Utilization of Digital Image and Convolution Neural Network Algorithm in Customer Satisfaction Survey with Facial Expressions

Tri Andre Anu[1*], Rika Rosnelly[1], Dedi Irawan[1], Progresif Bulolo[1], Ubaidullah Hasibuan[1]
[1]Universitas Potensi Utama, Indonesia

ABSTRACT

The human face provides us with a lot of information about a person, and arguably the two most important pieces of information in a face are a person's identity and their emotional state. Judgments of identity and emotion facilitate social interactions. Services are a crucial part of the activities of all organizations, especially those in the service sector. Good services support customer satisfaction and ultimately impact the progress of the organization. The Convolutional Neural Network algorithm has become the most widely used neural architecture in various tasks, including image classification, audio pattern recognition, machine translation of text, and speech recognition. The data groups (angry, fearful, happy, neutral, sad, and surprised) tested with a threshold value of 30 epochs achieved a loss (error) accuracy of 1.5146 on the test data. The accuracy on the test data is 0.61. The proposed Convolutional Neural Network algorithm and digital image utilization achieved high accuracy performance to assist in evaluating a service-related field.

Keywords: Digital Image; Convolutional Neural Network; Customer Satisfaction; Facial Expression.

Corresponding Author:
Tri Andre Anu,
Department of Computer Science
Universitas Potensi Utama
L. KL. Yos Sudarso Km. 6,5 No. 3-A, Tanjung Mulia, 20241, Indonesia.
Email : triandreanu07@gmail.com

Article history:
Received Jun 9, 2023
Revised Jul 7, 2023
Accepted Aug 30, 2023

1. INTRODUCTION

The human face provides us with a lot of information about someone, and it can be said that the two most important pieces of information in a face are identity and emotional state. Assessing identity and emotions facilitates social interactions (Noyes et al., 2021). Facial expression recognition has become important with the advancement of technology in computers, mobile phones, robots, and so on. This advancement has made human-technology interactions increasingly unavoidable. By using facial expressions, systems can be developed to recognize customer satisfaction, among other things (Alamsyah et al., 2020). Customer service is a crucial part of all organizations, especially those in the service sector, and good service supports customer satisfaction, ultimately leading to the progress of the respective organization. Consumer satisfaction as an implication of service maximization has been extensively researched both in service and non-service companies (Badar et al., 2021). Segmentation refers to partitioning an image into several parts based on similarity of characteristics or uniformity. Its usefulness is particularly significant in image analysis and digital image processing applications (Anwar et al., 2021). Digital Image Processing is a discipline that studies techniques for processing images, which can be images or videos (Anwar et al., 2021) Segmentation refers to partitioning an image into several parts based on similarity of characteristics or uniformity. Its usefulness is particularly significant in image analysis and digital image processing applications (Anwar et al., 2021). Digital Image Processing is a discipline that studies techniques for processing images, which can be images or videos (Fakultas et al., 2020) Classification is one of the techniques in data mining that involves the concept of neural networks, where the artificial neural network, often referred to as Convolutional Neural Network (CNN), undergoes validation processes with training and test data. The process of neural networks involves classification with guidance (Sari Hutagalung et al., 2023). Convolutional Neural Network (CNN) has become the most widely used neural architecture in various tasks, including image classification, audio pattern recognition, machine translation of texts, and speech recognition (Zhu et al., n.d.) Based on human knowledge in distinguishing facial expressions, this research will explore the utilization of digital image of human facial expressions with Convolutional Neural Network Algorithm in a survey of customer satisfaction towards a service.

2. RESEARCH METHOD

A. Convolution Neural Network (CNN)

CNN consists of an input layer, several hidden convolutional-pooling layers, and an output layer. Essentially, convolution is a mathematical operation on two functions to produce a third function that expresses a modified version of those functions CNN restructures a regular neural network but possesses intriguing characteristics comprising neurons with learnable weights and biases. Each neuron receives multiple inputs and then performs a dot product, optionally followed by non-linearity (Bhatt et al., 2021). CNN is more adept at reducing the number of parameters compared to classical Artificial Neural Networks (ANN). This enables researchers and developers to tackle tasks that were previously infeasible with traditional ANN, granting them access to larger models. CNN assumes that features have no spatial dependence in problem-solving (Kim et al., 2021)

B. Dataset

In this research, a dataset was collected from https://kaggle.com. The facial expressions used were (Angry, Fearful, Happy, Neutral, Sad, and Surprised), and the facial samples used had a size of 48x48 pixels. The dataset consists of grayscale digital images of facial expressions..

