</td></tr>
<tr><td>Conv2D-1024</td></tr>
<tr><td>Batch Norm</td><td>Batch Norm</td></tr>
<tr><td>MaxPool</td><td>MaxPool</td></tr>
<tr><td>Drop Out 0.25</td><td>Drop Out 0.25</td></tr>
<tr><td>FC 1024</td><td>FC 1024</td><td>FC 2048</td></tr>
<tr><td>Drop Out 0.5</td><td>Drop Out 0.5</td><td>Drop Out 0.5</td></tr>
<tr><td>FC 7</td><td>FC 7</td><td>FC 7</td></tr>
<tr><td>SoftmaxLoss</td><td>SoftmaxLoss</td><td>SoftmaxLoss</td></tr>
</table>

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

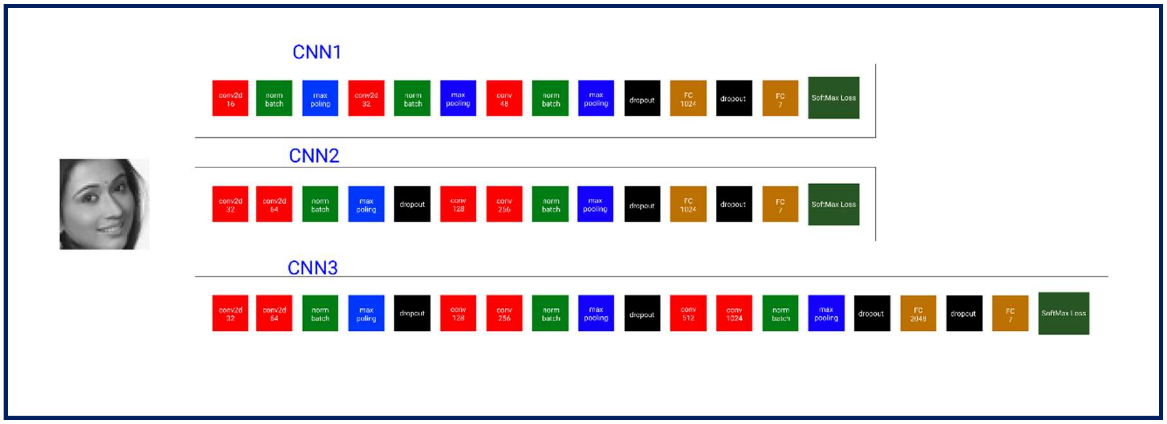


Figure 2: Architecture of 3 Model CNN:The black rectangles represent of each CNN architecture that we proposed. Red boxes represents convolutional layers, green box represents normalization layer, blue represents max pooling layers, black represents dropout layer, dark yellow represents fully connected layers, and the dark green represents softmax layers

b. *Training*: in the training process we used the Adam Optimizer with a learning rate of 0.0001 and also a decay of 1e-6 to help reduce losses during the latest training step, when the computed loss with the previously associated lambda parameter has stopped to decrease. The given batch size is fixed to 64 and the initial weight of the model is obtained through a random normal distribution. The number of epochs varies between 60 to 120.

5. EXPERIMENTAL RESULT

We use the FER2013 dataset which has been divided into a training set and a testing set where there are 28709 and 7178 data respectively, as previously mentioned, has an emotional label for each data. Before proceeding to the training process, the training set will be preprocessed in the form of scaling and data augmentation in the form of random zoom and horizontal flip. After that, each training and testing set is given a batch size of 64. Also, the training set is shuffled so that the training process will be evenly distributed. and then the training process for the three proposed models will be carried out. To evaluate the model created, we visualize it into several metrics. We use precision, recall, and F1 score to evaluate our model. Precision is used to see how precise/accurate our model is out of the predicted result, how many of them that has same actual and predicted output. Recall used for calculate how many of the actual results our model capture through labeling it as same results. F1 score used to check balance between precision and recall and if we want to seek there is an uneven class distribution.

Model	Training Accuracy	Validation Accuracy
CNN 1	78,71 %	62,73 %
CNN 2	**92,40 %**	65,42 %
CNN 3	86,29 %	**67,43 %**

Table III: Model Accuracy

We train the first model with 120 epochs, second model with 60 epochs, and the third model with 60 epochs. And the result is displayed in Table III. Our models overall accuracies with FER2013 datasets are around 65% which can be consider a an average human accuracy with 65 ± 5 %. Our third model get the highest validation accuracy. We can expect this because of the third model have more preprocessing layers (conv, norm, pooling, dropout), and the fully connected layers contain much more neuron than the first and second.

