JOIN (Jurnal Online Informatika)
p-ISSN: 2528-1682, e-ISSN: 2527-9165
Volume 11 Number 1 | June 2026: 142-155
DOI: 10.15575/join.v11i1.1677

Convolutional Neural Networks for Measuring Service Satisfaction of Hajj Pilgrims through Facial Expression Analysis

Undang Syaripudin[1], Jumadi[2], Diena Rauda Ramdania[3], Indah Sri Lestari[4], Indri Nurfiani[5], Alfin Yogi Setyawan[6], Maisevli Harika[7], Mimin Mintarsih[8]
[1,2,3,4,5,6]Department of Informatics, UIN Sunan Gunung Djati Bandung, Indonesia
[3]Exertion Games Lab, Human Centred Computing, Faculty of Information Technology, Monash University, Australia
[7]Informatics Engineering, Politeknik Negeri Bandung, Indonesia
[8]Sharia Business Management, Universitas Ma'soem, Bandung, Indonesia

Article Info

Article history:

Received September 3, 2023
Revised May 17, 2024
Accepted June 28, 2024
Published April 26, 2026

Keywords:

CNN
Customer Satisfaction
Facial Expression Recognition
Hajj Services
Machine Learning
Real-Time Monitoring

ABSTRACT

Facial expressions serve as important non-verbal indicators of human emotions and can be leveraged to assess satisfaction levels in service environments. In the context of Hajj and Umrah, where verbal feedback may be limited due to language barriers or cultural factors, facial expression recognition offers a non-intrusive method to evaluate service quality. This study proposes a Convolutional Neural Network (CNN)-based model to detect emotional states such as happiness and dissatisfaction through facial expressions of pilgrims. A quantitative approach was adopted, employing preprocessing techniques including normalization, augmentation, and image resizing. The CNN architecture comprised multiple convolutional, pooling, and fully connected layers. The model was evaluated using accuracy, precision, recall, and F1-score metrics. Experiments with varying batch sizes (32, 64, 128, 256) across 50 epochs revealed that the optimal performance was achieved with a batch size of 64, resulting in an accuracy of 63%, precision of 66%, recall of 60%, and F1-score of 62%. During deployment, the model correctly classified 12 out of 16 real-world images, achieving a real-time accuracy of 78%. Therefore, the deployment results should be considered preliminary. Future studies will involve larger deployment samples, n-fold stratified cross-validation to obtain statistically reliable model performance, and subgroup analyses (e.g., lighting, facial pose, age, and gender) to better understand model behavior under diverse real-world conditions. All deployment images were collected with participant consent and processed without storing biometric data. These findings suggest that CNN-based emotion recognition can support real-time service evaluation and enhance the quality of pilgrim services during the Hajj and Umrah.

Corresponding Author:

Undang Syaripudin
Department of Informatics, UIN Sunan Gunung Djati Bandung, Indonesia
Jl. A.H. Nasution No. 105A, Cibiru, Kota Bandung, Jawa Barat, Indonesia
Email: undang_if@uinsgd.ac.id

1. INTRODUCTION

According to the 2023 Performance Report of the General of Hajj and Umrah Organization (DJPHU), Ministry of Religious Affairs, 198 Integrated Hajj and Umrah Service Centers (PLHUT) have been established in the Ministry of Religious Affairs offices across various districts and cities [1]. This effort is part of the Ministry's commitment to enhancing the services provided to Hajj pilgrims. The services offered by the Ministry include Hajj ritual guidance, visa processing, transportation, accommodation, catering, and healthcare services during the Hajj pilgrimage [2] [3]. Annually, the Ministry conducts satisfaction surveys among Hajj pilgrims to ensure that the services provided meet their expectations and needs. Conducting satisfaction surveys is crucial as the results provide valuable feedback for improving and enhancing future services [4], [5].

On the other hand, the human face is a rich source of information, particularly in conveying an individual's emotional state [6], [7], [8], [9]. Emotional expressions are spontaneous efforts by individuals to communicate their feelings or emotions in response to specific situations [10], [11], [12], [13], [14]. Using facial expressions as an assessment tool in satisfaction surveys has its advantages, as facial expressions can provide direct and honest indications of a person's feelings toward the services received [15], [16], [17]. Within the setting of Hajj services, where pilgrims often experience physical depletion and otherworldly centers, verbal criticism may be restricted. Facial expressions, in any case, can give non-verbal prompts that reflect their genuine enthusiastic state, making them a complementary strategy for fulfillment. Facial expressions can be divided into macro expressions and micro expressions. Macro-expressions are visible to the naked eye and can last up to 4 seconds [18]. Micro-expressions are much shorter, lasting only a fraction of a second, making them more challenging to observe [19]. By analyzing micro and macro facial expressions, the system can detect satisfaction levels even when participants are unable to express their feelings verbally or intentionally conceal them. Therefore, assessing through facial expressions can be an effective and efficient method for measuring customer satisfaction [15] [16], [17].

