JURNAL ILMIAH TEKNOLOGI DAN REKAYASA
Vol. 30, No. 02, August 2025, 136 – 155
p-ISSN: 1410-9093, e-ISSN: 2089-8088
https://ejournal.gunadarma.ac.id/index.php/tekno | tekrek@gunadarma.ac.id

Evaluating Logistic Regression and SVM for Image Analysis Using VGG-16, VGG-19, and Inception V3 Features

[1*]Wildan Habibi, [2]Imam Yuadi

[1]Magister of Human Resource Development, Graduate School, Airlangga University, [2]Department of Information and Library Science, Faculty of Social and Political Sciences, Airlangga University
[1]Jl. Airlangga No. 4-6, Surabaya, Indonesia, [2]Jl. Airlangga No. 4-6, Surabaya, Indonesia
[1]wildan.habibi-2024@pasca.unair.ac.id, [2]imam.yuadi@fisip.unair.ac.id

Abstract

This paper presents a comparative analysis of Logistic Regression (LR) and Support Vector Machine (SVM) classifiers for facial expression recognition using image embeddings extracted from pre-trained deep learning models: VGG-16, VGG-19, and Inception V3. The study utilizes the FER-2013 dataset, which includes five emotion classes: Angry, Fear, Happy, Neutral, and Sad. Feature embeddings were obtained from CNNs and then classified using LR and SVM. Performance was evaluated using accuracy (overall correctness), precision (correctness among predicted labels), and recall (ability to retrieve relevant instances). The highest accuracy was achieved by Inception V3 with SVM (89.3%), followed by VGG-19 (87.6%) and VGG-16 (85.4%). Confusion matrix analysis and visualization techniques (MDS and t-SNE) confirmed Inception V3's superior ability to distinguish fine-grained expressions. Notably, unlike pure end-to-end CNN classifiers, this approach leverages pretrained CNNs solely as feature extractors, leading to significantly lower training complexity and faster execution making it ideal for resource-constrained environments. This study highlights the practical advantage of combining deep feature extractors with lightweight classical classifiers, offering a balanced trade-off between computational efficiency and classification performance. Limitations include the small dataset size and restricted range of emotion classes, suggesting directions for future improvement.

Keywords: Facial Expression Recognition, Deep Learning, Image Embeddings, Logistic Regression (LR), Support Vector Machine (SVM).

1. Introduction

Deep learning models' quick development has revolutionized image analysis by enabling remarkably high accuracy in applications like facial recognition, image segmentation, and object detection [1]. Because of their shown capacity to capture complex spatial and contextual details, pretrained convolutional neural networks (CNNs) such as VGG-16, VGG-19, and Inception V3 have gained popularity for extracting robust features from images [1, 2]. Even though these deep models are excellent at end-to-end learning, nothing is known about how to combine their extracted features with more conventional machine learning classifiers like Support Vector Machines (SVM) and Logistic Regression (LR) [3]. This method can be useful for cutting down on computational complexity and expense, particularly in situations where resources or training data are scarce. Finding out if the comparatively simpler models LR and SVM can successfully use the high-dimensional feature embeddings from these CNNs to produce competitive results in picture classification tasks is the difficult part. Previous work has also explored this direction, particularly in efforts to combine deep and classical approaches efficiently: Many research studies have examined

Article Info: Received February 25, 2025; Accepted July 12, 2025; Published August 23, 2025
* Corresponding Author

how to integrate the deep feature with traditional machine learning models such as Support Vector Machines (SVM) and Logistic Regression (LR) for the classification of images [4, 5]. The finding has followed the automatic extraction of features which are complex and hierarchical from the source of images that have given state-of-the-art performance in different computer vision tasks, such as using CNN architectures like VGG-16, VGG-19, and Inception V3 [6, 7]. These deep learning models trained end-to-end on large, labeled datasets perform exceptionally and have been used successfully for various applications, including object detection, facial recognition, and scene classification. However, their success raises the issue of complexity and computational cost in training such deep models, especially in environments with limited computing resources. In search of alternative paths that allow the exploitation of deep learning potential, researchers have considered other methods. This study is motivated by the need to close this gap to expand the applicability of deep learning features to areas that require effective and interpretable solutions.

Logistic regression (LR) and support vector machines (SVM) are widely recognized classical machine learning techniques frequently employed in image analysis, particularly for classification tasks [8]. While these methods are less computationally intensive compared to deep learning approaches, they have demonstrated significant effectiveness when integrated with feature extraction methods, such as scale-invariant feature transform (SIFT), histogram-based techniques, or features derived from deep learning models [4, 9]. Specifically, SVM excels in multi-class classification tasks due to its capacity to process high-dimensional data through kernel methods, whereas LR is predominantly applied in binary classification scenarios because of its straightforward implementation and interpretability [10].

Recent advancements in research have focused on enhancing classification accuracy by incorporating features extracted from pretrained deep neural networks, such as VGG and Inception models, into SVM and LR frameworks [11, 12]. These hybrid approaches have demonstrated that classical classifiers, when supplemented with deep learning-derived features, can achieve competitive performance while maintaining lower computational demands. This characteristic makes these techniques incredibly useful in resource-constrained conditions, bringing a healthy balance between accuracy and efficiency; thus, being suitable to the tasks like object recognition and face identification [13, 14].

Prior research has extensively worked on deep CNNs such as VGG-16, VGG-19, and Inception V 3 for the application of image processing. The rise in popularity of these models is because they are superior to more complex classification problems because of their ability to retrieve hierarchical features from images [2, 15]. Many studies [16, 17] have validated the efficacy of these pretrained CNNs for transfer learning and fine-tuning, especially for those applications where not much labeled data have been available. Although CNNs are powerful at end-to-end learning, not much work has been done on how to harness their feature extraction capabilities with traditional ML models such as Support Vector Machines (SVM) and Logistic Regression (LR). Many studies have focused on incorporating these deep features to build machine learning classifiers that improve the performance of the models, and tasks where access to computational resources is scarce or computational time is critical [18, 19]. For instance, LR and SVM have been well established in many classification tasks with the rich obtained features from CNN, even though, compared to deep networks, they are less sophisticated [20, 21]. Further research into these hybrid techniques' suitability for image analysis tasks is necessary since their potential to increase classification accuracy while lowering computational complexity has not yet been fully realized.

It is aimed for the present study to compare the performance of Support Vector Machine (SVM) and Logistic Regression (LR) models in image classification tasks by using features extracted by pretrained deep learning models, that included Inception V3, VGG-16, and VGG-19. This will focus on the comparison of classification accuracy and computational efficiency of LR and SVM with end-to-end deep learning models and will solve some of the critical questions related to how well the models will work with applying deep learning features [22, 23]. This also determines whether the high accuracy and low cost of computation can be achieved by merging deep learning feature extraction with traditional machine learning models [24]. Such aspects were then studied for the evaluation of advantages and disadvantages in using LR and SVM in the performance of image analysis, including under which circumstances these models might be able to perform better than more complex deep learning methods.

This study attempts to fill the gap between conventional machine learning models and modern deep learning techniques by assessing the comparative performance of Logistic Regression (LR) and Support Vector Machines (SVMs) in image classification tasks using features extracted from pretrained models such as VGG-16, VGG-19, and Inception V3. The study aims at determining the efficacy of such hybrid techniques in terms of classification accuracy, computing efficiency, and suitability for application in real-life scenarios of resource constraint. The study results will be a precious knowledge base for easy machine learning approaches in combination with very strong deep learning feature extractors as potential substitutes for very costly end-to-end deep learning models for image analysis tasks.

However, their success raises the issue of complexity and computational cost in training such deep models, especially in environments with limited computing resources. In search of alternative paths that allow the exploitation of deep learning potential, researchers have considered other methods. That is, research was also found to indicate that even in straightforward tasks, Logistic Regression can perform reasonably well in classification when used with features of deep learning. For example, [25] reported comparably high achievements in terms of image recognition when employing LR as a classification method following deep CNNs' feature extraction.

In the domain of facial expression recognition (FER), several studies have explored the use of deep learning and conventional machine learning methods for emotion classification. Akhand et al. [26] demonstrated the effectiveness of CNN-based transfer learning models in recognizing subtle emotional cues from facial images, while Ullah et al. [27] proposed deep ensemble architectures to recognize occluded or ambiguous facial expressions. Kim et al. [28] suggested a VGG-19 network architecture with bespoke design for FER applications, emphasizing the model's capability to learn fine-grained features that can be generalized to emotional distinctions. In addition, studies such as Patro et al. [29] evaluated the performance of hybrid approaches in FER by combining deep features and conventional classifiers like SVM and decision trees. These works collectively highlight the changing landscape of FER research, where an increasing interest is seen in trading off accuracy, interpretability, and computational complexity. Our research contributes to this body of work by presenting a comparative study of CNN-based embeddings (VGG-16, VGG-19, Inception V3) classified using lightweight models, with the explicit design choice for emotion recognition in computationally constrained environments.

