INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, Decembaer 2024, pp : 1572-1580
e-ISSN : 2583-1062
Impact Factor : 7.001
www.ijprems.com
editor@ijprems.com

ENHANCED EMOTION DETECTION USING FACIAL EXPRESSIONS: A DEEP LEARNING APPROACH

Sneha Suresh Shinde[1], Nimisha Vinesh Jethva[2], Deepti Nikumbh[3]
[1,2,3]Computer Engineering Shah and Anchor Kutchhi Engineering College Mumbai, India.
sneha.shinde16118@sakec.ac.in, nimisha.jethva16091@sakec.ac.in, deepti.nikumbh@sakec.ac.in
DOI: https://www.doi.org/10.58257/IJPREMS37870

ABSTRACT

The demand for emotion detection systems has grown, particularly in human-computer interaction and mental health monitoring. This study introduces a method for improving emotion recognition through facial expression analysis. Our model, utilizing Convolutional Neural Networks and advanced data pre-processing, achieved significant improvements in accuracy. Using a comprehensive facial expression dataset, we trained and evaluated the model, demonstrating its effectiveness in detecting emotions across multiple classes. The results highlight the potential of deep learning in emotion detection, with applications in healthcare and virtual reality.

Keywords: Emotion Detection, Facial Expressions, Convolutional Neural Network, Deep Learning

1. INTRODUCTION

Human-machine interactions are improved when emotions are recognized from facial expressions. The intricacies of facial expressions are difficult to represent by conventional techniques, which rely on handcrafted features and shallow models. The ability to recognize emotions has greatly increased with the advent of deep learning. This study investigates how Convolutional Neural Networks (CNNs), which are highly effective at picture identification tasks, can be used to enhance emotion detection by utilizing face expressions. Our objective is to create a trustworthy system that can recognize a wide variety of emotions, such as neutrality, fear, disgust, rage, surprise, sadness, and happiness. In order to train and assess our generative AI model and enable it to recognize subtle emotional cues, a complete dataset that reflects these emotions is essential. Through experiments, we show that our strategy works better than conventional approaches, providing better accuracy and generalization across various persons and settings.

The contributions of this research are:

- We present a CNN-based architecture with a pre-trained model, enhancing emotion detection accuracy across various emotional states.
- Our research explores deep learning approaches, including fine-tuning and augmentation, to optimize the emotion detection system.
- We evaluate different CNN configurations to identify the most effective methods for capturing emotion-related patterns.
- Our findings demonstrate the superiority of deep learning methods in generalization, robustness, and accuracy across multiple datasets and real-world scenarios.

This paper is organized as follows: Section 1 introduces the topic, Section 2 reviews the literature, with subsections on traditional approaches (2.1), deep learning approaches (2.2), and a comparative analysis (2.3). Section 3 details the methodology, covering architecture (3.1), algorithm (3.2), and implementation (3.3). Section 4 concludes with future directions, and Section 5 lists the references

2. LITERATURE REVIEW

Facial recognition technology has undergone a significant transformation, advancing from fundamental image processing techniques to the implementation of advanced deep learning models. This literature review examines both approaches, comparing their methodologies, performance, and the mathematical principles they are based on.

2.1. Traditional Approaches

Traditional facial recognition relies on handcrafted features and algorithms for face detection, alignment, and recognition. Key techniques include:

2.1.1. Principal Component Analysis (PCA)

By employing PCA, the dimensionality of facial images is reduced, ensuring that the crucial features needed for recognition are maintained [5]. It finds the eigenvectors (eigen faces) of the covariance matrix of the image set, projecting faces onto a lower-dimensional subspace [9]. The covariance matrix C is computed as given in Equation (1)

www.ijprems.com
editor@ijprems.com
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, Decembaer 2024, pp : 1572-1580
e-ISSN : 2583-1062
Impact Factor : 7.001

Formula:

$C = \frac{1}{M}\sum_{i=1}^{M}(x_i - \mu)(x_i - \mu)^T$ ……..……(1)

where x_i is an image vector and μ is the mean face.

