OMÜ Mühendislik Bilimleri ve Teknolojisi Dergisi,
OMU Journal of Engineering Sciences and Technology

e-ISSN: 2791-8858 OMUJEST,Eylül 2025, 5(2): 19-33

Deep Learning for Facial Emotion Recognition on the FER-2013 Dataset using the ResNet50v2 Model

ResNet50v2 Modelini Kullanarak FER-2013 Veri Setinde Yüz Duygularını Tanımak için Derin Öğrenme

Abdoul MALİK[1], Mohammad Adnan AYOUBİ[2], Mohammad ALSUEDANİ[3]

[1]Department of Intelligent Systems Engineering, Graduate Education Institute, Ondokuz Mayıs University, Samsun, Türkiye
· **aabdoulmalik99@gmail.com** · ORCİD > 0009-0006-6525-8346

[2]Department of Aerospace Engineering, Aerospace Faculty, Samsun University, Samsun, Türkiye
· **mohammadadnanayoubi@gmail.com** · ORCİD > 0009-0003-5647-5558

[3]Department of Mechanical Engineering, Engineering Faculty, Ondokuz Mayıs University, Samsun, Türkiye
· **Mohmadsw200184@gmail.com** · ORCİD > 0009-0002-4528-971X

Makale Bilgisi/Article Information

Makale Türü/Article Types: Araştırma Makalesi/Research Article
Geliş Tarihi/Received: 6 Ocak/January 2025
Kabul Tarihi/Accepted: 29 Eylül/September 2025
Yıl/Year: 2025 | **Cilt–Volume:** 5 | **Sayı–Issue:** 2 | **Sayfa/Pages:** 19-33

Atıf/Cite as: Malik, A., Ayoubi, M.A., Alsuedani, M. "Deep Learning for Facial Emotion Recognition on the FER-2013 Dataset using the ResNet50v2 Model" Ondokuz Mayis University Journal of Engineering Sciences and Technology 5(2), Eylül 2025: 19-33.

Sorumlu Yazar/Corresponding Author: Abdoul MALİK

DEEP LEARNING FOR FACIAL EMOTION RECOGNITION ON THE FER-2013 DATASET USING THE RESNET50V2 MODEL

ABSTRACT

Facial emotion recognition is a sophisticated approach that uses facial expression analysis to identify and understand human emotions. Its numerous applications in human-computer interaction, healthcare, and market research have recently received much attention. This technology aims to develop sophisticated algorithms and systems capable of accurately identifying and understanding an individual's emotional state by analysing facial features. This study focuses on classifying human emotions using a deep learning model based on the ResNet50v2 architecture. This work presents a comprehensive facial expression recognition model utilising the FER-2013 dataset, which includes thousands of images annotated with seven distinct emotions: happy, angry, neutral, sad, disgust, fear, and surprise. Our approach involves several key steps, including significant image preprocessing to improve the input data quality, image transformation to increase the diversity and robustness of the model, and implementing a modified ResNet50v2 architecture to improve recognition accuracy. Our model achieved a 69% accuracy on the test data, demonstrating competitive performance compared to existing models applied to FER-2013. The results of this study highlight the great potential of deep learning methods in precisely identifying and deciphering human emotions from facial expressions, paving the way for more emotionally intelligent and responsive human-computer interaction systems.

Keywords: Facial Emotion Recognition, Emotion Classification, FER-2013, ResNet50v2.