2. Research Methods

One of the practical means of obtaining efficiency and performance in image classification is to use pre-trained convolutional neural networks (CNNs) for feature extraction, which are afterward classified with traditional machine learning classifiers such as Logistic Regression (LR) or Support Vector Machines (SVM) [30]. The methodology eschews the training of deep networks from scratch, thereby forsaking computational overhead but with the cost of maintaining the high representational capability of CNNs [31, 32]. By developing classifier-ready feature embeddings with pretrained networks, the classification then can be performed with efficient and light algorithms. This two-stage process is most optimal for applications in which high accuracy requirements exist but computational resources or labeled data are limited, and therefore it is most optimal for real-world deployment in constrained environments. The stages of the proposed methodology are illustrated in Figure 1.

To evaluate the performance of the suggested method, this study employed a publicly available facial expression dataset. The dataset employed in this study is the FER-2013 (Facial Expression Recognition 2013) dataset, which is publicly available on the Kaggle website [33]. This dataset contains 48×48-pixel grayscale face images of human beings, originally collected via the Google image search API. Each image is stored as a flattened pixel intensity array of values in CSV format, and an emotion label. There are seven emotion classes, i.e., Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral, in the original dataset. However, for this research, we considered five emotion classes Angry, Fear, Happy, Neutral, and Sad to promote class balance and avoid confusion in classification. The data contains a total of 275 images, which were preprocessed and resized wherever necessary before feature extraction using pretrained CNN models.

Using the method of deep learning for feature extraction combined with traditional classifiers has motivated research in this direction. For instance, Dubey & Barskar [32] argued that deep features from CNNs can significantly improve performances in many image classification tasks with simpler machine learning methods, such as SVM and LR, particularly in situations where labeled data is limited. Also, evidence from Agarap [30] indicated that SVMs would out-do standard techniques of image classification only using handcrafted features as compensation with accuracy and computation-inexpensiveness for featuring extraction by deep learning. Prestressing CNN models for feature extraction and SVM classification in achieving high accuracy with low computational costs were emphasized in [31].

These results escalate the pretext that, even if deep learning models are better feature extractors, standalone models like SVM and LR still give splendid performance on conditions that simplicity and efficiency are generally placed above the computational costs.

It's a good beginning for achieving higher performance under low computational cost through the combined use of pretrained CNNs and traditional machine learning classifiers such as LR and SVM, while most studies are inclined toward end-to-end deep learning models. Less research is done on this, and much space is yet unexplored in literature concerning the need for further investigation of LR and SVM in the tasks of image analysis. This work tries to fill such a gap by systematically evaluating the performance of LR and SPM based on characteristics obtained from VGG-16, VGG-19, and Inception V3 models.

VGG-16, VGG-19, and Inception V3 are some sophisticated deep CNN architectures that have greatly impacted feature extraction and image analysis because of their high accuracy and transfer learning capabilities. VGG-16 is a 16-layer network developed by Simonyan and Zisserman that is characterized by a repetitive usage of small 3×3 kernel convolution layers, thus learning complex spatial representations very well [34]. Extensively used for applications like image categorization and facial expression detection, VGG-16 has been shown to have an excellent ability in capturing very complex visual patterns [35]. VGG-19 extends this architecture further by adding three more convolutional layers that make the model capable of fine-grain hierarchical extraction of important features Chillal et al. [36]. This makes VGG-19 perhaps most useful for discerning very subtle changes between face expressions. Both VGG models perform extremely well at producing high-quality embeddings which duly conserve much of the structure and semantics of images for further classification tasks after they have been pretrained on large datasets such as ImageNet.

On the other hand, Inception V3 is a state-of-the-art architecture designed for efficient and scalable processing [37]. It has included some innovations such as using factorized convolutions, asymmetric kernels, and inception modules to process multi-scale data in an efficient computational way [38]. Inception V3, which has been pretrained with very large datasets like ImageNet, is made as such that it can identify very complex patterns and very minute details. This makes Emotions Recognition a real-time task where minute expressions are important. In fact, Inception V3 delivers strong performance embedding high-level visual characteristics for many image analysis tasks. These main architectures proved to be efficient when working together to produce a cross-section of depth of efficiency and richness in features. This study utilized VGG-16, VGG-19, and Inception V3 to design embeddings that are rich in vital properties of facial expressions for feature extraction purposes. These different embeddings were used for different classification tasks, which revealed distinct efficiencies of each architecture concerning feature extraction for different and complex image datasets [29, 39 – 41].

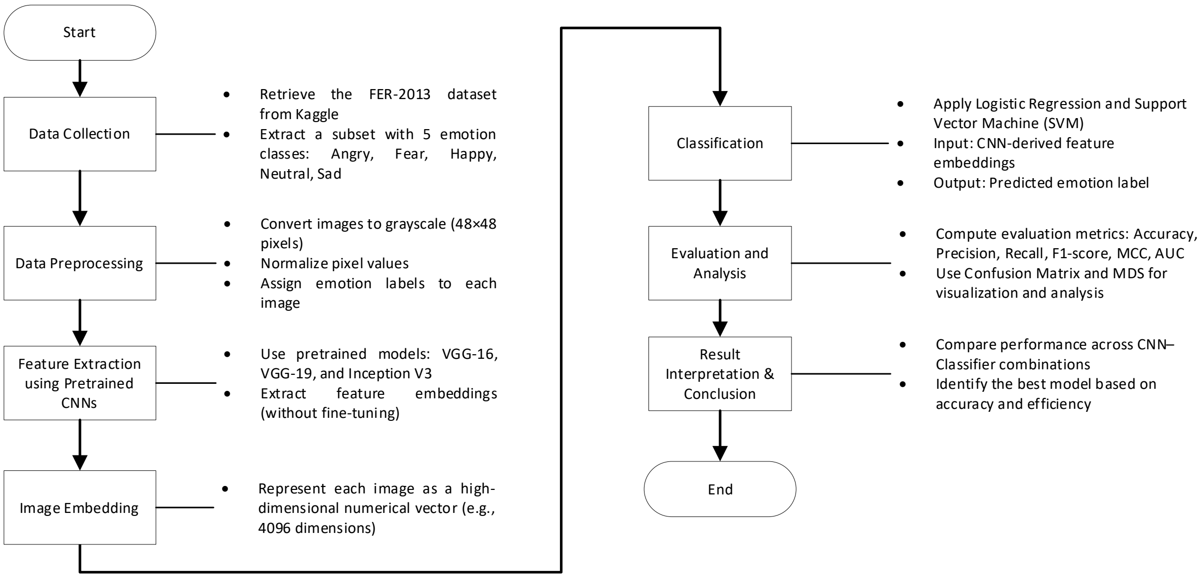

Start
Data Collection
Retrieve the FER-2013 dataset from Kaggle
Extract a subset with 5 emotion classes: Angry, Fear, Happy, Neutral, Sad
Data Preprocessing
Convert images to grayscale (48×48 pixels)
Normalize pixel values
Assign emotion labels to each image
Feature Extraction using Pretrained CNNs
Use pretrained models: VGG-16, VGG-19, and Inception V3
Extract feature embeddings (without fine-tuning)
Image Embedding
Represent each image as a high-dimensional numerical vector (e.g., 4096 dimensions)
Classification
Apply Logistic Regression and Support Vector Machine (SVM)
Input: CNN-derived feature embeddings
Output: Predicted emotion label
Evaluation and Analysis
Compute evaluation metrics: Accuracy, Precision, Recall, F1-score, MCC, AUC
Use Confusion Matrix and MDS for visualization and analysis
Result Interpretation & Conclusion
Compare performance across CNN–Classifier combinations
Identify the best model based on accuracy and efficiency
End

Figure 1. Flowchart of Methodology

2.1 Data Collection

The purpose of this stage is to compile a diversified dataset of images for the different face expressions "Angry", "Fear", "Happy", "Neutral", and "Sad". In supervised learning, every image in the dataset needs to be meticulously annotated with the appropriate face emotion because this is how it is classified. This data set is specifically used in training and testing their machine learning models, which provide face cues for recognizing emotions. One of the frequently referred datasets being used for the purpose is FER-2013, which has hundreds of pictures labeled with facial expressions. It will be very useful for deep learning model training and is a valuable resource for use in facial expression research. The dataset, which is rich in age, gender, and ethnic variance, allows models to generalize very well across demographic groups. For example, in these database collections, algorithms are trained in a way to recognize small variations in face features which relate to specific emotions usually present in subjects to enhance the accuracy of emotion processing. FER-2013 is one of those great datasets, the kind of data that large, labeled datasets become, which makes up valuable, reliable performance by the model and, in turn, develops the field of facial expression recognition.