Figure 3 shows a comparison of training loss and validation loss along with training accuracy and validation accuracy where as can be seen in the curve image, the loss of the three models does not experience overfitting and the accuracy of the three models also has a pretty good performance, especially in the model CNN 3 achieved validation accuracy of 67.43%. On the other hand, the accuracy of the CNN 1 model has an optimal curve between the training set and the validation set even though the accuracy obtained tends to be smaller than the other two models.

From the Table IV, we got the highest performance for Happy emotion class. The second and third model classes shows that the happy class got the highest in Precision, Recall, and F1. The first model shows us the Happy class emotion is the highest Recall and f1 metrics. While the highest precision for the first model is disgust model, because disgust class has the least test data. Overall, the reason why happy has the almost all best performance in 3 classed is due to it's amount of data more than the others.

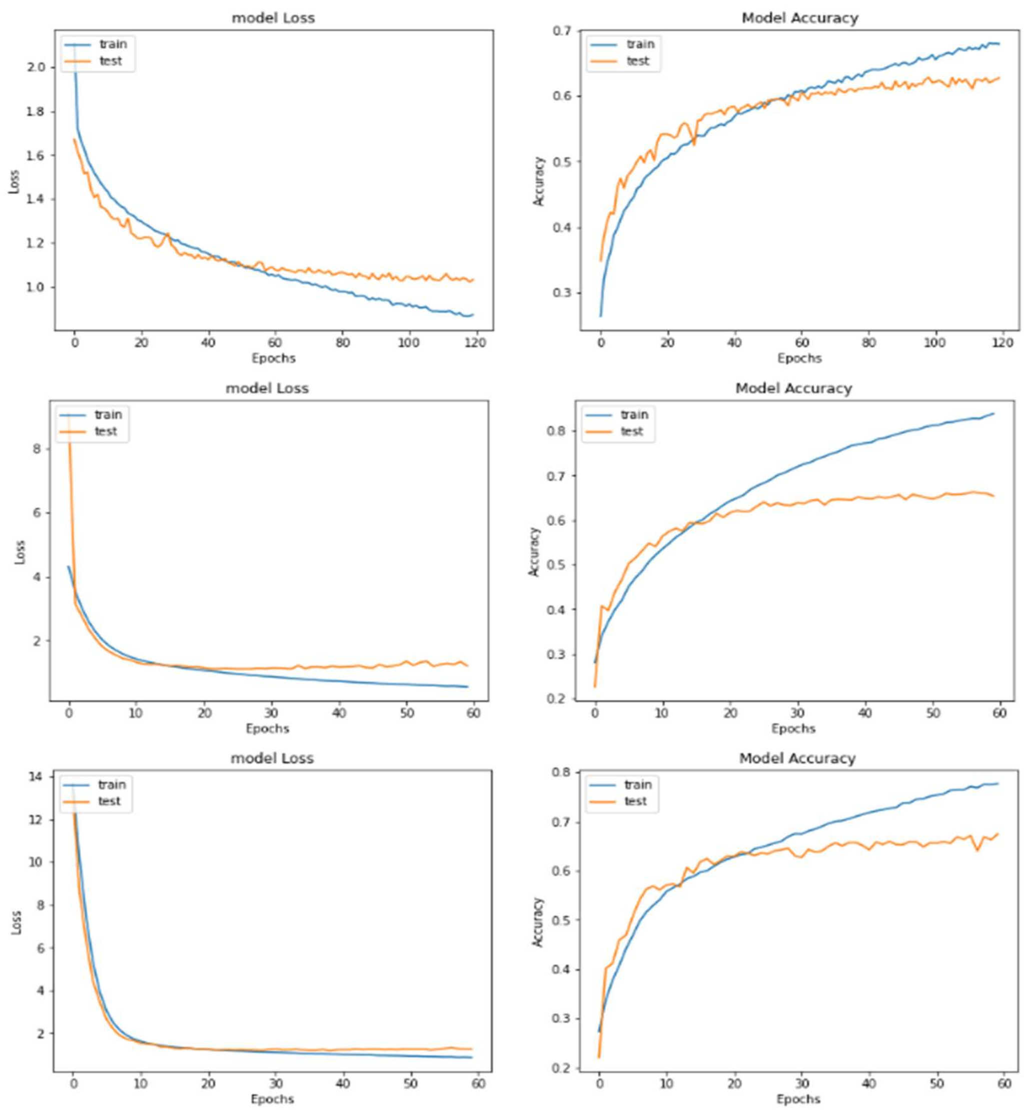


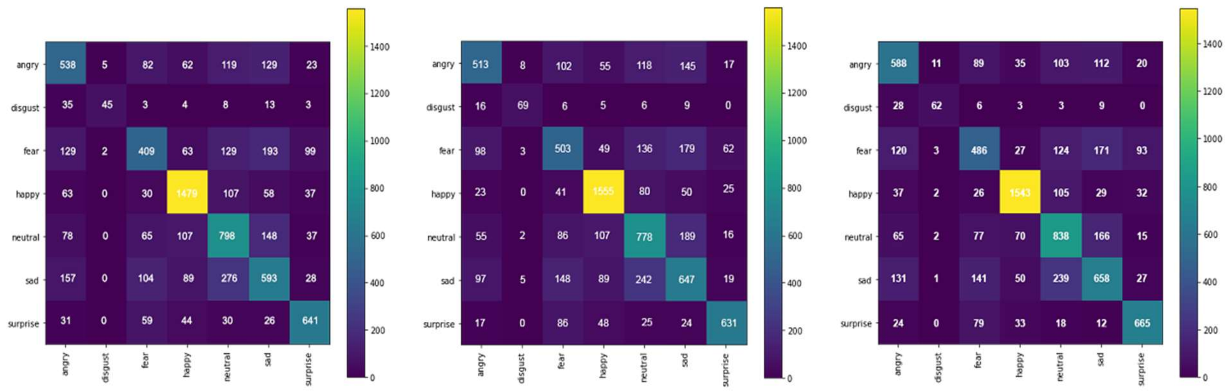
Figure 3: Loss and Accuracy curve on each model: CNN 1 (top), CNN 2 (middle), and CNN 3 (bottom)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table IV: Classification Report on each model (Precision, Recall, F1-Score)

Emotion	Model 1			Model 2			Model 3		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Angry	0,52	0,56	0,54	0,63	0,54	0,58	0,59	0,61	0,60
Disgust	**0,87**	0,41	0,55	0,79	0,62	0,70	0,77	0,56	0,65
Fear	0,54	0,40	0,46	0,52	0,49	0,50	0,54	0,47	0,50
Happy	0,80	**0,83**	**0,82**	**0,81**	**0,88**	**0,84**	**0,88**	**0,87**	**0,97**
Neutral	0,54	0,65	0,59	0,56	0,63	0,59	0,59	0,68	0,63
Sad	0,51	0,48	0,49	0,52	0,52	0,52	0,57	0,53	0,55

Figure 4: Confusion Matrix: CNN 1 (left), CNN 2 (middle), and CNN 3 (right)



