Journal of Applied Informatics and Computing (JAIC)
Vol.8, No.2, December 2024, pp. 257~264
e-ISSN: 2548-6861

Facial Expression Recognition using Convolutional Neural Networks with Transfer Learning Resnet-50

Annisa Ayu Istiqomah [1]*, Christy Atika Sari [2], Ajib Susanto [3], Eko Hari Rachmawanto [4]
Study Program in Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia
annisaayuistiqomah@gmail.com*[1], atika.sari@dsn.dinus.ac.id [2], ajib.susanto@dsn.dinus.ac.id [3], eko.hari@dsn.dinus.ac.id [4]

Article Info

Article history:

Received 2024-08-16
Revised 2024-08-26
Accepted 2024-08-29

Keyword:

Convolutional Neural Networks,
Confusion Matrix,
Facial Expression Recognition,
Image Classification,
Transfer Learning Resnet-50.

ABSTRACT

Facial expression recognition is important for many applications, including sentiment analysis, human-computer interaction, and interactive systems in areas such as security, healthcare, and entertainment. However, this task is fraught with challenges, mainly due to large differences in lighting conditions, viewing angles, and differences in individual eye structures. These factors can drastically affect the appearance of facial expressions, making it difficult for traditional recognition systems to consistently and accurately identify emotions. This study addresses these issues by employing transfer learning with ResNet-50 and effective pre-processing techniques, specifically image enhancement and resizing to 48 x 48 pixels. The dataset consists of grayscale images categorized into seven classes: anger, contempt, disgust, fear, happiness, sadness, and surprise, with a total of 680 samples. The dataset was divided so that 80% was allocated for training and 20% for testing to ensure robust model evaluation. The main contribution of this research is the demonstration that using ResNet-50 for transfer learning significantly improves the accuracy and reliability of facial expression recognition compared to traditional methods. The results show that the model achieved an exceptional performance level, with an accuracy of 99.49%, precision of 99.49%, recall of 99.71%, and an F1-score of 99.60%. These results highlight the superior performance of the ResNet-50 model in handling the challenges of varying lighting conditions and angles. Future research will focus on implementing real-time facial recognition systems and exploring other advanced transfer learning models such as GoogLeNet and VGGNet to further enhance accuracy and operational efficiency.

I. Introduction

Computer Vision is an area of rapid expansion and development of artificial intelligence that allows machines to decode and comprehend visual information from the environment, similar to how humans use their vision [1], [2], [3], [4]. The course includes the development of algorithms and models that can process images and videos to extract meaningful information, such as voice detection, image segmentation, and scene recognition [5]. Applications of computer vision cover many industries, from medical care and self-driving cars to safety and entertainment. The technology extensively utilizes both conventional image processing methods and contemporary machine learning approaches, with a particular emphasis on deep learning, to achieve high levels of accuracy and performance [6], [7], [8]. The integration of computer vision into various systems has revolutionized how machines interact with and interpret their surroundings, allowing for more sophisticated automation and decision-making processes [9]. As computer vision continues to evolve, it is increasingly capable of handling complex tasks such as recognizing human emotions, actions, and even generating visual content, pushing the boundaries of what machines can perceive and understand [10], [11]. This advancement has laid the foundation for specialized applications like facial expression recognition, which is a sophisticated and significant field of study within artificial intelligence and computer vision, focusing on the ability of systems to recognize and interpret human emotional expressions in facial images [12].

Facial expression recognition is an essential and intricate domain of research in artificial intelligence and computer vision, focusing on the ability of systems to recognize and interpret human emotional expressions in facial images [12]. In its early development, traditional methods such as feature-based computation and pixel-based methods were used to capture and classify facial expressions [13]. However, these methods are generally ineffective in dealing with existing differences, such as differences in lighting, viewing angles, and individual differences in facial features [14]. These challenges significantly hinder the reliability and robustness of facial expression recognition systems, leading to inaccuracies in emotion detection, particularly in diverse and uncontrolled environments. Moreover, the static nature of traditional methods limits their adaptability to dynamic changes in facial expressions over time, further complicating their practical application. As a result, there is a pressing need for more advanced techniques, such as those offered by machine learning and deep learning, to address these issues by enhancing the system’s ability to generalize across various conditions and individual differences, ultimately improving the accuracy and applicability of facial expression recognition technologies.

To overcome these limitations, machine learning and deep learning have become the dominant approaches [15]. Machine learning enables models to learn from data and discern patterns in facial expressions. Especially, deep learning approaches, including convolutional neural networks, have achieved positive results in overcoming this challenge. [16], [17]. CNNs can automatically extract useful features from raw data and create complex patterns [18], [19], [20], thereby increasing the accuracy of facial expression recognition. However, achieving good results requires large and diverse data sets, as well as high computing power to train complex models [21]. Furthermore, developing algorithms that can adapt to dynamic changes in human facial expression is also a challenge that needs to be addressed in this research.