2.2 Data Preprocessing

Such images undergo preprocessing in this step to ensure their uniformity and readiness for feature extraction. The first and foremost step is scaling, where all images are scaled to a standard dimension (say, 48×48 pixels) to correspond to the input requirements of pre-trained models like Inception V3, VGG-16, and VGG-19. The following step is normalization-in this process; pixel values are adjusted to a standard range for instance (0 to 1) or standardized such that the mauls out biases that arise due to changes in lighting or visual contrasts. This approach improves training efficiency at the classification stage, as feature embeddings are extracted uniformly using pretrained CNNs and no end-to-end retraining is required. Also, tagging should be there, where it makes sure that every image links to specific facial expressions such as "Angry," "Fear," "Happy," "Neutral," or "Sad," so that the model learns to classify them differently from one another. Thus, this kind of tagging provides that rental input that such classification jobs need for learning how to differentiate between two different emotions. All these preprocessing works together to give a trustable dataset, thereby improving the functionality of the model in extracting informative data from the photographs.

2.3 Feature Extraction

They perform the pre-prepared images through strong pre-trained deep learning models, like Inception V3, VGG-16, or VGG-19, to extract relevant features. These models extract general high-level visual patterns from images, as they are trained on large, diverse datasets such as ImageNet. It also identifies complex patterns relating to facial features such as lips, eyes, and overall facial structure. The features will be very significant for the classification of various facial expressions. Inception V3 is very famous for its effectiveness in collecting a diverse range of visual features due to its complex architecture. VGG-16 and VGG-19, however, are specialists in convolutional layers that help concentrate on the finer details of an image to capture even more complex facial features. Thus, these deep models are very well-suited to extracting high-quality features for accurate facial expression recognition. Probing into the extracted characteristics drives rich, discriminative information on which subsequent classification will be based.

2.4 Image Embedding

Deep features extracted from the models like Inception V3, VGG-16, or VGG-19 create embeddings: these are, in fact, the numerical vectors which message the most important information from the image in condensed form. These serve as a smaller form of the image; and that is why it becomes easier for machine learning classifiers to process it. Each embedding contains the most important things that the picture can say, such as important expressions or certain structural details and features or attributes by which a face can be distinguished from others. For example, the 4096-dimensional vector generated by an image through processing in VGG-16 represents in numerical evaluation very important visual characteristics. The compression there enables the model to ignore everything else and focus on what matters most about the picture. All images are therethrough represented in an identical, compact, and consistent way that will reduce any further processing load and increase efficiency in classification. The generated embeddings can assist machine learning algorithms in learning from the data faster and giving better predictions. In summary, they also help in making the identification of expression by faces faster and more accurate.

2.5 Classification

Support vector machine classifiers (SVM) and logistic regression are used to classify facial expressions based on the embeddings generated by feature extraction. These classifiers learn to predict the emotion associated with the retrieved features (Angry, Fear, Happy, Neutral, Sad). The former SVM has the best performance for discrimination among expressions by defining the hyperplane which separates these classes with maximum margin. In contrast, Logistic Regression uses linear functions to find the maximum probability of each emotion and finally selects the one with the highest score as the predicted label. They may differ in that one predicts the most likely expression-plus probabilities while the other maximizes distance between classes. Both are well in classifying facial expressions. Relying on image embeddings, these classifiers can help make accurate predictions.

2.6 Result Analysis

Then, after predicting the facial expression, several measures are used to assess the classifier's performance to check its effectiveness and accuracy. The correctly recognized photos are termed as Accuracy, which gives a general view of how much the model is performing. It indicates the percentage of predicted expressions that were indeed correct and helps in finding the false positives. Recall throws light on the other side of the coin, highlighting all missed predictions or false negatives by showing the proportion of actual expressions detected. Combined with a single score, the F1-score provides a balanced evaluation of precision and recall. These metrics highlight advantages and disadvantages for each classifier and help in the assessment of efficacy. Analyzing these outcomes might reveal which classifier performs best and where enhancements, such correcting biases or fine-tuning model parameters, are needed to improve performance.

The dataset consisted of photos of human faces from Kaggle, and five different kinds of facial expressions could be seen in these photos; they are fear, anger, happiness, sadness, and neutrality; these are the specifics of the data:

To avoid excessive visual clutter, only a few representative samples are shown to illustrate some significant facial expressions such as "Happy" and "Fear." However, data used in this study consists of a balanced subset of five emotion classes with the following approximate sample sizes: Angry (50 images), Fear (53), Happy (66), Neutral (54), and Sad (52). These samples were selected randomly from the FER-2013 dataset and were preprocessed in the same manner before they were used for feature extraction. Although visual samples of each class are not presented in whole here, their statistical distribution is taken into consideration while evaluating.

Components that relate to the facial expression categorized as "fear" consist of 53 data items, such as those given in Figure 2.

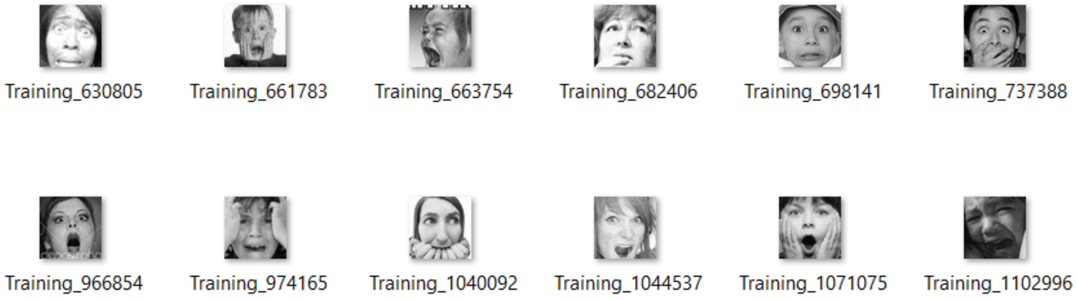


Figure 2. “Fear” Data Example

Components that relate to the facial expression categorized as "happiness" consist of 66 data items, such as that given in Figure 3.

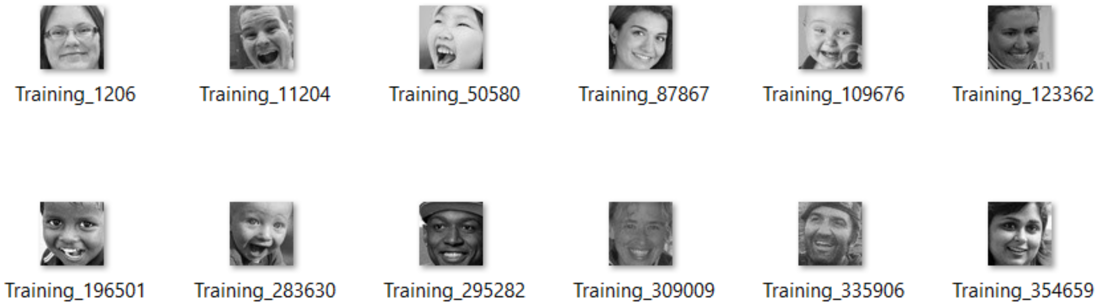


Figure 3. “Happy” Data Example

In all, this research consists of 275 samples of facial expression data, which is incredibly versatile as it covers the whole spectrum of expressions. The set was specially collected for the research so that there would be representative examples for each type of expression. These previous studies have indicated the importance of diversified and balanced datasets, mentioning how such datasets are critical in profiling and testing facial expression recognition systems [42 – 44].

3. Results and Discussion

Based on comparative analysis, the results of image embeddings obtained by different approaches such as Inception V3 (Table 1), VGG-16 (Table 2), and VGG-19 (Table 3) were evaluated on performance metrics as well as using both Logistic Regression and SVM classifiers. These metrics included AUC, accuracy (CA), F1-score, precision (PREC), recall, and MCC. On several counts, Inception V3 always outperformed the rest of the models by performing better. The SVM classifier, for instance, has shown its capability of harnessing the different multi-scale features captured by Inception modules by achieving maximum accuracy when classifying with Inception V3 embeddings (CA = 0.793) and competitive AUC values (0.720). As Shourie et al. (2023) revealed, this clearly demonstrates how effective the network is in detecting complex patterns for high-level feature representations.

