Identification of Customer Facial Expressions to Improve Service Experience Using Convolutional Neural Networks (CNN)

Icha Miranti Irzan[1], Sriani Sriani[2]
[12]Computer Science, Faculty of Science and Technology
[12]State Islamic University of North Sumatra, Medan, Indonesia
[1]ichamiranti66@gmail.com, [2]sriani@uinsu.ac.id

Abstract— Service experience is an important factor that determines the success and sustainability of a business. Understanding customer facial expressions can help business owners identify emotional responses that arise during the service process. This study develops a model for identifying customer facial expressions using a Convolutional Neural Network (CNN) with the MobileNetV2 architecture to enhance the customer service experience at Toko Saudara. This study aims to develop a system model that can identify customer facial expressions and measure the accuracy of the model in classifying facial expressions based on the dataset used. The model was trained using a transfer learning approach with training parameters of 50 epochs, a batch size of 32, and a learning rate of 0.0001 using the Adam optimizer. The dataset consisted of 2000 augmented images from customer CCTV recordings with two main classes, namely satisfied and dissatisfied. The dataset was split into 80% training, 10% validation, and 10% testing portions. The model achieved a training accuracy of 95.94% with a loss value of 0.1377, and testing performance reached 98% accuracy, with precision, recall, and F1-score all at 98%. The resulting model is able to accurately identify customer facial expressions so that it can be used by businesses to understand customer emotional responses and improve service quality.

Keywords— Identification, Facial Expression, Service Experience, Convolutional Neural Network, MobileNetV2

I. INTRODUCTION

In the modern business landscape, service experience has emerged as a critical determinant of business success. Customers no longer merely seek quality products or services; they demand pleasant and memorable interaction experiences throughout their customer journey [1]. This encompasses all touchpoints from initial information seeking to after-sales service [2]. Recent survey data from PwC Indonesia (2023) underscores this trend, revealing that 76% of Indonesian consumers are willing to pay up to 15% more for superior service experiences, while 62% would switch to competitors after just one or two disappointing encounters [3]. These statistics highlight the substantial impact of service experience on customer retention and business sustainability.

Despite its importance, accurately assessing customer satisfaction during service interactions remains challenging. Many retail businesses face a critical blind spot: store owners often assume customers are satisfied when, in reality, dissatisfaction exists but goes unexpressed. This silent dissatisfaction manifests through gradual customer attrition rather than explicit complaints, leaving businesses unable to identify and address service deficiencies. Traditional feedback methods such as surveys and comment cards often have low participation rates and provide delayed feedback, making them ineffective for obtaining real-time insights emotional responses during service encounters. Consequently, there is a pressing need for an automated, non-intrusive method to assess customer satisfaction in real-time.

Facial expressions offer a promising avenue for understanding customer emotions,

as they serve as a form of nonverbal communication conveying emotions such as happiness, sadness, anger, or dissatisfaction [4]. Recent advances in artificial intelligence, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable effectiveness in facial expression recognition. Previous research has successfully employed CNN architectures to classify customer satisfaction based on facial expressions, achieving accuracy rates of approximately 90.57% [5]. Although the research demonstrates the effectiveness of CNN in facial expression classification, most previous studies still focus on the technical aspects of pattern recognition without considering how the identification results can be fully integrated into strategies to improve the customer service experience.

To address these limitations, this study proposes a facial expression recognition model using MobileNetV2 architecture. MobileNetV2 is a lightweight and efficient CNN variant that employs depthwise separable convolutions, inverted residuals, and linear bottlenecks to achieve high accuracy with reduced computational complexity [6]. Unlike heavier CNN architectures, MobileNetV2 maintains competitive performance while being more efficient to train and deploy, making it particularly suitable for analyzing facial images captured from CCTV footage. This approach leverages existing surveillance infrastructure in retail environments, where customer facial images can be captured and subsequently analyzed to evaluate service quality and customer satisfaction.

The primary objective of this research is to develop and evaluate a MobileNetV2-based facial expression recognition model for assessing customer satisfaction from CCTV-captured images in retail service environments.

II. METHOD

This study uses an experimental approach to develop a customer facial expression identification system using a Convolutional Neural Network (CNN) with a MobileNetV2 architecture.