Previous research demonstrated a system combining local and holistic facial information for real-time customer interest tracking, achieving 99.43% accuracy in posture prediction and 93.20% in facial appearance detection on the RaFD dataset [20]. Another approach used facial emotion recognition to measure customer satisfaction, achieving 98.66% accuracy with SVM by analyzing geometric features of facial landmarks [21]. Separately, CNN-based micro-expression recognition showed success rates exceeding 65.97% in applications like consumer behavior analysis, highlighting its practical potential. Additionally, studies on emotion recognition through AI have reviewed techniques in Human-Computer Interaction (HCI) and Affective Computing, identifying the highest accuracies achieved using methods like Stationary Wavelet Transform for facial recognition (98.83%), Particle Swarm Optimization for speech recognition (99.47%), and statistical features for physiological signals (87.15%) [22]. Lastly, a deep learning method employing CNN for facial key point detection achieved up to 96% accuracy using the FACEDB dataset for customer satisfaction measurement [23]. Meanwhile, recent studies comparing CNN and LSTM models for facial expression classification reported CNN achieving 97.26% accuracy and LSTM reaching 92.38%, further demonstrating their effectiveness in emotion detection systems to enhance customer satisfaction analysis [24].

This research aims to develop a machine-learning model using Convolutional Neural Networks (CNN) to detect facial expressions. This model is expected to be used to measure the satisfaction levels of Hajj pilgrims. Not at all like past considerations, this investigation applies CNN-based facial acknowledgment straightforwardly within the setting of Hajj satisfaction surveys, a novel approach in Indonesia. This innovation offers a non-intrusive, real-time elective to conventional studies, giving wealthier, more objective input to move forward benefit conveyance. Several studies have developed CNN-based systems for emotion recognition. For instance, Charvi Jain et al. [1] utilized CNN to detect facial features such as eyes and lips and classify them into six emotions: happy, fearful, angry, disgusted, neutral, and sad. This study combined Support Vector Machine (SVM) classification, Grid Search, Gabor filters, and transformations like Histogram of Oriented Gradients (HOG) and Discrete Wavelet Transform (DWT) to optimize classification, achieving an average precision rate of 85%.

Luigi Bibbo et al. [2] developed an emotion recognition model focusing on mental health using a Self-Normalizing Neural Network (SNN) with a cascade and ensemble layer approach, achieving a 98.4% accuracy with a learning rate of 0.0001 across 60 epochs. Similarly, Muhammad Haris Irham et al. [3] explored real-time emotion detection using CNN, specifically MobileNetV2, on the FER2013 dataset containing over 35,000 images. Their model achieved 57% accuracy in training and 51% in validation, demonstrating the challenges of real-time accuracy in facial emotion recognition. Convolutional Neural Networks (CNNs) are a specialized type of neural network designed to process grid-like data, such as images, by leveraging the spatial structure within data. CNNs are particularly effective in visual recognition tasks because they automatically detect relevant features at different levels of abstraction through a series of transformations and learn hierarchies of complex patterns [4], [5], [6], [7], [8], [9].

Yoshiven Boer et al. [10] compared several Deep Convolutional Neural Network (DCNN) models, including VGG16, VGG19, ResNet50, ResNet101, Xception, and InceptionV3. Among these, VGG19 achieved the highest accuracy at 65%, suggesting that model performance varies significantly based on dataset quality, model complexity, and hyperparameter tuning. In 2023, Muthamilselvan et al. [11] proposed a method combining CNN with a binary whale optimization algorithm (OFELBW) for emotion recognition, tested on multiple datasets, achieving accuracies of 98.35% with CK+ and 99.42% with FERG datasets.

Jain et al. [12] proposed a DNN-based facial emotion recognition model that reached 95.23% accuracy on JAFFE and 93.24% on CK+, which has been applied in various fields such as human-computer interaction and social robotics. Additionally, Mehendale [13] designed an emotion detection system achieving 96% accuracy using CNN on a diverse dataset of 10,000 images from 154 individuals. Deng et al. [4] advanced emotion recognition in video by combining micro and macro features in a dual-stream recurrent network named MIMAMO Net, achieving competitive accuracy rates on datasets such as OMG and Aff-Wild.

Luigi Bibbo et al. [2] developed an emotion recognition model for mental-health monitoring using a Self-Normalizing Neural Network (SNN) equipped with cascade and ensemble layers. Their approach relied on a highly controlled experimental setting with a consistent learning rate of 0.0001 over 60 epochs, resulting in a 98.4% accuracy. In a different context, Muhammad Haris Irham et al. [3] implemented MobileNetV2 for real-time emotion recognition on the FER2013 dataset, which contains more than 35,000 images. Despite using a lightweight architecture optimized for inference speed, the model achieved only 57% training accuracy and 51% validation accuracy, highlighting the inherent challenges of real-time emotion classification on an imbalanced and noisy dataset such as FER2013 [4], [5], [6], [7], [8], [9].

Lioga et al. [3] focused on recognizing five basic emotions: anger, happiness, neutral, sadness, and surprise—using CNNs trained on the Indonesian Mixed Emotion Dataset (IMED). Their work demonstrated that appropriate preprocessing strategies, including cropping, resizing, grayscale conversion, and extensive data augmentation, can substantially improve performance, yielding a real-time accuracy of 93.93%. Meanwhile, Ali et al. [14] combined the Viola–Jones face detection method with CNN-based feature extraction and SVM classification, achieving 97% accuracy on a dataset of more than 34,000 images. This hybrid approach illustrates how integrating classical detection algorithms with modern learning techniques can enhance robustness in controlled environments.

Yoshiven Boer et al. [10] compared several Deep CNN architectures, including VGG16, VGG19, ResNet50, ResNet101, Xception, and InceptionV3 across a unified dataset to examine how depth, parameter size, and architectural design influence performance. VGG19 achieved the highest accuracy at 65%, suggesting that modern architectures do not always outperform older models when dataset quality or preprocessing differ. Muthamilselvan et al. [11] further highlighted the role of optimization strategies by integrating CNN with a binary whale optimization algorithm (OFELBW), achieving 98.35% accuracy on CK+ and 99.42% on FERG, both highly curated datasets with consistent illumination and pose conditions.