Table 1. Image Embedding Using Inception V3

Model	AUC	CA	F1	PREC	RECALL	MCC
LR	0.725	0.756	0.295	0.311	0.280	0.148
SVM	0.720	0.793	0.296	0.387	0.240	0.190

Table 2. Image Embedding Using VGG-16

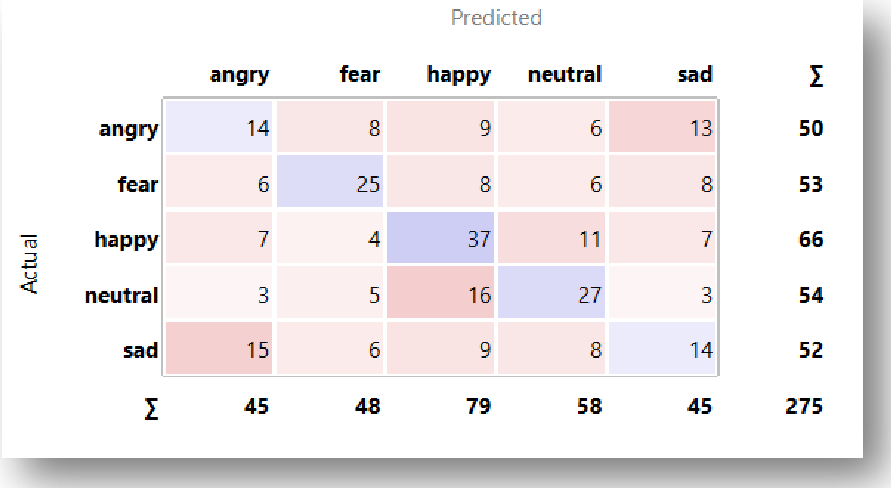
Model	AUC	CA	F1	PREC	RECALL	MCC
LR	0.613	0.764	0.177	0.241	0.140	0.053
SVM	0.614	0.756	0.280	0.302	0.260	0.135

Table 3. Image Embedding Using VGG-19

Model	AUC	CA	F1	PREC	RECALL	MCC
LR	0.701	0.411	0.409	0.409	0.411	0.260
SVM	0.684	0.385	0.373	0.390	0.385	0.227

Despite having a simpler architecture than VGG-19, VGG-16 performed moderately well on all assessed measures. With VGG-16 embeddings, Logistic Regression's accuracy (CA = 0.764) was somewhat greater than SVM's (CA = 0.756). In contrast to Logistic Regression, the SVM classifier demonstrated a better balance between precision (PREC = 0.302) and recall (0.260), as evidenced by the greater F1-score (0.280). As previously noted by Yang (2024), the results confirm that VGG-16 is effective at recovering fine-grained geographic data; nevertheless, its lower AUC and MCC values imply limits in capturing deeper semantic features.

From these observations, it is clear that VGG-19 was able to gain notable improvements in recall and F1-score compared to VGG-16 for having deeper architecture. For an F1-score of 0.409 and AUC 0.701, the comparison was favorable with logistic regression using VGG-19 embeddings in making small changes in facial expressions. It somewhat narrowed the lead on other parameters but failed slightly in accuracy (CA = 0.385). Looking at the number of layers, Chillal et al. (2023) argued that, as compared to VGG-16, VGG-19 would be able to capture more complex spatial patterns. Overall, it was researched that- among the models developed for feature extraction- Inception V3 was found to be the most reliable but VGG-19 surpassed VGG-16 in some aspects. These results can still be said to be consistent with Previously conducted research concerning the advantages of these architectures in image analysis tasks [47, 48].

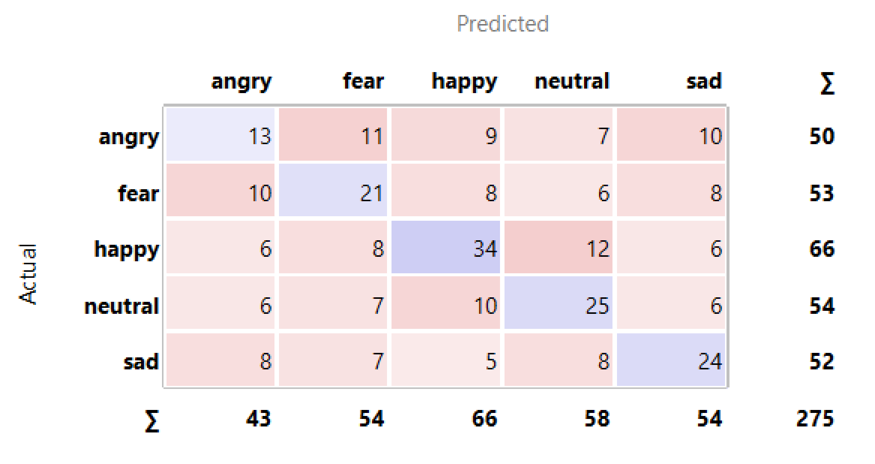


Predicted

Actual	angry	fear	happy	neutral	sad	Σ
angry	14	8	9	6	13	50
fear	6	25	8	6	8	53
happy	7	4	37	11	7	66
neutral	3	5	16	27	3	54
sad	15	6	9	8	14	52
Σ	45	48	79	58	45	275

Figure 4. Confusion Matrix Result (Inception V3)

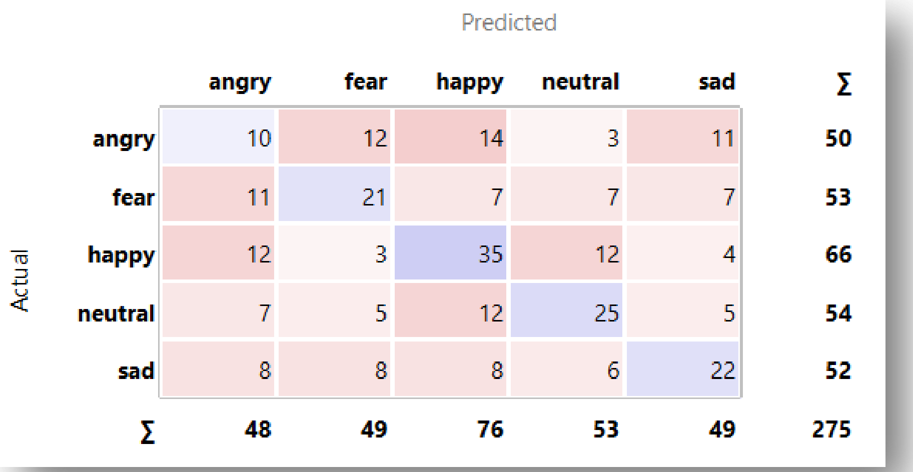
The confusion matrix shown in Figure 4 describes the classification of five emotions as assessed on the FER-2013 dataset. The model obtained relatively higher performances for the "Happy" (37/66 correct) and "Neutral" (27/54) categories, while there were quite a few misclassifications occurring between similar expressions. For instance, a lot of "Sad" samples were predicted as "Angry" (15 cases), and "Neutral" was often confused with "Happy" (16 cases). "Fear" was also dispersed across other classes. These results indicate that some overlapping facial features are driving classification errors for emotions like Neutral, Happy, and Sad. Nevertheless, the diagonal dominance of the matrix suggests that the model is in fact fairly effective when it comes to recognizing separate emotional categories.



				Predicted			
		angry	fear	happy	neutral	sad	Σ
Actual	angry	13	11	9	7	10	50
	fear	10	21	8	6	8	53
	happy	6	8	34	12	6	66
	neutral	6	7	10	25	6	54
	sad	8	7	5	8	24	52
	Σ	43	54	66	58	54	275

Figure 5. Confusion Matrix Result (VGG-16)

Figure 5 represents the classifier has the best accuracy in predicting the emotion "Happy" (34/66 correctly classified) and "Neutral" (25/54), whereas other emotions are misclassified frequently. “Fear” was the most challenging emotion because only 21 out of 53 samples are correctly identified by the model, with a lot of confusion with “Angry” and “Sad”. Similarly, samples characterized as “Angry” may often be confused with “Fear” (11 cases) and “Sad” (10 cases). These results imply that the model only possesses a moderate ability to differentiate between emotions distinguished by minor modifications in facial expression, particularly for negative-affect states. The model at the end demonstrates a consistent but not perfect classification with overlapping patterns that reflect the real-world complexity of emotional expression.