❁❁❁

RESNET50V2 MODELİNİ KULLANARAK FER-2013 VERİ SETİNDE YÜZ DUYGULARINI TANIMAK İÇİN DERİN ÖĞRENME

ÖZ

Yüz duygu tanıma, yüz ifadesi analizini kullanarak insan duygularını tanımlamak ve anlamak için kullanılan sofistike bir yaklaşımdır. İnsan-bilgisayar etkileşimi, sağlık hizmetleri ve pazar araştırması alanlarında sayısız uygulaması son zamanlarda büyük ilgi görmüştür. Bu teknoloji, yüz özelliklerini analiz ederek bir bireyin duygusal durumunu doğru bir şekilde tanımlayıp anlayabilen sofistike algoritmalar ve sistemler geliştirmeyi amaçlamaktadır. Bu çalışma, ResNet50v2 mimarisine dayalı bir derin öğrenme modeli kullanarak insan duygularını sınıflandırmaya odaklanmaktadır. Bu çalışma, mutlu, kızgın, nötr, üzgün, tiksinti, korku ve şaşkınlık olmak üzere yedi farklı duygu ile etiketlenmiş binlerce görüntüyü içeren FER-2013 veri setini kullanan kapsamlı bir yüz ifadesi tanıma modeli sunmaktadır. Yaklaşımımız, girdi veri kalitesini iyileştirmek için önemli görüntü ön işleme, modelin çeşitliliğini ve sağlamlığını artırmak için görüntü dönüştürme ve tanıma doğruluğunu iyileştirmek için değiştirilmiş bir ResNet50v2 mimarisi uygulamak gibi birkaç önemli adımı içermektedir. Modelimiz, test verilerinde %69 doğruluk oranına ulaşarak, FER-2013'e uygulanan mevcut modellere kıyasla rekabetçi bir performans sergilemiştir. Bu çalışmanın sonuçları, yüz ifadelerinden insan duygularını hassas bir şekilde tanımlama ve deşifre etme konusunda derin öğrenme yöntemlerinin büyük potansiyelini vurgulamakta ve daha duygusal zeka ve duyarlılığa sahip insan-bilgisayar etkileşim sistemlerinin önünü açmaktadır.

Anahtar Kelimeler: Yüz Duygu Tanıma, Duygu Sınıflandırma, FER-2013, ResNet50v2.

❁❁❁

Highlights

- ResNet50v2 adapted for facial emotion recognition using FER-2013 dataset
- Data augmentation improved model robustness and diversity
- Preprocessing normalized data and enhanced input quality
- Regularization techniques applied to avoid overfitting

1. INTRODUCTION

Interaction and communication are essential in the daily lives of all living beings, and emotions, such as body language and facial expressions, play a crucial role in enriching human exchanges [1]. Even a basic understanding of human emotions is enough to establish social connections. Detecting facial expressions is central to interpreting emotions, reflecting the innate human ability to decode subtle emotional signals [2]. Facial Emotion Recognition (FER) has become a flourishing field of study in computer vision and artificial intelligence research. This technology can identify emotions in real-time, offering applications in many fields, such as healthcare [3], where it can be used to diagnose emotional disorders or monitor the well-being of patients; in education, where it can improve teaching interactions by detecting student frustration or engagement; and in marketing, where emotion analysis can provide valuable information about consumer reactions to products or advertisements [4]. In addition, in surveillance and security, this technology makes it possible to analyze abnormal behaviour in real time, contributing to better crisis management [5].

The importance of facial expressions in non-verbal communication is also reflected in human-computer interaction (HCI). Modern interactive systems, such as virtual assistants, social robots and virtual reality platforms, increasingly incorporate emotion recognition to create more natural and personalized interactions [6]. By equipping these systems with the ability to understand and respond to users' emotions, interaction becomes more fluid and empathetic, enhancing the user experience [7]. This is particularly useful in mental health, where emotionally intelligent robots can support patients, or in e-learning environments, where interactive systems can adapt their pedagogical approach according to learners' emotional state. Facial emotion recognition aims to develop intelligent systems capable of understanding and interacting with individuals more intuitively and empathetically. This makes improving the efficiency and quality of services possible while offering more personalized and engaging experiences. Studies show that emotionally intelligent interfaces positively impact user satisfaction and engagement in various environments, from gaming platforms to telemedicine[8]. With the recent advances in machine learning and computer vision, facial emotion recognition is developing rapidly, attracting increasing attention [9]. Convolutional neural networks (CNNs) and transfer learning models are enabling impressive accuracy in detecting emotions, even in conditions such as poor lighting or partial facial expressions [10]. This technology promises to revolutionize human-machine interactions and transform sectors such as entertainment, security and healthcare by offering more human and empathetic solutions.