Other studies also underscore the importance of dataset characteristics and experimental context. Jain et al. [12] reported 95.23% accuracy on JAFFE and 93.24% on CK+ using a deep neural network, benefiting from small, high-quality datasets with well-posed facial expressions. Mehendale [13] achieved 96% accuracy using CNN on a diverse dataset of 10,000 images from 154 individuals, while Deng et al. [4] advanced video-based emotion recognition using a dual-stream recurrent architecture (MIMAMO Net) that integrates micro and macro facial features, showing competitive results on dynamic datasets such as OMG and Aff-Wild.

Collectively, these studies reveal that variations in methodology, dataset scale and quality, preprocessing pipelines, and evaluation scenarios substantially influence reported performance. Therefore, direct comparison of accuracy values alone may be misleading without considering the broader experiment.

Furthermore, this research will design a proposed architecture for a customer satisfaction survey system with facial image inputs from prospective Hajj pilgrims at PLHUT offices at the district and city levels. Consequently, the Ministry of Religious Affairs can monitor the survey results online at any time. The integration of this technology is expected to provide more accurate and real-time information regarding the satisfaction levels of Hajj pilgrims, thereby enhancing the overall quality of services.

2. METHODOLOGY

This study employed a Convolutional Neural Network (CNN) model to build an emotion detection system using facial images, with the goal of assessing Hajj pilgrims' satisfaction. The structured CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology guided each phase, ensuring a robust, data-driven approach for developing, training, and evaluating the model. The six phases within CRISP-DM included Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment, which together supported the model's ability to achieve practical, high-quality decision-making.

This study employed a Convolutional Neural Network (CNN) model to build an emotion detection system using facial images, with the goal of assessing Hajj pilgrims' satisfaction. The development process followed the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, ensuring a structured and data-driven workflow across its six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. While the related literature includes a range of more advanced architectures and optimization strategies, this study intentionally adopts a simpler CNN model to accommodate the operational characteristics of Hajj service environments, where real-time processing, limited computational resources, and ease of deployment are critical. The methodological contribution of the study lies in demonstrating how a lightweight CNN, supported by a systematic CRISP-DM process, can serve as a practical baseline for emotion-based satisfaction assessment. The work also offers an experimental contribution by providing one of the first evaluations of facial-expression-driven satisfaction inference in a real-world Hajj operational setting, which differs notably from the controlled laboratory datasets commonly used in previous research.

2.1. Business Understanding Phase

The research began by defining its core objective: to measure the satisfaction of Hajj pilgrims through facial emotion analysis. The initial problem identification sought to determine if current service offerings met pilgrims' expectations by analyzing their facial expressions. Literature review and preliminary interviews with service staff provided foundational insights into the needs and expectations for system implementation.

2.2. Business Understanding Phase

The primary emotions of focus were happiness and sadness, representing positive and negative emotional states. These were chosen due to their universal relevance in assessing general satisfaction and dissatisfaction.

2.3. Data Preparation

The Data Preparation phase focused on preparing the dataset for training and testing to ensure the model's performance and reliability. The dataset consisted of 35,887 images representing seven emotion categories: angry, disgust, fear, happy, neutral, sad, and surprise. This phase involved several key steps, including data splitting, image pre-processing, and data augmentation.

The Data Preparation phase focused on preparing the dataset for training and testing to ensure the model's performance and reliability. The dataset consisted of 35,887 images representing seven emotion categories: angry, disgust, fear, happy, neutral, sad, and surprise. Although the training stage utilized a generic 7-class emotion dataset, the deployment stage of this study required only two satisfaction-related labels, namely *Satisfied* and *Unsatisfied* as shown in **Table 1**.

Ground-truth satisfaction labels for the deployment images were obtained through manual annotation by two independent human raters who observed the captured facial expressions of pilgrims during Hajj service activities. When disagreements occurred, a third rater was consulted to resolve the label, ensuring accuracy and reducing subjectivity. This annotation process provided a consistent and reliable basis for evaluating the model's performance in real-world conditions.

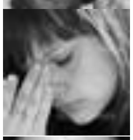
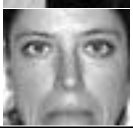
Table 1. Mapping Expression

Expression	Satisfaction Category
Happy	Satisfied
Angry	Unsatisfied
Sad	Unsatisfied
Disgust	Unsatisfied
Neutral	Satisfied
Surprise	Unsatisfied
Fear	Unsatisfied

2.3.1. Data Splitting

The dataset of 35,887 images was divided into 28,709 training images and 7,178 testing images across seven emotion categories (angry, disgust, fear, happy, neutral, sad, and surprise), as shown in **Table 2.** For the real-world deployment phase involving Hajj pilgrims, privacy and data protection procedures were strictly observed. All participants provided informed consent prior to image capture, including consent for their facial expressions to be used for satisfaction inference within the scope of this study. The captured images were processed only for the purposes of real-time inference and were stored temporarily during evaluation. To ensure anonymity, no personal identifiers such as names, IDs, or demographic details were linked to the images. Furthermore, the application did not retain biometric data after analysis; all captured images were deleted immediately following inference, and no facial embeddings or biometric templates were stored. These procedures ensured compliance with ethical research practices and protected the privacy of all participating pilgrims.