In addition, a confusion matrix is calculated to show the behavior of different emotional classes and also the classification performance of each class. While the task that we want to complete is a multi class classification, there are 7 labels to predict. To visualize the confusion matrix in the three models, see Figure 4. Also from Figure 3, training accuracy is the accuracy that is given from training proceess. This accuracy is obtained from the percentage of how many the prediction and the actual output match. This accuracy is not final because the training process can use the testing data repeately. And validation accuracy is the accuracy that given by matching the predictive output from the model with the result that neither in training nor testing data so can be referred as final result.

As mentioned earlier, we were inspired by the architectural model presented by [29] which also performs image recognition of facial expressions from the FER2013 dataset. The model presented in the study [29] consists of two stages: (1) The first stage takes a face image as input, and feeds it to the 3 CNN subnets. The 3 subnets contain 8 to 10 layers respectively which is compactly designed, and easy to train. They are the core components of the proposed architecture. (2) The second stage is responsible for predicting the expressions based on the previous stage output. The features extracted by these subnets are concatenated together by adding a fully connected layer at the end. Finally a softmax layer is used as the output layer of the whole network. Through this model, we also develop it to a higher level by using several additional layers such as dropout layers, regularizers layers, and batch normalization layers. From these additional layers, the accuracy that we get from the three models can get ± 62% where the last model, the CNN 3 model, gets the highest accuracy of 67.43% with 2.4% higher than the ensemble model presented in [29] by 65.03%. So it can be ascertained that the addition of the above layers will also improve the performance of the model that we built in solving the problem of the face expression recognition task.

