

Facial Expression for emotion recognition using Infrared Imaging

Rupali J.Dhabarde¹, Dr. D.V. Kodavade², Dr. Vijay Manwatkar³, Aditya Konnur⁴

Department of Technology, Shivaji University, Kolhapur^{1,4}, DKTE society's Textile & Engineering Institute Ichalkaranji², Vidya Vikas Arts, Commerce & Science College, Samudrapur Dist- Wardha, India³

Abstract: Facial expression recognition (FER) plays a crucial role in human-computer interaction, with applications ranging from healthcare to security systems. This study proposes a robust FER system leveraging thermal imaging for emotion recognition. Utilizing the Kaggle dataset "Chia theo nguoi_KTFEV2-7 emotions," our approach focuses exclusively on thermal facial images, comprising 2,538 labeled samples spanning seven emotional categories: happiness, sadness, anger, fear, surprise, disgust, and neutrality. The VGG-16 convolutional neural network architecture, combined with a Support Vector Machine (SVM) classifier, is employed to identify emotions. The model's performance is evaluated using accuracy, precision, recall, and F1-score. Experimental results demonstrate the system's effectiveness, with classification accuracies of up to 94% for neutral expressions and 93% for fear. Despite challenges in detecting surprise (52% accuracy), the approach proves promising for FER applications in thermal imaging, highlighting its potential for emotion analysis in non-ideal visual conditions.

Introduction: Facial expression recognition (FER) has become an integral aspect of human-computer interaction, with diverse applications ranging from healthcare and security systems to virtual reality and emotional intelligence analysis. By analyzing facial expressions, machines can interpret human emotions, enhancing their ability to respond appropriately in real-time scenarios.

Objectives: propose a robust FER system utilizing thermal images to classify facial expressions into seven distinct emotions: happiness, sadness, anger, fear, surprise, disgust, and neutrality.

Methods: The proposed model (as shown in figure 1) for facial expression recognition utilizes the VGG-16 architecture, a widely recognized convolutional neural network (CNN) designed for image classification tasks. VGG-16 is known for its simplicity, depth, and effectiveness in extracting hierarchical image features. Below, the architecture's components and their contributions are detailed

Results: The proposed model is trained with data as shown in table no. 1 and tested with 200 images. Figure 2 shows feature map generation for anger and happy emotion.

Conclusions: The model achieved an accuracy of 90.52%, which indicates its capability to correctly classify emotions. The precision of 95.59% which indicates the model's ability to identify true positive instances of emotion without being misled by false positives. And a recall of 93.44% says the effectiveness in capturing true positive emotions.

Keywords: FER, VGG, SVM, neural network, machine learning

1. Introduction

Facial expression recognition (FER) has become an integral aspect of human-computer interaction, with diverse applications ranging from healthcare and security systems to virtual reality and emotional intelligence analysis. By analyzing facial expressions, machines can interpret human emotions, enhancing their ability to respond appropriately in real-time scenarios. While conventional FER systems rely on visual images captured under optimal lighting conditions, their performance can degrade significantly in challenging environments, such as low light,

occlusion, or extreme weather. To address these limitations, thermal imaging has emerged as a promising alternative for facial expression analysis. Unlike visual imaging, thermal imaging captures heat signatures emitted by the human face, providing a reliable modality for emotion recognition regardless of external lighting conditions. This modality is particularly useful in scenarios like night-time surveillance, medical diagnostics, or environments where visual features are obscured.

In this study, we propose a robust FER system utilizing thermal images to classify facial expressions into seven

distinct emotions: happiness, sadness, anger, fear, surprise, disgust, and neutrality. The system employs the VGG-16 convolutional neural network (CNN) for feature extraction and a Support Vector Machine (SVM) for classification. The Kaggle dataset "Chia theo ngui_KTFEV2-7 emotions" is utilized, focusing exclusively on its thermal image directory, which contains 2,538 labeled samples of diverse emotional expressions.

The proposed system is evaluated based on accuracy, precision, recall, and F1-score. Experimental results highlight the model's effectiveness in recognizing emotions, with accuracies of up to 94% for neutral expressions and 93% for fear. However, the study also identifies challenges in detecting surprise expressions (52% accuracy), suggesting avenues for further improvement. By leveraging the unique strengths of thermal imaging, this work demonstrates a viable solution for robust emotion recognition in non-ideal visual conditions, paving the way for advanced FER systems in real-world applications.