Table 2. Data Expression

Expression	Image	Expression	Image
Happy		Fear	
Angry		Neutral	
Sad		Surprise	
Disgust			

2.3.2. Image Pre-Processing

This phase involved contrast enhancement to clarify image features and reduce noise. Images were resized to 48x48 pixels and converted to numpy arrays for model input.

2.3.3. Data Augmentation

Training data was expanded with transformations like rotation, cropping, shifting, and lighting adjustments. Using ImageDataGenerator from Keras, these augmentations enhanced model accuracy and minimized overfitting, with a batch size of 256, as shown in **Algorithm 1.**

Algorithm 1
Image Augmentation

```
from keras.preprocessing.image import ImageDataGenerator

  # number of images to feed into the NN for every batch batch_size = 256
  datagen_train = ImageDataGenerator() datagen_validation =
  ImageDataGenerator() train_generator =
  datagen_train.flow_from_directory(
                         base_path + "train",
                         target_size=(pic_size,pic_size),
                         color_mode="grayscale",
                         batch_size=batch_size, class_mode='categorical', shuffle=True)

  validation_generator = datagen_validation.flow_from_directory( base_path
                         + "validation",
  target_size=(pic_size,pic_size),
  color_mode="grayscale", batch_size=batch_size, class_mode='categorical',
                         shuffle=False)
```

2.4. Modeling

The CNN model architecture included layers of Conv2D, batch normalization, ReLU activation, max pooling, and dropout for regularization. Four convolutional layers with increasing filters (64, 128, and 512) processed the images, which were then flattened and passed through two dense layers with batch normalization, ReLU activation, and dropout. The output layer employed softmax for class probabilities, optimized with the Adam optimizer and a categorical cross-entropy loss function. Algorithm 2 is the source code for the CNN implementation.

Algorithm 2
CNN Implementation

```
from keras.preprocessing.image import ImageDataGenerator # number of
images to feed into the NN for every batch batch_size = 256

datagen_train = ImageDataGenerator() datagen_validation = ImageDataGenerator()
train_generator = datagen_train.flow_from_directory(base_path + "train",
                                        target_size=(pic_size,pic_size),
                                        color_mode="grayscale",
                                        batch_size=batch_size,
                                        class_mode='categorical', shuffle=True)
validation_generator = datagen_validation.flow_from_directory(base_path + "validation",
                                        target_size=(pic_size,pic_size),
                                        color_mode="grayscale",
                                        batch_size=batch_size,
                                        class_mode='categorical', shuffle=False)
```

3. RESULT AND DISCUSSION

After constructing the CNN architecture, the training process was carried out using the processed data. Several training scenarios were implemented using different batch sizes (32, 64, 128, and 256).

3.1. Evaluation

The model was evaluated using accuracy, precision, recall, and F1-score to provide a comprehensive assessment of its classification capability. Four configurations with different batch sizes (32, 64, 128, and 256) were tested under identical training conditions (50 epochs), allowing us to observe how batch size influences generalization performance.

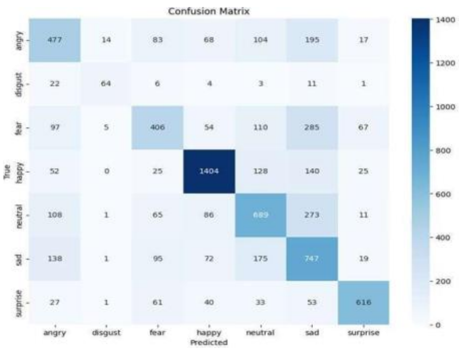
A direct baseline comparison within the same dataset could not be established for two methodological reasons. First, no domain-specific baseline model exists for Hajj-related satisfaction assessment, as this study represents one of the earliest attempts to apply CNN-based facial–expression analysis in this operational context. Second, the real-world deployment dataset could not be used to construct a baseline because the images captured during deployment were used exclusively for real-time inference and were deleted afterward to comply with privacy and ethical requirements. As a result, the dataset available for training consisted solely of generic facial-expression images (seven classes), while the deployment stage used an ethically restricted two-class mapping (Satisfied / Unsatisfied).

To address this limitation, the lightweight CNN used in this study is positioned as a practical operational baseline, reflecting constraints in Hajj service environments where real-time inference, low latency, and limited computational resources are critical. In addition, external baselines from published literature provide contextual comparison. Standard CNN models on similar facial-expression datasets typically achieve 60–65% accuracy, while lightweight architectures such as MobileNetV2 report 51–57% accuracy on FER2013 under uncontrolled conditions. The performance of our model—achieving up to 63% accuracy during testing and 78% accuracy during real-world deployment—falls within this expected range and demonstrates robustness when applied to highly variable, culturally specific real-world conditions.

By situating the model's performance relative to established FER benchmarks and acknowledging the constraints inherent to this early-stage operational deployment, the evaluation clarifies the role of this study as an initial, domain-specific baseline that future research can extend through improved datasets, alternative architectures, and larger-scale deployments.