Recent research indicates that advances in image processing and deep learning techniques have accelerated the machine's understanding of facial expressions. The latest study, however, has concentrated chiefly on enhancing the model to get advanced outcomes by increasing the overall precision [11]. The field of computer vision has made extensive use of deep learning algorithms [12]. Convolutional Neural Networks (CNN) are currently used to handle numerous picture categorization challenges. Convolutional, pooling, and fully connected layers are frequently used in CNNs to capture information and reduce complexity while preserving key features [13]. Various strategies are being adopted to improve performance; for example, ReLU (rectified linear unit) activation has replaced the Sigmoid activation function in an effort to avoid gradient divergence and speed up training, and it is widely used for image recognition applications [14].Other tapproach as Thermal imaging has also used to enhance the accuracy of facial emotion recognition [15]. Additionally, a variety of pooling techniques are used to help with generalization and decrease the input sample size [16]. On the ImageNet validation set, this study used ResNet, VGG, PReLu-Net, GoogLeNet, and BN-Inception [17]. The findings show that compared to other models, The ResNet architecture models had reduced error rates. The 50/101/152-layer ResNets outperform the 34-layer ones in terms of accuracy by significant margins. Despite having three more block layers, putting them deeper (11.3 billion FLOPs), ResNet 101-layer and 152-layer ResNets are still less complex than VGG 16/19 networks. According to the research from [18], MobileNetV1 and MobileNetV2 have fewer parameters than any of the models used on both datasets (FER2013, AffectNet), including the Xception and VGG models. An average accuracy of 67.4% was attained by Xception, 61.8% and 62.1% by MobileNet and MobileNetV2, and 66.6% by DenseNet-40, according to the findings of this investigation, which examined multiple models using the FER2013 database.

Using the FER2013 database, researchers from [19] tested training three different CNNs and combining them to improve efficiency. Their highest single-network total precision is 62.44%. A CNN and batch normalisation-based architecture was put out, with a fixed batch size of 256. Higher epochs were found to cause overfitting, hence the training procedure was limited to 8 epochs [20]. FER-2013's private test results showed a test accuracy of 60.12% thanks to this architecture. In [21], several designs with variable numbers of filters that combined batch normalisation and CNN were presented. The distinctive feature of both architectures is that they maintain a kernel size of 8 and lack any completely connected layers. Both architectures were able to reach test accuracy of 65%. Given their capacity to automatically extract local hierarchical features (edges, textures, and intricate patterns) using convolutional and pooling layers, convolutional neural networks (CNNs) have emerged as the core of computer vision [13]. Facial emotion recognition has demonstrated its effectiveness in processing subtle variations in expressions. Meanwhile, transfer learning allows us to leverage models pre-trained on large datasets such as ImageNet, thereby reducing the need for specific training data while improving generalisation, a decisive advantage for FER-2013, where certain classes, such as disgust, are underrepresented [22].

The motivation for this study is to leverage the immense power of facial emotion recognition to transform critical sectors, including market research, healthcare and human-computer interaction. The aim is to design advanced algorithms capable of accurately detecting and interpreting emotional states through the analysis of facial expressions, thereby improving the responsiveness and empathy of interactive systems while opening up new perspectives for more personalized and relevant interactions with users. Unlike previous work focusing on standard CNN architectures or model ensembles [19, 33], this study leverages the power of ResNet50v2 residual connections optimised through targeted fine-tuning to capture fine-grained discriminative features under realistic conditions (lighting variations, partial angles). Our contribution lies in an integrated pipeline combining adaptive pre-processing, advanced regularisation, and hyperparameter optimisation, establishing a new benchmark for ResNet50v2 on FER-2013.

This study introduces a facial expression recognition model built with the ResNet50V2 architecture and trained on the FER-2013 dataset. With thousands of facial pictures labelled with seven primary emotions, the FER-2013 dataset is a popular standard for emotion detection. Based on residual connections, ResNet50V2 is an improved version of the original ResNet50 (Residual Network) model. [15]. Compared to its predecessor, ResNet50V2 introduces improvements in the ordering of normalisation and activation layers, resulting in faster convergence and increased training stability, making it highly suitable for emotion recognition tasks [23].

Previous research on FER-2013 reported accuracies ranging from 62% to 67% accuracy using traditional CNNs, GoogleNet, or hybrid techniques [12,19,33-36]. This research emphasizes the difficulty of improving performance on this challenging dataset, due to class imbalance and low image resolution. These studies have been chosen for comparison because they include typical baselines CNNs and more sophisticated designs (Inception-based, GAP-based, and VGG-based), providing a representative benchmark of the current state of the art. Our solution surpasses these methods by 69% accuracy, illustrating the effectiveness of our preprocessing pipeline and model fine-tuning. This improvement validates the contribution of carefully designed data augmentation and regularization techniques in enhancing model generalization on FER-2013. The proposed methodology, including all processing steps, is detailed in the following section. The results are then presented and discussed in Section 3. Finally, the conclusion outlines future research directions.