6. CONCLUSION AND FUTURE WORK

In this paper, we solve Facial Expression Recognition problem with 3 CNN model with our own designed architecture. Then we train and evaluate the model's performance on the FER 2013 dataset. The experimental result of the CNN model

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

got 92.40% for the best train accuracy from the second model and 67.43% for the best validation accuracy from the third model.

From our project, there's some limitation such we haven't use pre trained model such as VGG19, ResNet40V2, or the others. We might get the more accuracy if we successfully fined tuned the pretrained models. Our dataset (FER2013) consists of 48 x 48 pixels grayscale images. Nowadays, most of peoples use the images with much higher resolutions and have 3 input channels (RGB images). Kuang Liu and friends [29] use the concept Ensemble model for their 3 models, yet we still not implement it.

In the future works, we would like to use the pretrained model to see if we can get more accuracy. We also want to make a model that can keep up with images with higher resolutions and has 3 input channels. We also want to Implement Ensemble model for our three models to overcome high variance, low accuracy, and feature noise and bias. So, we will get the better accuracy and the model can tolerate many features variance and noises. After that, we want to implement our model in a device to make a real time face scanner.

REFERENCES

[1] G. Yolcu, I. Oztel, S. Kazan, C. Oz, K. Palaniappan, T. E. Lever and F. Bunyak, "Facial expression recognition for monitoring neurological disorders based on convolutional neural network," Multimedia Tools and Applications, vol. 78, no. 22, pp. 31581-31603, 2019, doi: 10.1007/s11042-019-07959-6.

[2] J. Thevenot, M. B. López and A. Hadid, "A Survey on Computer Vision for Assistive Medical Diagnosis From Faces," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 5, pp. 1497-1511, Sept. 2018, doi: 10.1109/JBHI.2017.2754861.

[3] B. Jin, L. Cruz and N. Gonçalves, "Deep Facial Diagnosis: Deep Transfer Learning From Face Recognition to Facial Diagnosis," in IEEE Access, vol. 8, pp. 123649-123661, 2020, doi: 10.1109/ACCESS.2020.3005687.

[4] G. Sarolidou, J. Axelsson, T. Sundelin, J. Lasselin, C. Regenbogen, K. Sorjonen, J. N. Lundström, J. N. Lundström and M. J. Olsson, Emotional expressions of the sick face. Brain, behavior, and immunity, pp. 286-291, 2019.

[5] K. Mohan, A. Seal, O. Krejcar and A. Yazidi, "Facial Expression Recognition Using Local Gravitational Force Descriptor-Based Deep Convolution Neural Networks," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-12, 2021, Art no. 5003512, doi: 10.1109/TIM.2020.3031835.

[6] F. Pessanha, K. McLennan and M. Mahmoud, "Towards automatic monitoring of disease progression in sheep: A hierarchical model for sheep facial expressions analysis from video," 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), 2020, pp. 387-393, doi: 10.1109/FG47880.2020.00107.

[7] M. Jorquera-Chavez, S. Fuentes, F. R. Dunshea, R. D. Warner, T. Poblete1 and E. C. Jongman, "Modelling and validation of computer vision techniques to assess heart rate, eye temperature, ear-base temperature and respiration rate in cattle," Animals, pp. 9 - 21, 2019.

[8] Y. Wang et al., "Automatic Depression Detection via Facial Expressions Using Multiple Instance Learning," 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020, pp. 1933-1936, doi: 10.1109/ISBI45749.2020.9098396.

[9] M. Leo et al., "Towards the automatic assessment of abilities to produce facial expressions: The case study of children with ASD," 20th Italian National Conference on Photonic Technologies (Fotonica 2018), 2018, pp. 1-4, doi: 10.1049/cp.2018.1675.

[10] X. Li, W. Guo and H. Yang, "Depression severity prediction from facial expression based on the DRR_DepressionNet network," 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020, pp. 2757-2764, doi: 10.1109/BIBM49941.2020.9313597.

[11] G. Bargshady, J. Soar, X. Zhou, R. C. Deo, F. Whittaker and H. Wang, "A Joint Deep Neural Network Model for Pain Recognition from Face," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), 2019, pp. 52-56, doi: 10.1109/CCOMS.2019.8821779.