Numerous researchers have employed CNN for facial expression recognition, utilizing various approaches. CNN has become a popular choice because of its exceptional ability to derive visual features from images and recognize intricate patterns, which are crucial in detecting facial expressions. The approaches adopted by researchers vary, ranging from combining CNN with other models to optimizing CNN architectures to enhance accuracy and efficiency in facial expression recognition. This literature review will examine several recent studies that demonstrate how different combinations and adaptations of CNN have been applied to achieve varying results in facial expression recognition. Researcher [22] conducted a study that employed a hybrid approach by integrating Convolutional Neural Networks (CNN) with Convolutional Long Short-Term Memory (ConvLSTM) for recognizing facial expressions In this approach, CNN was used to derive spatial characteristics from facial images., while ConvLSTM was employed to capture temporal information or dynamics from image sequences. The conclusions of this study indicated that this combination achieved a very high accuracy rate of 95.10%. The effectiveness of this model emphasizes its notable potential for capturing the complexities of facial expressions by leveraging both spatial and temporal features simultaneously. Researcher [23] conducted a hybrid approach by integrating K-Nearest Neighbors (KNN) with CNN for facial expression recognition. In this approach, CNN was used for feature extraction, while KNN served as the classification method. However, the results of this study indicated that this combination yielded a lower accuracy rate of 75.26%.

Researcher [24] conducted a hybrid approach involving Maximum Boosted CNN and Long Short-Term Memory (LSTM) was utilized. This combination was designed to enhance performance by capturing critical features from facial images as well as the temporal dynamics associated with facial expressions. The study's findings indicated that the model reached an average accuracy of 83.125%. While not as high as some other models, these results demonstrate that the approach combining maximum boosting in CNN with LSTM can offer reasonably good performance in facial expression recognition.

Based on the above literature review, this study proposes the use of CNN combined with transfer learning using the Resnet-50 architecture for facial expression recognition. Transfer learning enables the model to utilize insights gained from extensive datasets, which can enhance accuracy and efficiency in the training process. By utilizing Resnet-50 as transfer learning, known for its capability to handle complex visual data, this research aims to realize a substantial enhancement in the accuracy of facial expression recognition, compared to previous approaches.

II. METHOD

The proposed method begins with data collection, where a diverse set of facial images is gathered to ensure robust model performance. This is followed by a pre-processing stage, where images are resized and enhanced to improve feature extraction. Edge detection techniques are then applied to emphasize the boundaries and contours within the images, which are crucial for identifying key facial regions. Subsequently, region detection is performed to isolate areas of interest, such as the eyes and mouth, which are integral to facial expression recognition. The ResNet-50 architecture is then initialized, with its layers designed to extract advanced features from these detected regions. The model uses the processed data for training, which helps it identify the detailed patterns associated with different facial expressions. Finally, the method is tested on single predicted images to assess its real-time applicability and effectiveness in recognizing facial expressions. The flow according to the proposed method is depicts in Figure 1.

Journal of Applied Informatics and Computing (JAIC)
Vol.8, No.2, December 2024, pp. 257~264
e-ISSN: 2548-6861

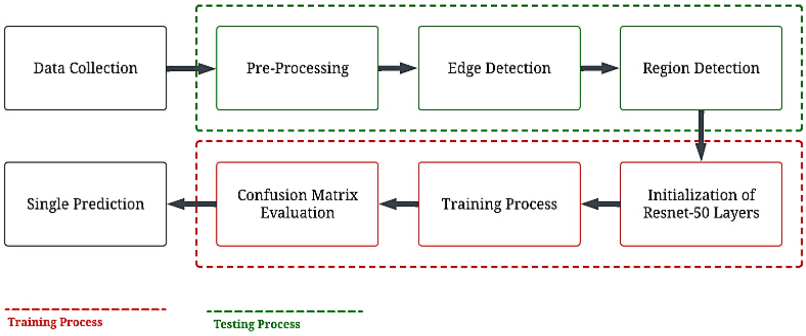


Figure 1. Proposed Method

A. Data Collection

In this section, the images in the data are grayscale with a resolution of 48 x 48 pixels and a single color channel (grayscale), resulting in a size of 48 x 48 x 1. The dataset was obtained from Kaggle [25] and categorized into seven distinct classes: anger with 35 samples, contempt with 54 samples, disgust with 177 samples, fear with 75 samples, happiness with 207 samples, sadness with 84 samples, and surprise with 249 samples. This distribution of samples indicates a significant class imbalance, with some classes such as anger and contempt having notably fewer samples compared to others like happiness and surprise. This imbalance could potentially lead to biased model training. To ensure a robust training and testing process, the dataset is divided so that 80% is used for training and 20% for testing, but additional techniques such as data augmentation or class rebalancing may be necessary to address the imbalance and improve model performance.