A. Data Collection

Essentially, a data collection method is a scientific technique for gathering data for a specific purpose [7]. The following are the data collection methods used in this study:

1. Literature Study

Facial expressions serve as a fundamental form of nonverbal communication that conveys human emotions during interactions. Facial expressions communicate 55% of emotional information, significantly more than voice (38%) and verbal language (7%) [8]. This predominance of facial cues in emotional communication makes facial expression analysis particularly valuable for understanding customer satisfaction in service contexts. There are six universal facial expressions that can be recognized across different cultures include anger, disgust, fear, happiness, sadness, and surprise [9]. These universal expressions form the foundation for automated facial expression recognition systems. In the context of customer service, the ability to recognize and interpret facial expressions becomes crucial, as facial expressions can serve as important indicators for understanding customers' emotional conditions, particularly in measuring their satisfaction levels with the services provided.

To analyze facial expressions automatically and accurately, technology capable of processing visual data effectively is required. Deep learning has emerged as a powerful branch of machine learning within computer vision, revolutionizing how machines process and interpret visual information. Image processing methods that utilize multiple layers of structure and involve multiple transformation processes are known as deep learning. deep learning, as a subfield of machine learning, leverages multi-layered neural networks to process and learn complex data representations simultaneously [10]. Deep learning methods are particularly effective for handling complex problems requiring large-scale data processing, such as image, audio, and text processing. Image classification represents a type of research

that uses artificial intelligence to aid in various identification processes, including disease diagnosis, object recognition, and emotion detection based on specific characteristics found in particular images.

Convolutional Neural Network (CNN) is a type of deep learning algorithm specifically designed to accept image input and identify objects within the image. CNN is a development of Multi-Layer Perceptron (MLP) that can process 2D image data and is a type of deep neural network frequently applied to image data due to its network depth [11]. Because of their ability to capture spatial patterns, Convolutional Neural Networks have become the primary method for feature extraction in images. CNN's ability to automatically extract features from data distinguishes it from traditional machine learning methods, where features must be manually extracted from images, which was one of the most difficult tasks in computer vision [12]. The layers that make up a CNN consist of several specialized components, each serving a specific purpose in the feature extraction and classification process. The fundamental layers comprising a CNN include the Convolution Layer, Activation ReLU Layer, Pooling Layer, and Fully Connected Layer [13].

While standard CNN architectures achieve high accuracy in facial expression recognition, their computational demands pose challenges for deployment on resource-constrained devices or real-time applications. MobileNetV2 is one of the latest highly effective deep learning models designed to address the problem of consuming large computational resources without sacrificing accuracy [14]. MobileNetV2 is a CNN architecture developed by Google AI in 2018 to support deep learning applications on devices with limited computational resources, such as IoT devices, smartphones, and edge computing [15]. MobileNetV2 uses two types of convolutions: depthwise and pointwise. MobileNetV2 has a new layer module with inverted residuals with a linear bottleneck. This significantly reduces the amount of memory required for processing. The advantages of using the MobileNetV2 architecture are high accuracy and a smaller number of training parameters compared to other CNN architectures, which reduces the amount of computation required. Furthermore, the MobileNetV2 model size is small but still provides good performance, making it ideal for efficient and accurate customer facial expression identification to improve service quality.

2. Observation

One of the data collection methods used was observation, which was conducted to obtain more in-depth information about customer facial expressions during service interactions. The observation was carried out at Toko Saudara, located in Jl. Melur Huta III Nagori Silau Manik, Siantar District, Simalungun Regency, North Sumatra, to understand the patterns of customer emotional responses and identify suitable moments for capturing facial expression data through CCTV recordings.

3. Data Collection

The total data used in this study amounted to 2,000 images, obtained from CCTV recordings at Toko Saudara. Data collection was conducted with the owner's permission, with agreements that CCTV recordings would only be used for academic purposes, customer facial images would not be published identically, and all data would be deleted after the research was completed. The dataset used in this study consists of 2,000 images evenly divided into two categories, namely 1,000 images of satisfied expressions and 1,000 images of dissatisfied expressions. The initial dataset comprised 372 facial images (186 satisfied and 186 dissatisfied) captured from CCTV footage. To meet the requirements for CNN training that necessitates large datasets, data augmentation was performed using techniques including rotation (±10°), brightness adjustment (0.8-1.2), horizontal flip, zoom (5-15%), and translation (±10 pixels), resulting in 1,000 images per class.

In this study, the process of classifying customer facial expressions using the CNN method with the MobileNetV2 architecture was divided into three data sets, namely

(80%) training data, (10%) validation data, and (10%) test data. Training data is used to build or train the model, validation data to optimize the training process, and test data to evaluate the resulting model. Details of the data sets can be seen in Table 1.