2. RELATED WORK

The paper reviews the known factors of thermal versus RGB imaging for facial emotion recognition, highlighting the ongoing research in thermal physiology since the late nineties and its intersections with various fields such as medicine, psychology, machine learning, optics, and affective computing. It discusses the limitations of collecting and developing thermal facial emotion recognition (FER) data for AI training, emphasizing the challenges in transitioning to thermal imagery as a reliable source for human-centered AI tasks, which requires high fidelity data sources across multiple demographics and thorough validation [1].

The literature review discusses various methodologies for recognizing human emotions from thermal images, highlighting the effectiveness of different deep learning models such as Faster R-CNN, SSD, and IRFacExNet. It notes that the Faster R-CNN model trained with ResNet50 achieved a mean average precision of 100 on the OSU thermal dataset, while IRFacExNet achieved a recognition accuracy of 81.16 for facial expressions in thermal images. The review emphasizes the challenges faced in human detection and emotion recognition due to environmental factors like lighting conditions, occlusions, and weather conditions. It also mentions the importance of performance metrics such as accuracy, precision, recall, and F1-score in evaluating

the effectiveness of the proposed methodologies in the context of thermal face recognition[2].

Emotions play a crucial role in mammals, providing vital information for survival and adaptation to the environment. The perception of emotions involves the ability to recognize and respond to emotional cues in oneself and others, highlighting the distinction between emotions, which are rapid and unconscious responses, and feelings, which are conscious interpretations of those emotions. The literature emphasizes that emotions manifest as physical responses with specific physiological activation patterns, but different emotions can exhibit similar physiological responses, such as increased heart rate in both fear and anger. This complexity illustrates the challenges in accurately recognizing and classifying emotions, which can be influenced by various contextual factors [3].

The paper [4] addresses the challenges of facial emotion identification due to factors such as darkness and lighting conditions, suggesting that thermal images can effectively overcome these issues while providing various benefits for emotion recognition. It introduces a novel approach that focuses on significant regions of the face, specifically the left eye, right eye, and lips, to enhance processing efficiency and accuracy, utilizing techniques like ten-folded cross-validation with Convolutional Neural Networks (CNN) and parallelism to reduce processing time by 50%.

The paper proposes a light-weight convolutional neural network (CNN) for real-time and extensive facial emotion recognition, enhancing classification performance by removing the fully connected layer, integrating residual modules, and utilizing depth-wise separable convolutions. This approach significantly reduces the number of parameters in the convolutional layer, making the model more portable while achieving good detection results on images taken outside of the dataset. Another study discussed in the literature review presents an automatic video-based facial expression detection system that classifies human facial expressions from image sequences. This system consists of two main components: feature extraction and peak expression detection from frames. The authors recommend using the Double Local Binary Pattern (DLBP) method to identify peak expression characteristics, which effectively reduces detection time and minimizes dimensional size [5].

The paper highlights the growing importance of facial emotion recognition (FER) in various applications such as surveillance, activity recognition, home automation, and human-computer interaction, indicating a significant interest in improving emotion analysis technologies. It addresses the challenges faced in emotion recognition due to occlusions, lighting fluctuations, and changes in physical appearance, emphasizing the need for improved research and methodologies to enhance the accuracy and reliability of FER systems [6].

The study highlights the challenges in detecting and analyzing facial expressions related to fear and deception, emphasizing the limitations of traditional lie detection methods which are often inaccurate and time-consuming. It discusses the application of Support Vector Machines (SVMs) for automatic recognition and classification of facial expressions, detailing how SVMs can effectively tackle both linear and non-linear problems, and how they create optimal boundaries for classifying data points based on facial features extracted from thermal images [11].

The paper discusses the significance of facial recognition technology, highlighting its applications in various fields such as cyber security, crime cases, and biometrics, which have attracted considerable research interest in computer vision. It details the methodology employed in the study, which involves the combination of Convolutional Neural Networks (CNN) and Support Vector Machines (SVM) to analyze two face datasets, ultimately achieving an accuracy of up to 58% in lie detection tasks [12].