This classification enables the model to learn from large amounts of data while leaving a small portion for performance evaluation, which ensures the model's ability to integrate new, unobserved data. Sample of dataset can be depicted in Figure 2.

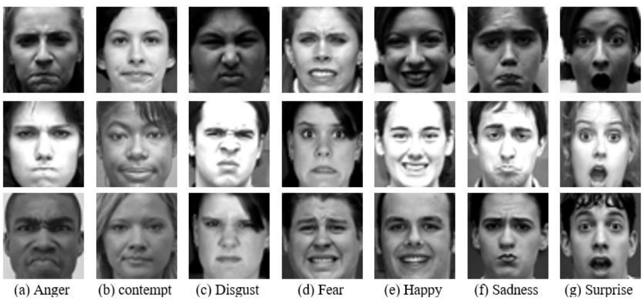


Figure 2. Sample of Datasets each Class

B. Pre-Processing

This section is crucial for ensuring that all images in the dataset conform to the required dimensions of 48 x 48 x 1. If any image in the dataset does not meet this size specification, it is resized accordingly during pre-processing [26], [27]. This resizing step ensures uniformity across all input images, which is essential for consistent feature extraction and model performance during training. Additionally, image enhancement techniques are employed in this research to clarify and improve the quality of the images before training. This enhancement process helps in accentuating important features, making the data more suitable for training the CNN model. By combining image enhancement with resizing, the pre-processing ensures that the dataset is uniformly prepared, facilitating more effective and accurate training outcomes.

C. Edge Detection

In this Edge Detection subsection, the Canny edge detection method [28], [29] is employed to identify the boundaries and contours within the facial images. This technique is known for its ability to detect edges with high accuracy by minimizing noise while preserving the essential structural features of the image. To further enhance the visibility of these edges, contrast improvement is applied using histogram equalization, which redistributes the intensity values to maximize the contrast across the image. This combination of Canny edge detection and histogram equalization ensures that the key facial features are clearly outlined, which is vital for subsequent region detection and feature extraction processes.

The results of the edge detection process can be depicted in Figure 3, while the detailed algorithm combining Canny edge detection and histogram equalization can be depicted in Algorithm 1 in the result and discussion section.

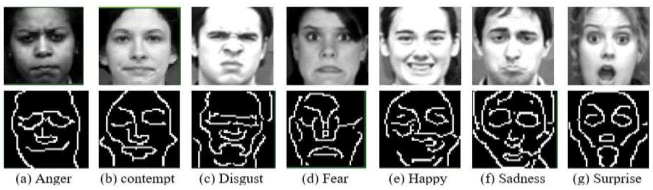


Figure 3. Results of Edge Detection each Class

D. Region Detection

In the next subsection, the Cascade Object Detector method [30] is utilized to accurately identify and isolate key facial regions, such as facial features like the eyes, nose, and mouth, within the images. This method leverages a cascade of classifiers, which are trained to detect specific features based on the structural characteristics learned from the training data [31]. By focusing on these critical regions, the detection process becomes more efficient and precise, allowing the model to concentrate on areas most relevant to facial expression recognition.

The results of the region detection process can be depicted in Figure 4, providing a visual representation of the successfully identified regions. The detailed algorithm incorporating Canny edge detection and histogram equalization in conjunction with Cascade Object Detector can be seen in Algorithm 2 in the result and discussion section.

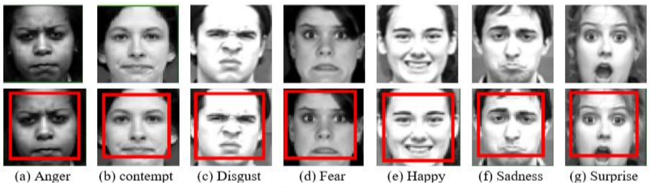


Figure 4. Results of Region Detection each Class

E. CNN Based on Transfer Learning Resnet-50

The structure of the CNN in this study is based on ResNet-50 [32], a deep neural network known for its ability to handle complex visual tasks effectively. The network starts with an input stage that receives an image with dimensions 48x48x1, where one channel represents a gray image.

ResNet-50 uses a series of convolutional layers, arranged in residual blocks, crafted to learn and extract detailed features from the input data while mitigating the gradient vanishing problem through cross-correlations. [33]. The program includes several stages of adaptation, each followed by a stage adaptation and ReLU activity, which is important to maintain stability and improve the learning process [17]. These layers are divided into four main branches or branches, each of which gradually releases higher-level structures. After the convolutional layers, we used a global averaging layer, which reduces the spatial dimension of the feature maps, transforming them into a single vector for each feature map. This is succeeded by a fully connected layer containing 1000 neurons, often used in ResNet-50 for classification tasks.