2.1.2. Linear Discriminant Analysis (LDA)

LDA improves the separation of classes by maximizing the ratio of variance between classes to variance within classes [7]. It projects data onto a subspace that maximizes class discrimination [10]. The objective function $J(\omega)$ is given in Equation (2)

Formula:

$J(\omega) = \frac{v^A T_B v}{v^A T_w v}$ ……..……..……(2)

Where the matrix representing between-class scatter is denoted as T_B while the within-class scatter matrix is referred to as T_w .

2.1.3. Local Binary Patterns (LBP)

LBP (Local Binary Patterns) extracts local texture features by applying thresholding to the surrounding pixels of each pixel, transforming the results into a binary number [8]. The histogram of these patterns forms the face descriptor [11]. The LBP operator is defined in Equation (3)

Formula:

$LBP(x,y) = \sum_{r=0}^{Q-1} v(g_k - g_c).2^r$……(3)

where g_c is the centre pixel value, g_k is the neighbour pixel value, and s(x) = 1 if x>=0, else 0.

2.1.4. Support Vector Machines (SVM)

SVM classifies faces by determining the optimal hyperplane that separates different classes [9]. It is particularly useful for face verification tasks [12]. The optimization problem for SVM is formulated as given in Equation (4)

Formula:

$min_{v,b}\ \frac{1}{2}\ ||v||^2 + C\ \sum_{i=1}^{M}\varepsilon_i$ ……(4)

Subjected to $y_i\ (v.x_i\ + b)\ \geq 1 - \varepsilon_i$ and $\varepsilon_i\ \geq 0$, where ε_i are slack variables.

2.2. Deep Learning Approaches

Convolutional neural networks (CNNs) have significantly advanced facial recognition by learning complex features from large datasets, improving accuracy across varying conditions like lighting and angles. This progress has expanded facial recognition's use in security, surveillance, and interactive user experiences [5][14][18].

2.2.1 Convolutional Neural Networks (CNNs)

CNNs have advanced facial recognition by learning features from large datasets, enhancing accuracy under varying conditions. This progress has broadened its use in security, surveillance, and user experiences [5][14][18].

Formulas:

- Convolution:
  $Conv\ (X,W) = \sum_{i=1}^{k}\sum_{j=1}^{k} X_{i,j}\ .\ W_{i,j}$ ……...(5) where X is the input image and W is the filter.
- ReLU:
  $ReLU\ (x) = \max\ (0,x)$ ……..……(6)
- Pooling:

$MaxPooling\ (X) = max_{(i,j)\ \in\ window} X_{i,j}$ ….(7)

2.2.2 Transfer Learning (Resnet pretrained Model)

In the context of transfer learning, a pre-trained ResNet model, such as ResNet-50 or ResNet-101, is utilized, having been previously trained on a comprehensive dataset like ImageNet [8],[11].

Res Net Architecture:

The ResNet framework consist of several residual blocks, each designed with multiple convolutional layers, batch normalization, and ReLU activation functions. Each block also incorporates a shortcut connection that bypasses the convolutional layers, allowing the input to be added directly to the outputs. The output of a residual block is mathematically represented as follows:

www.ijprems.com
editor@ijprems.com
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, Decembaer 2024, pp : 1572-1580
e-ISSN : 2583-1062
Impact Factor : 7.001

$y = F(x, \{W_i\}) + x$ .....(8)

where

- X is a input of residual block
- $F(x, \{W_i\})$ *represents the residual function*
- $\{W_i\}$ *weights of the convolutional layers*
- y represents the output of the block after the input x has been added.

2.2.3 Deep Metric Learning

Deep metric learning focuses on training models to create an embedding space in which similar faces are positioned closer together. Popular methods include Triplet Loss and Contrastive Loss [12],[16].

Accuracy, Precision, Recall F1 score.