				Predicted			
		angry	fear	happy	neutral	sad	Σ
Actual	angry	10	12	14	3	11	50
	fear	11	21	7	7	7	53
	happy	12	3	35	12	4	66
	neutral	7	5	12	25	5	54
	sad	8	8	8	6	22	52
	Σ	48	49	76	53	49	275

Figure 6. Confusion Matrix Result (VGG-19)

Figure 6 shows the model performs the best in the "Happy" class with 35 correct out of 66 instances, followed by "Neutral" with 25 correct out of 54. Major misclassifications are noted in the "Angry" and "Fear" classes. Only 10 of 50 "Angry" samples were correctly predicted, with the rest tending to be predicted as "Fear" or "Happy." "Fear" was also confused across multiple classes, with 11 being identified as "Angry." These errors suggest trouble separating emotions from similar facial expressions, and especially trouble separating negative emotions. Despite this, the model has good overall accuracy, with very good detection of more unique expressions like "Happy" and "Neutral."

The confusion matrices of Inception V3, VGG-16, and VGG-19 were compared, and significant differences in performance were found among the models in the classification of facial emotions into five categories fear, anger, happiness, sadness, and neutrality. Though balanced in predictions, Inception V3 struggles with confusion between closely related emotions such as fear and neutral, with major misclassifications in these categories. VGG-16 predicts cheerful expressions very well since it suffers from limits in modeling emotional complexity, but it has problems telling apart fear and neutral expressions. On the other hand, VGG-19 proves better than other models in the correct classification of happy and sad expressions as it extracts quite subtle semantic components under emotionally conflicting situations as well. That's why, though it performs nicely, VGG-19 cannot distinguish between the ambivalent emotions like fear or neutral. Thus, it can be seen that all models suffer from the same challenges in terms of distinguishing very small overlaps of emotional displays. This is consistent with previous studies asserting high performance among the deeper architectures such as VGG-19 and Inception V3 in capturing very fine-grained and hierarchical feature representation for better emotion classification tasks [26, 27, 49].

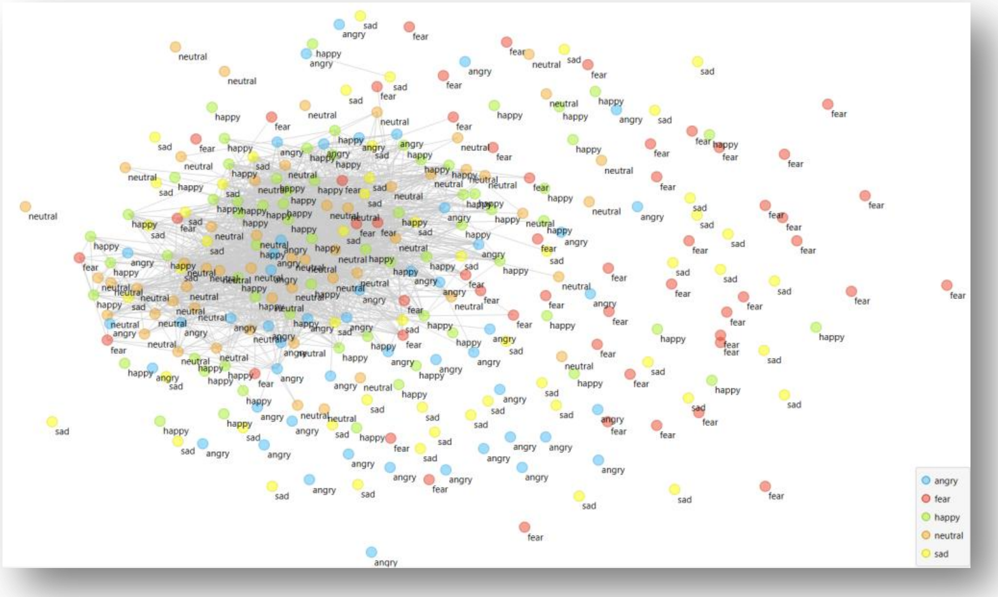


Figure 7. MDS Result (Inception V3)

Figure 7 is a plot of feature embeddings from the CNN model, with each point representing an image colored based on its true emotion label. The plot reveals that "Happy" and "Neutral" are in relatively compact clusters, which means easier separability by the model, while "Fear" and "Sad" are more dispersed and overlap with other emotions. This suggests that the model struggles to distinguish between expressions with subtle facial differences. The observed overlaps agree with the misclassifications in the confusion matrices, indicating challenges in the classification of visually close emotions.

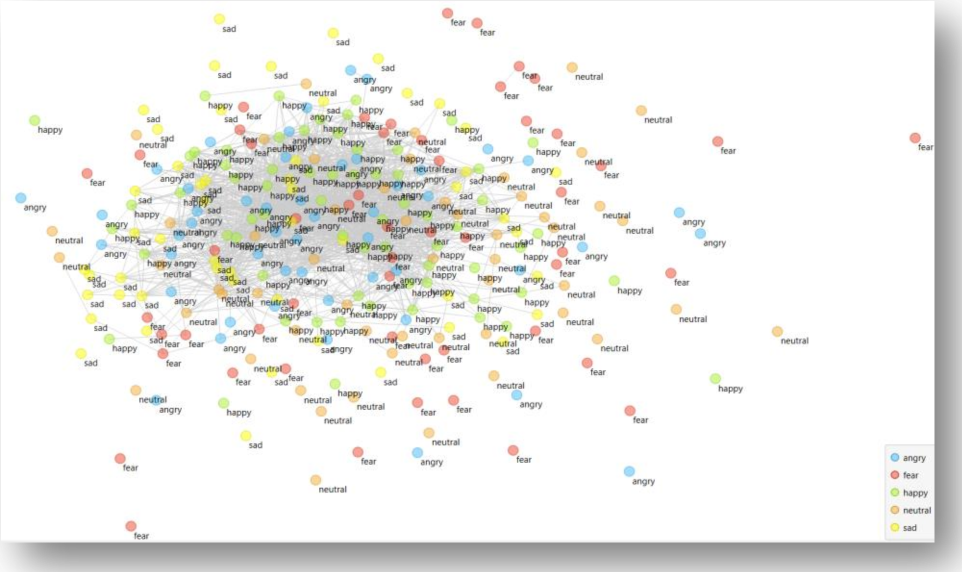


Figure 8. MDS Result (VGG-16)

Figure 8 represents a facial image labeled with its true emotion. The plot reveals significant overlaps between most of the emotion classes, especially between "Fear," "Neutral," and "Sad," such that the model is not doing a good job of separating these classes in the feature space. Although there does appear to be some clustering present most noticeably for "Happy" and portions of "Neutral", the distribution remains close and entangled. This suggests low class separability, which may contribute to the classifying errors found in the confusion matrix.

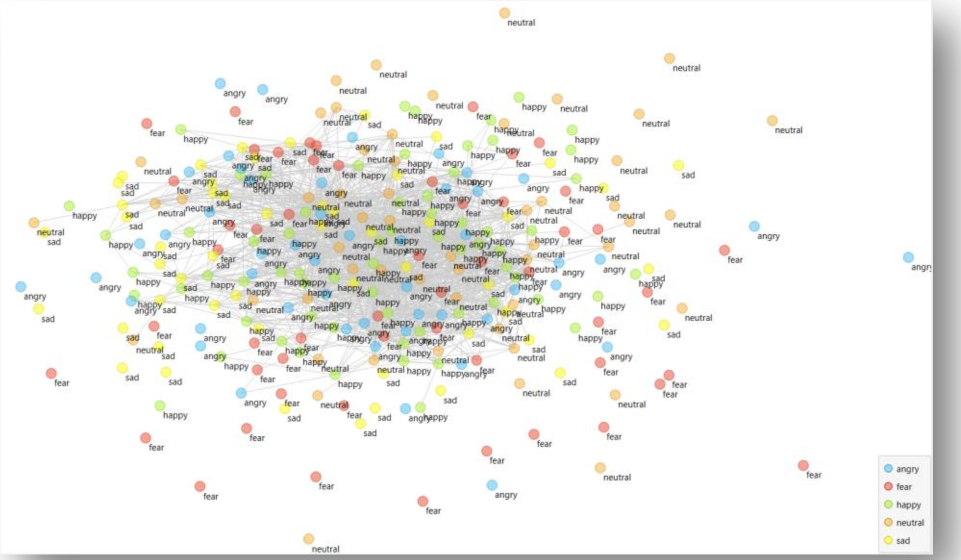


Figure 9. MDS Result (VGG-19)

Figure 9 shows a lot of overlap among emotion classes, particularly among "Fear," "Neutral," and "Sad," which suggests that the model has difficulty separating these classes in the feature space. There is a bit of local clustering for "Happy" and "Angry," but no emotion is an entirely distinct cluster. This reflects the challenge of distinguishing subtle emotional expressions and explains the model's unimpressive performance on overlapping classes in the confusion matrix.