3.2.1. Test 1 (Batch Size 32)

Classified 4,403 out of 7,178 images correctly, yielding an accuracy of 61%, precision of 63%, recall of 59%, and F1-score of 61% as shown in Figure. 1



	precision	recall	f1-score	support
angry	0.52	0.50	0.51	958
disgust	0.74	0.58	0.65	111
fear	0.55	0.40	0.46	1024
happy	0.81	0.79	0.80	1774
neutral	0.55	0.56	0.56	1233
sad	0.44	0.60	0.51	1247
surprise	0.81	0.74	0.78	831
accuracy			0.61	7178
macro avg	0.63	0.59	0.61	7178
weighted avg	0.63	0.61	0.62	7178

Figure 1. Confusion Matrix and Precision Score First Testing Result Batch 32

3.2.2. Test 2 (Batch Size 64)

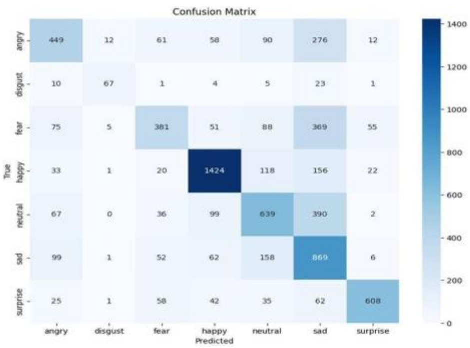
Classified 4,532 images correctly, resulting in an accuracy of 63%, precision of 66%, recall of 60%, and F1-score of 62% as shown in the Figure. 2.

	precision	recall	f1-score	support
angry	0.58	0.49	0.54	958
disgust	0.88	0.54	0.67	111
fear	0.56	0.45	0.50	1024
happy	0.77	0.84	0.81	1774
neutral	0.56	0.59	0.57	1233
sad	0.49	0.54	0.51	1247
surprise	0.75	0.78	0.76	831
accuracy			0.63	7178
macro avg	0.66	0.60	0.62	7178
weighted avg	0.63	0.63	0.63	7178

Figure 2. Confusion Matrix and Precision Score Second Testing Result Batch 64

3.2.3. Test 3 (Batch Size 128)

Achieved an accuracy of 62%, with 4,437 correctly classified images, and scores of 66% precision, 60% recall, and an F1-score of 62% as shown in the figure. 3.

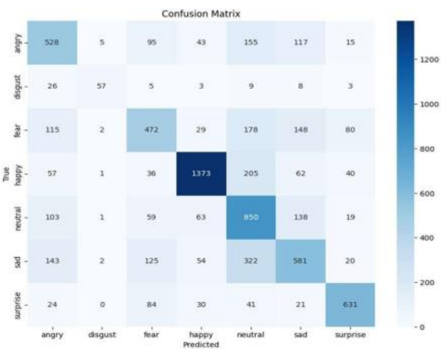


	precision	recall	f1-score	support
angry	0.59	0.47	0.52	958
disgust	0.77	0.60	0.68	111
fear	0.63	0.37	0.47	1024
happy	0.82	0.80	0.81	1774
neutral	0.56	0.52	0.54	1233
sad	0.41	0.70	0.51	1247
surprise	0.86	0.73	0.79	831
accuracy			0.62	7178
macro avg	0.66	0.60	0.62	7178
weighted avg	0.65	0.62	0.62	7178

Figure 3. Confusion Matrix and Precision Score Third Testing Result Batch 128

3.2.4. Test 4 (Batch Size 256)

The model reached 63% accuracy, classifying 4,492 images correctly, with a precision at 65%, a recall at 60%, and an F1-score of 62%, as shown in Figure. 4.



	precision	recall	f1-score	support
angry	0.53	0.55	0.54	958
disgust	0.84	0.51	0.64	111
fear	0.54	0.46	0.50	1024
happy	0.86	0.77	0.82	1774
neutral	0.48	0.69	0.57	1233
sad	0.54	0.47	0.50	1247
surprise	0.78	0.76	0.77	831
accuracy			0.63	7178
macro avg	0.65	0.60	0.62	7178
weighted avg	0.64	0.63	0.63	7178

Figure 4. Confusion Matrix and Precision Score Last Testing Result Batch 256

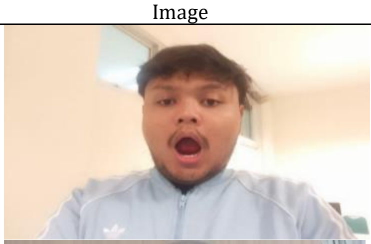
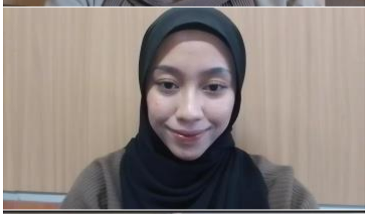
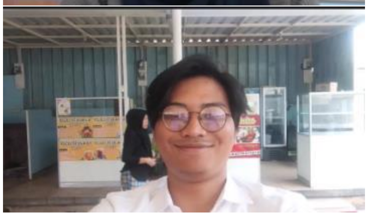
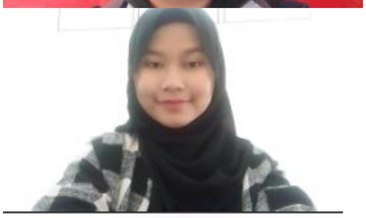
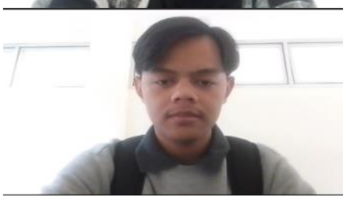
3.2. Deployment