Triplet loss:

$L = \max\left(0, b\left(g(x_a), g(x_p)\right) - b\left(g(x_a), g(x_n)\right) + \alpha\right)$.....(9)

Where x_a is an anchor, x_p is a positive sample, x_n is a negative samples, and α is margin.

Contrastive loss:

$L = \frac{1}{2}(1 - x)c^2 + \frac{1}{2}x \max(0, m - c)^2$.....(10)

where x is the label (0 or 1) and c is the distance between embeddings.

2.2.4. Face Net and Deep Face

FaceNet and DeepFace are landmark models that utilize deep architectures to achieve high accuracy in face recognition tasks. FaceNet uses a deep network with a triplet loss function, while DeepFace uses a 3D model for face alignment and a deep network for feature extraction [30],[21].

2.3. Comparative Analysis

2.3.1. Feature Representation

- Traditional methods rely on manually crafted features (e.g., edges, textures) [5],[7].
- Deep learning models learn features automatically from data [22],[30].

2.3.2. Performance

- Deep learning models generally outperform traditional methods in both accuracy and resilience, especially when applied to large datasets and complex conditions. [37],[21].

2.3.3. Computational Complexity

- Traditional methods often require less computational power but may be less effective on large datasets [7].
- Deep learning models demand substantial computational power but are able to utilize parallel processing effectively (e.g., GPUs) [22],[37].

2.3.4. Adaptability

- Deep learning models are more adaptable to new data and variations in lighting, pose, and occlusions than traditional methods. [21],[22].

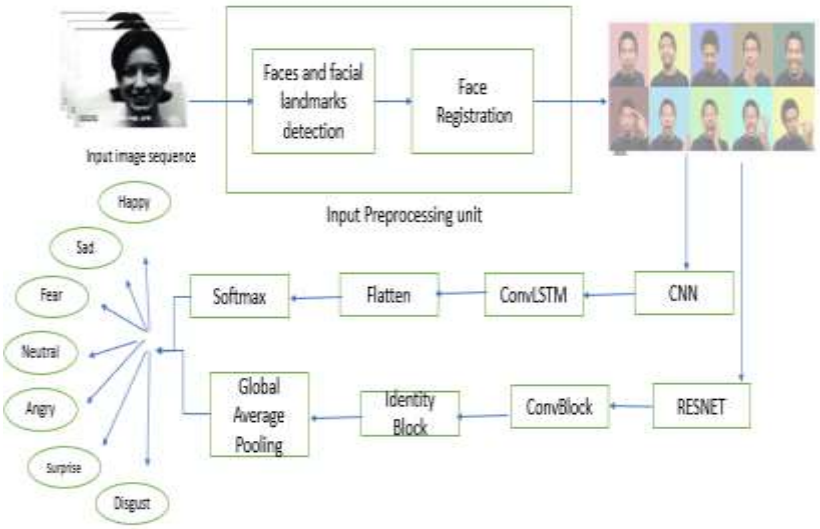
Data Augmentation: This technique expands datasets by generating variations of existing data, improving model performance and generalization. In facial expression detection, common methods include rotation, flipping, scaling, adding noise, and adjusting brightness or contrast. These modifications help the model resist overfitting and enhance its ability to handle diverse real-world scenarios [13].

3. METHODOLOGY

3.1 Architecture

Our proposed emotion detection system employs two distinct methods to enhance emotion recognition accuracy: Convolutional Neural Networks (CNNs) and a pre-trained ResNet model. Each methodology is executed separately to evaluate their individual performances. Figure 1 details the architecture and components of both approaches.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, Decembaer 2024, pp : 1572-1580
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001



3.1.1 Data Pre-processing:

Face Detection:

An algorithm for face detection is applied to locate and isolate faces within images [37].

Normalization:

The input images are standardized to a fixed size, with pixel values scaled to a normalized range for uniformity [36].