Table 1. Dataset Distribution

Class	Training Data	Validation Data	Testing Data	∑
Satisfied	800	100	100	1000
Dissatisfied	800	100	100	1000
Total	1600	200	200	2000
Percentage	80%	10%	10%	1000 %

The following is a visualization of a sample facial expression image for the satisfied class as shown in Figure 1.

Figure 1. Sample Dataset of Satisfied Expression

The following is a visualization of an example facial expression for the dissatisfied class as shown in Figure 2.

Figure 2. Sample Dataset of Dissatisfied Expression

B. Data Preprocessing

Data preprocessing is crucial to ensure the quality and effectiveness of the developed classification model. This process begins with the collection of images of customer facial expressions from CCTV footage of varying quality, achieved by standardizing the image processing to ensure data consistency. Image preprocessing is performed prior to model training to improve the quality of the input data and ensure that the CNN model can effectively learn relevant characteristics in customer facial expression images to produce accurate satisfaction classification.

The initial stage of data processing in this study involves:

Image Resizing**:** All images were resized to 224×224 pixels to match the standard input dimension of MobileNetV2 architecture, ensuring uniformity across the dataset and optimizing processing speed.

After resizing is performed, the next step is to load the dataset. At this stage, the dataset is entered, separated as needed, and then directories are determined for training, validation, and testing data. This process also includes labeling and data division (dataset splitting) with a proportion of 80% for training data, 10% for validation data, and 10% for testing data.

C. Data Augmentation

Data augmentation is performed to increase the number and variety of images in a dataset, allowing the model to learn various image characteristics. This technique is used to prevent overfitting, a condition where a model performs well on training data but fails to generalize to new data. In this study, augmentation was applied to the model through several image transformations, including rotation (±10°), brightness adjustment (0.8-1.2), horizontal flip, zoom (5-15%), and translation (±10 pixels). The augmentation process was conducted offline by applying random combinations of these techniques to each original image until reaching 1,000 images per class, producing an average of 4-5 augmentation variations per image.

Figure 3. Results of Satisfied Expression Augmentation

ISSN 2442-7888 (online) DOI 10.24167/sisforma.v12i2.14314

D. CNN Model Design

The facial expression classification model in this study was developed using the MobileNetV2 architecture through a transfer learning approach. MobileNetV2 was selected due to its efficiency in extracting image features with minimal computational resources, achieved through depthwise separable convolution and inverted residual blocks. This architecture is particularly suitable for facial expression recognition tasks that require both accuracy and computational efficiency.