[12] G. Yolcu et al., "Deep learning-based facial expression recognition for monitoring neurological disorders," 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2017, pp. 1652-1657, doi: 10.1109/BIBM.2017.8217907.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[13] Peng Wang F. B., Automated video-based facial expression analysis of neuropsychiatric disorders, 2008. DOI: https://doi.org/10.1016/j.jneumeth.2007.09.030
[14] Andreia Matos, Vítor Filipe, and Pedro Couto. 2016. Human-Computer Interaction Based on Facial Expression Recognition: A Case Study in Degenerative Neuromuscular Disease. In Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2016). Association for Computing Machinery, New York, NY, USA, 8–12. DOI:https://doi.org/10.1145/3019943.3019945
[15] G. Muhammad, M. Alsulaiman, S. U. Amin, A. Ghoneim and M. F. Alhamid, "A Facial-Expression Monitoring System for Improved Healthcare in Smart Cities," in IEEE Access, vol. 5, pp. 10871-10881, 2017, doi: 10.1109/ACCESS.2017.2712788.
[16] F. Wang, H. Chen, L. Kong and W. Sheng, "Real-time Facial Expression Recognition on Robot for Healthcare," 2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR), 2018, pp. 402-406, doi: 10.1109/IISR.2018.8535710.
[17] S. A. Khan, S. Hussain, S. Xiaoming and S. Yang, "An Effective Framework for Driver Fatigue Recognition Based on Intelligent Facial Expressions Analysis," in IEEE Access, vol. 6, pp. 67459-67468, 2018, doi: 10.1109/ACCESS.2018.2878601.
[18] T. Altameem and A. Altameem, "Facial expression recognition using human machine interaction and multi-modal visualization analysis for healthcare applications," Image and Vision Computing, vol. 103, p. 104044, 2020, doi: 10.1016/j.imavis.2020.104044.
[19] P. Wei and C. Wang, "Low-cost multi-person continuous skin temperature sensing system for fever detection," In Proceedings of the 18th Conference on Embedded Networked Sensor Systems, pp. 705-706, 2020.
[20] T. H. Trinh, M.-T. Luong and Q. V. Le, "Self-supervised pretraining for image embedding," pp. 29-40, 2019.
[21] A. Fathallah, L. Abdi and A. Douik, "Facial Expression Recognition via Deep Learning," 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), 2017, pp. 745-750, doi: 10.1109/AICCSA.2017.124.
[22] C. Jonitta Meryl, K. Dharshini, D. Sujitha Juliet, J. Akila Rosy and S. S. Jacob, "Deep Learning based Facial Expression Recognition for Psychological Health Analysis," 2020 International Conference on Communication and Signal Processing (ICCSP), 2020, pp. 1155-1158, doi: 10.1109/ICCSP48568.2020.9182094.
[23] D. Ahmedt-Aristizabal, C. Fookes, K. Nguyen, S. Denman, S. Sridharan and S. Dionisio, "Deep facial analysis: A new phase I epilepsy evaluation using computer vision," pp. 17-24, 2018.
[24] V. Wiley and T. Lucas, "Computer vision and image processing: a paper review," International Journal of Artificial Intelligence Research, pp. 29-36, 2018.
[25] T. Zhao, H. Zhang and J. Spoelstra, "A Computer Vision Application for Assessing Facial Acne Severity from Selfie Images," 2019.
[26] A. Alotaibi and A., "Enhancing computer vision to detect face spoofing attack utilizing a single frame from a replay video attack using deep learning," In 2016 International Conference on Optoelectronics and Image Processing (ICOIP), pp. 1-5, 2016.
[27] D.S. W. Ting, L. R. Pasquale, L. Peng, J. PeterCampbell, A. Y. Lee, R. Raman, G. S. W. Tan, L. Schmetterer, P. A. Keane and T. Y. Wong, "Artificial intelligence and deep learning in ophthalmology," British Journal of Ophthalmology, pp. 167-175, 2019.
[28] M. Mohammadpour, H. Khaliliardali, S. M. R. Hashemi and M. M. AlyanNezhadi, "Facial emotion recognition using deep convolutional networks," *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, 2017, pp. 0017-0021, doi: 10.1109/KBEI.2017.8324974.\
[29] K. Liu, M. Zhang and Z. Pan, "Facial Expression Recognition with CNN Ensemble," *2016 International Conference on Cyberworlds (CW)*, 2016, pp. 163-166, doi: 10.1109/CW.2016.34.