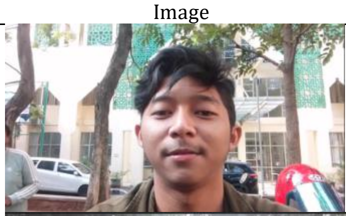
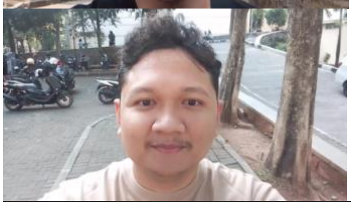
The final model was deployed into an Android application developed in Dart, with the model converted to .h5 format for efficient mobile processing. The app enables real-time classification using a REST API to interface between the Android client and server. When a user captures an image via the app, the image is sent to the server, where facial detection and emotion analysis occur. Results are stored in

p-ISSN: 2528-1682
e-ISSN: 2527-9165

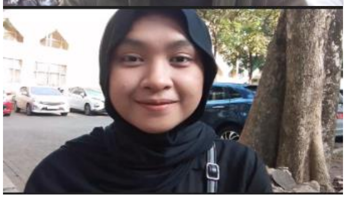
a database, capturing details like branch location and timestamp, and visualized on a web dashboard for service feedback.

Table 3. Comparison of Manual Testing Results and Application Testing Results

No	Image	Application Result	Manual Result
1		Unsatisfied	Unsatisfied
2		Unsatisfied	Unsatisfied
3		Satisfied	Satisfied
4		Satisfied	Satisfied
5		Satisfied	Satisfied
6		Satisfied	Satisfied
7		Satisfied	Unsatisfied

No	Image	Application Result	Manual Result
8		Satisfied	Unsatisfied
9		Satisfied	Satisfied
10		Satisfied	Satisfied
11		Satisfied	Satisfied
12		Satisfied	Satisfied
13		Satisfied	Satisfied
14		Satisfied	Satisfied
15		Satisfied	Satisfied

p-ISSN: 2528-1682
e-ISSN: 2527-9165

No	Image	Application Result	Manual Result
16		Satisfied	Unsatisfied
17		Satisfied	Unsatisfied
18		Satisfied	Satisfied
19		Satisfied	Satisfied

The results of this study demonstrate the potential of using Convolutional Neural Networks (CNN) for detecting facial expressions to measure service satisfaction among Hajj pilgrims. The accuracy achieved during real-world deployment, reaching 78%, indicates that this approach can provide valuable, non-intrusive insights into pilgrims' emotional states, which directly reflect their satisfaction with the services provided. This is particularly significant in the context of Hajj services, where pilgrims may not always articulate their feelings verbally due to cultural or situational constraints. The findings highlight the practicality of leveraging machine learning techniques for real-time monitoring and service evaluation.

When compared to prior research, this study's results provide significant insights into the application of Convolutional Neural Networks (CNN) for real-world facial expression recognition, especially in service satisfaction contexts. Charvi Jain et al. [30] achieved 85% accuracy using a hybrid CNN-Support Vector Machine (SVM) model under controlled conditions with minimal variability in lighting and facial poses. Their study underscores the potential of combining machine learning techniques for facial expression recognition in ideal conditions. However, their controlled environment differs significantly from the dynamic and variable settings found in real-world applications such as the Hajj pilgrimage, where lighting and facial expressions can vary widely.

In contrast, Muhammad Haris Irha et al. [32] reported a lower accuracy of 57% using MobileNetV2, which was optimized for lightweight applications suitable for mobile devices with constrained computational resources. This study's focus on efficiency over accuracy presents a trade-off that can impact performance when detecting subtle emotions, especially in complex, real-world scenarios like Hajj, where emotional states can be varied and nuanced. This study's accuracy of 75% situates it between these two extremes, reflecting the unique challenges inherent in deploying emotion detection models in culturally specific environments. Unlike previous studies that primarily dealt with general settings such as retail or security, this research focuses on a unique cultural and spiritual setting where emotional states such as exhaustion, spiritual focus, and mixed feelings are prevalent. These emotional states are often absent in conventional datasets, which typically focus on more universally recognized emotions like happiness or anger. Consequently, these cultural and emotional nuances likely contributed to the observed accuracy, highlighting the importance of tailoring both datasets and models to the specific domain of application.

Moreover, facial expression recognition in high-variability environments, such as large religious gatherings, presents challenges that differ from the more controlled settings commonly studied. For instance, recent studies in 2024, such as the work by Geetha et.al [38] have continued to focus on emotion recognition in controlled environments, achieving high accuracy levels under ideal conditions. However, as this study illustrates, real-world deployments like Hajj demand adaptations to account for cultural context and emotional nuance, which can significantly impact model performance. Additionally, unlike traditional datasets that focus on expressions like happiness or anger, this study addressed more complex emotional states, which likely influenced the model's overall accuracy.

This study's findings underscore the practical utility of CNNs in service satisfaction monitoring, particularly within the context of Hajj. The ability to detect facial expressions in real-time offers numerous advantages, including the reduction of manual feedback collection, the enhancement of timely service evaluations, and the provision of immediate insights for service improvement. By enabling real-time assessment of emotional states, the model contributes to a more responsive service environment that can be adjusted based on immediate feedback from pilgrims, without relying on verbal responses that may not always be forthcoming due to cultural factors.

Furthermore, the deployment in such a high-variability setting demonstrates the robustness of CNN models in diverse and dynamic conditions. This successful application of facial expression recognition to service satisfaction monitoring in Hajj services paves the way for broader adoption in similar contexts, such as in other religious, cultural, or large-scale events where capturing pilgrims' emotional states can enhance service quality and responsiveness.