The model implementation utilized pre-trained MobileNetV2 weights from the ImageNet dataset as a fixed feature extractor. The base model layers were frozen (trainable=False) to preserve the learned representations from large-scale image data. Custom classification layers were then added on top of the base model to adapt it for binary facial expression classification. The complete architecture consists of: (1) MobileNetV2 base model with input shape 224×224×3, producing feature maps of size (7, 7, 1280); (2) GlobalAveragePooling2D layer to reduce spatial dimensions into a 1280-dimensional feature vector; (3) Dense layer with 128 neurons and ReLU activation for learning task-specific representations; and (4) Output Dense layer with 2 neurons and softmax activation for binary classification into satisfied and dissatisfied classes.

```
model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	Param #
mobilenetv2_1.00_224 (Functional)	(None, 7, 7, 1280)	2,257,984
global_average_pooling2d (GlobalAveragePooling2D)	(None, 1280)	0
dense (Dense)	(None, 128)	163,968
dense_1 (Dense)	(None, 2)	258

Total params: 2,422,210 (9.24 MB)
Trainable params: 164,226 (641.51 KB)
Non-trainable params: 2,257,984 (8.61 MB)

Figure 4. Summary Model MobileNetV2

As shown in Figure 4, the model contains 2,422,210 total parameters, with 164,226 trainable parameters (641.51 KB) in the classification layers and 2,257,984 non-trainable parameters (8.61 MB) in the frozen MobileNetV2 base. This parameter distribution demonstrates the efficiency of transfer learning, where extensive pre-learned features are leveraged while only a small portion requires training on the specific facial expression dataset.

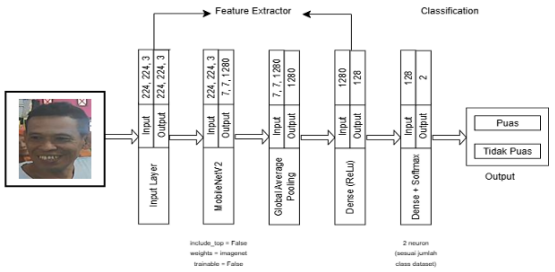

Figure 5. MobileNetV2 Architecture

Figure 5 illustrates the complete processing pipeline of the model. Input facial images pass through the MobileNetV2 feature extraction stage, followed by GlobalAveragePooling2D, Dense layers with ReLU activation, and finally the output layer with softmax activation to produce classification probabilities for satisfied and dissatisfied expressions.

E. Testing

The model's generalization ability was tested using pre-prepared test data comprising 200 images (100 satisfied and 100 dissatisfied) that were never seen during the training process. The testing procedure involved feeding each image from the test dataset into the trained MobileNetV2-CNN model to generate classification predictions. For each input image, the model produces probability scores for both classes using the softmax activation function, and the class with the highest probability is selected as the final prediction. The testing process was implemented using Python programming language on the Google Colaboratory platform, utilizing the model.predict() function to generate predictions and model.evaluate() to calculate overall test accuracy and loss.

F. Model Evaluation

The goal of the evaluation phase of this research is to determine how well a Convolutional Neural Network (CNN) model with MobileNetV2 architecture classifies

customer facial expressions. Part of this process involves using test data that differs from the data used to train the model. To assess model accuracy, the trained CNN model predicts facial expression satisfaction based on the test data. These predictions are then compared with the labels on the actual test data. Model performance is quantitatively measured using metrics such as accuracy, precision, recall, and F1-Score. The evaluation in this study used a confusion matrix to calculate the accuracy, precision, recall and F1-Score values shown in equations 1-4.

Accuracy is the percentage of correct predictions out of the total predictions made by the model. It is calculated using the formula

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} x100\% \quad (1)$$

Precision shows how many positive predictions are correct out of all positive predictions made. The formula is

$$Precision = \frac{TP}{TP+FP} x100\% \quad (2)$$

Recall measures the model's ability to find all positive instances. The formula is

$$Recall = \frac{TP}{TP+FN} x100\% \quad (3)$$

F1-score is the harmonic mean between precision and recall, providing an overview of the balance between the two. The formula is

$$F1 - Score = 2x \frac{Precision\ x\ Recall}{Precision+Recall} \quad (4)$$

III. RESULTS AND DISCUSSION

In this study, a series of tests were conducted to obtain the best results from a customer facial expression classification model for satisfaction assessment, based on a Convolutional Neural Network (CNN) algorithm with a MobileNetV2 architecture. The dataset was divided into three parts: training dataset (80%), validation dataset (10%), and testing dataset (10%). The training dataset was used to train the model to recognize patterns in facial expression images, while the validation dataset was used during the training process to monitor model performance and prevent overfitting. The testing dataset was used in the final stage to evaluate the model's performance on new, previously unseen data, ensuring that the results reflect the model's overall generalization capability. The results of the training process are then analyzed through the accuracy and loss values obtained at each epoch for the training and validation data, and then reinforced with evaluation using a confusion matrix to observe the distribution of predictions on the testing data.

A. Model Training Results

The training process of the MobileNetV2 model was conducted over 50 epochs using a dataset that had been divided into training data (1600 images) and validation data (200 images) with a batch size of 32 and a learning rate of 0.0001 using the Adam optimizer. Each epoch produced accuracy and loss values for both datasets, which were then used to monitor the model's performance over time. The validation data served as an early stopping indicator to ensure the model did not overfit during the training process. Detailed results of the training process can be seen in the following output.