2. MATERIALS AND METHODS

Various Python libraries were used to build and train our model. OpenCV was used for image processing, and Keras and TensorFlow were used to build and train the CNN model based on ResNet50V2. A complete pipeline for facial expression recognition was created using these libraries, from data preparation to training and model evaluation. Each library played a specific and complementary role, allowing us to perform this experiment efficiently and robustly. The experiments were conducted on google Collaboratory.

2.1. Dataset

Initially presented by Pierre-Luc Carrier and Aaron Courville in 2013, the Facial Expression Recognition 2013 (FER-2013) dataset was made available on Kaggle. There are 48x48 pixel grayscale pictures of faces in this dataset. The 35,887 photos in the FER-2013 dataset are grouped into seven distinct emotional categories. Each image is labelled in accordance with these classifications, which range from 0 to 6.

There are two sections to the dataset. The table below (Figure 1 and Table 1) describes the test and training data.

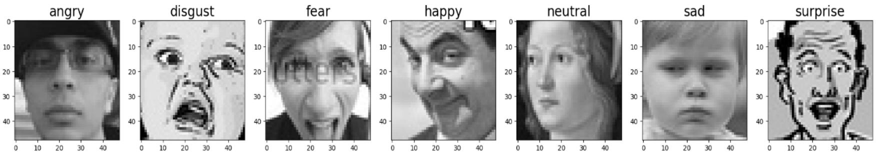

Figure 1. Sample of each class from the dataset.

Table 1. Dataset Distribution

Facial Expression (Class)	Test Data	Training Dataset	Total
Angry	958	3995	4953
Disgust	111	436	547
Fearful	1024	4097	5121
Happy	1774	7215	8989
Sad	1247	4830	6077
Surprise	831	3171	4002
Neutral	1233	4965	6198
Total	**7178**	**28709**	**35887**

2.2. Image Preprocessing

Data preprocessing is a primordial step in improving model performance. To enhance the diversity of training data and avoid overlearning, we applied different picture transformations using Keras' ImageDataGenerator class.

Preprocessing for training data includes:

- Normalization of pixel values
- Rotation of images (rotation_range =10)
- Zooming of images (zoom_range=0.2)
- Random horizontal and vertical translation
- Random horizontal translation
- Fill mode for transformations

For the test data, preprocessing is limited to normalizing pixel values. We then loaded the images from the training and test directories using these generators, defining the class parameters, target image size of 400x400 pixels, colour mode, and data distribution method (random for training and fixed for testing).

This study's key contribution is using the ResNet50V2 model with a properly constructed preprocessing and transformation methodology. It boosted the training data's diversity and robustness using normalization and modifications such as rotation, zoom, translation, flipping, and adaptive fill mode. This approach directly impacts the model's ability to generalize, since it allows the network to learn invariant properties despite differences in lighting, position, and scale. This pipeline is the foundation for our contribution, directly influencing the model's ability to generalize in real-world settings.

2.3. ResNet50V2 Architecture

Convolutional neural networks (CNNs) are deep learning architectures designed to process grid data, such as images. They consist of convolution layers, pooling layers and fully connected layers. Their ability to learn hierarchical representations, from simple edges to complex semantic features, makes them very effective for image classification tasks. Figure 2 depicts a simple CNN architecture to help understand the main components of a CNN. Convolutional layers extract features, pool layers reduce dimensionality and fully connected layers for classification. This broad framework supports the ResNet50V2 model utilised in this study, whose specific architecture is discussed further below.