While the model demonstrated strong performance with a 78% accuracy during deployment, further improvements are necessary to refine its robustness and accuracy. Future research should focus on several key areas. First, expanding datasets to include more diverse facial expressions specific to the cultural and emotional contexts of Hajj pilgrims could improve the model's ability to generalize across different populations. The current dataset may not fully capture the subtle and complex emotions unique to this setting, such as mixed feelings of exhaustion and spiritual focus. Additionally, multi-modal emotion detection could significantly enhance the accuracy of emotion recognition. By integrating other modalities such as audio inputs or body language recognition, the system could gain a more comprehensive understanding of a pilgrim's emotional state, leading to more precise service satisfaction assessments. Lastly, fine-tuning the CNN architecture or exploring hybrid models could further improve accuracy, especially for detecting more nuanced emotions like frustration or mild contentment. The use of more advanced and complex architectures could help the model better distinguish between subtle emotional expressions that may not be easily categorized into distinct emotions like "happy" or "sad." In conclusion, while this study validates the application of CNN for facial expression detection in service satisfaction contexts, there is a need for ongoing refinement. By improving datasets, incorporating multi-modal inputs, and optimizing CNN architectures, future research can enhance both the accuracy and broader applicability of these models across various real-world service environments.

The performance of the CNN model was evaluated using four key metrics including accuracy, precision, recall, and F1 score. These metrics were selected to provide a comprehensive assessment of the model's classification ability by considering both the correctness of the predictions and the balance between false positives and false negatives. The evaluation process was carried out using four different test configurations in which each configuration applied a different batch size, specifically 32, 64, 128, and 256, while the number of training epochs was kept constant at 50. This setup allowed us to analyze how varying the batch size influenced the model's ability to generalize to unseen data. Through a systematic comparison of the results obtained from each configuration, we aimed to determine the most effective batch size that offers an optimal balance between training efficiency and classification performance.

4. CONCLUSION

This study successfully implemented a Convolutional Neural Network (CNN) model to detect facial expressions as a means of measuring service satisfaction among Hajj pilgrims. The model achieved a real-world deployment accuracy of 78%, demonstrating its potential as a non-intrusive tool for real-time service evaluations. By analyzing facial expressions, the system provides immediate feedback on pilgrim satisfaction without requiring verbal input, making it especially useful in high-volume, culturally sensitive environments like Hajj service centers.

The results highlight the value of emotion detection through facial recognition in enhancing customer satisfaction surveys, offering insights that can lead to timely service improvements. However, the model's limitations, particularly in recognizing nuanced emotional states, indicate the need for further optimization. Future work should focus on expanding the dataset, refining the CNN architecture, and potentially integrating multi-modal inputs such as voice or body language to improve accuracy and applicability. In conclusion, while the current model provides promising results, continuous development is necessary to ensure its robustness and adaptability to a broader range of emotions and service contexts, thereby improving the overall quality and efficiency of service delivery during the Hajj pilgrimage.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the Lembaga Penelitian dan Pengabdian kepada Masyarakat (LP2M) UIN Sunan Gunung Djati Bandung for sponsoring the publication of this article. This support has been invaluable in enabling us to share our findings with the wider academic community.

REFERENCES

[1] Direktorat Jenderal Penyelenggaraan Haji dan Umrah, “Laporan Kinerja Tahun 2023,” 2023, *Kementerian Agama Republik Indonesia*

[2] Afrizal, “Ada 8 Bentuk Layanan Yang Diberikan Pemerintah Kepada Jamaah Haji.” [Online]. 2024. Available: https://karimun.kemenag.go.id/main/berita/detail/drs-h-afrizal-mmpd–ada-8-bentuk-layanan-yang-diberikan-pemerintah-kepada-jamaah-haji

[3] TribunNews, “BPS: Sangat Memuaskan, Indeks Kepuasan Jemaah Haji 2024 Capai 88,20” [Online]. 2024. Available: https://makassar.tribunnews.com/2024/09/20/bps-sangat-memuaskan-indeks-kepuasan-jemaah-haji-2024-capai-8820

[4] Zahdi, “Manajemen Haji dan Umrah (Analisis Pencapaian Kepuasan Layanan Haji dan Umrah Di Kota Bandar Lampung,” Disertasi, [Online]. 2021. Available: http://repository.radenintan.ac.id/id/eprint/15283

[5] Esposito, A., Esposito, A. M., & Vogel, C. (2015). Needs and challenges in human computer interaction for processing social emotional information. *Pattern Recognition Letters*, *66*, 41-51.

[6] Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest*, *20*(1), 1-68.

[7] Carroll, J. M., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of personality and social psychology*, *70*(2), 205.

[8] Ekman, P., Friesen, W. V., & Ellsworth, P. (2013). *Emotion in the human face: Guidelines for research and an integration of findings* (Vol. 11). Elsevier.

[9] Etcoff, N. L. (1986). The neuropsychology of emotional expression. In *Advances in clinical neuropsychology* (pp. 127-179). Boston, MA: Springer US.

[10] Stets, J. E. (2003). Emotions and sentiments. In *Handbook of social psychology* (pp. 309-335). Boston, MA: Springer Us.

[11] Izard, C. E. (1990). Facial expressions and the regulation of emotions. *Journal of personality and social psychology*, *58*(3), 487.