```
Epoch 1/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 132s 2s/step - accuracy: 0.5445 - loss: 0.7912 - val_accuracy: 0.6500 - val_loss: 0.6020
Epoch 2/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 126s 2s/step - accuracy: 0.6674 - loss: 0.6109 - val_accuracy: 0.7100 - val_loss: 0.5477
Epoch 3/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 110s 2s/step - accuracy: 0.7301 - loss: 0.5561 - val_accuracy: 0.7200 - val_loss: 0.5174
Epoch 4/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 113s 2s/step - accuracy: 0.7572 - loss: 0.5085 - val_accuracy: 0.7750 - val_loss: 0.4713
Epoch 5/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 138s 2s/step - accuracy: 0.7791 - loss: 0.4949 - val_accuracy: 0.8000 - val_loss: 0.4551
Epoch 6/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 120s 2s/step - accuracy: 0.8042 - loss: 0.4389 - val_accuracy: 0.7900 - val_loss: 0.4465
Epoch 7/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 131s 2s/step - accuracy: 0.8244 - loss: 0.4149 - val_accuracy: 0.7950 - val_loss: 0.4454
Epoch 8/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 142s 2s/step - accuracy: 0.8083 - loss: 0.4310 - val_accuracy: 0.8300 - val_loss: 0.3975
Epoch 9/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 107s 2s/step - accuracy: 0.8361 - loss: 0.3834 - val_accuracy: 0.8450 - val_loss: 0.3789
Epoch 10/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 119s 2s/step - accuracy: 0.8526 - loss: 0.3706 - val_accuracy: 0.8450 - val_loss: 0.3809
Epoch 11/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 117s 2s/step - accuracy: 0.8595 - loss: 0.3570 - val_accuracy: 0.8700 - val_loss: 0.3472
Epoch 12/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 111s 2s/step - accuracy: 0.8612 - loss: 0.3414 - val_accuracy: 0.8700 - val_loss: 0.3435
Epoch 13/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 141s 2s/step - accuracy: 0.8681 - loss: 0.3289 - val_accuracy: 0.8700 - val_loss: 0.3325
Epoch 14/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 110s 2s/step - accuracy: 0.8839 - loss: 0.3253 - val_accuracy: 0.8650 - val_loss: 0.3260
Epoch 15/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 142s 2s/step - accuracy: 0.8788 - loss: 0.3089 - val_accuracy: 0.9000 - val_loss: 0.2918
Epoch 16/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 151s 2s/step - accuracy: 0.8969 - loss: 0.2919 - val_accuracy: 0.8300 - val_loss: 0.3458
Epoch 17/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 133s 2s/step - accuracy: 0.8671 - loss: 0.3064 - val_accuracy: 0.8900 - val_loss: 0.3005
Epoch 18/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 140s 2s/step - accuracy: 0.8904 - loss: 0.2882 - val_accuracy: 0.9100 - val_loss: 0.2705
Epoch 19/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 145s 2s/step - accuracy: 0.8863 - loss: 0.3126 - val_accuracy: 0.9050 - val_loss: 0.2622
Epoch 20/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 149s 2s/step - accuracy: 0.8987 - loss: 0.2777 - val_accuracy: 0.9000 - val_loss: 0.2690
Epoch 21/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 144s 2s/step - accuracy: 0.8928 - loss: 0.2771 - val_accuracy: 0.9050 - val_loss: 0.2642
Epoch 22/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 108s 2s/step - accuracy: 0.9076 - loss: 0.2559 - val_accuracy: 0.9100 - val_loss: 0.2502
Epoch 23/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 143s 2s/step - accuracy: 0.9136 - loss: 0.2370 - val_accuracy: 0.9150 - val_loss: 0.2468
Epoch 24/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 143s 2s/step - accuracy: 0.9158 - loss: 0.2451 - val_accuracy: 0.9250 - val_loss: 0.2373
Epoch 25/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 116s 2s/step - accuracy: 0.8926 - loss: 0.2556 - val_accuracy: 0.9050 - val_loss: 0.2477
Epoch 26/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 109s 2s/step - accuracy: 0.8948 - loss: 0.2641 - val_accuracy: 0.9100 - val_loss: 0.2326
Epoch 27/50
50/50 ━━━━━━━━━━━━━━━━━━━━ 108s 2s/step - accuracy: 0.9207 - loss: 0.2352 - val_accuracy: 0.9550 - val_loss: 0.2139
```

Figure 6. Model Training Log

The training log shows that the model started with relatively low performance in the early epochs, where training accuracy reached around 54% and validation accuracy 65%. As the epochs progressed, both metrics showed consistent and gradual improvement. In the middle phase of training (epochs 10-30), the model demonstrated stable learning with training accuracy increasing from 70% to 90%, while validation accuracy followed a similar pattern without showing significant signs of overfitting. The loss on both training and validation decreased consistently from around 0.7 in the early epochs to 0.2-0.3 in the final epochs.