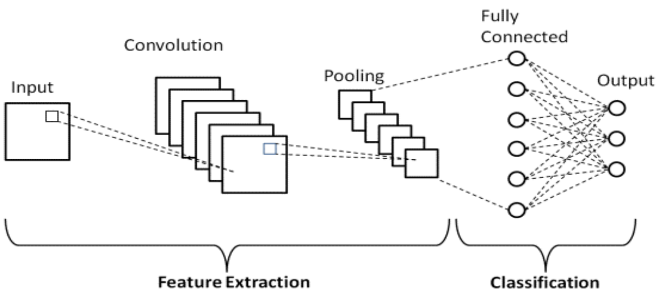

Figure 2. Basic CNN Architecture

In this study, the ResNet50V2 architecture was used, a convolutional neural network architecture deeply optimized for image classification tasks. ResNet50V2, an improved version of ResNet50, leverages advanced normalization techniques and jump connections, alleviating the problem of vanishing gradients and enabling intense networks to be trained more efficiently. This model suits our facial expression recognition problem well because it can extract complex features from images. ResNet50V2 is a widely used CNN architecture in computer vision, known for its improved ability to learn detailed feature representations[24]. Building on the original ResNet, it incorporates 50 layers, including convolution layers, batch normalization, and ReLU activation functions. The core component, the residual block, allows the network to bypass degradation issues common in deep networks by creating shortcuts or "skip connections" [25]. These connections enable more efficient learning by focusing on differences or additional features in the input data, and they smooth gradient flow during training, addressing vanishing or exploding gradients. The Residual structure for ResNet-50 is shown in Figure 3.

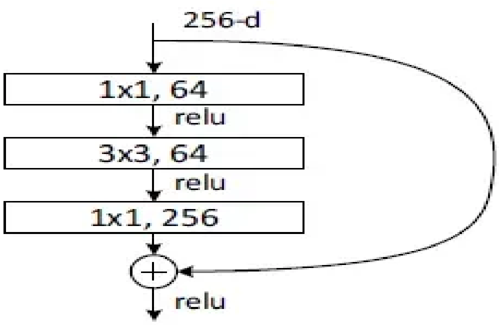

Figure 3. Residual Block for ResNet-50 [17]

To improve feature learning and generalisation, more layers were added to the ResNet50V2 backbone, such as batch normalisation, dropout, and a dense layer with 64 units. The last softmax layer guaranteed appropriate multi-class classification for all seven emotions. Despite testing several loss functions and deeper variants, these changes did not result in appreciable performance gains. To balance stability, efficiency, and generalization, categorical cross-entropy with this slightly expanded design was ultimately retained.

Batch normalization is another critical element of ResNet50V2, enhancing training stability and convergence by normalizing data in each batch, which helps reduce overfitting[26]. Additionally, the architecture includes a bottleneck design that reduces the computational load while maintaining the ability to extract deep and abstract representations from visual data [27]. The dropout technique is also employed to combat overfitting by randomly ignoring some units during training, improving the model's ability to generalize to the new data. A notable feature of the model is the pre-activation block, where batch normalization and ReLU are applied before convolution, further stabilizing training[28]. Dimensionality reduction is achieved through 1x1 convolutions, which decrease feature dimensions while preserving essential information and reducing computational complexity [29]. ResNet50V2 is highly effective for complex image-related tasks such as classification, object detection, and segmentation[30]. Its ability to overcome training challenges like gradient issues and performance degradation makes it a go-to choose for building deep learning models in diverse computer vision applications. Our model consists of the following layers:

- ResNet50V2: for extracting features from 224x224 pixel images.
- Dropout (0.25): For normalization and to prevent overlearning.
- Batch Normalization: to normalize activations and speed up training.
- Flatten: To convert 2D feature maps into 1D vectors.
- Dense (64 units, activation 'real'): A fully connected layer for feature learning.
- Batch Normalization and Dropout (0.5): For additional regularisation.
- Dense (7 units, activation 'softmax'): Output layer for classifying seven emotions.

The model is built with Adam optimizer, known for its ability to adapt to the learning rate, and the loss function "categorical_crossentropy" adapted to multi-class classification problems. The metrics tracked include accuracy in evaluating the performance of the model. Figure 3 present the model architecture.

Layer (type)	Output Shape	Param #
resnet50v2 (Functional)	(None, 7, 7, 2048)	23564800
dropout (Dropout)	(None, 7, 7, 2048)	0
batch_normalization (Batch Normalization)	(None, 7, 7, 2048)	8192
flatten (Flatten)	(None, 100352)	0
dense (Dense)	(None, 64)	6422592
batch_normalization_1 (BatchNormalization)	(None, 64)	256
dropout_1 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 7)	455

Total params: 29996295 (114.43 MB)
Trainable params: 22779527 (86.90 MB)
Non-trainable params: 7216768 (27.53 MB)

Figure 3. ResNet50V2-based model architecture used for facial emotion classification

2.4. Model Training

To train the ResNet50V2 model, we used regularization strategies such as Dropout and Batch Normalization to avoid overlearning and improve model generalization.