Figure 1. Image traning data

C. CNN Architecture

CNN is a development of Multilayer Perceptron (MLP) designed to process two-dimensional data. It belongs to the category of Deep Neural Networks due to its high network depth and has been widely applied to image data (Fu'adah et al., 2020). CNN has an architecture similar to neural networks in general. Neurons in CNN have weights, biases, and activation functions. The CNN architecture, as shown in Figure 2, consists of convolutional layers with ReLU activation, pooling layers as feature extraction layers, and fully connected layers with softmax activation as the classification layer, as illustrated in Figure 2.

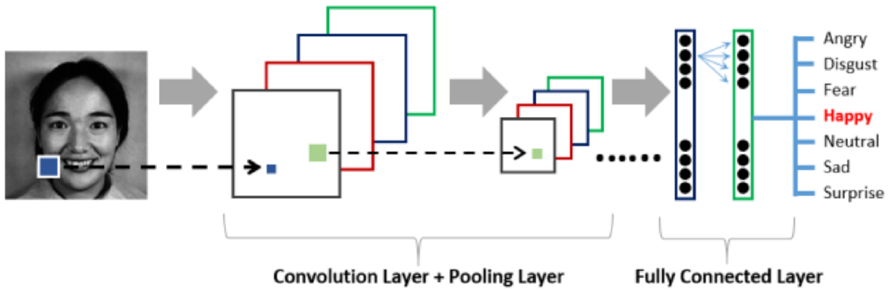


Figure 2. CNN Architecture (Rere et al., 2019)

- Convolution Layer: The Convolution Layer is a crucial part of CNN because most of the computations in CNN are performed in this layer (Alamsyah et al., 2020). The operation performed is similar to the standard convolution operation used in image processing, where there is a kernel and sub-images. The kernel used in CNNs is generally sized 64x64. Then, for each sub-image that is the same size as the kernel, a convolution operation is applied.
- Pooling Layer: The Pooling Layer is used to reduce the size of the image (downsampling) to facilitate the convolution operation in the next layer. This process can be performed using various techniques, such as max pooling (taking the maximum value from the sub-image) and average pooling (taking the average value from the sub-image) (Ahlawat et al., 2020). The pooling layer also has another impact, which is triggering translational invariance (Khan et al., 2021).
- The Fully Connected Layer is the layer at the end of the architecture used in multilayer perceptrons. This layer connects all neurons from the previous activation layer. At this stage, all neurons in the input layer need to be transformed into one-dimensional data (flatten process) (Fu'adah et al., 2020). Fully connected layers are added to the network to summarize and combine the learned features (Said et al., 2020).

3. RESULTS AND DISCUSSION

A. Pre-processing

The preprocessing process in Convolutional Neural Network (CNN) is a crucial stage in data processing before the data is given as input to the neural network. To ensure fair comparison and facilitate reproducibility, all models are trained using the same data preprocessing process, training protocol, and evaluation protocol (Chen et al., n.d.). he data graph to be trained is shown in Figure 3.

Figure 3. Train Data Distribution

Preprocessing aims to prepare data in a suitable way so that neural networks can better understand and extract important features from the data. The following are common steps in the preprocessing process for CNN:

- Shuffling the data:
  Shuffling data is an important step in data preprocessing before training a machine learning or deep learning model. Data shuffling involves randomly changing the order of data in the training set, so each data example is shuffled without changing its label. This process ensures that the original ordering of data does not influence the pattern formation by the model during training. Data shuffling should be done before each epoch (iteration through the entire training data) during training to ensure that the model experiences different data variations in each iteration. This helps the model to be more general and avoid overfitting, where the model memorizes the training data and does not generalize well to new data.
- One Hot Encoding:
  One Hot Encoding is a data preprocessing technique used to transform categorical or label data into binary form. This technique is commonly used in machine learning or deep learning when dealing with multi-class classification problems. In One Hot Encoding, each categorical value is represented as a binary vector with a length equal to the number of classes or labels present. This vector will have a value of 1 at the position corresponding to the represented class or label, and a value of 0 at other positions.
- Standardization:
  Standardization is a frequently used technique in data preprocessing to ensure that data has a uniform scale or to normalize the data to have a mean of zero and a variance of one. The goal is to improve the performance of machine learning models or other algorithms that are sensitive to data scale. The result of the standardization process is data with a more uniform distribution, making it easier for machine learning models to process. Standardization is generally used in algorithms that use distance or gradient as a metric to measure similarity or dissimilarity between data.
- Reshaping the data:
- Reshaping the data involves changing the dimensions or shape of the data. In the context of images with a size of (48, 48), reshaping means changing the dimension of the image to a format suitable for the input of the model or algorithm to be used. By performing this reshaping process, facial images originally sized (48, 48) are transformed into a shape that is compatible with the input of the model, allowing them to be used as input for predictions.

- Train-test-validation split:
  This process involves dividing the dataset into three different subsets for training, testing, and validating the machine learning model. Separating the dataset into these subsets is crucial for measuring the model's performance fairly and avoiding overfitting. Now, you have 35,887 images, each containing 64x64 pixels. The data is split into training, testing, and validation data with a ratio of 10%.
  a) Training data: (29,068, 48, 48, 1)
  b) Testing data: (3,589, 48, 48, 1)
  c) Validation data: (3,230, 48, 48, 1)

  You can visualize some training data containing one example from each class using the code provided in Figure 4.

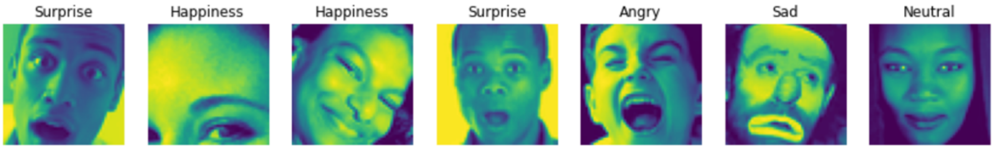
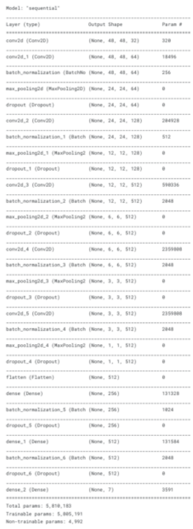


Figure 4 images from each class

- Data augmentation using ImageDataGenerator
  We can perform data augmentation to have more data used in model training and validation to prevent overfitting. Data augmentation can be applied to both the training and validation sets as it helps the model become more generalized and robust to data variations. Data augmentation is a technique that involves applying small transformations to the original data to generate new variations. For example, in the context of image processing, we can perform rotations, shifts, zooms, or flips on images to produce slightly different versions of the originals. By applying data augmentation, we can generate a greater variety of data, allowing the model to learn from different conditions and become more effective at generalizing to new, unseen data.

B. CNN model
In Figure 5, the CNN model architecture is presented. It provides a summary of information for each layer in the model, the number of trainable parameters during training, and the output size of each layer.



Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 48, 48, 32)	320
conv2d_1 (Conv2D)	(None, 48, 48, 64)	18496
batch_normalization (BatchNo	(None, 48, 48, 64)	256
max_pooling2d (MaxPooling2D)	(None, 24, 24, 64)	0
dropout (Dropout)	(None, 24, 24, 64)	0
conv2d_2 (Conv2D)	(None, 24, 24, 128)	204928
batch_normalization_1 (Batch	(None, 24, 24, 128)	512
max_pooling2d_1 (MaxPooling2	(None, 12, 12, 128)	0
dropout_1 (Dropout)	(None, 12, 12, 128)	0
conv2d_3 (Conv2D)	(None, 12, 12, 512)	590336
batch_normalization_2 (Batch	(None, 12, 12, 512)	2048
max_pooling2d_2 (MaxPooling2	(None, 6, 6, 512)	0
dropout_2 (Dropout)	(None, 6, 6, 512)	0
conv2d_4 (Conv2D)	(None, 6, 6, 512)	2359808
batch_normalization_3 (Batch	(None, 6, 6, 512)	2048
max_pooling2d_3 (MaxPooling2	(None, 3, 3, 512)	0
dropout_3 (Dropout)	(None, 3, 3, 512)	0
conv2d_5 (Conv2D)	(None, 3, 3, 512)	2359808
batch_normalization_4 (Batch	(None, 3, 3, 512)	2048
max_pooling2d_4 (MaxPooling2	(None, 1, 1, 512)	0
dropout_4 (Dropout)	(None, 1, 1, 512)	0
flatten (Flatten)	(None, 512)	0
dense (Dense)	(None, 256)	131328
batch_normalization_5 (Batch	(None, 256)	1024
dropout_5 (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 512)	131584
batch_normalization_6 (Batch	(None, 512)	2048
dropout_6 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 7)	3591

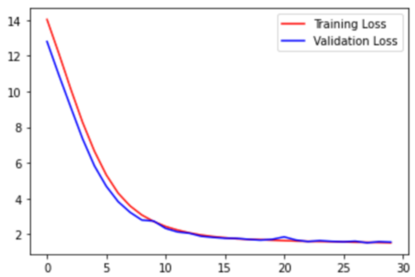
Total params: 5,810,183
Trainable params: 5,805,191
Non-trainable params: 4,992

Figure 5 CNN Model Architecture

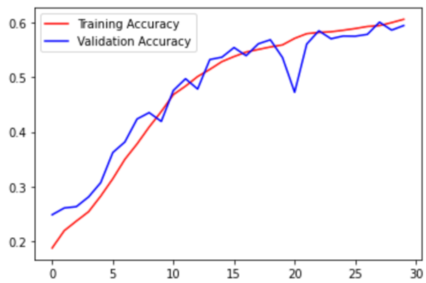
- Early stopping
  Additionally, there is a callback used to stop the training early if the model's performance does not improve for several consecutive epochs. This callback is monitored based on the 'val_accuracy' metric, which represents the accuracy on the validation data. If the accuracy does not improve for 5 consecutive epochs, the training will be stopped. The parameter 'restore_best_weights=True' restores the model to the best weights when the training stops, ensuring the model retains the best performance on the validation data. This callback also monitors the 'val_accuracy' metric. The model is trained by calling the 'fit' method on the model object. The training process continues for 30 epochs with a batch size of 64. The 'checkpointer' callback is used to enable EarlyStopping and ModelCheckpoint

during training. The training data used is from the 'train_generator', and the validation data is from the 'val_generator'. During the CNN model training for 30 epochs, the 'val_accuracy' metric did not improve from the value of 0.61889. This indicates that the model failed to increase its performance or accuracy on the validation data after the last 30 epochs.

- Visualizing results.