However, in this study, fully connected layers are fine-tuned or modified to fit a number of facial expression classes. The combination of these layers in ResNet-50 enables robust feature extraction and segmentation, making it very suitable for face recognition tasks. The layer of resnet-50 as trained using CNN can be depicted in Figure 5.

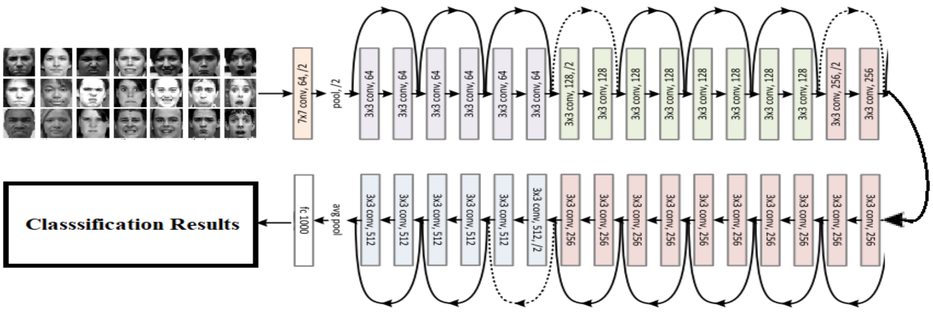


Figure 5. Layers of Resnet-50

F. Confusion Matrix Performance

In evaluating facial recognition across the seven categories—anger, contempt, disgust, fear, happiness, sadness, and surprise—the model's effectiveness is assessed using a confusion matrix. This matrix offers a thorough overview of the model's accuracy, precision, recall, and F1 score. It provides details on true positives, false positives, true negatives, and false negatives for each category, facilitating a precise evaluation of these performance metrics [34].

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

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

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

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

Where, accuracy evaluates the overall accuracy of the model’s predictions, whereas precision determines the ratio of accurate predictions to the total number of predictions for a specific class. Recall, or sensitivity, evaluates the model’s capability to recognize all pertinent examples within a class. The F1 score, which merges precision and recall, gives an overall view of the model’s performance, especially in cases of imbalanced datasets. The confusion matrix equations are detailed in eq (1) – (4).

III. RESULTS AND DISCUSSION

In the Results section, the initial phase of this research is based on a program executed using MATLAB 2024a, which facilitates the analysis of the implemented algorithms. The outcomes of the edge detection and region detection processes, as described in Algorithm 1 and Algorithm 2, are shown in Table 1.

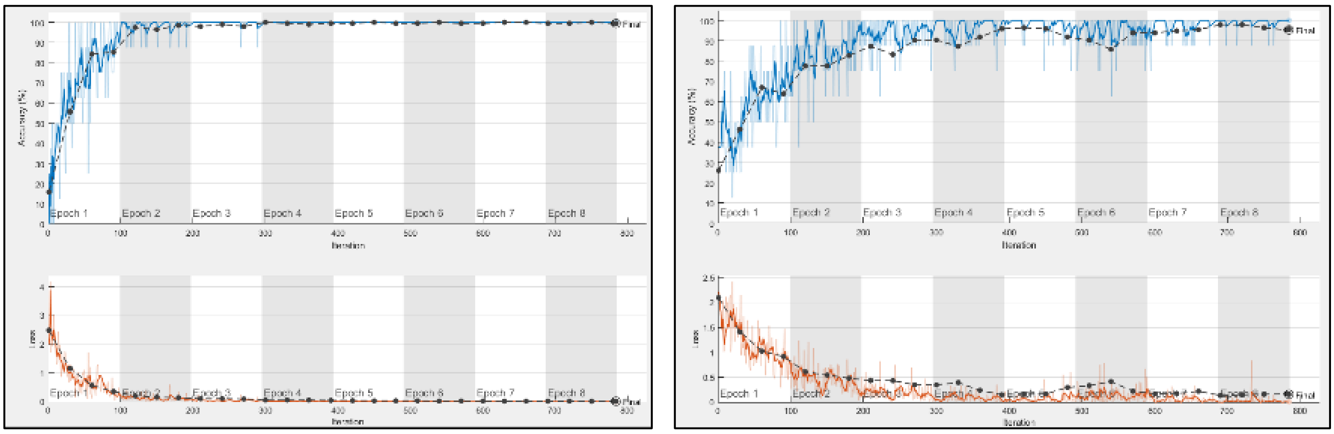
Figures 3 and 4 show the visual representation of the results from these algorithms. Figure 3 illustrates the effects of the edge detection process, showcasing how the Canny method combined with histogram equalization delineates the edges within the facial images. Meanwhile, Figure 4 displays the results of the region detection, where the vision Cascade Object Detector method identifies and isolates key facial region.