Data Augmentation:

To enrich the training dataset, this approach involves random transformations, such as rotation, scaling, and translation, to increase data variety [8]

3.1.2 Feature Extraction:

Convolutional Layers :

Convolutional layers in CNNs automatically learn and extract hierarchical features from images by applying filters to produce feature maps that highlight patterns like edges, textures, and shapes. Deeper layers build on earlier features, merging simple patterns to identify complex ones, such as object parts or entire objects [2], [6], [10]. This hierarchical learning allows CNNs to capture spatial relationships and enhance performance in complex image recognition tasks [11], [12].

Pooling Layers:

Pooling layers in CNNs reduce the spatial dimensions of feature maps while preserving key information. Max pooling selects the maximum value from localized regions, and average pooling calculates the average, both downsampling the data. This dimensionality reduction decreases computational load and memory requirements, improving efficiency. Pooling also introduces translation invariance, helping the network generalize better to variations in input images [4], [18], [22].

3.1.3 Classification:

Fully Connected Layers: The extracted features are connected to fully connected layers for classification. [3], [9].

Softmax Layer: To generate the probabilities for each emotion class in the output layer, a softmax activation function is applied.[3],[7],[16]

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, Decembaer 2024, pp : 1572-1580
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

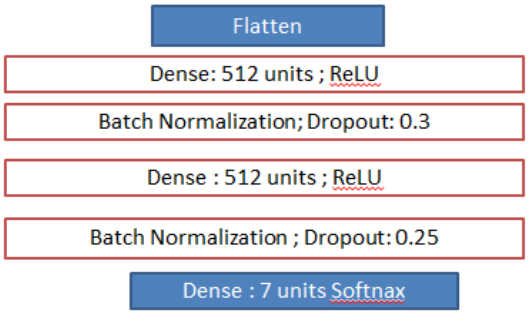


Fig 2: Proposed Custom designed CNN

3.2 Algorithm

3.2.1 Convolutional Neural Networks (CNNs):

Convolutional Neural Networks (CNNs) are perfect for applications like facial expression-based emotion detection because they automatically learn hierarchical representations of spatial data from images [2], [24]. Convolutional layers create feature maps that emphasize various elements, including edges, textures, and patterns, by using learnable filters (kernels) that move across the image [22], [19].

In order to generate an output value in the feature map, these filters multiply the image element-by-element and then add up the results [36]. CNNs that use many filters are able to collect a wide variety of features, producing detailed feature maps that depict different facets of the image [15], [30].

3.2.1.2 Activation function: Activation functions, like the Rectified Linear Unit (ReLU), introduce nonlinearity to neural networks, enabling them to learn complex patterns. Defined as ReLU(x) = max(0, x), ReLU preserves positive values and sets negative ones to zero. This function also helps address the vanishing gradient problem, which can hinder deep network training.

3.2.1.3 Pooling Layers: Pooling layers reduce the spatial dimension of feature maps, lowering the computational load and the number of parameters [4], [3]. Max-pooling, the most common method, selects the maximum value from each feature map patch. This improves the model's resilience to small translations and distortions while preserving important features and reducing spatial resolution [5].

3.2.1.4 Fully connected layer: The last step, fully connected (dense) layers, is in charge of classification. To create predictions, they employ high-level information from pooling and convolutional layers [3], [4]. A probability distribution across emotion classes is produced by running the output through a softmax algorithm [3].

3.2.1.5 Softmax Layer: The softmax layer transforms the output scores from the fully connected layers into probabilities that sum to 1, representing the likelihood of each emotion class. The softmax function is defined as follows:

$$(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}},$$

where xi are the scores, and K is the number of classes. This layer allows the model to makes a final predictions by choosing the class with the highest probabilities [1]..

3.2.2 Pre-trained Model (ResNet):

Residual Networks (ResNet) represent a CNN architecture that incorporates shortcut connections to address the vanishing gradient issue, facilitating the efficient training of much deeper networks [36].