- Number of epochs: The model was trained over a number of epochs to ensure proper convergence. However, to avoid overtraining, callbacks such as early stops are used to stop training when performance on validation data starts to deteriorate.
- Batch Size: A chunk size of 64 was used, balancing training speed and gradient stability.
- Image size: Images were resized to 224x224 pixels to accommodate the ResNet50V2 architecture and to utilize pre-trained weights on large images.
- Callbacks: Several callbacks were used, including "ModelCheckpoint" to save the best model based on validation accuracy, "EarlyStopping" to stop training if validation accuracy stops improving, and "ReduceLROnPlateau" to reduce the learning rate if performance stagnates.

2.5. Performance Metrics

This study used F1-Score, recall, precision and accuracy metrics to evaluate the performance of the model. These metrics provide a comprehensive and nuanced assessment of Deep Learning models, guaranteeing their robustness and reliability [31]. The formulas for the evaluation metrics are given in Equations (1-4).

$$Accuracy=(TP+TN) / (TP+FP+TN+FN) \qquad (1)$$

$$Recall = TP / (TP + FN) \qquad (2)$$

$$Precision = TP/ (TP + FP) \qquad (3)$$

$$F1\text{-} Score = (2 * Precision * Recall) / (Precision + Recall) \qquad (4)$$

TP denotes the number of true positives in the model, while TN represents the number of true negatives false negatives (FN) and false positives (FP) [32].

3. RESULT AND DISCUSSION

When we looked at the model's accuracy rates during the training and testing stages, we found that the model achieved a 77% accuracy rate on the training data over 30 epochs (Figure 3). This implies that the model's performance on the training set is satisfactory. However, the model's actual accuracy percentage on test data is 69%. This suggests that the model's generalizability and applicability are lower on real-world data than on training data.

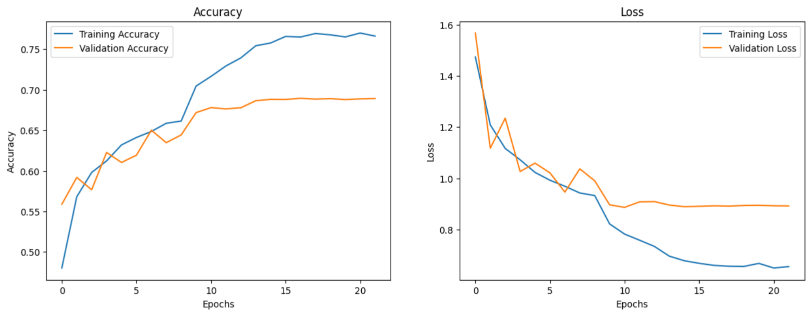

Figure 4. Training accuracy and training loss of the model.

The model has been evaluated on test data to measure its generalizability. Evaluation metrics included additional metrics such as precision, loss and weighted precision, recall, F1-score and AUC. A confusion matrix (Figure 4) was used to visualize the model's classification performance for each emotion class. A detailed classification report on the evaluation metrics for each emotion class. This report provides an overview of the model's performance on different classes and allows for detecting imbalances in predictions.

Figure 4. Test data confusion matrix.

Table 2. Test data classification results.

Emotion	Precision	Recall	F1 Score	Number of Images
Angry	0.61	0.64	0.63	917
Disgust	0.72	0.51	0.60	109
Fear	0.55	0.46	0.50	1000
Happy	0.88	0.90	0.89	1763
Neutral	0.62	0.70	0.66	1215
Sad	0.57	0.57	0.57	1228
Surprise	0.80	0.78	0.79	815

The observed difference in the number of images between classes in the training and test datasets is due to removing images that do not represent faces before training the model. This was done to improve data quality, train the model more accurately, and make more reliable predictions based on the test data. This process resulted in a significant reduction in the number of images for some classes, resulting in a difference between the number of images in the test dataset and the original test dataset. On the other hand, fine-tuning improves the model's stability and reduces overlearning during the training process, thereby improving the model's performance. Overall, the results indicate that the model performs well, though there is still room for improvement. The accuracy of 69%, precision of 68.76%, recall of 68.98%, and F1 score of 68.74% are all closely aligned, reflecting the model's ability to make balanced and reliable predictions. Furthermore, the AUC score of 81.44% demonstrates strong discriminative ability. Nevertheless, additional data analysis and model refinement techniques could be explored to enhance the model's performance further. The Comparaison with different other studies is given in Table 3.

Table 3. Comparaison with different studies in literature on fer2013.