TABLE I
INITIALIZATION OF ALGORITHM

Algorithm 1: **Edge Detection (Canny)**	**Algorithm 2**: **Region Detection (Cascade Object Detector)**

BEGIN
$INPUT \leftarrow img$
IF img has 3 $channels$ THEN
CONVERT img to $grayscale$ $(grayImg)$
ELSE
USE img as $grayscale$ $(grayImg)$
END IF
ENHANCE contrast of $grayImg$ using $histogram\ equalization$ $(enhancedImg)$
DETECT edges in $enhancedImg$ using $Canny\ method$ $(edgeImg)$
$OUTPUT \leftarrow edgeImg$
END
	BEGIN
	$INPUT \leftarrow img$
	IF img has 3 $channels$ THEN
	CONVERT img to $grayscale$ $(grayImg)$
	ELSE
	USE img as $grayscale$ $(grayImg)$
	END IF
	INITIALIZE face detector using $vision.CascadeObjectDetector()$
	DETECT faces in $grayImg$ using the face $detector$ $(faces)$
	DEBUG: PRINT $'Faces\ detected:'$ and the faces data
	COPY img to $imgWithBoxes$
	IF $faces$ are $detected$ THEN
	DRAW rectangles around faces in $imgWithBoxes$ with red color and line width 2
	DISPLAY $imgWithBoxes$
	ELSE
	DISPLAY $'No\ faces\ detected.'$ as message
	END IF
	END

After initializing the pre-processing steps for the testing data, the next phase involves setting up the parameters for the training process. The parameters consist of employing the Adam optimizer, conducting training over 8 epochs, starting with a learning rate of 0.0001, and using a mini-batch size of 8. Following this setup, the training phase commences. As outlined in the methods section, the dataset is partitioned into 80% for training and 20% for testing to ensure a robust evaluation of the model. The results obtained from the training are illustrated through accuracy and loss graphs displayed in Figure 6. Specifically, Figure 6(a) presents the graph depicting the performance with the implementation of transfer learning using ResNet-50, while Figure 6(b) shows the graph without transfer learning. These visualizations provide insight into the effectiveness of transfer learning in improving model performance compared to training without this technique.

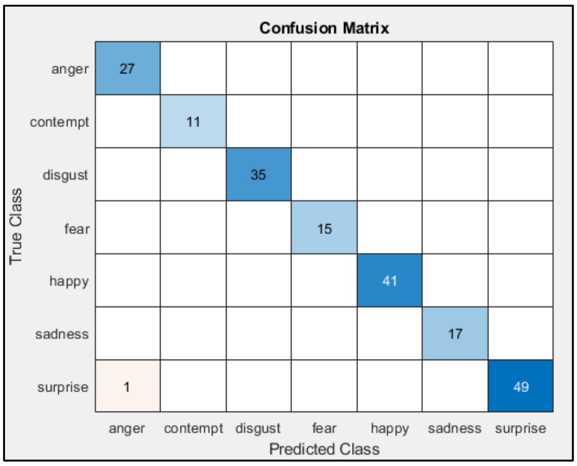


(a) Training and Loss Graph with Transfer Learning Resnet-50

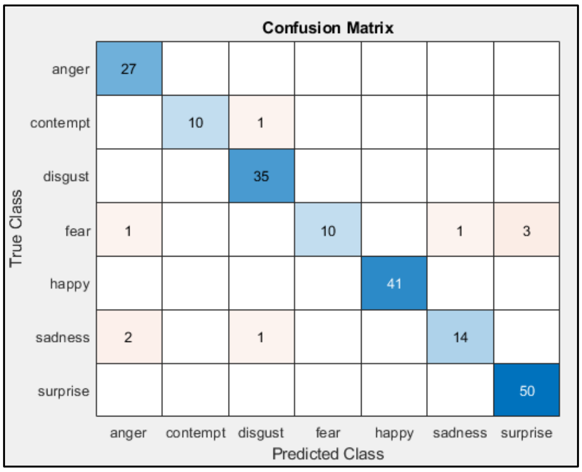
(b) Training and Loss Graph witout Transfer Learning Resnet-50

Figure 6. Training and Loss Graph with and without Transfer Learning Resnet-50

The results from the images presented in Figure 6 were also analyzed with the confusion matrices shown in Figure 7. Specifically, Figure 7(a) displays the confusion matrix of the model using transfer learning resnet-50, while Figure 7(b) presents the confusion matrix for the model created without transfer learning resnet-50. These fuzzy matrices provide a detailed view of the model's classification performance across all facial expression categories. In addition, the evaluation metrics—precision, accuracy, recall, and F1 score—are summarized in Table 2. This table shows the quantitative performance metrics obtained from the confusion matrices, giving a thorough evaluation of the model’s performance in facial expression recognition through transfer learning.