By the final epochs (40-50), the model reached convergence with training accuracy stable at around 95.94% and validation accuracy reaching 97%. This pattern indicates that the model successfully learned well without experiencing overfitting, as evidenced by the validation accuracy that remained close to or even slightly higher than the training accuracy. The final training results showed an accuracy of 95.94% with a loss of 0.1377.

Figure 7 displays the comparison between training and validation metrics over 50 epochs. The "Model Accuracy" graph shows that both training accuracy (blue) and validation accuracy (orange) curves started from around 60% in the early epochs and increased consistently to reach 95-97% in the final epochs. Both curves followed similar patterns without a significant gap, indicating no serious overfitting.

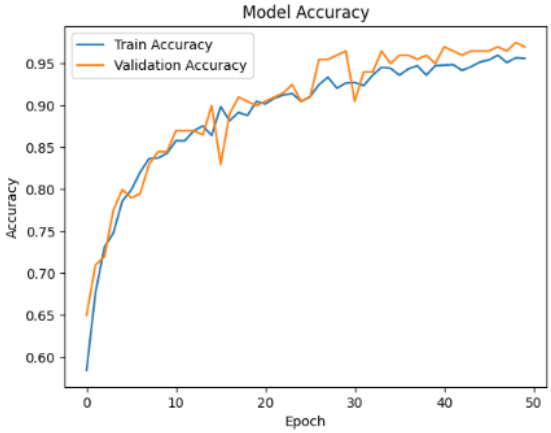

Figure 7. Model Accuracy Graph

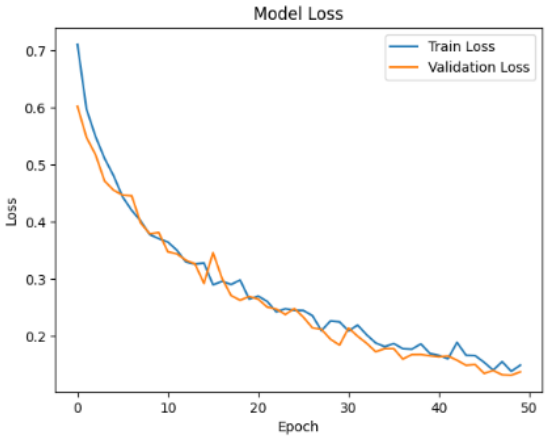

Figure 8. Model Loss Graph

The "Model Loss" graph displays consistent loss reduction from around 0.7 in the first epoch to stabilize at around 0.15-0.18 in the last epoch. Training loss (blue) and validation loss (orange) showed parallel decreasing patterns, confirming that the model successfully learned optimally and could generalize well to unseen data. The small gap between training and validation curves throughout the training process indicates good model stability and generalization capability.

B. Data Testing

The following shows the results of testing the CNN model on the test data with two classes: satisfied and dissatisfied. The image displays example outputs, showing pictures of customers' facial expressions along with the predicted class. These predictions are displayed directly on the image with the labels

ISSN 2442-7888 (online) DOI 10.24167/sisforma.v12i2.14314

"Prediction" and "Confidence", indicating the model's classification results. This output demonstrates the model's ability to distinguish between two levels of customer satisfaction based on features extracted from the input facial images. In this study, MobileNetV2 based on Convolutional Neural Network (CNN) was used for customer facial expression classification. The testing phase involves several systematic procedures to evaluate the model's performance. First, the model's generalization ability is tested with test data that has been prepared and never seen during the training process. The MobileNetV2-CNN model that has been trained to generate satisfaction classification predictions is used to process facial expression images in this test data.

Figure 9. Data Test Results

The test results in Figure 9 show that the CNN model based on MobileNetV2 is capable of performing classification with very high accuracy. The model consistently predicts customer facial expressions with a high level of confidence, demonstrating its ability to effectively recognize the characteristics of satisfied and dissatisfied expressions.

C. Model Evaluation

The evaluation phase of this study aims to measure the effectiveness of the Convolutional Neural Network (CNN) model with the MobileNetV2 architecture in classifying customer facial expressions. The evaluation is conducted using test data that is different from the training data to ensure objectivity. The trained CNN model is used to predict the satisfaction levels in the test data, and the prediction results are then compared with the actual labels. The model's performance is quantitatively assessed using evaluation metrics such as accuracy, precision, recall, and F1-score.

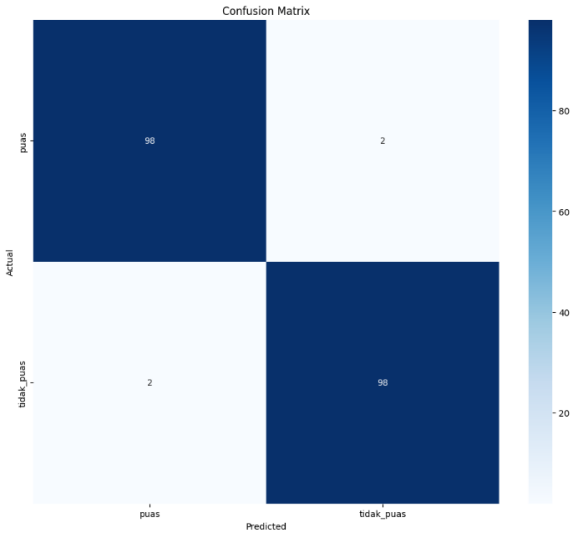

Figure 10. Confusion Matrix

Figure 10 shows the confusion matrix results from the model evaluation on the testing data. Out of a total of 200 samples, there are 98 images of the "satisfied" class that were correctly predicted (True Positive) and 98 images of the "dissatisfied" class that were also correctly predicted (True Negative). Conversely, 2 images of the "satisfied" class were incorrectly predicted as "dissatisfied" (False Negative) and 2 images of the "dissatisfied" class were incorrectly predicted as "satisfied" (False Positive).