3.2.2.1 Residual Learning: Learning residual functions in relation to layer inputs is the main goal of residual learning. A residual block that includes a shortcut connection that avoids layers is expressed as y = F(x, {Wi}) + x. The utilization of very deep networks is made possible by these shortcut connections, which maintain gradient flow and alleviate the vanishing gradient issue [36].

Transfer learning applies pre-trained models to new tasks, leveraging existing feature representations to reduce the need for extensive training data. ResNet, pre-trained on large datasets like ImageNet, provides learned features that support emotion detection. By replacing the final classification layer and fine-tuning the model, we adapt ResNet to our specific emotion classes. The early layers are frozen, while later layers are fine-tuned to our dataset, retaining general features while adapting to facial expressions [30].

Using ResNet's deep architecture and pre-trained weights enhances accuracy and generalization in emotion detection, helping distinguish subtle emotional differences. Transfer learning accelerates training and improves performance, especially when labeled data is scarce [34].

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, Decembaer 2024, pp : 1572-1580
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

3.3. Implementation:

1. Experimental setup:

For our facial expression detection, we utilized two deep learning models: a Convolutional Neural Network (CNN) and a pre-trained ResNet. The CNN consists of convolutional layers (128–512 filters), combined with Max Pooling, Batch Normalization, and Dropout layers to aid in feature extraction and reduce overfitting. Fully connected layers with ReLU activation and a softmax layer classify images into seven emotion categories [11]. The ResNet model uses residual blocks with shortcut connections to preserve gradient flow in deeper networks, with global average pooling before the final dense layer [36]. During training, image data augmentation was applied, with a 20% validation split. Both models used Adam optimizers with categorical cross-entropy loss and were monitored with Early Stopping and Reduce LR On Plateau callbacks. Each model trained for up to 30 epochs, after which performance was evaluated on a test set, and the best-performing model was saved for future use [17].

2. Datasets used in study:

For both CNN and ResNet models, selecting the right dataset and managing its characteristics are crucial for model performance.

- **Image Resolution**: Higher resolution images provide more detailed features but require more computational resources. CNN and ResNet models perform well with standard resolutions like 48x48, 64x64, or 224x224 pixels [14].
- **Dataset Size**: Larger datasets help create more robust models and reduce overfitting. For ResNet, datasets with tens to hundreds of thousands of images are ideal for training [17].

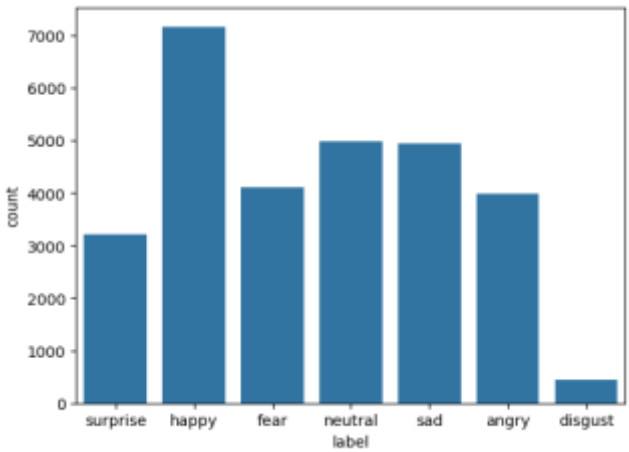
Table:1

Datasets	Categories						
	Happy	Fear	Sad	Neutral	Angry	Surprise	Disgust
Train Data	7215	4097	4830	4965	3995	3171	436
Test Data	1774	1024	1247	1233	958	831	111