3. System Methodology

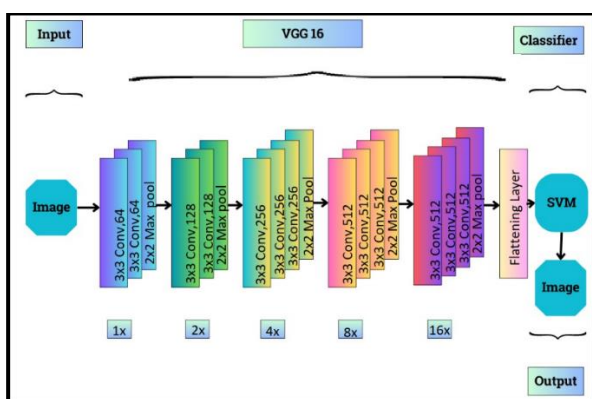


Fig. 1. Model Architecture

The proposed model (as shown in figure 1) for facial expression recognition utilizes the VGG-16

architecture, a widely recognized convolutional neural network (CNN) designed for image classification tasks. VGG-16 is known for its simplicity, depth, and effectiveness in extracting hierarchical image features. Below, the architecture's components and their contributions are detailed.

Dataset

We evaluate the proposed approach on dataset for facial expression recognition, the dataset used in this research is taken from Kaggle (Chia theo nguoi_KTFEV2-7 emotions) [7] which is carefully curated to include a diverse range of examples that cover various emotions. The dataset contains two directories, one for thermal images and other for normal/visual images of which we have used the thermal directory. The datasets consist of 2538 labeled thermal facial images corresponding to different emotional expressions such as happiness, sadness, anger, fear, surprise, and disgust. We compare the performance of our method with accuracy, precision, recall, and F1-score.

Preprocessing

The thermal images are carefully processed to ensure consistency in dimensions and quality. Techniques like resizing, normalization, and augmentation are applied to improve the robustness of the model. The dataset contains 2,538 labeled thermal facial images.

Model Architecture

Backbone Model: The architecture is based on VGG-16, a widely used convolutional neural network (CNN) for image classification. The model begins with the input layer, which accepts preprocessed thermal facial images. These images are resized to a fixed dimension to ensure compatibility with the network and maintain uniformity during processing. The thermal nature of the input emphasizes heat signatures rather than visual features, making the method robust against challenges such as lighting variations and occlusion [8,9]

Convolutional Layers: The network includes 13 convolutional layers with 3x3 filters, followed by ReLU activation functions to extract spatial and feature hierarchies. The architecture includes 13 convolutional layers, each using 3x3 filters with a stride of 1. These filters are designed to capture spatial patterns such as edges, textures, and higher-level features. The convolutional layers are grouped into blocks, with the number of filters doubling in successive blocks (e.g., 64, 128, 256, 512). This gradual increase in filter count

enhances the network's capacity to learn complex features from the input images. Each convolutional operation is followed by a ReLU (Rectified Linear Unit) activation function, which introduces non-linearity, enabling the model to learn intricate patterns.

Pooling Layers: Max-pooling layers are added after convolution blocks to reduce the spatial dimensions while retaining essential features. After each convolutional block, a max-pooling layer is applied. The pooling layers use a 2x2 filter to reduce the spatial dimensions of the feature maps while retaining the most significant features. This process not only minimizes computational complexity but also makes the model invariant to minor translations in the input images. The pooling operation is essential for progressively condensing the feature maps into compact representations [10].

Flattening Layer: The final convolutional output is flattened to prepare the feature map for classification. The output from the final convolutional block is a multidimensional feature map. This is passed through a flattening layer, which converts the feature map into a one-dimensional vector. This step bridges the convolutional feature extraction layers with the fully connected layers or classifier.

Classifier: The extracted features are passed to a Support Vector Machine (SVM), which is used to classify the input images into the seven emotion categories. The SVM classifier is chosen for its ability to handle high-dimensional data and binary/multiclass problems effectively. Instead of traditional fully connected layers for classification, the model employs a Support Vector Machine (SVM) as the final classifier. The SVM is adept at handling high-dimensional data, making it an ideal choice for the feature-rich output from the VGG-16 network. The SVM classifies the flattened features into one of the seven emotion categories: happiness, sadness, anger, fear, surprise, disgust, and neutrality. Its robust decision boundaries ensure precise categorization of thermal features.

Training the Model

- The VGG-16 model, pre-trained on large-scale datasets, is fine-tuned with the thermal facial expression dataset. This transfer learning approach leverages pre-learned features while adapting to the specific characteristics of thermal imaging.