```
Classification Report (Sklearn):
              precision    recall  f1-score   support

        puas       0.98      0.98      0.98       100
  tidak_puas       0.98      0.98      0.98       100

    accuracy                           0.98       200
   macro avg       0.98      0.98      0.98       200
weighted avg       0.98      0.98      0.98       200
```

Figure 11. Classification Report

Based on these results, the precision, recall, and F1-score values are 0.98 for both classes. A precision of 0.98 indicates that 98% of all positive predictions are correct. A recall of 0.98 shows that 98% of all actual positive data were correctly identified by the model. An F1-score of 0.98 also indicates a good balance between precision and recall.

The overall classification report shows an accuracy of 98%, with 196 correct predictions out of a total of 200 samples. These results indicate that the CNN model with MobileNetV2 architecture is capable of classifying customer facial expressions with excellent performance, and the prediction errors are relatively small.

D. Detailed Analysis of Test Results

Testing on 200 test samples showed that the model achieved 98% accuracy with 196 correct predictions and 4 misclassifications (2 false negatives and 2 false positives). The prediction confidence ranged from 50.51% to 99.97%, with the majority of predictions having confidence above 80%, indicating that the model is capable of effectively learning the differences between satisfied and dissatisfied expressions.

Error Case Analysis

Four error cases were identified with the following patterns:

1. poor/blurry image quality causing facial features to be suboptimally detected (aug_0116_1844.jpg, confidence = 77%),
2. potential augmentation bug that shuffled the labels (aug_0101_1102.jpg, confidence = 76%),
3. feature detection issues due to extreme lighting or suboptimal angles (aug_0053_6241.jpg, confidence = 62%), and
4. ambiguous expressions without clear indicators (ori_0167.jpg, confidence = 59%).

Overall, the results of this testing confirm that the MobileNetV2 model is capable of achieving a high accuracy of 98% with a low error rate. Nevertheless, the model's robustness against variations in real-world conditions, such as suboptimal lighting, tilted facial angles, and expressions resembling other classes, can still be improved. Therefore, the proposed method has proven effective for identifying customer facial expressions and can be relied upon for classification processes in the context of customer service under real-world conditions.

IV. CONCLUSION

This study shows that the CNN architecture based on MobileNetV2 is capable of delivering excellent performance in classifying customer facial expressions into two categories, namely satisfied and dissatisfied. The model, trained with a dataset of 2,000 images and split using a stratified approach (80% training, 10% validation, 10% testing), achieved a training accuracy of 95.94% with a loss of 0.1377, and a testing accuracy of 98.00% with a loss of 0.1285. The confusion matrix results indicate that out of 200 testing samples, only 4 samples were misclassified, with precision, recall, and F1-score values all being 0.98 for both classes. This confirms that MobileNetV2 with a transfer learning approach using pre-trained ImageNet weights, combined with optimal hyperparameter settings (learning rate 0.0001, batch size 32, Adam optimizer, and 50 epochs), has proven to be efficient and effective for facial expression classification tasks in the context of customer service.