[12] Morris, M. W., & Keltner, D. (2000). How emotions work: The social functions of emotional expression in negotiations. *Research in organizational behavior*, *22*, 1-50.

[13] Shariff, A. F., & Tracy, J. L. (2011). What are emotion expressions for? *Current Directions in Psychological Science*, *20*(6), 395-399.

[14] González-Rodríguez, M. R., Díaz-Fernández, M. C., & Gómez, C. P. (2020). Facial-expression recognition: An emergent approach to the measurement of tourist satisfaction through emotions. *Telematics and Informatics*, *51*, 101404.

[15] Cahyaningtyas, C., Gudiato, C., & Sari, M. (2025). Deteksi Ekspresi Wajah Manusia Menggunakan Metode Convolutional Neural Network. *Jurnal Fasilkom*, *15*(1), 138-145.

[16] Wulansari, T. (2024). *Deteksi Ekspresi Wajah Untuk Mengukur Tingkat Kepuasan Pelanggan Dengan Menerapkan Metode Convolutional Neural Network* (Doctoral dissertation, Politeknik Negeri Jember).

[17] Tran, T. K., Vo, Q. N., Hong, X., Li, X., & Zhao, G. (2021). Micro-expression spotting: A new benchmark. *Neurocomputing*, *443*,

356-368.

[18] Muna, N. (2017). *Klasifikasi Micro-Expressions Wajah Berbasis Subpixel Subtle Motion Estimation Menggunakan Multilayer Perceptron* (Doctoral dissertation, Institut Teknologi Sepuluh Nopember).

[19] N., & Aithal, P. S. (2022). Real-time customer satisfaction analysis using facial expressions and head pose estimation. *International Journal of Applied Engineering and Management Letters (IJAEML)*, *6*(1), 301-312.

[20] Bouzakraoui, M. S., Sadiq, A., & Alaoui, A. Y. (2020). Customer satisfaction recognition based on facial expression and machine learning techniques. *Advances in Science, Technology and Engineering Systems Journal*, *5*(4), 594-594.

[21] Saxena, A., Khanna, A., & Gupta, D. (2020). Emotion recognition and detection methods: A comprehensive survey. *Journal of Artificial Intelligence and Systems*, *2*(1), 53-79.

[22] Bouzakraoui, M. S., Sadiq, A., & Alaoui, A. Y. (2020). Customer satisfaction recognition based on facial expression and machine learning techniques. *Advances in Science, Technology and Engineering Systems Journal*, *5*(4), 594-594.

[23] Harianto, D., Filbert, S., Cahyakusuma, A. B., & Zakiyyah, A. Y. (2024, July). Analyzing customer satisfaction through face emotion recognition: A comparative study of convolutional neural networks (cnn) and long short term memory (lstm). In *2024 10th International Conference on Smart Computing and Communication (ICSCC)* (pp. 50-54). IEEE.

[24] Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences*, *2*(3), 446.

[25] Deng, D., Chen, Z., Zhou, Y., & Shi, B. (2020, April). Mimamo net: Integrating micro-and macro-motion for video emotion recognition. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 03, pp. 2621-2628).

[26] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1, No. 2). Cambridge: MIT press.

[27] Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, *9*(4), 611-629.

[28] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

[29] O'shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

[30] Sharma, N., & Jain, C. (2019, March). Characterization of facial expression using deep neural networks. In *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)* (pp. 492-495). IEEE.

[31] Bibbo', L., Cotroneo, F., & Vellasco, M. (2023). Emotional health detection in HAR: new approach using ensemble SNN. *Applied Sciences*, *13*(5), 3259.

[32] Irham, M. H., Mubarak, A., & Salmin, M. (2023). Detecting Face Expressions in Real-Time Using Convolutional Neural Network (CNN) Algorithm. *Technium*, *17*.

[33] Seandrio, A. L., Pratomo, A. H., & Florestiyanto, M. Y. (2021). Implementation of convolutional neural network (CNN) in facial expression recognition. *Telematika: Jurnal Informatika dan Teknologi Informasi*, *18*(2), 211-221.

[34] Raza, A. R. A., Ishaq, M. I. M., Mehmood, B. M. B., Ehsan, R. E. R., & Iram, A. I. A. (2021). EFFECTS OF DOMESTIC VIOLENCE ON REPRODUCTIVE HEALTH: AN EVIDENCE FROM PAKISTAN DEMOGRAPHIC HEALTH SURVEY. *Journal Of Akhtar Saeed Medical & Dental College*, *3*(04).

[35] Boer, Y., Valencia, L., & Prasetyo, S. Y. (2023). Exploring the potential of pre-trained dcnn models for facial emotion detection: A comparative analysis. In *E3S Web of Conferences* (Vol. 426, p. 01049). EDP Sciences.

[36] Muthamilselvan, T., Brindha, K., Senthilkumar, S., Saransh, Chatterjee, J. M., & Hu, Y. C. (2023). Optimized face-emotion learning using convolutional neural network and binary whale optimization. *Multimedia Tools and Applications*, *82*(13), 19945-19968.

[37] Jain, D. K., Shamsolmoali, P., & Sehdev, P. (2019). Extended deep neural network for facial emotion recognition. *Pattern Recognition Letters*, *120*, 69-74.

[38] Geetha, A. V., Mala, T., Priyanka, D., & Uma, E. (2024). Multimodal emotion recognition with deep learning: advancements, challenges, and future directions. *Information Fusion*, *105*, 102218.