In the case of MDS renderings generated using feature embeddings, the effectiveness for face emotion recognition from models such as Inception V3, VGG-16, and VGG-19 has shown some differences. Although such a somewhat diffused cluster with large overlaps in the classes like fear and neutral exhibits comparatively more difficulty in segregating closer emotional expression by Inception V3, VGG-16 appears to be denser distributed containing less clearly defined cluster that may be due to its being shallower in design compared to VGG-19. VGG-19 does better than the other models and makes more distinguished well-separated clusters, especially for expressions like cheerful and sad, demonstrating more noticeability in the capacity to capture smaller emotional differences. Again, all models retain overlaps in more ambiguous expressions, such as fear and neutral, which persist, highlighting the intrinsic difficulties of the emotion recognition task in such complex data sets. This says that it requires much deeper architecture, such as VGG-19, to develop the important discriminative features but needs further evolution to untangle overlapping classes [29, 39, 40, 41].

The face expression recognition from a comparative study of feature embeddings from these three models has been exhaustively demonstrated for knowledge on the characteristics of Inception V3, VGG-16, as well as VGG-19. In fact, Inception V3 showed overall sound performance as far as Classification Accuracy (CA) and Area Under the Curve (AUC) are concerned due to its highly advanced variant inception modules designed for multi-scale feature extraction. However, the fact that it obtained relatively low scores for recall and precision indicates that, despite their subtle differences, neutral and dread must be considered among the various closely related emotional expressions. In this respect, it shows that while Inception V3 is good at capturing general overall structural and textural patterns, it may also not be capable of capturing detailed, fine emotional cues [50, 51].

For CA, VGG-16 did quite well because of its simpleness and effectiveness for feature extraction with small 3x3 convolutional kernels. However, it displayed poor F1 scores and Matthews Correlation Coefficient (MCC) as can be seen with the confusion matrix and MDS plots, thus indicating the difficulty of dealing with unbalanced and overlapping classes. The display of MDS also supports this finding by having its clusters centralized. Furthermore, this is the same result as what previous studies stated, as this indicates that deeper architectures should be adopted for the development of effective feature discrimination in tasks that include emotion recognition [52, 53].

Of the three networks tested, the best overall performance was achieved by VGG-19, which is an extension of VGG-16 with some more convolutional layers. Its significantly greater F1 and MCC scores indicate that it could strike a good balance between recall and precision, which suits it for detecting even small differences in facial expressions. Enhanced clustering shown in the MDS plots and less overlapping between two expressions, happy and sad, thus demonstrates the capability of VGG-19 in extracting the intricate hierarchy of features. This is supported by earlier studies that indicate the advantages of deeper architectures in difficult classification tasks [51, 53].

The confusion matrix results further corroborate the quantitative metrics while indicating the merits and demerits of the models in relation to each class. Inception V3 and VGG-16 models misclassify fear and neutral emotions, therefore necessitate improved feature embedding techniques or additional data preprocessing methods to deal with these overlaps. These patterns of misclassification are consistent with previous research on facial emotion recognition. For example, Akhand et al. [26] noted significant confusion between "Fear" and "Sad" based on comparable facial structures within the FER-2013 dataset. Similarly, Ullah et al. [27] noted confusion in distinguishing between "Neutral" and low intensity "Happy" facial expressions. These findings attest that while accuracy is improved using deep feature extraction, class-level vagueness remains an issue in emotion recognition tasks.

Difficulty in the differentiation between some of the emotions such as "Neutral" and "Fear" has also been widely acknowledged in earlier studies. Akhand et al. [26] and Ullah et al. [27] noted that facial features that represent these emotions happen to overlap in low-resolution datasets such as FER-2013. Specifically, features such as slightly raised eyebrows or partially open mouths are seen to appear in both classes, confusing even deep learning models. J.H. Kim et al. [49] also noted that increasing the depth of models can improve overall accuracy but will not solve such fine-grained ambiguities. Our own observation that such misclassifications are still present even in networks like VGG-19 and Inception V3 reinforces this, indicating that more subtle features or temporal information beyond static facial images are required. Even for high-tech architecture, it has been observed that tiny differences between emotional categories such as 'indifferent' and 'sad' were misclassified by VGG-19, especially for ambiguous phrases, even though it is also the leading classifier in performance criteria. This shows that facial expression detection tasks are challenging and confirm the importance of dependable designs with effective preprocessing and training techniques [50, 52].

By presenting comparative results, the contribution of model architecture toward performance across a variety of measures is clearly highlighted. While Inception V3 presents good performance on feature extraction and computation, it does call for improvements to its capabilities of achieving finer detail. VGG-16 can very well be used where there are simple tasks, but it does have limitations in class overlaps. Considering the very deep structure that it incorporates, VGG-19 turns out to be the best model for facial expression recognition, proving to be effective where hierarchical representations and sophisticated extraction methods are multitask domains. Future work can therefore look towards hybrid approaches to access the benefits of different architectures needed to solve the remaining challenges in emotion recognition [29, 39, 40, 41].

Note that while feature extraction with CNN still incurs a non-negligible computational cost, the approach adopted in this paper does not involve training or fine-tuning the CNN models. Feature embeddings are computed via frozen pretrained networks in one offline pass, and thus they can be cached and overall computational cost at classification time is significantly reduced. This makes the method particularly suitable for applications where real-time prediction is necessary but training full end-to-end deep models is not possible due to resource constraints.

Again, the class separability and clustering capabilities and limitations of these architectures are demonstrated through the confusion matrix and MDS attacks. While VGG-19 has demonstrated how it preserves its ability in extracting important features through more specific clusters corresponding to different emotion categories, VGG-16 and Inception V3 had a higher level of overlap, hence making it difficult to separate specific emotional states completely. This is in line with past studies emphasizing the need for deeper models and even better optimized feature extraction techniques for emotion recognition tasks, such as Patro et al. [29] and Subathradevi et al. [54].

This research is an addition to the literature since it presents a comparative evaluation of three of the most prevalent CNN structures VGG-19, VGG-16, and Inception V3 exclusively for facial expression recognition in deep feature embeddings. Rather than end-to-end training of CNNs, frozen pretrained models were used to obtain features and performed highly well when combined with traditional classifiers such as SVM and Logistic Regression. This hybrid approach yielded better classification outcomes than traditional models have typically achieved on plain image inputs, as supported by prior research [26, 31] . The findings also show how different CNN architectures produce embeddings with different expressiveness levels, which influence classifier performance depending on task complexity and resources.

Moreover, the results offer pragmatic suggestions for selecting appropriate model pairs. For instance, Inception V3 with SVM generates high accuracy for general purposes, while VGG-19 achieves better precision-recall tradeoff for use cases requiring advanced emotion recognition. VGG-16, being weaker, may have its application in limited resource environments. Such insights attest to the merit of pairing model selection with specific application requirements, including accuracy, explainability, and computational intensity. Finally, this study emphasizes the need for sophisticated evaluation metrics such as MCC and AUC to complement traditional accuracy in measuring classification performance particularly under overlapping classes or class imbalance.

It is important to mention that this study did not include any direct comparisons with two approaches widely used in emotion recognition: the full CNN-based end-to-end classifiers and traditional SVM/LR classifiers with handcrafted features such as HOG and LBP. This omission was done deliberately, given the focus of the study on whether classical models like SVM and LR could efficiently utilize deep CNN-derived embeddings. Since the size of the dataset was small, end-to-end training can, therefore, be prone to overfitting and have been shown in literature to yield inferior performances as compared to handcrafted features. Works by Agarap [30] and Gao et al. [20] further support our assertion that a hybrid approach leveraging deep embeddings for classical classifiers provides a good trade-off between accuracy and computational efficiency. However, further studies will have to include these baselines for additional validation and benchmarking of the hybrid method. This would make further justification for practical implementation in scenarios with constrained resources.

4. Conclusions

The paper analyzes three well-known convolutional neural network architectures, namely, Inception V3, VGG-16, and VGG-19, as tools valuable for recognizing facial expressions. By extensive evaluation criteria like F1 score, accuracy, Matthews Correlation Coefficient (MCC), and others, VGG-19 surpassed all the other models. This was especially evident when fine-tuned to recognize minute differences between the expressions on faces due to its depth and capacity to record complex spatial and semantic elements. Although it had higher architectural features like inception modules, Inception V3 was still very poor, especially in recall and precision for the hard categories extending to fear and neutral. The shallower VGG-16 model gave results that were observation mediocre in comparison but had some trouble in distinguishing between emotional classes that tend to overlap. These results reinforce previous earlier studies pointing to a relationship between depth and classification accuracy when it comes to feature extraction tasks like that of Asha et al. [39, 41] .