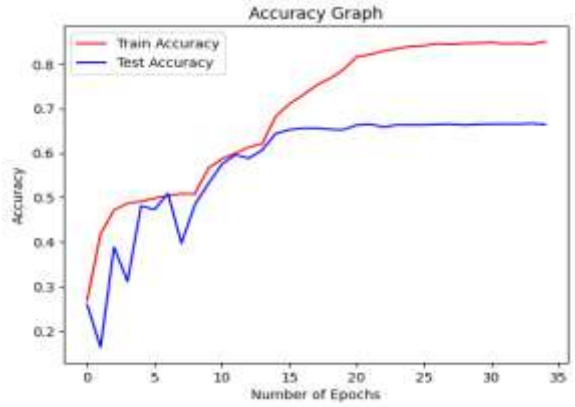
4. RESULT ANALYSIS

Table:2

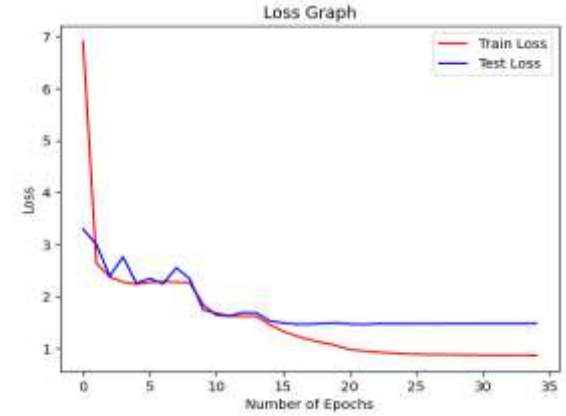
Algorithms	CNN	Resnet
Accuracy	0.6650155675063685	0.5424568355505236
Precision	0.6640840468657843	0.5469521741149242
Recall	0.6335062476927534	0.48058634730594807
F1- Score	0.6440903976383091	0.4865741830674475
Error Rate	0.3349844324936315	0.4575431644494764

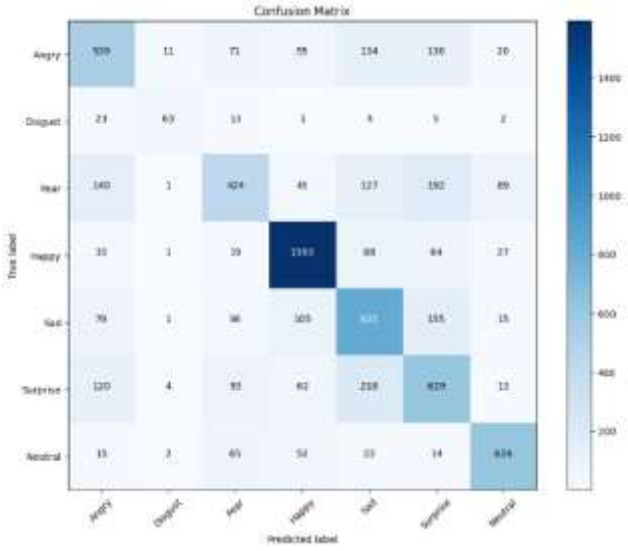


INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, Decembaer 2024, pp : 1572-1580
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

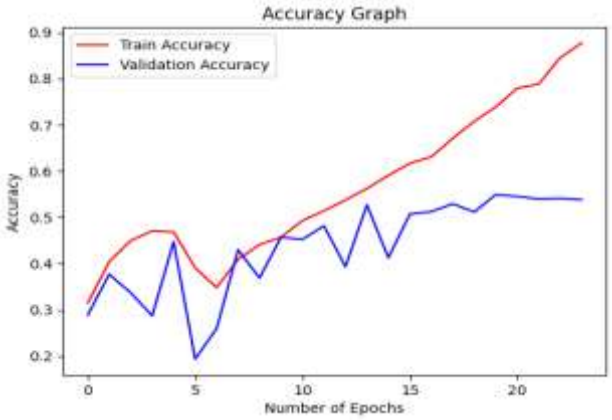


CNN:

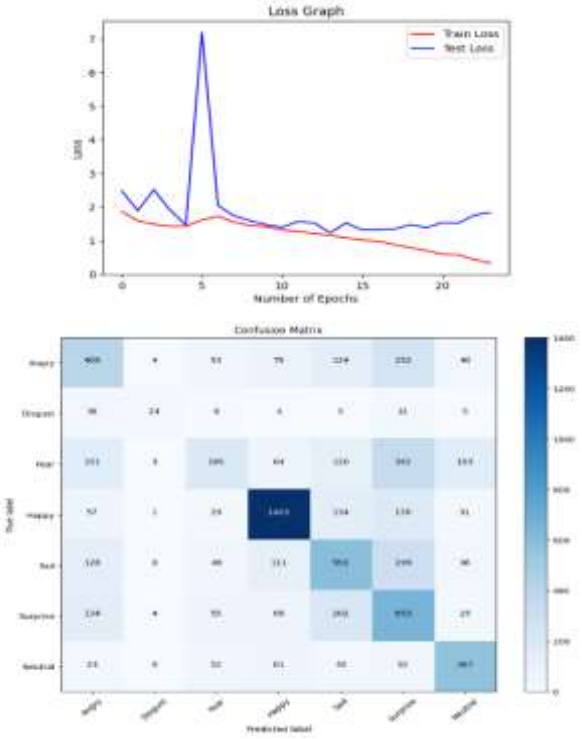




RESNET



INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, Decembaer 2024, pp : 1572-1580
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001



5. CONCLUSION/FUTURE SCOPE

In this implementation, both Convolutional Neural Networks (CNN) and Residual Networks (ResNet) were employed for facial expression recognition. The CNN model extracts hierarchical features using convolutional layers, max-pooling, and dropout, similar to Arriaga et al. (2019)[3] for emotion classification. The ResNet model, addressing challenges in deeper networks like gradient vanishing, uses residual connections as proposed by He et al. (2016)[36].

To improve generalizability, expanding the dataset and utilizing advanced data augmentation, as in Shi et al. (2021), could enhance performance. Transfer learning with pre-trained models, effective in facial recognition tasks, may also accelerate training. This model has potential for real-time emotion detection, inspired by Amal et al. (2022) using the FER2013 dataset[6]. To ensure fairness and accuracy in diverse real-world applications, evaluating with additional metrics and addressing biases, as noted by Li and Deng (2018), would be essential for strengthening its performance[14].

6. REFERENCES

[1] L. Pham , T. H. Vu, & T. A. Tran, "Facial Expression Recognition Using Residual Masking Network", 2020 25th International Conference on Pattern Recognition (ICPR), 4513–4519, (2021).

[2] C. Shi, C. Tan, & L. Wang, "A Facial Expression Recognition Method Based on a Multibranch Cross-Connection Convolutional Neural Network", IEEE Access, 9, 39255–39274 (2021).

[3] O. Arriaga, H. Bonn-Rhein-Sieg, & M. Valdenegro, "Realtime Convolutional Neural Networks for Emotion and Gender Classification" (2019).

[4] A. Vulpe-Grigoraşi, O. Grigore, "Convolutional Neural Network Hyperparameters optimization for Facial Emotion Recognition", 2021 12th International Symposium on Advanced Topics in Electrical Engineering (ATEE), (2021).

[5] Pandey, A., Kumar, "A. Facial Emotion Intensity: A Fusion Way", SN COMPUT. SCI. 3, 162 (2022).

[6] Amal, V. S., Suresh, S., & Deepa, G. ,"Real-time emotion recognition from facial expressions using convolutional neural network with Fer2013 dataset", In Ubiquitous Intelligent Systems (pp. 541-551). Springer, Singapore (2022).

[7] Dino, H. I., Abdulrazzaq, M. B., "Facial expression classification based on SVM, KNN and MLP classifiers", 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 70- 75). IEEE . (2019).

[8] Shima, Y., Omori, Y, "Image augmentation for classifying facial expression images by using deep neural network pre-trained with object image database". (2018).

[9] Liu, J., Wang, H., Feng, Y., "An End-to-End Deep Model With Discriminative Facial Features for Facial Expression Recognition", IEEE Access, 9 (2021),

[10] N. Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)," SN Applied Sciences, vol. 2, no. 3, pp. 1-8, (2020).