(a) Table of Confusion Matrix with Transfer Learning Resnet-50



(b) Table of Confusion Matrix witout Transfer Learning Resnet-50
Figure 7. Confusion Matrix with and without Transfer Learning Resnet-50

Based on analysis of Figure 6. reveals significant differences between the performance of the models with and without the use of transfer learning resnet-50. Figure 6 (a) Highlights the accuracy and loss curves for the model employing transfer learning with ResNet-50, while Figure 6 (b) presents the same metrics for the model trained from scratch. Notably, the graphs in Figure 6 (a) demonstrate a more rapid convergence, with the model achieving stable performance from epoch 4 onwards. In contrast, Figure 6 (b) exhibits slower convergence and more fluctuation throughout the training process. These results underscore the superiority of transfer learning, as the pre-trained ResNet-50 model effectively accelerates the training phase and enhances model stability compared to the training without transfer learning.

TABLE II
RESULTS EVALUATION

Performance	Results with Transfer Learning Resnet-50	Results without Transfer Learning Resnet-50
Accuracy	99.49%	95.41%
Precision	99.49%	96%
Recall	99.71%	91.41%
F1-Score	99.60%	93.11%

Fig. 7 provides further insights through confusion matrices, illustrating the performance differences between models with and without transfer learning resnet-50. Figure 7 (a) presents the confusion matrix for the model utilizing transfer learning, which indicates a higher number of true positive detections across the facial expression classes. In contrast, Figure 7 (b) reveals a larger number of true negatives (TN), false positives (FP), and false negatives (FN). The increased occurrence of true positives in Figure 7 (a) highlights the model's enhanced capability to correctly identify facial expressions, while the higher TN, FP, and FN in Figure 7 (b) suggest a less accurate performance. This comparison reaffirms that transfer learning offers a considerable boost to the model's classification accuracy.

The evaluation results presented in Table 2 further emphasize the benefits of using transfer learning resnet-50. The table compares performance metrics for models with and without transfer learning resnet-50, showing that the model with transfer learning achieved an impressive accuracy of 99.49%, compared to 95.41% for the model without transfer learning. Precision and recall also demonstrate superior performance with transfer learning, at 99.49% and 99.71% respectively, compared to 96% and 91.41% for the non-transfer learning model. The F1-score, which strikes a balance between precision and recall, is also notably high at 99.60% with transfer learning versus 93.11% without. These results collectively affirm that transfer learning not only enhances overall accuracy but also significantly improves the model’s precision, recall, and F1 score.

IV. CONCLUSIONS

This study proposed a method for facial expression recognition that effectively combines pre-processing techniques with advanced deep learning models. The results demonstrated that using transfer learning with ResNet-50 significantly enhanced the model's performance compared to training from scratch. Specifically, the model with transfer learning reached an accuracy of 99.49%, with precision at 99.49%, recall at 99.71%, and an F1-score of 99.60%. In contrast, the model without transfer learning showed lower performance metrics, with an accuracy of 95.41%, precision at 96%, recall at 91.41%, and an F1-score of 93.11%. The integration of transfer learning not only improved convergence rates but also resulted in more accurate detection of facial expressions and overall better evaluation metrics, making it a highly effective approach for high-performance facial expression recognition systems. However, while these results are impressive, it is important to discuss how they compare to other existing methods. For instance, although the ResNet-50 model demonstrated superior accuracy and consistency, there may be trade-offs such as increased inference time or higher computational requirements, which could affect the model's applicability in real-time systems. Additionally, comparing these results with other state-of-the-art models, such as GoogLeNet or VGGNet (Next Research), could provide a clearer picture of where this approach stands in the broader context of facial expression recognition technology. Future work should consider these factors to optimize both performance and efficiency, particularly in applications requiring real-time processing.

REFERENCES

[1] A. A. Khan, A. A. Laghari, and S. A. Awan, "Machine Learning in Computer Vision: A Review," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 8, no. 32, pp. 1–11, 2021, doi: 10.4108/eai.21-4-2021.169418.

[2] A. F. A. Fernandes, J. R. R. Dórea, and G. J. de M. Rosa, "Image Analysis and Computer Vision Applications in Animal Sciences: An Overview," Oct. 21, 2020, *Frontiers Media S.A.* doi: 10.3389/fvets.2020.551269.

[3] W. Sterkens *et al.*, "Computer vision and optical character recognition for the classification of batteries from WEEE," in *Procedia CIRP*, Elsevier B.V., 2022, pp. 110–115. doi: 10.1016/j.procir.2022.02.019.

[4] V. Kakani, V. H. Nguyen, B. P. Kumar, H. Kim, and V. R. Pasupuleti, "A critical review on computer vision and artificial intelligence in food industry," Dec. 01, 2020, *Elsevier B.V.* doi: 10.1016/j.jafr.2020.100033.

[5] D. Bhatt *et al.*, "Cnn variants for computer vision: History, architecture, application, challenges and future scope," Oct. 01, 2021, *MDPI*. doi: 10.3390/electronics10202470.