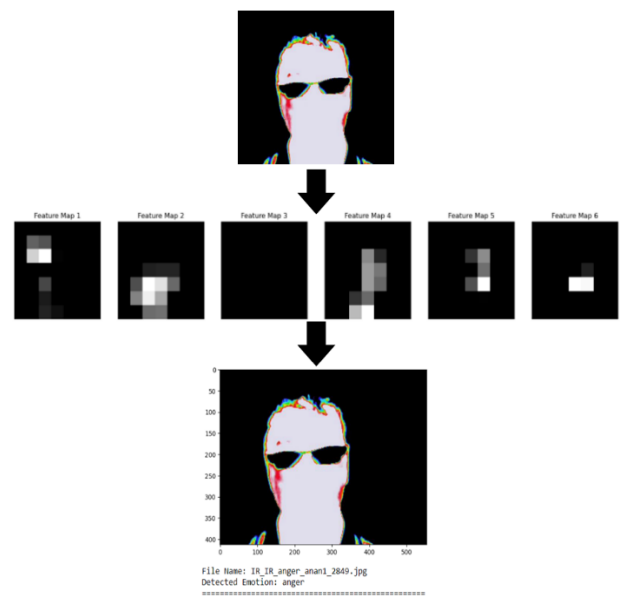
- The dataset is split into training, validation, and testing subsets. The model is trained using a categorical cross-entropy loss function and optimized using an Adam optimizer to ensure efficient convergence.

4. Experimental Result And Discussion

The proposed model is trained with data as shown in table no. 1 and tested with 200 images. Figure 2 shows feature map generation for anger and happy emotion.

Table no. 1 Number of images trained per emotion

Emotion	No. Of Images
Anger	238
Disgust	178
Fear	364
Happy	400
Neutral	336
Sadness	398
Surprise	224
Total	2138



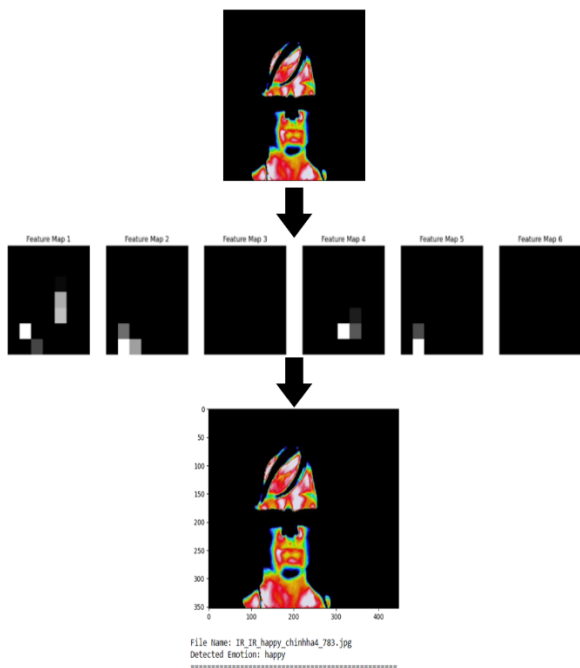


Fig. 2 Feature Map generation for Anger and Happy Emotion

It shows the prediction performance of the VGG16_SVM model for anger and happy emotion.

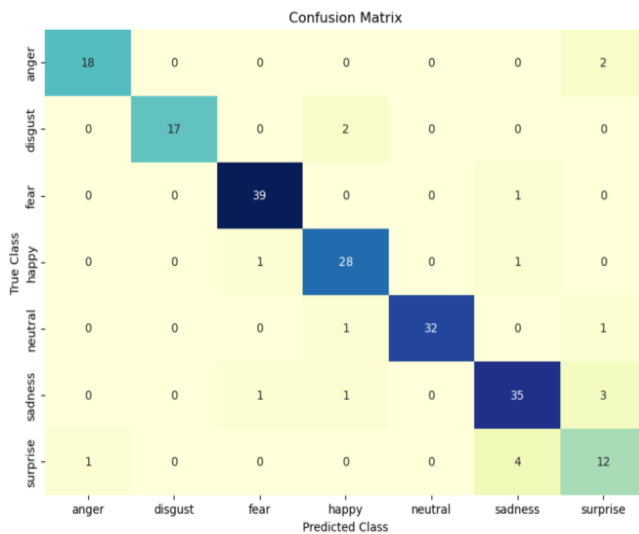


Fig. 3 Confusion Matrix

The confusion matrix shown in fig. 3 figures out which classes had, how many correct predictions and how many were wrongly predicted. The diagonal of the confusion matrix, indicated in blue in the heatmap, represents the correct predictions. Each element on the diagonal corresponds to the number of emotion that were correctly classified for each class. The cells outside the diagonal, displayed in yellow in the heatmap, represent the wrong predictions. Each element outside the diagonal corresponds to the number of instances that were misclassified for a specific class.