Gambar 6 Plot Tranning Loss dan validation Los

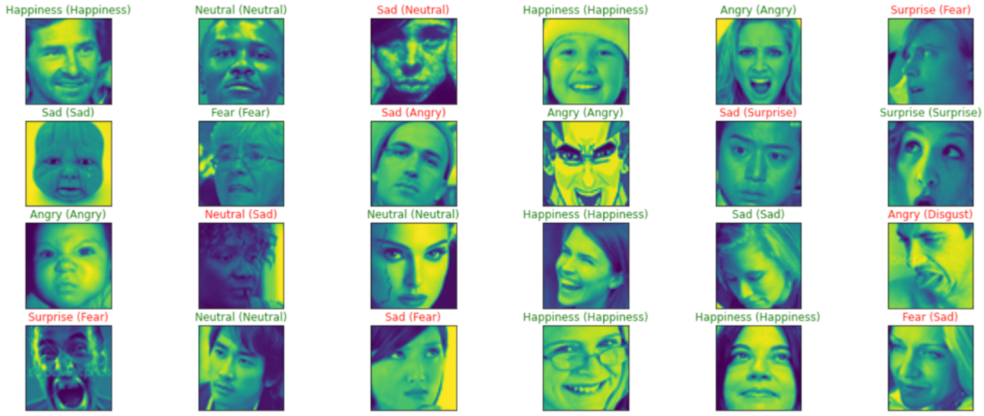


Gambar 7 Plot Tranning Accuracy dan Validation Accuracy

Figure 6 shows the changes in Training Loss (error) represented by the red line and Validation Loss represented by the blue line. Training Loss measures how well the model adapts to the training data. Validation Loss, on the other hand, measures how well the model generalizes to validation data that was not used during training. The objective of training is to obtain steadily decreasing curves for Training Loss and Validation Loss, enabling the model to learn effectively and generalize to new data. Image 7 displays the changes in Training Accuracy represented by the red line and Validation Accuracy represented by the blue line. Training Accuracy measures how accurately the model recognizes the training data, while Validation Accuracy measures the model's accuracy on the validation data that was not part of the training set. The goal of training is to achieve increasing and stable curves for Training Accuracy and Validation Accuracy, indicating that the model is learning effectively and capable of generalizing to new data. This allows us to evaluate the model's performance, identify overfitting, and ensure that it learns well and can be used on previously unseen data. The model's performance on the test data is evaluated, and the results are as follows:

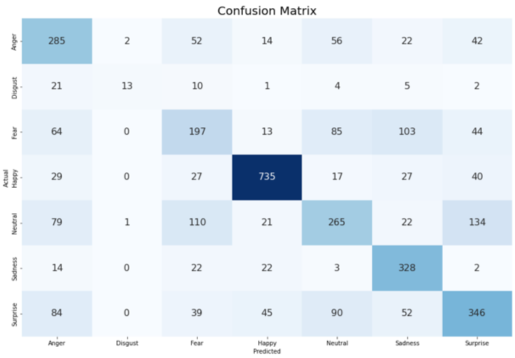
a) The Loss (error) on the test data is 1.5146.
b) The Accuracy on the test data is 0.61 or approximately 61%.

Hasil prediksi dari model pada data uji (test data). Setelah melakukan prediksi pada data uji, hasil prediksi dan label sebenarnya digunakan untuk menampilkan gambar-gambar wajah beserta label prediksi dan label sebenarnya di bawah setiap gambar.



Gambar 7 Hasil Uji Data

The confusion matrix that is generated from the results of the data test can be seen in Figure 8.



	precision	recall	f1-score	support
0	0.495	0.603	0.543	473
1	0.812	0.232	0.361	56
2	0.431	0.389	0.409	506
3	0.864	0.840	0.852	875
4	0.510	0.419	0.460	632
5	0.587	0.839	0.691	391
6	0.567	0.527	0.547	656
accuracy			0.604	3589
macro avg	0.609	0.550	0.552	3589
weighted avg	0.607	0.604	0.599	3589

Gambar 8. Confusion Matrix

4. CONCLUSION

The conclusions that can be drawn from this study are as follows: Some of the data groups (angry, scared, happy, neutral, sad and afraid) tested with an epoch threshold value of 30. have an accuracy of:

	Precision	**recall**
Anger	0.495	0.603
Disgust	0.812	0.232
Fear	0.431	0.389
Happy	0.864	0.840
Neutral	0.510	0.419
Sadness	0.587	0.839
Surprise	0.567	0.527

Conducting a customer satisfaction survey by detecting facial expressions is one of the most important tasks for measuring a service. In this study, the use of digital images of human facial expressions using the Convolution Neural Network Algorithm in surveying customer satisfaction with a service. The Convolution Neural Network algorithm and utilizing the proposed digital image have succeeded in achieving high accuracy performance to assist in evaluating in a service sector.

ACKNOWLEDGEMENTS

We acknowledge the researchers, developers and open source community who have encouraged the use of digital imagery and the Convolutional Neural Network algorithm in customer satisfaction surveys with facial expressions. Their contribution was instrumental in increasing understanding of customer facial expressions and improving the quality of satisfaction surveys more effectively. We also respect the data contributed by users and concern for ethics and privacy in the use of this technology. Hopefully the continued development of this technology will provide greater benefits to society.