Reference	Method	Accuracy (%)
[19]	CNN	62.44
[33]	GoogleNet	65.2
[12]	VGG+SVM	66.31
[34]	Custom CNN	66.67
[35]	CNN+ GAP	66
[36]	Conv + Inception layer	66.40
This work	**Resnet50V2**	**69**

4. CONCLUSION

This study used the Facial Expression Recognition 2013 dataset, which consists of 35,887 grayscale images, each labelled according to seven emotion categories. Our main goal was to develop a model to accurately and efficiently classify these facial expressions. For this purpose, the ResNet50V2 model was adapted with regularization techniques to improve generalization. Training was performed using preprocessing and data augmentation techniques to increase the robustness of the model. The results show that the model can effectively recognize facial expressions, although it needs improvement. An accuracy of 69% was achieved on the test data. This study highlights a model's potential for facial expression recognition while presenting challenges and prospects for future improvements. While the model shows promising results, future work should include architecture refinement, cross-dataset testing, and model explainability. These improvements could facilitate real-world deployment in healthcare or human-machine interaction systems.

Author Contribution Rates

Design of Study: AM(%33,3), MAA(%33,3), MA(%33,3)

Data Acquisition: AM(%33,3), MAA(%33,3), MA(%33,3)

Data Analysis: AM(%33,3), MAA(%33,3), MA(%33,3)

Writing Up: AM(%33,3), MAA(%33,3), MA(%33,3)

Submission and Revision: AM(%33,3), MAA(%33,3), MA(%33,3)

REFERENCES

[1] R. Buck, "Social and emotional functions in facial expression and communication: The readout hypothesis," *Biological psychology,* vol. 38, no. 2-3, pp. 95-115, 1994.

[2] K. R. Scherer, A. Dieckmann, M. Unfried, H. Ellgring, and M. Mortillaro, "Investigating appraisal-driven facial expression and inference in emotion communication," *Emotion,* vol. 21, no. 1, p. 73, 2021.

[3] T. Altameem and A. Altameem, "Facial expression recognition using human machine interaction and multi-modal visualization analysis for healthcare applications," *Image and Vision Computing,* vol. 103, p. 104044, 2020.

[4] E. Arnaud, A. Dapogny, and K. Bailly, "Thin: Throwable information networks and application for facial expression recognition in the wild," *IEEE Transactions on Affective Computing,* vol. 14, no. 3, pp. 2336-2348, 2022.

[5] M. Sajjad, M. Nasir, F. U. M. Ullah, K. Muhammad, A. K. Sangaiah, and S. W. Baik, "Raspberry Pi assisted facial expression recognition framework for smart security in law-enforcement services," *Information Sciences,* vol. 479, pp. 416-431, 2019.

[6] X. Zhao, J. Zhu, B. Luo, and Y. Gao, "Survey on facial expression recognition: History, applications, and challenges," *IEEE multimedia,* vol. 28, no. 4, pp. 38-44, 2021.

[7] M. Sajjad *et al.*, "A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines," *Alexandria Engineering Journal,* vol. 68, pp. 817-840, 2023.

[8] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE transactions on affective computing,* vol. 13, no. 3, pp. 1195-1215, 2020.

[9] A. A. Pise, M. A. Alqahtani, P. Verma, P. K, D. A. Karras, and A. Halifa, "Methods for facial expression recognition with applications in challenging situations," *Computational intelligence and neuroscience,* vol. 2022, no. 1, p. 9261438, 2022.

[10] S. M. S. Abdullah and A. M. Abdulazeez, "Facial expression recognition based on deep learning convolution neural network: A review," *Journal of Soft Computing and Data Mining,* vol. 2, no. 1, pp. 53-65, 2021.

[11] I. J. Goodfellow *et al.*, "Challenges in representation learning: A report on three machine learning contests," in *Neural information processing: 20th international conference, ICONIP 2013, daegu, korea, november 3-7, 2013. Proceedings, Part III 20*, 2013: Springer, pp. 117-124.

[12] M.-I. Georgescu, R. T. Ionescu, and M. Popescu, "Local learning with deep and handcrafted features for facial expression recognition," *IEEE Access,* vol. 7, pp. 64827-64836, 2019.

[13] H. Ding, S. K. Zhou, and R. Chellappa, "Facenet2expnet: Regularizing a deep face recognition net for expression recognition," in *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)*, 2017: IEEE, pp. 118-126.