From this confusion matrix, the classes that had the most correct predictions and the least correct predictions have been identified. The class with the most correct predictions is fear, On the other hand, the class with the least correct predictions is surprise.

Table 2 Outcome Measure Parameters

Emotion	True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)
Anger	18	0	1	2
Disgust	17	0	0	2
Fear	39	0	2	1
Neutral	28	0	4	2
Happy	32	0	0	2
Sadness	35	0	6	5
Surprise	12	0	6	5

Table 2 indicates the model's performance for each emotion. It shows the number correctly predicted, incorrectly predicted and false negative values. We observe the strengths and weaknesses of our model from these values. For example, it performed well in detecting fear but struggled with surprise. The overall accuracy and other metrics have also been calculated using these values to provide a more comprehensive assessment of the model's performance.

In the context of emotion detection, where the dataset is labeled and each instance belongs to one specific emotion class, there cannot be true negatives for any particular emotion class. The absence of true negatives in the provided confusion matrix indicates that the model did not mistakenly classify any instances as negative (non-emotion) for any emotion class. The concept of true negatives typically applies to binary classification problems, where one class is considered positive and the other negative. In such cases, true negatives represent instances correctly classified as the negative class. However, in emotion detection, the focus is on correctly identifying instances belonging to specific emotion classes, rather than considering a separate negative class. Therefore, in the given confusion matrix, the absence of true negatives is expected and does not indicate any issues with the model or the evaluation. The focus should be on the true positives, false positives, and false negatives, which reflect the model's performance in correctly

identifying instances of each emotion and any misclassifications made.

Table 3 Precision, Recall and F1-Score

Emotion	Anger	Disgust	Fear	Happy	Neutral	Sadness	Surprise
Precision	0.95	1.00	0.95	0.8	1.00	0.85	0.67
Recall	0.9	0.89	0.97	0.93	0.94	0.88	0.71
F1 Score	0.92	0.94	0.96	0.9	0.97	0.86	0.69

The VGG16_SVM model achieved high precision values for most emotions, ranging from 0.88 to 1.00. For example, it achieved a precision of 0.95 for anger, indicating that 95% of instances predicted as anger were actually correct. The VGG16_SVM model obtained high recall values, ranging from 0.89 to 0.97, for different emotions. For instance, it achieved a recall of 0.93 for happy, indicating that it correctly identified 93% of all actual instances of happiness. The VGG16_SVM model achieved high F1 scores, ranging from 0.86 to 0.97, for different emotions. For example, the F1 score for neutral is 0.97, indicating a balance between precision and recall for that emotion class.

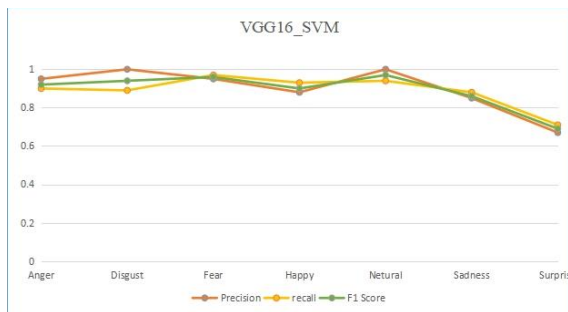


Fig.4 evaluation metrics of VGG16_SVM Model

These evaluation metrics from Fig. 4 provide the information of VGG16_SVM model's performance for each emotion category. High precision values indicate accurate predictions, high recall values signify the model's ability to capture instances of a specific emotion, and the F1 score provides an overall assessment of the model's performance by considering both precision and recall. The VGG16_SVM model demonstrates strong performance across various emotions, as evidenced by the high precision, recall, and F1 score values.