For further development, it is recommended to increase the dataset variety with more diverse lighting conditions, implement a real-time system on edge devices, and integrate it with an automatic notification mechanism for service staff.

REFERENCES

[1] A. Wardhana, *Consumer Behavior In The Digital Era 4.0*. Jawa Tengah: Eureka Media Aksara, 2024.

[2] Admin, “Pengertian Pengalaman

Pelanggan: Panduan Lengkap untuk Meningkatkan Kepuasan Pelanggan," *Bidang Usaha*, 2024. https://bidangusaha.co.id/pengertian-pengalaman-pelanggan/

[3] P. Indonesia, "Consumer Insights Survey 2023," 2023. https://www.pwc.com/id/en/consumer-industrial-products-services/indonesia-gcis-2023-placemat.pdf (accessed Apr. 02, 2025).

[4] D. P. Anggraini, D. E. Fitriani, F. Ulfah, and T. Agustin, "Klasifikasi Ekspresi Wajah Menggunakan Convolutional Neural Network (CNN) Dengan Perbandingan Dua Model Dimodifikasi," *Semin. Nas. Amikom Surakarta*, pp. 160–171, 2024.

[5] D. Prasetyawan and R. Gatra, "Model Convolutional Neural Network untuk Mengukur Kepuasan Pelanggan Berdasarkan Ekspresi Wajah," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 3, pp. 661–673, 2022, doi: 10.28932/jutisi.v8i3.5493.

[6] F. Ramadhani, S. Rahardiantoro, and M. Masjkur, "Acne Severity Classification Study Using Convolutional Neural Network Algorithm with MobileNetV2 Architecture," *Indones. J. Stat. Its Appl.*, vol. 8, no. 2, pp. 112–128, 2024, doi: 10.29244/ijsa.v8i2p112-128.

[7] O. Yuliardi, I. Zufria, and M. D. Irawan, "Sistem Pakar Mendiagnosis Penyakit pada Tanaman Jeruk Menggunakan Metode Dempster Shafer," *J. Compr. Sci.*, vol. 2, no. 1, pp. 389–397, 2023, doi: 10.59188/jcs.v2i1.224.

[8] A. T. Akbar, S. Saifullah, and H. Prapcoyo, "Klasifikasi Ekspresi Wajah Menggunakan Covolutional Neural Network," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 6, pp. 1399–1412, 2024, doi: 10.25126/jtiik.2024118888.

[9] Lina, A. M. Adhitya, Wasino, and D. Ajienegoro, "Identifikasi Emosi Wajah Pengguna Konferensi Video Menggunakan Convolutional Neural Network Dengan Arsitektur Vgg-16," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9(5), pp. 1047–1054, 2022.

[10] M. Diqi, "Mengenal Deep Learning," *ResearchGate*, 2023. https://www.researchgate.net/publication/367636088_MENGENAL_DEEP_LEARNING

[11] M. N. Kholis, "Mengenal Convolutional Neural Network (CNN)," *Medium*, 2023. https://muchamadnurkholis.medium.com/mengenal-convolutional-neural-network-cnn-36969caa709 (accessed Mar. 13, 2025).

[12] Supiyandi, W. E. Judistira, S. Nurliani, R. S. Darmono, and I. Putri, "Penerapan Deep Learning dalam Analisis Citra Gigi," *J. Pendidik. Dan Ilmu Sos.*, vol. 2, no. 4, pp. 117–128, 2024, doi: 10.54066/jupendis.v2i4.2165.

[13] D. Alamsyah and D. Pratama, "Implementasi Convolutional Neural Networks (CNN) untuk Klasifikasi Ekspresi Citra Wajah pada FER-2013 Dataset," *J. Teknol. Inf.*, vol. 4, no. 2, pp. 350–355, 2020, doi: 10.36294/jurti.v4i2.1714.

[14] M. S. A. M. Al-Gaashani, W. Xu, and E. Y. Obsie, "MobileNetV2-Based Deep Learning Architecture With Progressive Transfer Learning For Accurate Monkeypox Detection," *Appl. Soft Comput.*, vol. 169, 2025, doi: https://doi.org/10.1016/j.asoc.2024.112553.

[15] A. Hadhiwibowo, S. R. Asri, and R. A. Dinata, "Penerapan Convolutional Neural Network dengan Arsitektur Mobilenetv2 Pada Aplikasi Penerjemah dan Pembelajaran Bahasa Isyarat," *TIN Terap. Inform. Nusant.*, vol. 4, no. 8, pp. 518–523, 2024, doi: 10.47065/tin.v4i8.4879.