REFERENCES

Ahlawat, S., Choudhary, A., Nayyar, A., Singh, S., & Yoon, B. (2020). Improved handwritten digit recognition using convolutional neural networks (Cnn). *Sensors (Switzerland)*, *20*(12), 1–18. doi: 10.3390/s20123344

Akhand, M. A. H., Roy, S., Siddique, N., Kamal, M. A. S., & Shimamura, T. (2021). Facial emotion recognition using transfer learning in the deep CNN. *Electronics (Switzerland)*, *10*(9). doi: 10.3390/electronics10091036

Alamsyah, D., & Pratama, D. (2020). DATASET. *Jurnal Teknologi Informasi*, *4*(2).

Anwar, K., & Setyowibowo, S. (2021). Segmentasi Kerusakan Daun Padi pada Citra Digital. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *7*(1), 39. doi: 10.26418/jp.v7i1.42331

Badar, D. S., & Karsiwan, W. (2021). Survey Kepuasan Mahasiswa Atas Layanan Manajemen Pendidikan di STKIP Muhammadiyah Bogor Tahun 2020. *Jurnal Educatio FKIP UNMA*, *7*(1), 182–188. doi: 10.31949/educatio.v7i1.910

Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., Modi, K., & Ghayvat, H. (2021). Cnn variants for computer vision: History, architecture, application, challenges and future scope. In Electronics (Switzerland) (Vol. 10, Issue 20). MDPI. doi: 10.3390/electronics10202470

Chen, C.-F., Panda, R., Ramakrishnan, K., Feris, R., Cohn, J., Oliva, A., & Fan, Q. (n.d.). *Deep Analysis of CNN-based Spatio-temporal Representations for Action Recognition*. Retrieved from https://github.com/IBM/action-

Fakultas, S. R., Informasi, T., Islam, U., Muhammad, K., & Al Banjari, A. (2020). PENGOLAHAN CITRA DIGITAL DAN HISTOGRAM DENGAN PHYTON DAN TEXT EDITOR PHYCHARM. In Technologia (Vol. 11, Issue 3).

Fu'adah, Y. N., Pratiwi, N. C., Pramudito, M. A., & Ibrahim, N. (2020). Convolutional Neural Network (CNN) for Automatic Skin Cancer Classification System. *IOP Conference Series: Materials Science and Engineering*, *982*(1). doi: 10.1088/1757-899X/982/1/012005

Khan, M. H. M., Boodoo-Jahangeer, N., Dullull, W., Nathire, S., Gao, X., Sinha, G. R., & Nagwanshi, K. K. (2021). Multi-class classification of breast cancer abnormalities using Deep Convolutional Neural Network (CNN). *PLoS ONE*, *16*(8 August 2021). doi: 10.1371/journal.pone.0256500

Kim, G., Choi, I., Li, Q., & Kim, J. (2021). Article a cnn-based advertisement recommendation through real-time user face recognition. *Applied Sciences (Switzerland)*, *11*(20). doi: 10.3390/app11209705

Noyes, E., Davis, J. P., Petrov, N., Gray, K. L. H., & Ritchie, K. L. (2021). The effect of face masks and sunglasses on identity and expression recognition with super-recognizers and typical observers. *Royal Society Open Science*, *8*(3). doi: 10.1098/rsos.201169

Rere, L. M. R., Usna, S., & Soegijanto, D. (2019). Studi Pengenalan Ekspresi Wajah Berbasis Convolutional Neural Network. *Seminar Nasional Teknologi Informasi Dan Komunikasi STI&K (SeNTIK)*, *3*.

Said, Y., Barr, M., & Ahmed, H. E. (2020). Design of a Face Recognition System based on Convolutional Neural Network (CNN). In Technology & Applied Science Research (Vol. 10, Issue 3). Retrieved from www.etasr.com

Sari Hutagalung, F., & Maulana, H. (2023). THE ACCURACY OF MAPE MODEL ON CNN TO CLASSIFICATION OF MOVING OBJECT. *Journal of Theoretical and Applied Information Technology*, *15*(3). Retrieved from www.jatit.org

Zhu, F., Gong, R., Yu, F., Liu, X., Wang, Y., Li, Z., Yang, X., & Yan, J. (n.d.). *Towards Unified INT8 Training for Convolutional Neural Network*.