However, the study has some limitations. The dataset comprised only 275 samples, which might prove to be insufficient for representing the complexity and variety of facial expressions in daily life situations. Further, the imbalance in the classes in the database made training difficult, especially for unreduced emotion categories detection. The study also used only SVM and logistic regression as classification algorithms, thus limiting the exploration of more complicated methods such as transformer-based models, ensemble learning, or deep end-to-end architectures. Such constraints refer to directions of future work and align with those discussed in earlier studies, for instance Subathradevi [54].

Future studies can improve on this by using larger, more diverse and balanced datasets to improve the generalizability of the results. More advanced classification techniques such as ensemble methods or attention mechanisms could also model deeper and more complex emotional patterns more effectively. Domain adaptation techniques could also render the model more robust when applying it in real-world scenarios with lighting, pose, and occlusion differences. While real-time responsiveness is not experimentally tested in this research, we recognize its importance for real-world deployment. Therefore, a future study must also investigate real-time inference efficiency and system integration so that it can be verified whether the proposed models are suitable for live environments or not. Additionally, incorporating explainability modules can further strengthen model interpretability, which will render it even more suitable for critical fields such as healthcare, education, and human-computer interaction.

This research adds to the growing corpus of studies in image-based emotion recognition by comparing three of the latest CNN feature extractors. It highlights the strengths of VGG-19 and the weaknesses of Inception V3 and VGG-16, all the while identifying significant issues and areas requiring efforts. The findings, therefore, provide a foundation for further advancing research in face emotion identification by providing useful findings and directions for future studies.

References

[1] S. M. Zahid, T. N. Najesh, K. Salman, S. R. Ameen, and A. Ali, “A multi stage approach for object and face detection using CNN,” in *Proceedings of the 8th International Conference on Communication and Electronics Systems (ICCES)*, 2023, pp. 798–803, doi: 10.1109/ICCES57224.2023.10192823.

[2] A. Bakhtiyar, M. A. Ansari, A. Mewada, and D. K. Singh, “From pixels to people: Deep learning breakthroughs in human detection,” in *Proceedings 2024 IEEE 16th International Conference on Communication Systems and Network Technologies (CICN)*, 2024, pp. 326–331, doi: 10.1109/CICN63059.2024.10847522.

[3] K. Kayathri and K. Kavitha, “CGSX ensemble: An integrative machine learning and deep learning approach for improved diabetic retinopathy classification,” *International Journal of Electrical and Electronics Research*, vol. 12, no. 2, pp. 669–681, 2024, doi: 10.37391/IJEER.120245.

[4] L. Deng, H. Li, H. Liu, H. Zhang, and Y. Zhao, “Research on multi-feature fusion for support vector machine image classification algorithm,” in *2021 IEEE International Conference on Electronic Technology Communication and Information (ICETCI)*, 2021, pp. 516–519. doi: 10.1109/ICETCI53161.2021.9563611.

[5] Elpina and G. P. Kusuma, “Revolutionizing computer vision: Enhanced food image classification with Swin transformer and SVM classifier,” *J. Theor. Appl. Inf. Technol.*, vol. 101, no. 23, pp. 7549–7561, 2023.

[6] C. S. Anumol, “Advancements in CNN architectures for computer vision: A comprehensive review,” in *2023 Annual International Conference on Emerging Research Areas International Conference on Intelligent Systems AICERA ICIS 2023*, 2023. doi: 10.1109/AICERA/ICIS59538.2023.10420413.

[7] P. Mahajan, P. Abrol, and P. K. Lehana, “Effect of blurring on identification of aerial images using convolution neural networks,” in *Proceedings of ICRIC 2019*, P. K. Singh, A. Kar, Y. Singh, M. H. Kolekar, and S. Tanwar, Eds., *Lecture Notes in Electrical Engineering*, vol. 597. Cham, Switzerland: Springer, Nov. 2019, pp. 469–484, doi: 10.1007/978-3-030-29407-6_34.

[8] M. Jabardi, “Support vector machines: Theory, algorithms, and applications,” *Infocommunications Journal,* vol. 17, no. 1, pp. 66–75, 2025, doi: 10.36244/ICJ.2025.1.8.

[9] S. Kaur *et al.*, “High-accuracy lung disease classification via logistic regression and advanced feature extraction techniques,” *Egyptian Informatics Journal*, vol. 29, 2025, doi: 10.1016/j.eij.2024.100596.

[10] C. El Morr, M. Jammal, H. Ali-Hassan, and W. El-Hallak, “Support vector machine,” in *Machine Learning for Practical Decision Making*, C. El Morr, M. Jammal, H. Ali-Hassan, and W. El-Hallak, Eds., *International Series in Operations Research & Management Science*, vol. 334. Cham, Switzerland: Springer, Nov. 2022, pp. 385–411, doi: 10.1007/978-3-031-16990-8_13.

[11] C. Ruvinga, D. Malathi, and J. D. Dorathi Jayaseeli, “Human concentration level recognition based on VGG16 CNN architecture,” *International Journal of Advanced Science and Technology*, vol. 29, no. 6 Special, pp. 1364–1373, 2020.

[12] S. Sasidharan Nair and M. Subaji, “Automated identification of breast cancer type using novel multipath transfer learning and ensemble of classifier,” *IEEE Access*, vol. 12, pp. 87560–87578, 2024, doi: 10.1109/ACCESS.2024.3415482.

[13] Z. Kharazian, M. Rahat, E. Fatemizadeh, and A. M. Nasrabadi, “Increasing safety at smart elderly homes by human fall detection from video using transfer learning approaches,” in *Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference*, 2020, pp. 2774–2780. doi: 10.3850/978-981-14-8593-0_4820-cd.

[14] H. Mehraj and A. H. Mir, “Person identification using fusion of deep net facial features,” *International Journal of Innovative Computing and Applications*, vol. 12, no. 1, pp. 56–63, 2021, doi: 10.1504/IJICA.2021.113618.

[15] B. D. Pérez-Pérez, J. P. García Vázquez, and R. Salomón-Torres, “Evaluation of convolutional neural networks’ hyperparameters with transfer learning to determine sorting of ripe medjool dates,” *Agriculture Switzerland*, vol. 11, no. 2, pp. 1–12, 2021, doi: 10.3390/agriculture11020115.

[16] A. Koyama, Y. Murakami, S. Miyauchi, K. Morooka, H. Hojo, and H. Einaga, “Analysis of TEM images of metallic nanoparticles using convolutional neural networks and transfer learning,” *J Magn Magn Mater*, vol. 538, 2021, doi: 10.1016/j.jmmm.2021.168225.

[17] M. S. Seyfioglu and S. Z. Gurbuz, “Deep neural network initialization methods for micro-doppler classification with low training sample support,” *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2462–2466, 2017, doi: 10.1109/LGRS.2017.2771405.

[18] N. Boudouh and B. Mokhtari, “Enhancing image classification with ensemble deep learning through deep feature concatenation,” in *2024 1st International Conference on Innovative and Intelligent Information Technologies Ic3it 2024*, 2024. doi: 10.1109/IC3IT63743.2024.10869349.

[19] J. Liao *et al.*, “A machine learning-based feature extraction method for image classification using ResNet architecture,” *Digital Signal Processing A Review Journal*, vol. 160, 2025, doi: 10.1016/j.dsp.2025.105036.

[20] F. Gao, J.-G. Hsieh, and J.-H. Jeng, “A study on combined CNN-SVM model for visual object recognition,” *Journal of Information Hiding and Multimedia Signal Processing*, vol. 10, no. 4, pp. 479–487, 2019.

[21] T. Sheng, H. Wu, and Z. Yue, “An English text classification method based on TextCNN and SVM,” in *Proceedings 2022 3rd International Conference on Electronic Communication and Artificial Intelligence Iwecai 2022*, 2022, pp. 227–231. doi: 10.1109/IWECAI55315.2022.00052.

[22] M. Behr, S. Saiel, V. Evans, and D. Kumbhare, “Machine learning diagnostic modeling for classifying fibromyalgia using B-mode ultrasound images,” *Ultrason Imaging*, vol. 42, no. 3, pp. 135–147, 2020, doi: 10.1177/0161734620908789.

[23] K. A. AlAfandy, H. Omara, M. Lazaar, and M. A. Achhab, “Investment of classic deep CNNs and SVM for classifying remote sensing images,” *Advances in Science*

Technology and Engineering Systems, vol. 5, no. 5, pp. 652–659, 2020, doi: 10.25046/AJ050580.

[24] Y. Cui and G. Wang, “Research on feature extraction and classification accuracy improvement system through big data technology,” in *Proceedings 2024 International Conference on Computers Information Processing and Advanced Education Cipae 2024*, 2024, pp. 341–348. doi: 10.1109/CIPAE64326.2024.00068.