Table 4 Accuracy (%) score of All Emotions

Emotions	Accuracy (%)
Anger	86
Disgust	89
Fear	93
Happy	82
Neutral	94
Sadness	76
Surprise	52

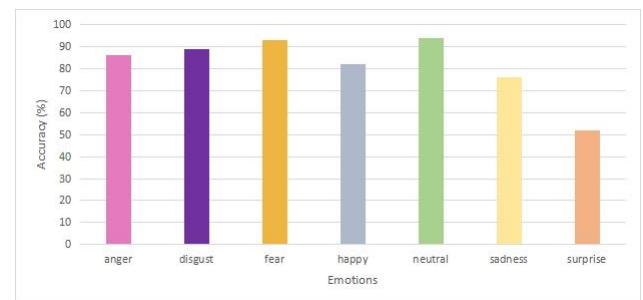


Fig. 5 Model's Achievement Level

CONCLUSION

The model achieved an accuracy of 90.52%, which indicates its capability to correctly classify emotions. The precision of 95.59% which indicates the model's ability to identify true positive instances of emotion without being misled by false positives. And a recall of 93.44% says the effectiveness in capturing true positive emotions. This high F1-score i.e. 0.97, indicates that this model maintains a strong balance between identifying true emotions and avoiding false detections. These results suggest that thermal imaging, when combined with advanced deep learning techniques like VGG16 and good classifiers like SVM, gives an effective solution for expression recognition. In future work, we plan to extend this research by exploring more advanced deep learning architectures and ensemble methods to further enhance the accuracy and robustness of emotion detection systems.

Conflict of Interest: There is no conflict of interest regarding the publication of this paper.

Data Availability: All data utilized in this study are publicly available and can be accessed on Kaggle at the following URL: <https://www.kaggle.com/datasets/ngkhang/ktfe2-7-emotions>.

References

- [1] Ordun, E. Raff, and S. Purushotham, "The Use of AI for Thermal Emotion Recognition: A Review of Problems and Limitations in Standard Design and Data," Department of Information Systems, University of Maryland, Baltimore County, Booz Allen Hamilton, Department of Computer Science, University of Maryland, Baltimore County, 2020.
- [2] S., Babu, Rajendra, Prasad., B., Sai, Chandana. (2023). 1. Human Face Emotions Recognition from Thermal Images Using DenseNet. International journal of electrical and computer engineering systems, doi: 10.32985/ijeces.14.2.5
- [3] Jesús, A., Ballesteros., Gabriel, M., Ramírez, V., Fernando, Moreira., Andrés, Solano., Carlos, Peláez. (2024). 2. Facial emotion recognition through artificial intelligence. Frontiers in computer science, doi: 10.3389/fcomp.2024.1359471
- [4] Basem, Assiri., Mohammad, Alamgir, Hossain. (2023). 3. Face emotion recognition based on infrared thermal imagery by applying machine learning and parallelism.. Mathematical Biosciences and Engineering, doi: 10.3934/mbe.2023042
- [5] P., Kanagaraj., M., A., Ranjith., K., Vijayasarathy. (2022). 4. Emotion detection from facial expression using image processing. International Journal of Health Sciences (IJHS), doi: 10.53730/ijhs.v6ns6.9748
- [6] Awais, Salman, Qazi., Muhammad, Farooq., Furqan, Rustam., Mónica, Gracia, Villar., Carmen, Lili, Rodríguez., Imran, Ashraf. (2022). 5. Emotion Detection Using Facial Expression Involving Occlusions and Tilt. Applied Sciences, doi: 10.3390/app122211797
- [7] <https://www.kaggle.com/datasets/ngkhang/ktfe2-7-emotions>.
- [8] Hanzhao, Pan., Huibin, Luo., Zhansen, Jiao., Jianxin, Deng. (2024). Facial Expression Recognition with an Improved VGG16 Network Based on SE Modules and Residual Connections. doi: 10.1109/mlise62164.2024.10674455
- [9] Mohammed, Nashat, Sadat. (2024). Facial Emotion Recognition using Convolutional Neural Network. Indian Scientific Journal Of Research In Engineering And Management, doi: 10.55041/ijsrem33503
- [10] Sakshi, Gupta., Anwasha, Sengupta. (2023). Unlocking Emotions Through Heat: Facial Emotion Recognition via Thermal Imaging. doi: 10.1109/icefeet59656.2023.10452206
- [11] R. Dhabarde, D. V. Kodavade, R. Deshmukh, and S. S. Zalte-Gaikwad, "Recognition of Facial Expressions in Infrared Images for Lie Detection with the Use of Support Vector Machines," in Synergistic Interaction of Big Data with Cloud Computing for Industry 4.0, 1st ed., CRC Press, 2022, pp. 163-173.
- [12] R. J. Dhabarde, D. V. Kodawade, and S. Zalte, "Hybrid Machine Learning Model for Lie-Detection," in 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), 2023, pp. 1-5.