[11] S. Minaee, A. Abdolrashidi, "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network" (2019).
[12] A. Saravanan, G. Perichetla, and D. K. Gayathri, "Facial emotion recognition using convolutional neural networks," arXiv preprint arXiv:1910.05602, (2019)
[13] Shorten, C., Khoshgoftaar, T. M., "A survey on image data augmentation for deep learning". Journal of big data, 6 (2019).
[14] S. Li and W. Deng, "Deep facial expression recognition: A survey," arXiv preprint arXiv:1804.08348 (2018).
[15] Burkert, P., Trier, F., Afzal, M. Z., Dengel, A., Liwicki, M. "Dexpression: Deep convolutional neural network for expression recognition" (2015).
[16] Y. Khaireddin, Z. Chen, Facial Emotion Recognition: State of the Art Performance on FER2013, (2021).
[17] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I.Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression.
[18] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E.Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. (2012).
[19] Xu, Q., Wang, C., Hou, Y. Attention Mechanism and Feature Correction Fusion Model, for Facial Expression Recognition. In 2021 6th International Conference on Inventive Computation Technologies (ICICT) (pp. 786-793). IEEE, (2021).
[20] S. Setty et al., "Indian movie face database: a benchmark for face recognition under wide variations," in 2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), (2013).
[21] Kollias D., & Zafeirio S., "Expression, affect, action unit recognition: Aff-wild2, multi-task learning and arcface." In IEEE transactions on Pattern Analysis and Machine Intelligence, (2019).
[22] I. Goodfellow, Y. Bengio, & A. Courville, "Deep Learning", MIT Press, (2016).
[23] A. Kumar & R. Vohra, "Advances in Deep learning techniques for Facial Expression Recognition", Journal of Ambient Intelligence and Humanized Computing, (2021)
[24] C. J. Huang and S. W. Wang, "Facial Expression Recognition using Deep Learning: A survey", IEEE Access, (2022).
[25] A. Gurel & H. Senoh, "A hybrid deep learning model for facial expression recognition", Journal of Computational Science, (2022).
[26] J. M. Valstar, M. Mehu, B. Jiang, M. Pantic, and K. Scherer, "Meta-analysis of the first facial expression recognition challenge," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 42, no. 4, pp. 966-979, (2012).
[27] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Facial landmark detection by deep multi-task learning," in European Conference on Computer Vision (ECCV), pp. 94-108. Springer, Cham, (2014).
[28] S. Wu, S. E. McKeown, Q. Zhang, M. W. Lin, R. J. Sclabassi, and M. F. Abdel-Mottaleb, "Real-time facial expression recognition from thermal infrared video based on a deep neural network," IEEE Transactions on Affective Computing, vol. 10, no. 4, pp. 438-450, (2019).
[29] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1-10, (2016).
[30] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815-823, (2015).
[31] Z. Huang, Y. Qin, and F. R. Chung, "Facial expression recognition using improved deep CNNs with hybrid preprocessing techniques," in IEEE Transactions on Multimedia, vol. 22, no. 1, pp. 188-198, (2020).
[32] M. Liu, S. Li, S. Shan, and X. Chen, "AU-inspired deep networks for facial expression feature learning," Neurocomputing, vol. 159, pp. 126-136, (2015).
[33] Y. Xie, R. Zheng, and Y. Zhang, "Facial expression recognition based on stacked autoencoder and extreme learning machine," Cognitive Computation, vol. 11, no. 4, pp. 562-571, (2019).
[34] C. Cao, Y. Liu, and M. G. Liu, "Cross-database facial expression recognition via transferable feature subspace learning," IEEE Access, vol. 6, pp. 58146-58154, (2018).
[35] X. Tang, and M. Wang, "Deeply-supervised facial expression recognition: A new benchmark and evaluation metrics," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 3, pp. 754-765, (2020).
[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, (2016).
[37] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multi-task Cascaded Convolutional Networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, (2016).