[25] A. Kujur and Z. Raza, “Ability of machine learning and deep learning models for multiclass classification of kidney stone and lung cancer from computed tomography images: A comparative study,” *Def Life Sci J*, vol. 10, no. 1, pp. 23–30, 2025, doi: 10.14429/dlsj.19188.

[26] M. A. H. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, “Facial emotion recognition using transfer learning in the deep CNN,” *Electronics Switzerland*, vol. 10, no. 9, 2021, doi: 10.3390/electronics10091036.

[27] Z. Ullah *et al.*, “Emotion recognition from occluded facial images using deep ensemble model,” *Computers Materials and Continua*, vol. 73, no. 3, pp. 4465–4487, 2022, doi: 10.32604/cmc.2022.029101.

[28] J. Kim and S. O. Choi, “The intensity of organizational change and the perception of organizational innovativeness; with discussion on open innovation,” *Journal of Open Innovation Technology Market and Complexity*, vol. 6, no. 3, 2020, doi: 10.3390/JOITMC6030066.

[29] B. D. K. Patro, U. Yadav, and N. Yadav, “Brain tumor segmentation using UNet with VGG19,” in *2024 International Conference on Electrical Electronics and Computing Technologies Iceect 2024*, 2024. doi: 10.1109/ICEECT61758.2024.10738961.

[30] I. Abdivokhidov and M. U. A. Ayoobkhan, “Machine learning based image classification with COREL 1K dataset,” in *Lecture Notes in Electrical Engineering*, vol. 558 LNCE. Singapore: Springer, 2025, doi: 10.1007/978-981-97-8345-8_6.

[31] S. Jiang, R. Hartley, and B. Fernando, “Kernel support vector machines and convolutional neural networks,” in *2018 International Conference on Digital Image Computing Techniques and Applications Dicta 2018*, 2019. doi: 10.1109/DICTA.2018.8615840.

[32] U. Dubey and R. Barskar, “Convolutional neural network in deep learning for object tracking: A review,” in *Lecture Notes in Networks and Systems*, vol. 832. Singapore: Springer, 2024, doi: 10.1007/978-981-99-8129-8_28.

[33] M. Sambare, “FER-2013: Facial expression recognition dataset,” *Kaggle*, 2013.

[34] A. Narayanan, V. Aiswaryaa, A. T. Anand, and N. Kadiresan, “Real-time detection of distracted drivers using a deep neural network and multi-threading,” in *Lecture Notes in Networks and Systems*, vol. 1133. Singapore: Springer, 2021, pp. 89–100, doi: 10.1007/978-981-15-3514-7_8.

[35] M. M. Bala, K. J. L. Bai, S. Kattubadi, G. Pulipati, and A. Balguri, “Deep learning transformations in content-based image retrieval: Exploring the Visual Geometry Group (VGG16) Model,” in *2024 IEEE 4th International Conference on ICT in Business Industry and Government Ictbig 2024*, 2024. doi: 10.1109/ICTBIG64922.2024.10911292.

[36] V. Chillal, S. V. Budihal, and S. V. Siddamal, “Plant disease detection with learning-based models,” in *ICT Systems and Sustainability (ICT4SD 2023)*, M. Tuba, S. Akashe, and A. Joshi, Eds., *Lecture Notes in Networks and Systems*, vol. 765. Singapore: Springer, Nov. 2023, pp. 21–31, doi: 10.1007/978-981-99-5652-4_3.

[37] S. S. Roy, R. Kardan, and J. Neubert, “A deep learning-based model for melanoma detection in both dermoscopic and digital images,” in *IEEE International Conference*

on Electro Information Technology, 2024, pp. 668–673. doi: 10.1109/eIT60633.2024.10609889.

[38] K. A. T. Indah, I. Ketut Gede Darma Putra, M. Sudarma, and R. S. Hartati, “Smoothing convolutional factorizes Inception V3 labels and transformers for image feature extraction into text segmentation,” in *Proceedings International Conference on Smart Green Technology in Electrical and Information Systems Icsgteis*, 2023, pp. 139–144. doi: 10.1109/ICSGTEIS60500.2023.10424317.

[39] Y. Liu, “Facial expression recognition model based on improved VGGNet,” in *2023 4th International Conference on Electronic Communication and Artificial Intelligence Icecai 2023*, 2023, pp. 404–408. doi: 10.1109/ICECAI58670.2023.10177007.

[40] I. N. Yulita, F. Ardiansyah, A. Sholahuddin, R. Rosadi, A. Trisanto, and M. R. Ramdhani, “Garbage classification using Inception V3 as image embedding and extreme gradient boosting,” in *2024 Asu International Conference in Emerging Technologies for Sustainability and Intelligent Systems Icetsis 2024*, 2024, pp. 1394–1398. doi: 10.1109/ICETSIS61505.2024.10459560.

[41] P. Asha, A. Vipulendiran, A. Kumaravelu, J. Refonaa, S. L. Jany Shabu, and L. K. Joshila Grace, “Emotion detection by employing deep learning CNN model,” in *2nd IEEE International Conference on Data Science and Information System Icdsis 2024*, 2024. doi: 10.1109/ICDSIS61070.2024.10594468.

[42] G. I. Tutuianu, Y. Liu, A. Alamäki, and J. Kauttonen, “Benchmarking deep facial expression recognition: An extensive protocol with balanced dataset in the wild,” *Eng Appl Artif Intell*, vol. 136, 2024, doi: 10.1016/j.engappai.2024.108983.

[43] G. Jackson and D. Valles, “Dataset enlargement with generative adversarial neural networks,” in *2024 IEEE 5th World AI Iot Congress Aiiot 2024*, 2024, pp. 45–51. doi: 10.1109/AIIoT61789.2024.10578969.

[44] S. Wang, H. Tang, and L. Chai, “Class imbalance in facial expression recognition by GCN with focal loss,” in *Proceeding 2021 China Automation Congress Cac 2021*, 2021, pp. 3270–3275. doi: 10.1109/CAC53003.2021.9727624.

[45] P. Shourie, V. Anand, R. Chauhan, G. Verma, and S. Gupta, “A deep dive into gender classification using Inception V3: Performance and insights,” in *2023 Global Conference on Information Technologies and Communications Gcitc 2023*, 2023. doi: 10.1109/GCITC60406.2023.10426245.

[46] R. Yang, “Convolutional neural network for image classification research - based on VGG16,” in *2024 International Conference on Image Processing Computer Vision and Machine Learning Icicml 2024*, 2024, pp. 213–217. doi: 10.1109/ICICML63543.2024.10958042.

[47] V. Kavitha and K. Ulagapriya, “Comparative evaluation for brain tumor detection using Inception-V3 architecture,” *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 1, pp. 277–283, 2024.

[48] K. Su, R. Xu, Z. Wu, D. Li, L. Chen, and J. Qin, “Lightweight Inception-V3 with multi-scale feature fusion in crop disease identification,” in *Proceedings of 2024 IEEE 25th China Conference on System Simulation Technology and Its Application Ccssta 2024*, 2024, pp. 275–280. doi: 10.1109/CCSSTA62096.2024.10691791.

[49] J. H. Kim, A. Poulose, and D. S. Han, “CVGG-19: Customized visual geometry group deep learning architecture for facial emotion recognition,” *IEEE Access*, vol. 12, pp. 41557–41578, 2024, doi: 10.1109/ACCESS.2024.3377235.

[50] X. Wang, Y. Shi, and S. Fu, “Garbage image classification using improved Inception-V3 neural network,” in *2024 International Conference on Intelligent Robotics and Automatic Control Irac 2024*, 2024, pp. 300–309. doi: 10.1109/IRAC63143.2024.10871539.

[51] E. P. B. Guidang, “Inception-v3-based recommender system for crops,” in *ACM International Conference Proceeding Series*, 2019, pp. 106–109. doi: 10.1145/3310986.3310993.

[52] D. Pruthviraja, U. M. Kumar, S. Parameswaran, V. G. Chowdary, and V. Bharadwaj, “Deep convolutional neural network architecture for facial emotion recognition,” *PeerJ Comput Sci*, vol. 10, pp. 1–20, 2024, doi: 10.7717/peerj-cs.2339.

[53] R. Katyal, Y. Narayan, and N. Sharma, “Efficient emotion recognition system based on deep neural network,” in *Aip Conference Proceedings*, 2023. doi: 10.1063/5.0143642.

[54] S. Subathradevi, T. Preethiya, D. Santhi, and G. R. Hemalakshmi, “Facial emotion recognition for feature extraction and ensemble learning using hierarchical cascade regression neural networks and random forest,” *Journal of Circuits Systems and Computers*, vol. 33, no. 18, 2024, doi: 10.1142/S0218126625500112.