[6] J. Gupta, S. Pathak, and G. Kumar, "Deep Learning (CNN) and Transfer Learning: A Review," in *Journal of Physics: Conference Series*, Institute of Physics, 2022. doi: 10.1088/1742-6596/2273/1/012029.

[7] M. Ahmad, S. Abbas, A. Fatima, G. F. Issa, T. M. Ghazal, and M. A. Khan, "Deep Transfer Learning-Based Animal Face Identification Model Empowered with Vision-Based Hybrid Approach," *Applied Sciences (Switzerland)*, vol. 13, no. 2, Jan. 2023, doi: 10.3390/app13021178.

[8] E. A. Aldakheel, M. Zakariah, G. A. Gashgari, F. A. Almarshad, and A. I. A. Alzahrani, "A Deep Learning-Based Innovative Technique for Phishing Detection in Modern Security with Uniform Resource Locators," *Sensors*, vol. 23, no. 9, May 2023, doi: 10.3390/s23094403.

[9] K. Kuru, S. Clough, D. Ansell, J. McCarthy, and S. McGovern, "WILDetect: An intelligent platform to perform airborne wildlife census automatically in the marine ecosystem using an ensemble of learning techniques and computer vision," *Expert Syst Appl*, vol. 231, p. 120574, Nov. 2023, doi: 10.1016/j.eswa.2023.120574.

[10] O. A. Montesinos López, A. Montesinos López, and J. Crossa, *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer International Publishing, 2022. doi: 10.1007/978-3-030-89010-0.

[11] K. M. O. Nahar *et al.*, "Recognition of Arabic Air-Written Letters: Machine Learning, Convolutional Neural Networks, and Optical Character Recognition (OCR) Techniques," *Preprints (Basel)*, Sep. 2023, doi: 10.20944/preprints202309.1806.v1.

[12] V. Maheswari, C. A. Sari, D. R. I. M. Setiadi, and E. H. Rachmawanto, "Study Analysis of Human Face Recognition using Principal Component Analysis," in *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2020, pp. 55–60. doi: 10.1109/iSemantic50169.2020.9234250.

[13] I. William, D. R. I. M. Setiadi, E. H. Rachmawanto, H. A. Santoso, and C. A. Sari, "Face Recognition using FaceNet (Survey, Performance Test, and Comparison)," in *2019 Fourth International Conference on Informatics and Computing (ICIC)*, 2019, pp. 1–6. doi: 10.1109/ICIC47613.2019.8985786.

[14] I. U. W. Mulyono, D. R. I. M. Setiadi, A. Susanto, E. H. Rachmawanto, A. Fahmi, and Muljono, "Performance Analysis of Face Recognition using Eigenface Approach," in *2019 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2019, pp. 1–5. doi: 10.1109/ISEMANTIC.2019.8884225.

[15] D. del-Pozo-Bueno, D. Kepaptsoglou, F. Peiró, and S. Estradé, "Comparative of machine learning classification strategies for electron energy loss spectroscopy: Support vector machines and artificial neural networks," *Ultramicroscopy*, vol. 253, Nov. 2023, doi: 10.1016/j.ultramic.2023.113828.

[16] N. R. D. Cahyo, C. A. Sari, E. H. Rachmawanto, C. Jatmoko, R. R. A. Al-Jawry, and M. A. Alkhafaji, "A Comparison of Multi Class Support Vector Machine vs Deep Convolutional Neural Network for Brain Tumor Classification," in *2023 International Seminar on Application for Technology of Information and Communication (iSemantic)*, IEEE, Sep. 2023, pp. 358–363. doi: 10.1109/iSemantic59612.2023.10295336.

[17] M. M. I. Al-Ghiffary, C. A. Sari, E. H. Rachmawanto, N. M. Yacoob, N. R. D. Cahyo, and R. R. Ali, "Milkfish Freshness Classification Using Convolutional Neural Networks Based on Resnet50 Architecture," *Advance Sustainable Science Engineering and Technology*, vol. 5, no. 3, p. 0230304, Oct. 2023, doi: 10.26877/asset.v5i3.17017.

[18] Y. A. Nisa, C. A. Sari, E. H. Rachmawanto, and N. Mohd Yaacob, "Ambon Banana Maturity Classification Based On Convolutional Neural Network (CNN)," *sinkron*, vol. 8, no. 4, pp. 2568–2578, Oct. 2023, doi: 10.33395/sinkron.v8i4.12961.

[19] N. R. D. Cahyo and M. M. I. Al-Ghiffary, "An Image Processing Study: Image Enhancement, Image Segmentation, and Image Classification using Milkfish Freshness Images," *IJECAR) International Journal of Engineering Computing Advanced Research*, vol. 1, no. 1, pp. 11–22, 2024.