[14] Y. Khaireddin and Z. Chen, "Facial emotion recognition: State of the art performance on FER2013," *arXiv preprint arXiv:2105.03588,* 2021.

[15] A. Atchogou and C. Tepe, "ROBUST EMOTION RECOGNITION IN THERMAL IMAGING WITH CONVOLUTIONAL NEURAL NETWORKS AND GRAY WOLF OPTIMIZATION," *Signal Processing: Image Communication,* p. 117363, 2025.

[16] K. O'Nell, R. Saxe, and S. Anzellotti, "Recognition of identity and expressions as integrated processes," 2019.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[18] K. Kimberly Octavina and G. Putra Kusuma, "Real-time Facial Expression Recognition using Convolutional Neural Network on Mobile Device," *International Journal of Computing and Digital Systems,* vol. 15, no. 1, pp. 1-9, 2024.

[19] K. Liu, M. Zhang, and Z. Pan, "Facial expression recognition with CNN ensemble," in *2016 international conference on cyberworlds (CW)*, 2016: IEEE, pp. 163-166.

[20] I. Talegaonkar, K. Joshi, S. Valunj, R. Kohok, and A. Kulkarni, "Real time facial expression recognition using deep learning," in *Proceedings of international conference on communication and information processing (ICCIP)*, 2019.

[21] A. Agrawal and N. Mittal, "Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy," *The Visual Computer,* vol. 36, no. 2, pp. 405-412, 2020.

[22] Z. Zhao, L. Alzubaidi, J. Zhang, Y. Duan, and Y. Gu, "A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations," *Expert Systems with Applications,* vol. 242, p. 122807, 2024.

[23] P. Kaushik and S. Choudhary, "Enhanced Breast Cancer Detection using ResNet50V2-based Convolutional Neural Networks," in *2024 4th International Conference on Soft Computing for Security Applications (ICSCSA)*, 2024: IEEE, pp. 374-379.

[24] L. Du, Z. Lu, and D. Li, "Broodstock breeding behaviour recognition based on Resnet50-LSTM with CBAM attention mechanism," *Computers and Electronics in Agriculture,* vol. 202, p. 107404, 2022.

[25] W. Boulila, A. Alzahem, A. Koubaa, B. Benjdira, and A. Ammar, "Early detection of red palm weevil infestations using deep learning classification of acoustic signals," *Computers and Electronics in Agriculture,* vol. 212, p. 108154, 2023.

[26] A. Panthakkan, S. Anzar, S. Jamal, and W. Mansoor, "Concatenated Xception-ResNet50—A novel hybrid approach for accurate skin cancer prediction," *Computers in Biology and Medicine,* vol. 150, p. 106170, 2022.

[27] A. K. Sharma *et al.*, "Brain tumor classification using the modified ResNet50 model based on transfer learning," *Biomedical Signal Processing and Control,* vol. 86, p. 105299, 2023.

[28] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2," *Informatics in medicine unlocked,* vol. 19, p. 100360, 2020.

[29] H. C. Reis and V. Turk, "Detection of forest fire using deep convolutional neural networks with transfer learning approach," *Applied Soft Computing,* vol. 143, p. 110362, 2023.

[30] D. Hindarto, "Use ResNet50V2 Deep Learning Model to Classify Five Animal Species," *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi),* vol. 7, no. 4, pp. 758-768, 2023.

[31] A. Tharwat, "Classification assessment methods," *Applied computing and informatics,* vol. 17, no. 1, pp. 168-192, 2020.

[32] A. Doğru, S. Buyrukoğlu, and M. Arı, "A hybrid super ensemble learning model for the early-stage prediction of diabetes risk," *Medical & Biological Engineering & Computing,* vol. 61, no. 3, pp. 785-797, 2023.
[33] P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on FER-2013," in *Advances in hybridization of intelligent methods: Models, systems and applications*: Springer, 2017, pp. 1-16.
[34] M.-A. Quinn, G. Sivesind, and G. Reis, "Real-time emotion recognition from facial expressions," *Standford University,* 2017.
[35] O. Arriaga, M. Valdenegro-Toro, and P. Plöger, "Real-time convolutional neural networks for emotion and gender classification," *arXiv preprint arXiv:1710.07557,* 2017.
[36] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *2016 IEEE Winter conference on applications of computer vision (WACV)*, 2016: IEEE, pp. 1-10.