[20] F. Farhan, C. A. Sari, E. H. Rachmawanto, and N. R. D. Cahyo, "Mangrove Tree Species Classification Based on Leaf, Stem, and Seed Characteristics Using Convolutional Neural Networks with K-Folds Cross Validation Optimalization," *Advance Sustainable Science Engineering and Technology*, vol. 5, no. 3, p. 02303011, Oct. 2023, doi: 10.26877/asset.v5i3.17188.

[21] I. P. Kamila, C. A. Sari, E. H. Rachmawanto, and N. R. D. Cahyo, "A Good Evaluation Based on Confusion Matrix for Lung Diseases Classification using Convolutional Neural Networks," *Advance Sustainable Science, Engineering and Technology*, vol. 6, no. 1, p. 0240102, Dec. 2023, doi: 10.26877/asset.v6i1.17330.

[22] R. Singh, S. Saurav, T. Kumar, R. Saini, A. Vohra, and S. Singh, "Facial expression recognition in videos using hybrid CNN & ConvLSTM," *International Journal of Information Technology (Singapore)*, vol. 15, no. 4, pp. 1819–1830, Apr. 2023, doi: 10.1007/s41870-023-01183-0.

[23] M. N. Ab Wahab, A. Nazir, A. T. Z. Ren, M. H. M. Noor, M. F. Akbar, and A. S. A. Mohamed, “Efficientnet-Lite and Hybrid CNN-KNN Implementation for Facial Expression Recognition on Raspberry Pi,” *IEEE Access*, vol. 9, pp. 134065–134080, 2021, doi: 10.1109/ACCESS.2021.3113337.

[24] S. Rajan, P. Chenniappan, S. Devaraj, and N. Madian, “Novel deep learning model for facial expression recognition based on maximum boosted CNN and LSTM,” *IET Image Process*, vol. 14, no. 7, pp. 1227–1232, May 2020, doi: 10.1049/iet-ipr.2019.1188.

[25] J. Oheix, “Face Expression Recognition Datasets.” Accessed: Aug. 20, 2024. [Online]. Available: https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset

[26] M. A. Khayer, M. S. Hasan, and A. Sattar, “Arabian date classification using CNN algorithm with various pre-trained models,” in *Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2021*, Institute of Electrical and Electronics Engineers Inc., Feb. 2021, pp. 1431–1436. doi: 10.1109/ICICV50876.2021.9388413.

[27] Q. A. Putra, C. A. Sari, E. H. Rachmawanto, N. R. D. Cahyo, E. Mulyanto, and M. A. Alkhafaji, “White Bread Mold Detection using K-Means Clustering Based on Grey Level Co-Occurrence Matrix and Region of Interest,” in *2023 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2023, pp. 376–381. doi: 10.1109/iSemantic59612.2023.10295369.

[28] A. A. Mohammmed, E. Elbasi, and O. M. Alsaydia, “An Adaptive Robust Semi-blind Watermarking in Transform Domain Using Canny Edge Detection Technique,” in *2021 44th International Conference on Telecommunications and Signal Processing, TSP 2021*, Institute of Electrical and Electronics Engineers Inc., Jul. 2021, pp. 10–14. doi: 10.1109/TSP52935.2021.9522657.

[29] U. A. Nnolim, “Automated crack segmentation via saturation channel thresholding, area classification and fusion of modified level set segmentation with Canny edge detection,” *Heliyon*, vol. 6, no. 12, Dec. 2020, doi: 10.1016/j.heliyon.2020.e05748.

[30] S. N. Shivappriya, M. J. P. Priyadarsini, A. Stateczny, C. Puttamadappa, and B. D. Parameshachari, “Cascade object detection and remote sensing object detection method based on trainable activation function,” *Remote Sens (Basel)*, vol. 13, no. 2, p. 200, 2021.

[31] A. Luo, X. Li, F. Yang, Z. Jiao, H. Cheng, and S. Lyu, “Cascade graph neural networks for RGB-D salient object detection,” in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16*, Springer, 2020, pp. 346–364.

[32] X. X. Li, D. Li, W. X. Ren, and J. S. Zhang, “Loosening Identification of Multi-Bolt Connections Based on Wavelet Transform and ResNet-50 Convolutional Neural Network,” *Sensors*, vol. 22, no. 18, Sep. 2022, doi: 10.3390/s22186825.

[33] G. S. Nugraha, M. I. Darmawan, and R. Dwiyansaputra, “Comparison of CNN’s Architecture GoogleNet, AlexNet, VGG-16, Lenet -5, Resnet-50 in Arabic Handwriting Pattern Recognition,” *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, May 2023, doi: 10.22219/kinetik.v8i2.1667.

[34] S. C. Kim and Y. S. Cho, “Predictive System Implementation to Improve the Accuracy of Urine Self-Diagnosis with Smartphones: Application of a Confusion Matrix-Based Learning Model through RGB Semiquantitative Analysis,” *Sensors*, vol. 22, no. 14, Jul. 2022, doi: 10.3390/s22145445.